

# **ANALYSE SCIENTOMETRIQUE D'UN CORPUS D'ETUDES SUR LES EFFETS DES PARTICULES DE L'AIR AMBIANT EXTERIEUR SUR LA SANTE**

**David Demortain (UMR LISIS, INRA)**

**Marc Barbier (UMR LISIS, INRA)**

**RAPPORT D'ETUDE POUR L'AGENCE NATIONALE DE SECURITE SANITAIRE  
Laboratoire Interdisciplinaire Sciences Innovations Sociétés**

**INRA, UPEM, CNRS, ESIEE**

Convention de Recherche & Développement n° 2017-CRD-11 du 31 Octobre 2018

21 décembre 2018

# Décembre 2018

## Table des matières

<b>INTRODUCTION</b>	<b>5</b>
<b>1. CONTEXTE DE L'ETUDE</b>	<b>5</b>
<b>2. ENJEU DE L'ETUDE</b>	<b>5</b>
<b>3. RESSOURCES POUR L'ETUDE</b>	<b>6</b>
<b>4. ORGANISATION DE L'ETUDE</b>	<b>6</b>
<b>5. RESTITUTION DES RESULTATS</b>	<b>7</b>
<b>PARTIE 1. ANALYSE STATISTIQUE DU CORPUS</b>	<b>9</b>
<b>1. INTRODUCTION</b>	<b>9</b>
<b>2. CORPUS DES ETUDES EPIDEMIOLOGIQUES</b>	<b>10</b>
▫ LISTE DES PRINCIPAUX AUTEURS DES ETUDES EPIDEMIOLOGIQUES	10
▫ LISTE DES INSTITUTIONS	10
▫ PAYS D'APPARTENANCE DES AUTEURS DU CORPUS	11
▫ LISTE DES MOTS-CLES DES ARTICLES PUBLIES, DEFINIS PAR LES AUTEURS	12
▫ LISTE DES REVUES DANS LESQUELLES LES ARTICLES DU CORPUS ONT ETE PUBLIES	13
▫ AUTEURS LES PLUS CITES DANS LES ARTICLES DU CORPUS	13
▫ ARTICLES LES PLUS CITES DANS LES DIFFERENTES PUBLICATIONS CONSTITUANT LE CORPUS	14
▫ REVUES LES PLUS CITEES DANS LE CORPUS	15
<b>3. CORPUS DES ETUDES TOXICOLOGIQUES ANIMALES</b>	<b>16</b>
▫ LISTE DES INSTITUTIONS	16
▫ LISTE DES PAYS	17
▫ LISTE DES AUTEURS	17
▫ LISTE DES REVUES	18
▫ LISTE DES MOTS-CLES	18
▫ LISTE DES AUTEURS LES PLUS CITES	19
▫ LISTE DES REVUES LES PLUS CITEES	19
▫ LISTE DES ARTICLES LES PLUS CITES	20
<b>PARTIE 2. ANALYSE DE LA STRUCTURATION DU DOMAINE DE RECHERCHE</b>	<b>22</b>
<b>1. CORPUS DES ETUDES EPIDEMIOLOGIQUES</b>	<b>22</b>
▫ RESEAU DE CO-PUBLICATION	22
▫ CARTES DE COOCCURRENCE DES TERMES UTILISES DANS LES RESUMES DES ARTICLES	23
▫ CARTE DES LIENS ENTRE AUTEURS ET MOTS-CLES	24
▫ CARTE DES REFERENCES CITEES FREQUEMMENT ENSEMBLE DANS LES ARTICLES	24
▫ AUTEURS, SELON LES REFERENCES QU'ILS CITENT LE PLUS FREQUEMMENT	25
<b>2. CORPUS DES ETUDES ANIMALES</b>	<b>26</b>
▫ RESEAU DE CO-PUBLICATION	26
▫ CARTES DE COOCCURRENCE DES TERMES DANS LES RESUMES DES ARTICLES	27
▫ CARTE DES LIENS ENTRE AUTEURS ET MOTS-CLES	28
▫ ETUDE DU FINANCEMENT DU CHAMP DE RECHERCHE SUR CORPUS EPIDEMIOLOGIQUE ET TOXICOLOGIQUE COMBINE	29
<b>PARTIE 3. ANALYSE DE LA DISTRIBUTION DES TERMES DE LA SAISINE DANS LE DOMAINE DE RECHERCHE</b>	<b>32</b>
<b>1. DISTRIBUTION DANS LE CORPUS D'ETUDES EPIDEMIOLOGIQUES</b>	<b>32</b>
▫ CARTE DE COOCCURRENCE DES TERMES DANS LES ETUDES EPIDEMIOLOGIQUES	33

▫ RESEAU DES AUTEURS EN FONCTION DE LEUR PROXIMITE AVEC LES TERMES DEFINIS PAR L'ANSES	35
▫ RESEAU DES RELATIONS ENTRE INSTITUTIONS ET TERMES ANSES	36
▫ RESEAU DES RELATIONS ENTRE PAYS D'ORIGINE DES AUTEURS ET TERMES ANSES	37
<b>2. DISTRIBUTION DANS LE CORPUS D'ETUDES TOXICOLOGIQUES</b>	<b>38</b>
▫ COOCCURRENCE DES TERMES DANS LES ABSTRACTS DES ETUDES TOXICOLOGIQUES	38
▫ RESEAU DES RELATIONS ENTRE PAYS D'ORIGINE DES AUTEURS ET TERMES ANSES	39
▫ RESEAU DES RELATIONS ENTRE INSTITUTIONS ET TERMES ANSES	40
▫ RESEAU DES AUTEURS EN FONCTION DE LEUR PROXIMITE AVEC LES TERMES DEFINIS PAR L'ANSES	42
▫ POSITIONNEMENT DES TERMES ANSES DANS LE CORPUS ETENDU	43
<b>CONCLUSION</b>	<b>47</b>
<b>ANNEXES</b>	<b>49</b>
<b>ANNEXE 1 : LISTE DE TERMES DECRIVANT SOURCES, COMPOSES ET EFFETS COUVERTS PAR LA SAISINE</b>	<b>49</b>
<b>ANNEXE 2</b>	<b>53</b>

Ce rapport présente une série d'analyses scientométriques d'un corpus d'études toxicologiques et épidémiologiques sur les effets de l'exposition à la pollution de l'air. Il contient des tableaux statistiques décrivant les auteurs, institutions, mots-clés (etc.) les plus fréquents à travers les études constituant le corpus. Il comporte des analyses visuelles de différents réseaux structurant le domaine de recherche que compose ce corpus. Il présente par ailleurs une analyse de la pertinence du corpus par rapport aux centres d'intérêt spécifiques de l'agence. Le rapport été réalisé par David Demortain et Marc Barbier, membres du LISIS, en mobilisant l'application CorText-Manager de la plateforme d'analyse quantitative de corpus textuels CorText (<https://www.cortext.net/>)

# Introduction

## 1. Contexte de l'étude

Le domaine de recherche sur les effets sanitaires de la présence dans l'air ambiant des particules fines est à la fois ancien, et très évolutif. Au fil des années, les recherches sur la pollution de l'air se sont en effet portées sur différentes substances, et se sont intéressées à des particules de taille de plus en plus réduite. Bien que pourvu d'une ontologie partagée sur les substances ou les indicateurs de composition de l'air ambiant, ce domaine est différencié selon les pays, car les polluants d'intérêt ne sont pas les mêmes, les équipes sont plus ou moins spécialisées sur l'étude de populations cibles ou de cohortes habitant des régions variées quant aux facteurs déterminant la composition de l'air ambiant. Les enjeux politiques liés à ces recherches peuvent donc être importants, car les mesures de présence de différents polluants — et plus encore les analyses attribuant à l'exposition à l'un ou autre de ces polluants une causalité dans la survenue de maladies — peuvent être à l'origine de réglementations sur les technologies de transport, les émissions industrielles ou d'autres sources, ou bien encore peuvent inspirer des prescriptions de comportement individualisées.

## 2. Enjeu de l'étude

Les agences sanitaires mobilisant l'expertise scientifique pour leur mission d'évaluation des risques, opèrent dans un contexte où les attentes sont élevées (que ce soit en termes d'objectivité, d'indépendance, de transparence, de complétude des informations scientifiques considérées). Les choix faits en amont du travail d'un groupe d'expert chargés d'un avis donné, qu'ils concernent la définition de l'objet de l'expertise, ou la sélection des connaissances jugées pertinentes pour en faire l'évaluation, font partie des éléments qui influencent la réception finale de l'avis. Il y a donc un intérêt particulier pour l'ANSES de disposer d'un regard réflexif sur les champs de connaissance scientifique dans lesquels sont sélectionnées les études scientifiques examinées pour rédiger les avis scientifiques. Les outils d'extraction d'information existants pour constituer des synthèses de connaissance ou des *systematic review*, ainsi que ceux provenant des études scientométriques et des sciences sociales computationnelles sont clairement des ressources pour conduire une réflexion sur cette pratique normalisée qui conduit à identifier le champ pertinent des connaissances certifiées disponibles lors d'une saisine. De ce point de vue, les analyses scientométriques et les études de corpus textuels peuvent constituer un appui complémentaire au travail conduit au cours des expertises, avec d'autant plus de pertinence que celles-ci sont conduites par des chercheurs en prises sur la recherche en sciences sociales de l'expertise scientifique, comme c'est le cas au LISIS.

### **3. Ressources pour l'étude**

La sélection des études scientifiques est menée grâce à la connaissance globale de la littérature scientifique des membres des comités d'experts spécialisés, des groupes de travail et du personnel scientifique de l'agence. Elle est fondée sur les connaissances des experts, mais aussi sur l'exploration des revues de littérature publiées, et des bases de données numériques de publications scientifiques, spécialisés ou généralistes, telles que Scopus ou le Web of Science, ou d'autres bases de données de domaine comme PubMed. La scientométrie développée à partir des années 1960, pour mesurer la production de la connaissance scientifique, opère de façon assez similaire mais avec des objectifs différents. Son développement sur des bases de données bibliographiques a rencontré des déploiements massifs d'instruments et de méthodes d'extraction d'information et de visualisation de connaissances. Cette étude repose sur des analyses statistiques diverses permettant d'analyser sous différents angles la production scientifique à différentes échelles et dans le temps. Elle est structurée par la sociologie des sciences, et la compréhension des comportements et des stratégies collectives des scientifiques en matière de publication et plus généralement en matière de construction d'orientations de recherche, de financement, etc. L'étude de domaine de recherche remobilise toujours les bonnes pratiques de la scientométrie (reposant sur la connaissance des bases de données recensant les publications, et la composition des champs auteurs, citations de références, etc.), mais elle tend à introduire des nouvelles sources d'information sur les financements, les projets, les données du web. Les efforts importants des sciences du langage et de l'informatique textuelle ont également changé la donne de l'analyse des contenus de la production scientifique à travers des capacités robustes d'extraction terminologique, de détection d'entités nommées et des approches statistiques sur les termes extraits dans les titres, résumés ou dans le corps des articles. Enfin, les efforts en théorie des graphes et en analyse des réseaux aboutissent à mettre à disposition des algorithmes puissants de calcul de similarité et de clusterisation de tout type de liens établis dans les corpus. Cela se traduit par une capacité d'analyse et de visualisation des relations entre chercheurs (de co-publication, de co-citation, etc.), et de reconstruction des communautés structurant les champs de recherche (algorithmes de détection de communauté). Ce sont ces ressources qui sont agencées et mises à disposition dans l'application en ligne CorText Manager de la plateforme du CorText du LISIS, et qui ont été mobilisées dans le cadre de la présente étude.

### **4. Organisation de l'étude**

Une réunion initiale entre l'équipe du LISIS et l'agence à l'automne 2017 a abouti au choix du sujet des effets sanitaires des particules, pour expérimenter la réalisation d'une telle étude scientométrique. Conformément à l'idée selon laquelle la scientométrie peut compléter la connaissance acquise des champs de recherche par les experts des comités, groupes de travail ou de l'agence elle-même, ce rapport vise à explorer le corpus des études sélectionnées par les experts de l'agence et du groupe de travail pour la production d'un avis sur les effets sanitaires liés à l'exposition à la pollution atmosphérique par les particules. La saisine n°2014-SA-0156 sous-tendant cet avis, pose les questions suivantes :

- Existe-il des données concluantes sur la différence de toxicité selon la composition et/ou les sources de particules ?
- Le cas échéant, quelles conclusions peut-on tirer sur la toxicité des particules selon la composition et/ou les sources de particules ?

Les experts ont effectué une revue de la littérature épidémiologique et toxicologique disponible. Le rapport d'expertise collective résultant de ce travail a abouti à l'identification de 162 études pertinentes publiées depuis 2013, dont 129 publications pour les études chez l'Homme (122 études observationnelles, 8 études cliniques et 5 études quasi-expérimentales) et 33 publications pour les études chez l'animal d'expérimentation.

## 5. Restitution des résultats

Ce rapport présente les résultats de l'application des méthodes scientométriques à ce corpus d'études et déployées à l'issue de trois réunions de travail avec les équipes concernées de l'Anses<sup>1</sup>. L'ensemble du travail a été réalisé par Marc Barbier et David Demortain, chercheurs INRA, membres du Laboratoire Interdisciplinaire Sciences Innovations Sociétés, sur la plateforme d'analyse quantitative de textes CorText. La première partie propose des tableaux statistiques simples décrivant les auteurs, institutions, mots-clés (etc.) les plus fréquents à travers les études constituant le corpus. La seconde partie comporte des analyses visuelles de différents réseaux structurant le corpus. La troisième partie présente une analyse de la structuration du corpus en tant qu'il porte sur une liste de mots-clés d'intérêt pour l'agence. La conclusion du rapport présente des éléments de synthèses issus de ces trois séries d'analyse.

---

<sup>1</sup> Matteo Redaelli, Guillaume Boulanger et Valérie Pernelet-Joly pour l'UERA, Olivier Merckel pour l'UERAP, Sophie Guitton pour la cellule veille, Thomas Bayeux et Benoit Vergriette pour la MISSES.

## Note sur l'interprétation des cartes

CorText Manager permet d'analyser des réseaux d'entités par la visualisation des partitions optimales (calculé ici avec l'algorithme de Louvain). Trois éléments sont à prendre en compte pour interpréter les réseaux représentés : les nœuds ou points ; les liens entre ces nœuds ; les clusters.

La taille des points varie en fonction du nombre d'occurrences dans le corpus (comptage simple). Par exemple, si le point représente un auteur, le point sera gros si l'auteur a publié un grand nombre d'articles du corpus.

S'agissant des liens : deux points sont reliés par un trait si leur cooccurrence dans le corpus dépasse un certain seuil (choisi par l'analyste). Si la cooccurrence est élevée, le lien sera à la fois plus foncé et plus épais. Des liens peuvent exister entre des points selon le corpus, mais ne pas être représenté sur la carte, du fait du seuil choisi. Leur absence signifie simplement que ce lien est plus faible que les autres qui sont représentés sur la carte.

Les clusters rassemblent des points qui ont des relations fortes entre eux suivant la mesure de similitude choisie (directe ou indirecte). La taille d'un cluster (surface) est proportionnelle aux nombres d'individus qui le composent. La position spatiale des clusters ne doit pas être sur-interprétée : ils sont positionnés par le logiciel de visualisation de manière à rapprocher les clusters qui partagent des points, mais aucune métrique ne s'applique à la longueur des liens. En cas de clusters trop nombreux et trop rapprochés, il convient d'examiner les liens qualitativement et systématiquement. .

# Partie 1. Analyse statistique du corpus

## 1. Introduction

Pour cette première partie de l'analyse, les études composant le corpus ont été extraites du Web of Science une par une (accompagnées de leur résumé, des mots-clés définis par les auteurs, des mots-clés associés par le WOS, des références citées, du nombre de citations, notamment). Les références ainsi collectées ont été assemblées en un corpus, pendant numérique du corpus des 162 études sélectionnées par le groupe de travail (GT) de l'ANSES. Ce corpus numérique a ensuite été traité par la plateforme CorText pour produire un ensemble de tableaux statistiques :

- Auteurs
- Institutions
- Pays
- Mots-clés
- Revues
- Auteurs cités
- Articles cités
- Revues citées

Les analyses ont été conduites séparément sur les 132 études épidémiologiques et les 30 études toxicologiques expérimentales. Les deux domaines étant largement distincts du point de vue des auteurs, institutions, revues, il n'est pas apparu judicieux de les traiter conjointement. Les cartes présentées en seconde partie de ce rapport auraient systématiquement affiché une grande polarisation entre les deux champs. Les tableaux statistiques et les cartes portant sur les études toxicologiques ne doivent pas être sur-interprétés: la « population » d'études étant très limitée, les analyses statistiques faites sur celle-ci produisent nécessairement des résultats quelque peu artificiels, puisqu'en statistique la signification des résultats augmente avec la taille des populations étudiées.

Les comptages simples au niveau des auteurs, publication par institutions, etc., laisse apparaître le poids important dans le corpus des publications produites à travers les projets européens, notamment les derniers projets ESCAPE, à travers lesquels sont étudiées différentes cohortes dans différents pays. Le projet ESCAPE se décline en de multiples publications, portant chaque fois sur un effet différent (par comparaison avec d'autres programmes d'étude des effets de la pollution atmosphérique, comme le programme américain du *Health Effects Institute*, qui produit des rapports contenant de multiples chapitres couvrant tous les aspects d'un effet sur la santé, par exemple les problèmes respiratoires).

## 2. Corpus des études épidémiologiques

### ▫ Liste des principaux auteurs des études épidémiologiques

Le corpus d'études épidémiologiques est constitué d'articles co-écrits. Il compte donc un grand nombre d'auteurs : 897 auteurs en tout ont été répertoriés. Un tiers de ces auteurs a publié (ou co-publié) plus d'une des 130 références du corpus. On peut donc dire, dans l'ensemble, que c'est un domaine de travail scientifique collectif, structuré par des collectifs assez larges. Les auteurs dominant le tableau sont principalement européens, reflétant le fait que la saisine de l'ANSES sur les particules représente un agenda globalement européen. Comme on le verra plus bas, c'est également le reflet de l'investissement en Europe sur la mise en place de nombreuses grandes cohortes, et de l'analyse collective, transeuropéenne, des résultats concernant ces cohortes. Les premiers auteurs font partie d'un groupe d'enseignants-chercheurs basés à l'université d'Utrecht notamment et qui relèvent à la fois du département d'épidémiologie, et de ce que cette université appelle le *Institute for Risk Assessment Science*, centre très mobilisé par l'organisme RIVM, qui produit les évaluations de risque pour le gouvernement.

years	total
brunekreef, b (uni utrecht, nl)	31
beelen, r (utrecht and rivm, nl)	27
hoek, g (utrecht and rivm, nl)	25
perschagen, g (imm, karolinska institute, se)	21
hoffmann, b (uni dusseldorf, ge)	20
de hoogh, k (imperial college, uk)	20
forastiere, f (rome regional health authority, it)	18
korek, m (imm, karolinska institute, se)	17
tsai, my (washington school of public health, seattle, us)	17
peters, a (harvard uni, us)	15
cyrys, j ( augsburg and munich uni, ge)	15
eefkens, m (uni utrech, nl and basel, sw)	15
kramer, u (envi hygiene institute, dusseldorf, ge)	15
stafoggia, m (Lazio regio health service, it)	14
penell, j (IUF Leibniz center, dusseldorf, ge)	14
wang, m (uni utrecht, nl)	14
heinrich, j (epidemio institute, munich, ge)	14
lanki, t (department of public health, fi)	13
de faire, u (karolinska institute, se)	13
cesaroni, g (lazio health regional authority, it)	13

### ▫ Liste des institutions

Ce tableau recense les institutions d'affiliation que les auteurs des articles constituant le corpus ont déclarées lors de la publication des articles constituant le corpus. Au total, le nombre d'institution atteint 370, avec 174 institutions impliquées dans plus d'un article. Le domaine est largement dominé par l'université d'Utrecht en Hollande, suivi d'un groupe

de 5 instituts ou universités, impliqués dans plus d'une trentaine d'études. Ces instituts européens disposent d'un assez grand nombre d'auteurs d'études. Ils abritent de véritables équipes, qui publient ensemble. En dessous d'un seuil de 20 publications environ, les contributions des instituts sont davantage liées à un chercheur unique. Les 4<sup>e</sup> et 5<sup>e</sup> institutions les plus représentées dans le corpus sont américaines. La première et seule institution française représentée est l'INSERM.

Institution	total
univ utrecht (brunekreef b, hoek g)	69
karolinska inst (penell j, cesaroni g, forastiere f, staffoglia m, de faire u)	53
univ london imperial coll (de hoog)	44
univ washington (tsai)	38
harvard univ (peters)	33
univ dusseldorf (hoffmann, penell)	25
univ basel	24
german res ctr environm hlth	23
swiss trop & publ hlth inst	22
Inserm	20
univ groningen	18
ctr res environm epidemiol creal	18
helmholtz zentrum munchen	17
natl inst hlth & welf	17
peking univ	16
lazio reg hlth serv	14
univ ulm	14
natl inst publ hlth & environm (rivm)	14
univ augsburg	13
univ michigan	13
iuf leibniz res inst environm med	13
univ athens	12
univ copenhagen	12
kings coll london	11
univ turin	11
fudan univ	11
norwegian inst publ hlth	11
univ manchester	11
reg agcy environm prevent emilia romagna	10

#### ▫ **Pays d'appartenance des auteurs du corpus**

Contrairement à ce qu'indique le tableau des institutions, c'est bien les Etats-Unis qui dominant le classement des pays d'origine des auteurs du corpus, et ce assez largement. Les 7 pays suivant dans le classement sont européens (Allemagne, Pays-bas, Espagne, Royaume-Uni, Italie, Suède et France), et cumulent cependant un total de 741 mentions (à 741 reprises, les auteurs se sont déclarés comme ayant ces nationalités). Etant donné que les auteurs européens publient largement ensemble, et travaillent dans des projets

européens transnationaux (voir plus bas), on peut dire que le corpus est globalement européen.

Pays		total
Usa		250
Germany	= 741	159
Netherlands		137
Spain		104
Uk		98
italy		96
sweden		80
france		67
china		56
switzerland	54	
denmark	39	
finland	34	
greece	26	
canada	24	
norway	21	
japan	18	
austria	7	
belgium	7	
thailand	5	
taiwan	5	

▫ **Liste des mots-clés des articles publiés, définis par les auteurs**

Le tableau des mots-clés les plus fréquemment employés donne un aperçu des thématiques qui dominent le corpus. C'est l'intérêt pour les particules fines et ultrafines qui dominent le corpus choisi, et leur effet de l'exposition aux particules fines en termes de mortalité, proportionnellement plus que d'autres effets chroniques comme l'asthme ou la pression sanguine.

terme	total
particulate matter	25
air pollution	22
mortality	10
pm2.5	7
time series	6
epidemiology	6
fine particulate matter	5
inflammation	5
asthma	4
oxidative stress	4
chemical constituents	4
fine particles	4

ultrafine particles	3
blood pressure	3
atherosclerosis	3
metals	3
hypertension	3
coarse particles	3
particle size	3
heart rate variability	3

▫ **Liste des revues dans lesquelles les articles du corpus ont été publiés**

Le corpus est très largement dominé par la revue *Environmental Health Perspectives*, qui précède une série très cohérente de titres relevant de la santé et épidémiologie environnementales. Les revues PLOS One et Lancet apparaissent un peu plus généralistes, ont publié 5 des 130 articles du corpus.

revue	total
environmental health perspectives	26
environmental research	8
epidemiology	8
environment international	8
journal of exposure science and environmental epidemiology	8
environmental health	6
occupational and environmental medicine	5
environmental science & technology	5
international journal of hygiene and environmental health	4
science of the total environment	4
american journal of epidemiology	4
european respiratory journal	4
particle and fibre toxicology	4
plos one	3
environmental pollution	3
journal of the air & waste management association	3
inhalation toxicology	2
lancet	2
thorax	2
international journal of environmental research and public health	2

▫ **Auteurs les plus cités dans les articles du corpus**

Contrairement à ce que l'on pourrait attendre d'un corpus de petite taille comme celui-ci, les références sont assez largement distribuées. Elles ne font pas apparaître de centralité excessive d'un ou de quelques auteurs. Les cartes présentées plus bas le confirment. La surprise réside dans le contraste entre cette liste et la liste des auteurs ayant produit le plus d'études dans le corpus : les auteurs les plus représentés ne sont pas les plus cités, et inversement, Maroes Eeftens apparaît comme l'auteur le plus cité, mais ne figure pas dans

les 20 auteurs publiant le plus. Pope, un économiste, 6<sup>e</sup> auteur le plus cité dans le corpus, est le premier non épidémiologiste de la liste. Dans l'ensemble, et alors que les publications reflètent le poids du projet européen ESCAPE, ce sont bien les auteurs américains qui, dans l'ensemble, dominent le tableau des auteurs les plus cités.

Auteur	total
eefkens m (uni utrecht, nl and swiss tropical and public health institute, basel, sw)	85
bell ml (uni yale, us)	78
brook rd (michigan uni)	76
beelen r (uni utrecht)	67
ostro b (CAL EPA)	61
pope ca (brigham uni, us)	58
delfino rj (uni californie)	55
zanobetti a (harvard uni)	51
peng rd (john hopkins uni)	45
cesaroni g (lazio regio authority)	36
brunekreef b (uni utrecht)	36
dominici f (harvard uni)	32
ito k (new york university)	29
cyrys j (helmholtz zentrum münchen, munich)	27
schwartz j (harvard)	26
lippmann m (new york university)	25
atkinson rw (st george uni, london)	25
mostofsky e (harvard uni)	24
hoek g (uni utrecht)	22
ghio aj (environmental protection agency, us)	22

▫ **Articles les plus cités dans les différentes publications constituant le corpus**

Le tableau présente les 20 articles les plus cités par les publications constituant le corpus. Il donne un aperçu de l'univers de références dans lequel se situent les auteurs du corpus, et le champ de discussions et de problématiques auquel ils souhaitent contribuer. Les effets aigus de la pollution atmosphérique, mesurés par les admissions à l'hôpital, dominent les articles américains qui sont cités dans le corpus, tandis que les articles européens principalement cités, portent eux sur la distribution géographique de la pollution atmosphérique.

référence	thème	total
brook rd_2010_circulation	particular matter air pollution and cardiovascular disease	45
eefkens m_2012_environ sci technol	land use regression models in 20 EU areas	31
eefkens m_2012_atmos environ	Spatial variation in EU areas	25
eefkens m_2011_occup environ med	Stability of measured and modelled spatial contrasts in NOx	25
peng rd_2009_environ health persp	Emergency admissions after exposure to fine particles	24

mostofsky e_2012_am j epidemiol	Modelling association between particles and health outcomes	24
beelen r_2013_atmos environ	No2 and Nox land use regress models	24
cesaroni g_2012_environ health-glob	Nitrogen dioxide levels from land use regression models	21
cyrus j_2012_atmos environ	Variation in NOx and NO2 concentrations in 30 areas	21
bell ml_2009_am j resp crit care	Hospital admissions and particle composition	20
ostro b_2007_environ health persp	Components of particles and mortality in California	20
dersimonian r_1986_control clin trials	Meta-analysis in clinical trials	17
ito k_2011_environ health persp	Fine particle constituents and hospitalizations	17
zhou ja_2011_environ health persp	Time series analysis of mortality of fine particles	16
zanobetti a_2009_environ health-glob	Fine/coarse particles and health, a national analysis	16
kelly fj_2012_atmos environ	size, source and chemical composition as determinants of toxicity	15
higgins jpt_2002_stat med	inconsistency in meta-analysis	15
brunekreef b_2002_lancet	Air pollution and health	15
brunekreef b_2005_eur respir j	Epidemiological evidence of effects of coarse particles on health	14
dominici f_2006_jama-j am med assoc	Fine particle air pollution and hospital admission	14

#### ▫ **Revues les plus citées dans le corpus**

En scientométrie, la liste des revues ayant publié les articles les plus cités est en général un bon indicateur de l'univers de référence dans lequel les auteurs du corpus cherchent à se situer, et dans lequel ils cherchent à produire une contribution ou à avoir une influence. Le tableau issu du corpus examiné ici est intéressant, puisque la revue *Environmental Health Perspectives* (également la revue dans laquelle a été publié le plus grand nombre d'études présentes dans le corpus) est plus de trois fois plus citée que la seconde revue la plus citée. C'est la revue de l'institut de recherche américain en santé environnementale, le NIEHS – montrant là aussi, comme avec la liste des auteurs les plus cités, que le champ est structuré par des références américaines.

revue	total
environ health persp	876
atmos environ	266
epidemiology	226
am j resp crit care	195
environ sci technol	174
inhal toxicol	160
am j epidemiol	137
occup environ med	129
circulation	128
environ health-glob	101

lancet	99
environ res	95
j air waste manage	80
eur respir j	76
j expo sci env epid	76
sci total environ	74
jama-j am med assoc	60
new engl j med	60
part fibre toxicol	54
thorax	41

### 3. Corpus des études toxicologiques animales

Les conclusions à tirer des comptages sur ce corpus sont à prendre avec précaution, puisque le corpus ne compte que trente études.

#### ▫ Liste des institutions

La liste des institutions fait apparaitre, au contraire des études épidémiologiques, une domination d'organismes hors Europe, japonais pour le premier, américains ensuite.

institution	total
natl inst environm studies	6
us epa	5
univ so calif	4
univ porto	4
univ calif davis	4
lovelace resp res inst	3
nyu	3
united arab emirates univ	3
ohio state univ	2
cnr	2
univ calif los angeles	2
german res ctr environm hlth	2
columbia univ	2
harvard univ	2
saitama univ	2

## ▫ Liste des pays

Le décompte par pays traduit ce que le précédent tableau indique, notamment la domination américaine, et la présence d'une plus grande diversité de pays hors-Europe.

pays	total
usa	46
germany	9
japan	8
australia	6
italy	4
portugal	4
china	4
u arab emirates	3
israel	3
greece	3
india	2
france	2
uk	2
spain	2
canada	1
mexico	1
oman	1
south korea	1

## ▫ Liste des auteurs

La liste des auteurs fait apparaitre le poids relatif de quelques équipes, particulièrement une équipe japonaise du National Institute for Environmental Studies, de la New York University.

auteurs	total
win-shwe, tt (national Institute for Environmental Studies, jp)	3
hirano, s (national Institute for Environmental Studies, jp)	3
chen, lc (envi medicine, NYU, US)	3
fujitani, y (national Institute for Environmental Studies, jp)	3
gordon, t (envi medicine, NYU, US)	2
galdanes, k (envi medicine, NYU, US)	2
gilmour, mi (environmental protection agency, us)	2
araujo, ja (david geffen school of medicine, los angeles, us)	2
guimaraes, l (University of Porto, pt)	2
dye, ja (environmental protection agency, us)	2
lippmann, m (envi medicine, NYU, us)	2
krantz, qt (environmental protection agency, us)	2
grimalt, jo (Institute of Environmental Assessment and Water Research, barcelona, es)	2
furuyama, a (national Institute for Environmental Studies, jp)	2
horton, l (envi medicine, NYU, us)	2

chillrud, s (columbia university, us)	2
kyi-tha-thu, c (national Institute for Environmental Studies, jp)	2
mauderly, jl (Inhalation Toxicology Research Institute, Albuquerque, us)	2
mesquita, sr (Institute of Environmental Assessment and Water Research, barcelona, es)	2
barata, c (University of Porto, pt)	2

▫ **Liste des revues**

La liste des revues est dominée par deux types de revue, du domaine de la toxicologie environnementale et de l'ingénierie environnementale.

revue	total
inhalation toxicology	4
environmental science & technology	3
science of the total environment	3
particle and fibre toxicology	2
plos one	2
atmospheric environment	2
environmental pollution	2
cellular physiology and biochemistry	2
cardiovascular toxicology	1
toxicological sciences	1
journal of lipid research	1
journal of toxicological sciences	1
air quality atmosphere and health	1
journal of toxicology and environmental health-part a-current issues	1
frontiers in neuroscience	1
international journal of environmental research and public health	1

▫ **Liste des mots-clés**

La liste des mots-clés indique, comparativement aux études épidémiologiques, une entrée davantage par sources ou par composés, que par effets.

terme	total
air pollution	7
diesel exhaust	4
particulate matter	4
oxidative stress	4
mice	4
secondary organic aerosol	3
atherosclerosis	3
brain	3
endothelial progenitor cells	2
diesel exhaust particles	2
reactive oxygen species	2
in vitro exposure	2

gasoline exhaust	2
inflammation	2
lung	2
coal emissions	2
wood smoke	2
pm	2
ahr	1

▫ **Liste des auteurs les plus cités**

Les auteurs appartiennent à quelques institutions ou groupes clés, par exemple le centre de recherche en *inhalation toxicology* d'Albuquerque aux Etats-Unis, ou l'école de santé publique de Harvard. Le groupe des auteurs cités est davantage international que celui des études épidémiologiques, traduisant le fait que ces dernières sont (logiquement) plus souvent réalisées dans les pays européens où le problème de la pollution de l'air est posé, et où se trouvent les cohortes humaines étudiées.

Auteur	total
win-shwe tt (national Institute for Environmental Studies, jp)	24
brook rd (michigan uni)	19
pope ca (brigham uni, us)	18
kodavanti up (University of Mississippi, us)	14
mcdonald jd (Lovelace Respiratory Research Institute, Albuquerque, us)	13
calderon-garciduenas l (University of Montana, us and university of mexico)	12
mauderly jl (Inhalation Toxicology Research Institute, Albuquerque, us)	12
peters a (harvard uni, us)	11
sun qh (Ohio State University Wexner, us)	10
reed md (Lovelace Respiratory Research Institute, Albuquerque, us)	10
mills nl (University of Edinburgh, UK)	10
nemmar a (College of Medicine and Health Sciences, United Arab Emirates University)	9
campen mj (Lovelace Respiratory Research Institute, Albuquerque, us)	8
dockery dw (harvard school of public health, us)	8
ghio aj (environmental protection agency, us)	8
araujo ja (david geffen school of medicine, los angeles, us)	7
godleski jj (harvard uni, us)	7
lund ak (Lovelace Respiratory Research Institute, Albuquerque, us)	6
happo ms (Department of Environmental Health, fi)	6
brunekreef b (uni utrecht, nl)	5

▫ **Liste des revues les plus citées**

revue	total
inhal toxicol	120
environ health persp	98
environ sci technol	52
circulation	48

part fibre toxicol	38
toxicol appl pharm	34
toxicol sci	33
am j resp crit care	30
atmos environ	29
j toxicol env heal a	24
arterioscl throm vas	21
environ res	20
circ res	18
jama-j am med assoc	18
new engl j med	18
sci total environ	17
toxicol lett	13
epidemiology	13
eur respir j	11
chemosphere	11

▫ **Liste des articles les plus cités**

Le premier article cité est le même que celui que lorsque l'on considère les études épidémiologiques. Les références épidémiologiques sont très présentes, montrant que les études toxicologiques se placent dans le contexte de l'étude du problème de la pollution atmosphérique comme problème de santé publique.

références	thème	total
brook rd_2010_circulation	particular matter air pollution and cardiovascular disease	9
pope ca_2002_jama-j am med assoc	Lung Cancer, Cardiopulmonary Mortality, and Long-term Exposure to Fine Particulate Air Pollution	7
sun qh_2005_jama-j am med assoc	Long-term air pollution exposure and acceleration of atherosclerosis and vascular inflammation in an animal model.	6
dockery dw_1993_new engl j med	An association between air pollution and mortality in six U.S. cities	5
godleski jj_2011_inhal toxicol	Toxicological evaluation of realistic emission source aerosols (TERESA): summary and conclusions.	5
araujo ja_2008_circ res	Ambient particulate pollutants in the ultrafine range promote early atherosclerosis and systemic oxidative stress	5
ntziachristos leonidas_2007_part fibre toxicol	Relationship between redox activity and chemical speciation of size-fractionated particulate matter.	5
brook rd_2004_circulation	Air pollution and cardiovascular disease: a statement for healthcare professionals	4
brunekreef b_2005_eur respir j	Epidemiological evidence of effects of coarse airborne particles on health.	4
dergham m_2012_chem res toxicol	Prooxidant and proinflammatory potency of air pollution particulate matter (PM <sub>2.5-0.3</sub> ) produced in rural, urban, or industrial surroundings in human	4

	bronchial epithelial cells (BEAS-2B).	
mauderly jl_2014_inhal toxicol	The National Environmental Respiratory Center (NERC) experiment in multi-pollutant air quality health research	4
pope ca_2004_circulation	Cardiovascular mortality and long-term exposure to particulate air pollution: epidemiological evidence of general pathophysiological pathways of disease.	4
hoffmann b_2007_circulation	Residential exposure to traffic is associated with coronary atherosclerosis.	3
calderon-garciduenas l_2008_toxicol pathol	Long-term air pollution exposure is associated with neuroinflammation, an altered innate immune response, disruption of the blood-brain barrier, ultrafine particulate deposition, and accumulation of amyloid beta-42 and alpha-synuclein in children and young adults.	3
campen mj_2010_toxicol appl pharm	Inhaled diesel emissions alter atherosclerotic plaque composition in ApoE(-/-) mice.	3
happo ms_2010_inhal toxicol	Inflammation and tissue damage in mouse lung by single and repeated dosing of urban air coarse and fine particles collected from six European cities.	3
farina f_2011_toxicol lett	The acute toxic effects of particulate matter in mouse lung are related to size and season of collection.	3
fujitani y_2009_inhal toxicol	Characterization of dilution conditions for diesel nanoparticle inhalation studies.	3
gilmour mi_2007_inhal toxicol	Comparative toxicity of size-fractionated airborne particulate matter obtained from different cities in the United States.	3
cho sh_2009_enviro health persp	Comparative toxicity of size-fractionated airborne particulate matter collected at different distances from an urban highway	3

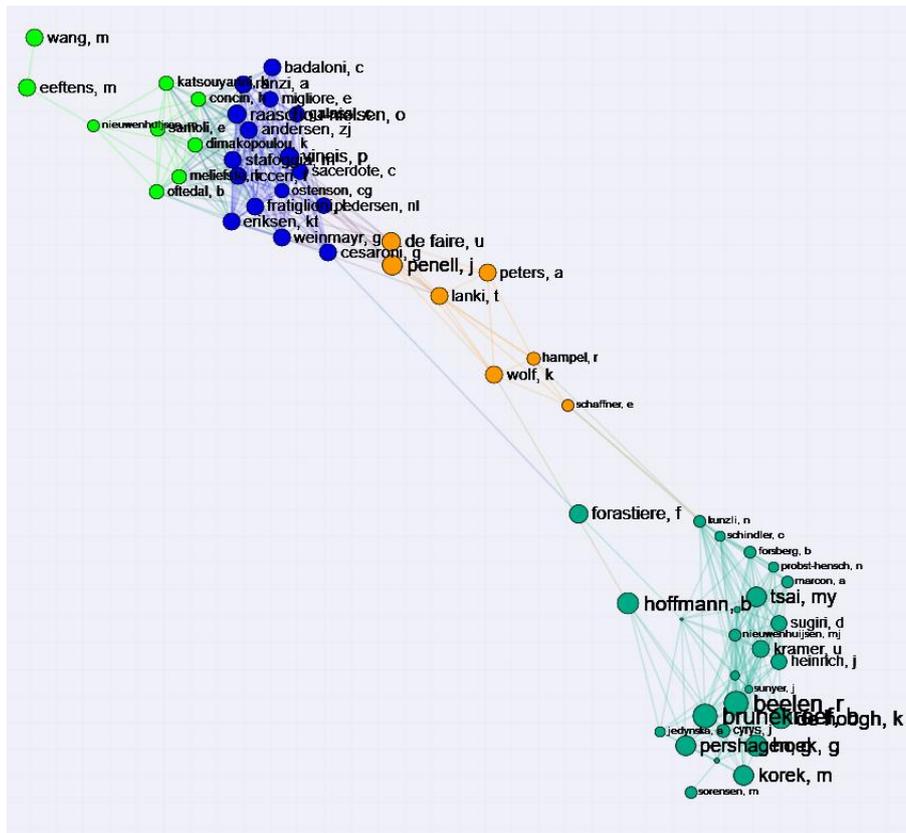
# Partie 2. Analyse de la structuration du domaine de recherche

La structuration d'un champ de recherche peut être saisie par l'analyse de réseaux de relations entre ses auteurs, ses institutions, les thématiques de recherche traitées dans les articles, etc. Ici, on présente une série de visualisations de réseaux : réseau de co-publication (représentation de qui publie avec qui, le plus fréquemment), réseau d'inter-citation (quels sont les auteurs, articles, revues ou institutions le plus fréquemment cités ensemble dans le domaine, ou les sous-champs de références), réseau de mots-clés utilisés dans les abstracts des études (*clusterisation* des mots-clés apparaissant ensemble le plus fréquemment), et divers réseaux dits « hétérogènes » (associations entre auteurs et mots-clés notamment : quels ensembles d'auteurs se constituent autour de quels intérêts scientifiques ?).

## 1. Corpus des études épidémiologiques

### ▫ Réseau de co-publication

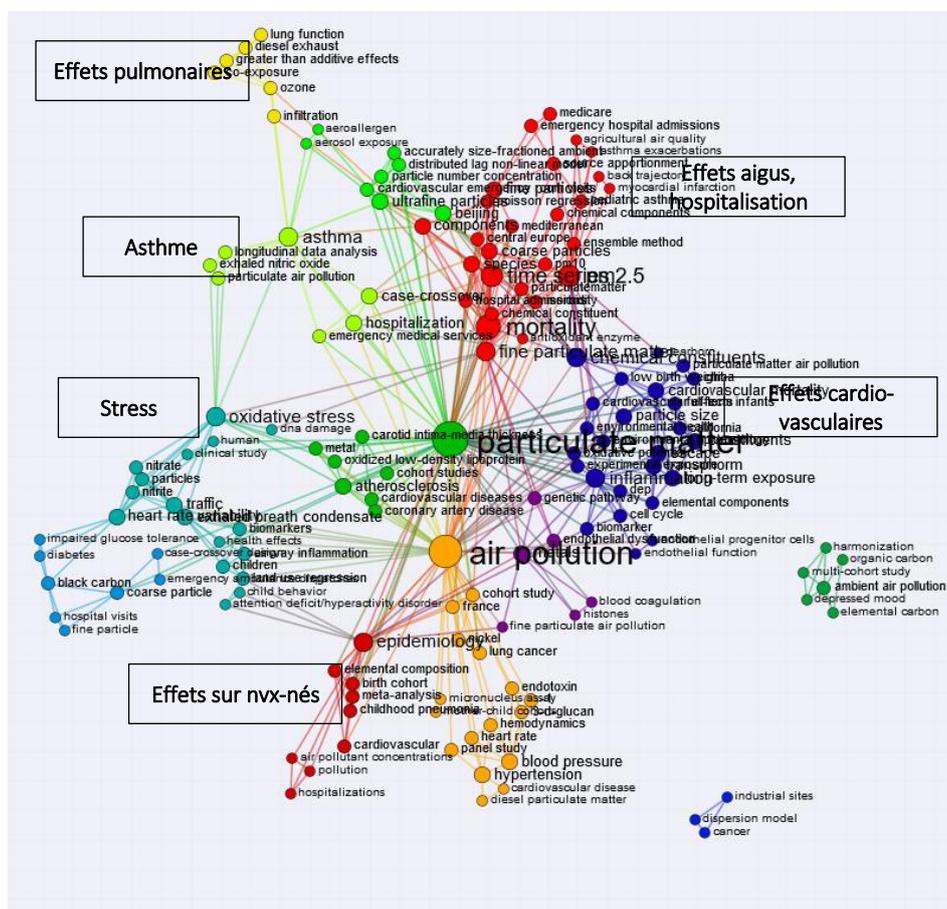
Quatre groupes d'auteurs apparaissent sur cette carte, en deux pôles principaux. Les deux pôles sont constitués d'auteurs publiant collectivement des articles dans le cadre du projet européen ESCAPE.



Le corpus d'études épidémiologiques contient plus de 20 articles issus de ce projet. Les clusters sont constitués de nœuds ou points chaque fois d'une assez grosse taille, indiquant que l'on parle d'un grand nombre de publications impliquant chaque fois un grand nombre d'auteurs. Le cluster du haut à gauche, en bleu, fait plus de place à des auteurs italiens, mais reste transeuropéen, impliquant chercheurs danois, hollandais, allemands. Il matérialise une série de publications issues d'ESCAPE, examinant l'incidence de la pollution atmosphérique du point de vue d'une série d'effets (cancer du sein, cancer du poumon, mortalité naturelle, tumeur cérébrale, évènements cardiovasculaires...), chaque fois signées par l'ensemble du réseau d'auteurs. Le cluster en bas à droite couvre une autre série de publications, globalement des mêmes auteurs, mais plutôt focalisé sur l'étude des composés et des sources de la pollution atmosphérique à travers l'Europe.

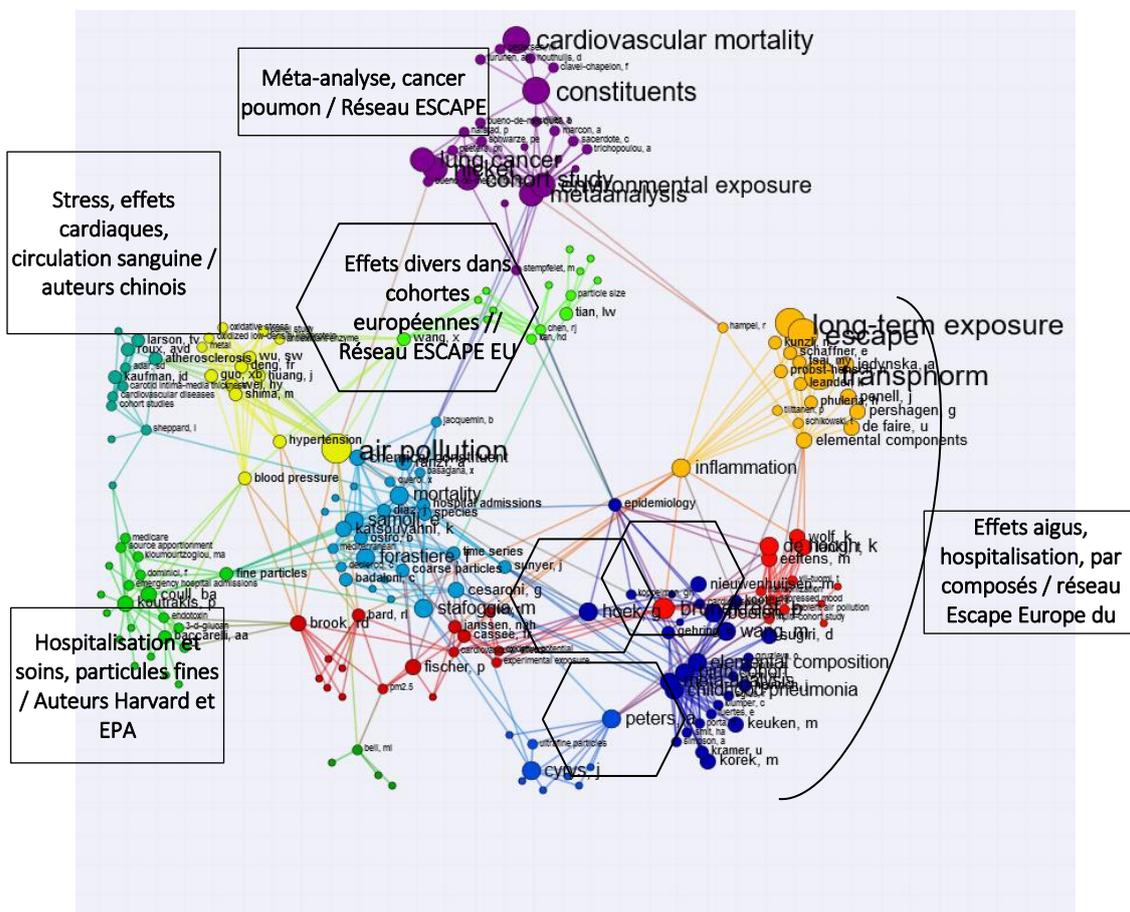
▫ **Cartes de cooccurrence des termes utilisés dans les résumés des articles**

La carte rassemblant les termes les plus fréquemment utilisés conjointement dans les résumés des articles montrent, comme il a été indiqué en réunion de travail par les scientifiques de l'agence et comme on pouvait attendre d'un corpus d'études épidémiologiques, que ce sont les effets qui structurent thématiquement la recherche dans le domaine, davantage que les sources de pollution atmosphérique ou la recherche sur différents composés.



▫ **Carte des liens entre auteurs et mots-clés**

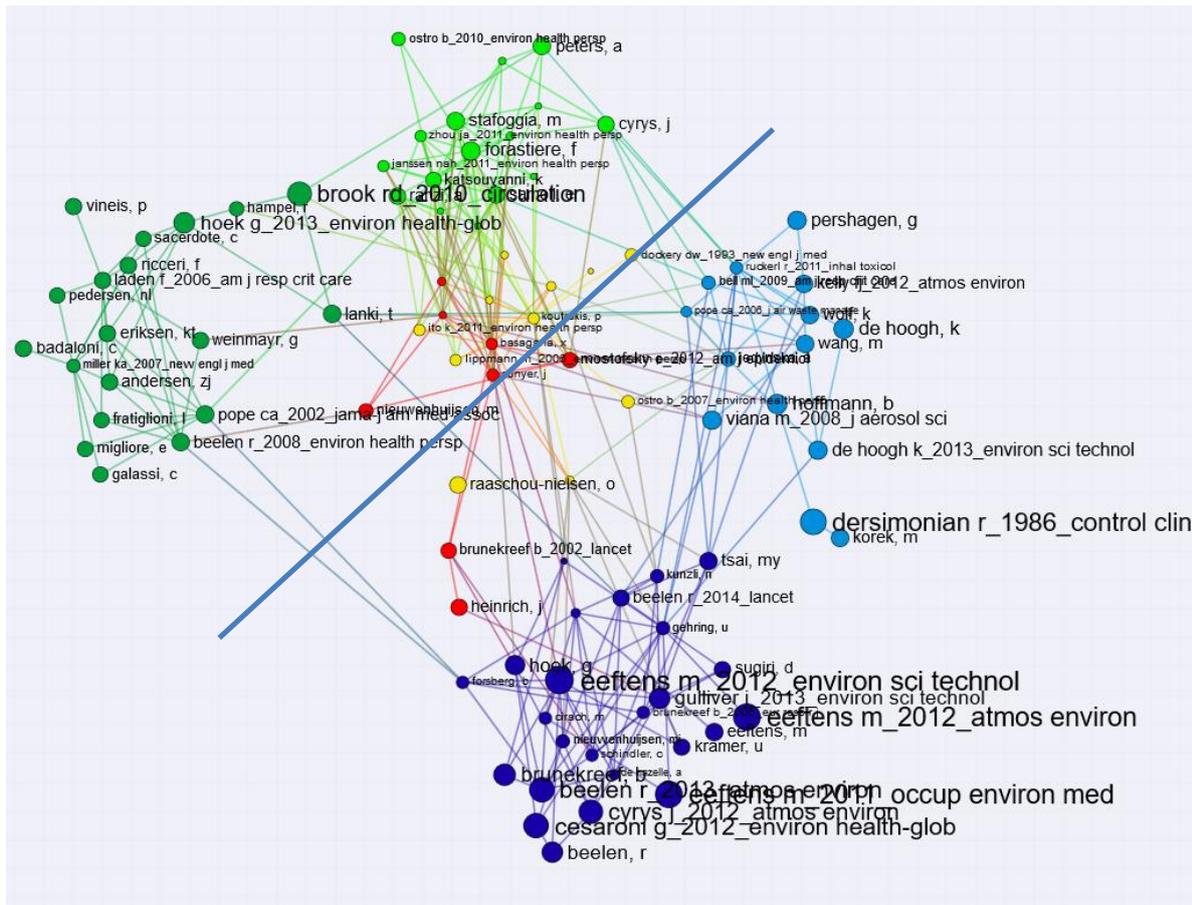
La carte rapprochant les auteurs utilisant de manière fréquente les mêmes mots-clés dans les résumés des articles ne laisse pas apparaître de très grande spécialisation thématique, au vu des liens longs et épais qui relient les clusters, qui sont par ailleurs souvent assez bien délimités. On peut supposer que différents thèmes génèrent des séries de publications collectives, et que des auteurs-clés font le pont entre ces séries de publications thématiques, structurant le champ dans son ensemble, tels un wang, x (boston medical center), un hoek, g (utrecht), un brunekreef, b (utrecht), un peters a. (harvard). Avec sa séries de publication examinant la probabilité de différents effets sur la santé à travers les cohortes d'ESCAPE, le réseau trans-européen « sud » de Staffoglia, Cesaroni ou Forastiere (le cluster bleu foncé de la carte de copublication, bleu ciel ici) se positionne très centralement, et se lie à un ensemble de mots-clés génériques comme *air pollution, fine particles, epidemiology*.



▫ **Carte des références citées fréquemment ensemble dans les articles**

Cette carte fait apparaître six clusters de références citées conjointement dans les études épidémiologiques, et constituant donc autant d'univers de références dans lesquels les auteurs de ces études se situent ou cherchent à se situer. Quatre sous-champs se dégagent. En bas, on identifie un cluster où se situent les articles d'Eeftens étudiant la distribution spatiale de la pollution atmosphérique en Europe par le biais de land use regression models. Le cluster vert en haut à gauche regroupe des références essentiellement américaines sur les liens entre particules et problèmes de santé, notamment problèmes cardiovasculaires. Le cluster bleu ciel





## 2. Corpus des études animales

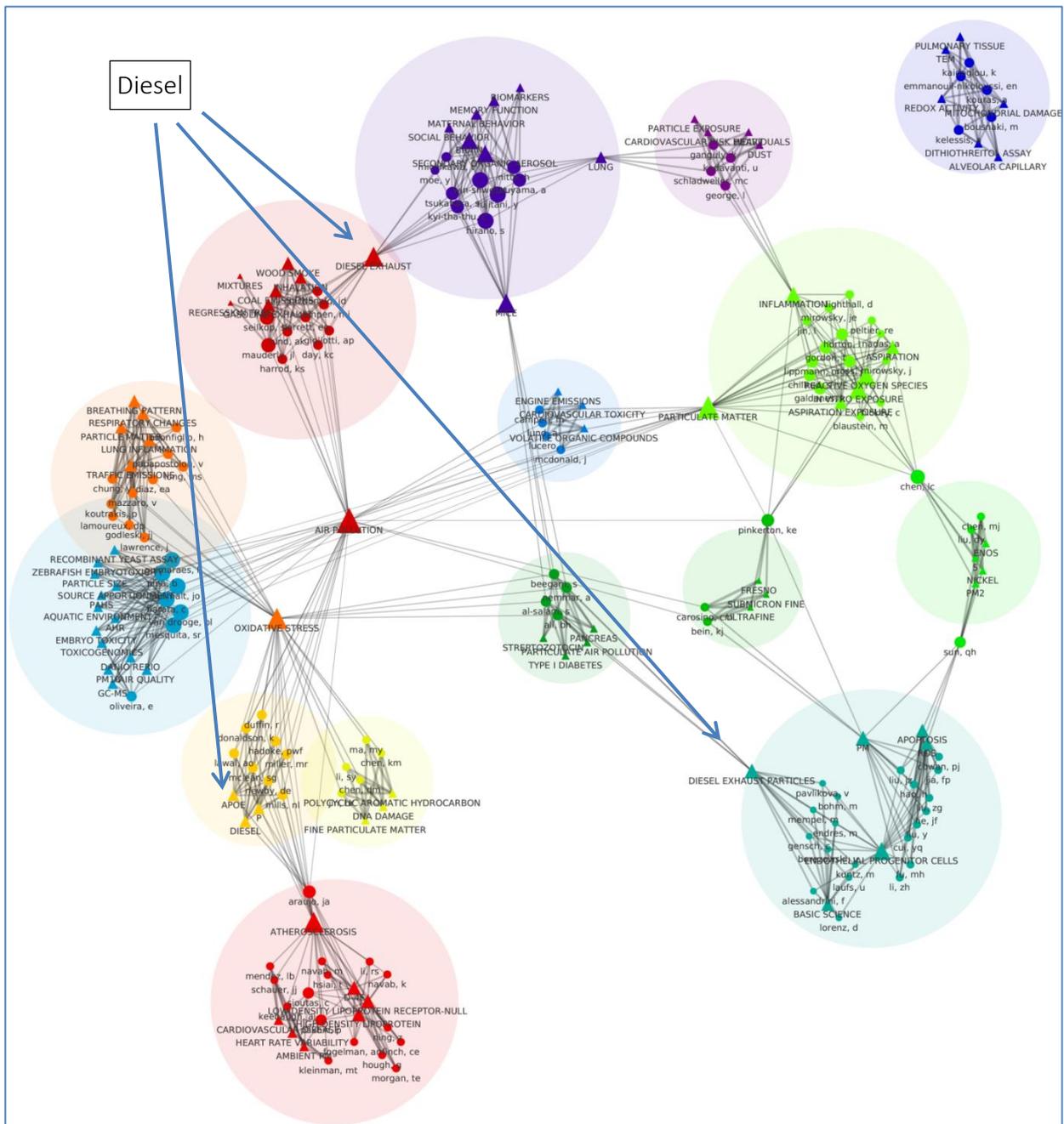
Les interprétations tirées des cartes de réseaux sont limitées dans leur pertinence, par la taille du corpus. Ces interprétations peuvent même être en un sens artefactuelles. Etant donné cette limite, seules quelques pistes d'interprétation sont proposées, à prendre comme autant d'hypothèses à vérifier sur des corpus plus larges.

- **Réseau de co-publication**

Le réseau de co-publication laisse apparaître une série de petits îlots d'auteurs ayant publié ensemble. Le peu de liens entre ces groupes est notable, mais n'est pas surprenant étant donné le peu d'articles concernés, dans un champ très international. A noter, deux personnes, le professeur Chen, de la New York University, et Arajaou, de la David Geffen School of Medicine, Los Angeles, font le pont entre plusieurs équipes. Une structuration trans-équipes et trans-expérimentations pourrait tout à fait donc apparaître avec un échantillon plus large.







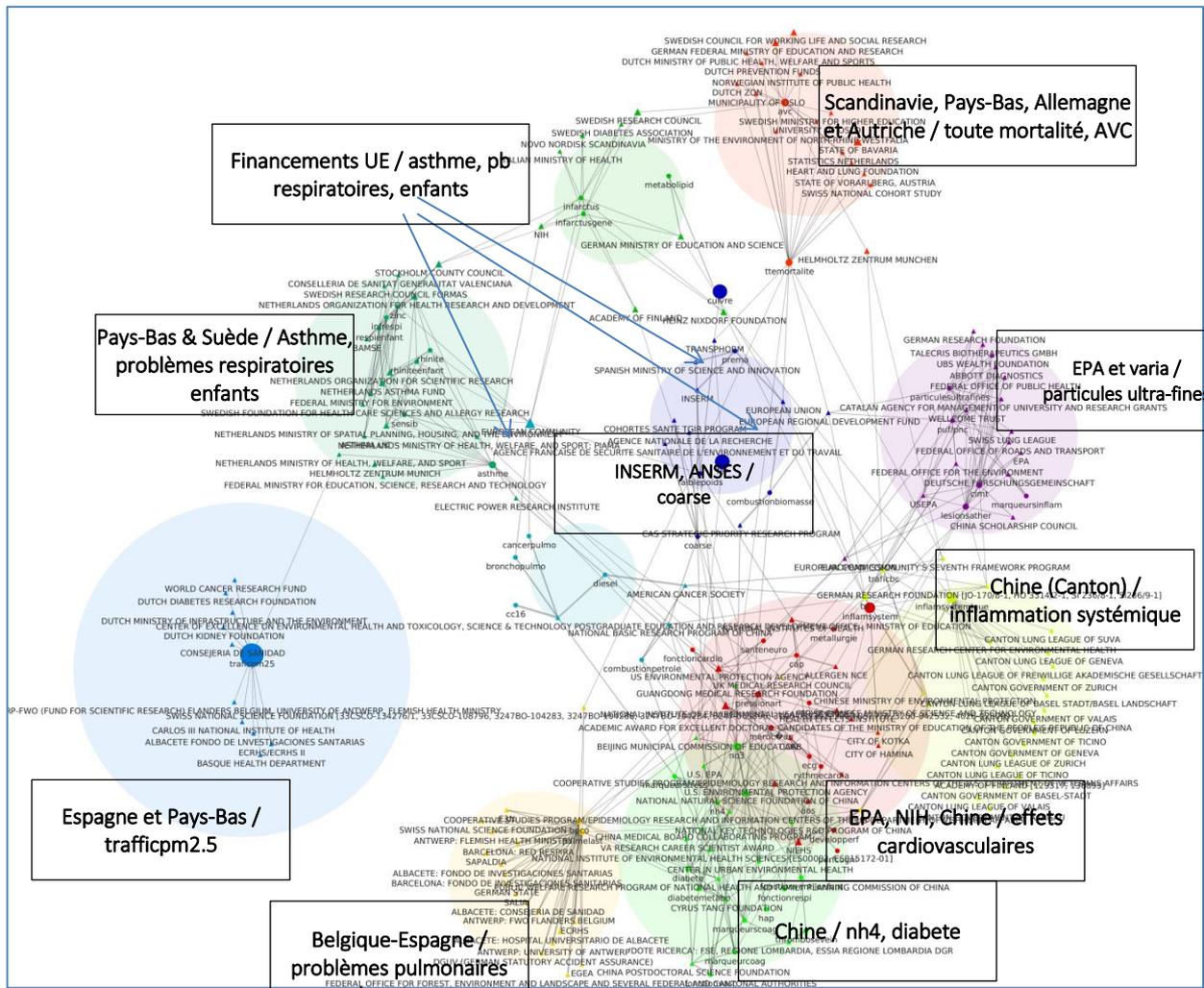
▫ **Etude du financement du champ de recherche sur corpus épidémiologique et toxicologique combiné**

La source du financement, lorsqu'elle est indiquée dans l'article, donne une prédominance des financements originaires de l'institut national de recherche en santé ou en santé environnementale américaine, suivi des institutions européennes et de l'agence environnementale américaine. Un grand nombre d'organismes d'Europe du nord est présent dans le tableau également. Ici, les 20 premiers financeurs sur 350 environ sont donnés.

Source de financement citée	Nbr de documents
-----------------------------	------------------

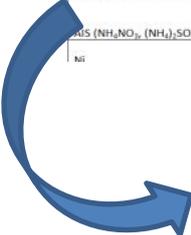
NIH-NIEHS	30
European Community	25
US EPA	21
Swedish Environmental Protection Agency	11
National Natural Science Foundation of China	11
State of Bavaria	9
Swedish Heart-Lung Foundation	9
Swedish Council for Working Life and Social Research	7
German Federal Ministry of Education and Research	7
Swedish Research Council	7
Academy of Finland	6
Netherlands Organization for Scientific Research	6
Stockholm County Council	6
Health Effects Institute	6
German Ministry of Education and Science	5
Netherlands Organization for Health Research and Development	5
UK Medical Research Council	5
Helmholtz Zentrum Munchen	5
Netherlands Asthma Fund	5
Heinz Nixdorf Foundation	5

Lorsque l'on relie ces financeurs aux intitulés de catégories de mots-clés sélectionnés par l'ANSES, on obtient la carte suivante. Elle laisse apparaître, de nouveau, que les financeurs chinois et américains financent des articles caractérisés par les mêmes thématiques, et que l'Europe finance des études qui portent particulièrement sur les effets sanitaires affectant les enfants, notamment effets pulmonaires et respiratoires. Les financements semblent rarement s'ordonner par composés, mais bien plus par investigation d'effets ou ensembles d'effets. Seules quelques associations financeurs-composés apparaissent : les financements français INSERM et ANSES sont associés à la recherche sur les particules épaisses. Un cluster de financeurs belges et espagnols est associé tout particulièrement aux particules 2.5, tandis que plusieurs fondations et instituts chinois sont reliés au nh4.



# Partie 3. Analyse de la distribution des termes de la saisine dans le domaine de recherche

La saisine de l'ANSES concernant les effets sanitaires de la pollution de l'air porte sur la contribution des particules aux problèmes de santé, décomposé par substances composant les particules (nickel, métaux, etc.), sources (trafic routier, biomasse, activité agricole) et effets sanitaires ou endpoints. Pour examiner dans quelle mesure le champ de recherche est structuré par ces objets, l'équipe de l'ANSES a fourni une liste de mots-clés décrivant chacune de ces dimensions (voir en annexe 1). La liste permet de positionner les mots-clés d'intérêt pour l'ANSES dans le domaine de recherche, en montrant à quels auteurs, institutions, et autres mots-clés est-ce qu'ils sont plus fréquemment associés. La liste a été reformatée pour pouvoir être traitable par la plateforme. Pour ce travail, les mots-clés relevant d'un même thème (« forms ») ont été regroupés et associés à une racine (« stem »)(voir en annexe 2). Chaque fois que la plateforme a trouvé une occurrence d'un des mots-clés, elle le signale sur les cartes en positionnant le terme racine.

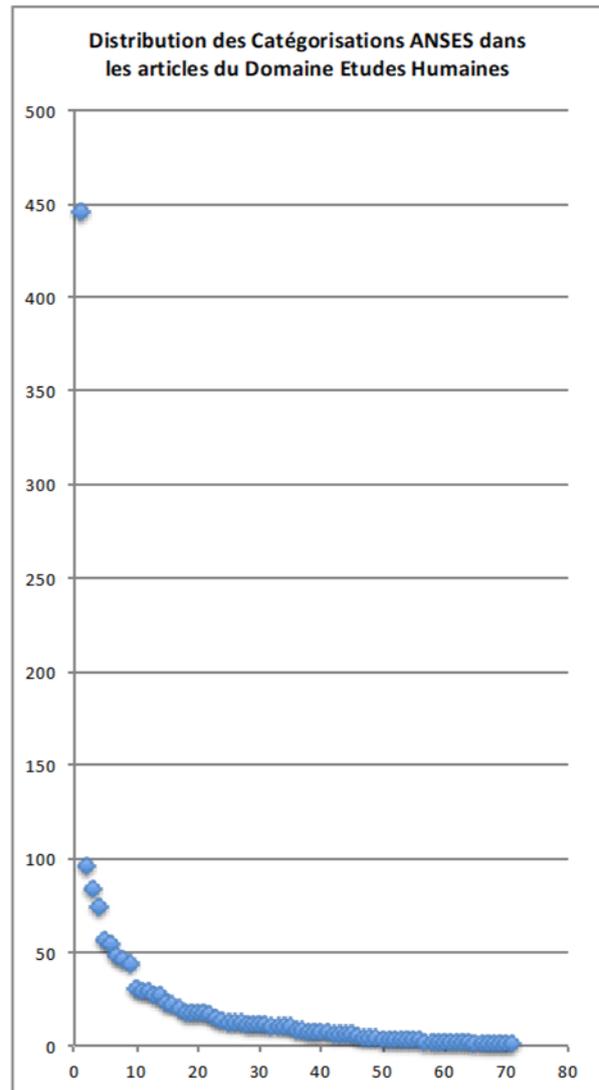


stem	main	forms
<b>Agri</b>	Agri	Agriculture & Rural
<b>SIA</b>	SIA	SIA & secondary inorganic aerosol & inorganic aerosol & inorganic particle
<b>Albu</b>	Albu	broncho-alveolar lavage proteins  & albumin
<b>AOS</b>	AOS	SOA & secondary organic aerosol & organic aerosol & organic particle
<b>ParticuleSecond</b>	secondary particle	secondary particle
<b>Asthme</b>	Asthme	asthma & wheezing
<b>Autisme</b>	autisme	autism
<b>AVC</b>	AVC	Stroke & cerebrovascular
<b>BlackCarbon</b>	BC	Black Carbon & BC & Elementary Carbon & EC & soot & PM Absorbance & black smoke
<b>BPCO</b>	BPCO	COPD &  chronic bronchitis & chronic pulmonary disease
<b>Brakedust</b>	Brakedust	Brake dust
<b>CancerDigestif</b>	Cancerautredig	Digestive cancer
<b>CancerAutres</b>	Cancerautres	cancer
<b>CancerCerveau</b>	Cancercerveau	brain cancer
<b>CancerColon</b>	Cancercol	colon cancer & rectum cancer & colorectal cancer
<b>CancerHemato</b>	Cancerhemato	hematological cancer & blood cancer
<b>Cancerpulmo</b>	Cancerpulmo	bronchopulmonary carcinoma & lung cancer
<b>CancerSein</b>	Cancersein	breast cancer
<b>CancerVessie</b>	Cancervessie	bladder cancer
<b>CAP</b>	CAP	CAP & concentrated ambient particle
<b>CardioCongestion</b>	Cardio	heart failure & congestive heart failure
<b>CC16</b>	CC16	CC16
<b>Cellularite</b>	Cellularite	broncho-alveolar lavage cellularity /neutrophil & macrophage
<b>CIMT</b>	CIMT	CIMT & CAC & coronary artery calcification & carotid intima-media thickness

## 1. Distribution dans le corpus d'études épidémiologiques

Un tri des termes en fonction de leur fréquence d'apparition dans les abstracts des articles composant le corpus permet de montrer qu'une poignée de termes sont très présents dans le corpus. Cela concerne notamment les particules fines (2.5) présentes dans le trafic routier.

Catégorie ANSES	NB d'Articles	Occurrence
TraficPM25	113	445
BC	34	96
Santecardio	43	83
Sulfates	41	74
NO3	37	56
Inflamsystem	21	54
Asthme	15	48
OC	22	46
Coarse	15	44
Nickel	18	30
Fer	16	29
Zinc	19	29
Pressionart	14	27
Particulesultrafines	9	27
Cuivre	16	23
Diesel	13	22
NH4	12	20
Marqueurstress	7	17
CIMT	4	17
Vanadium	15	17
Merocean	12	17
FENO	6	16
AVC	3	14
Combustionpetrole	11	13
Infrespi	4	12
Faiblepoids	3	12
Cancerpulmo	3	12
BPCO	3	11
Roaddust	7	11
Marqueurscoag	3	11
ttmortalite	8	11
lesionsather	7	10
PO	4	10
Industrie	7	10
Infarctus	2	10

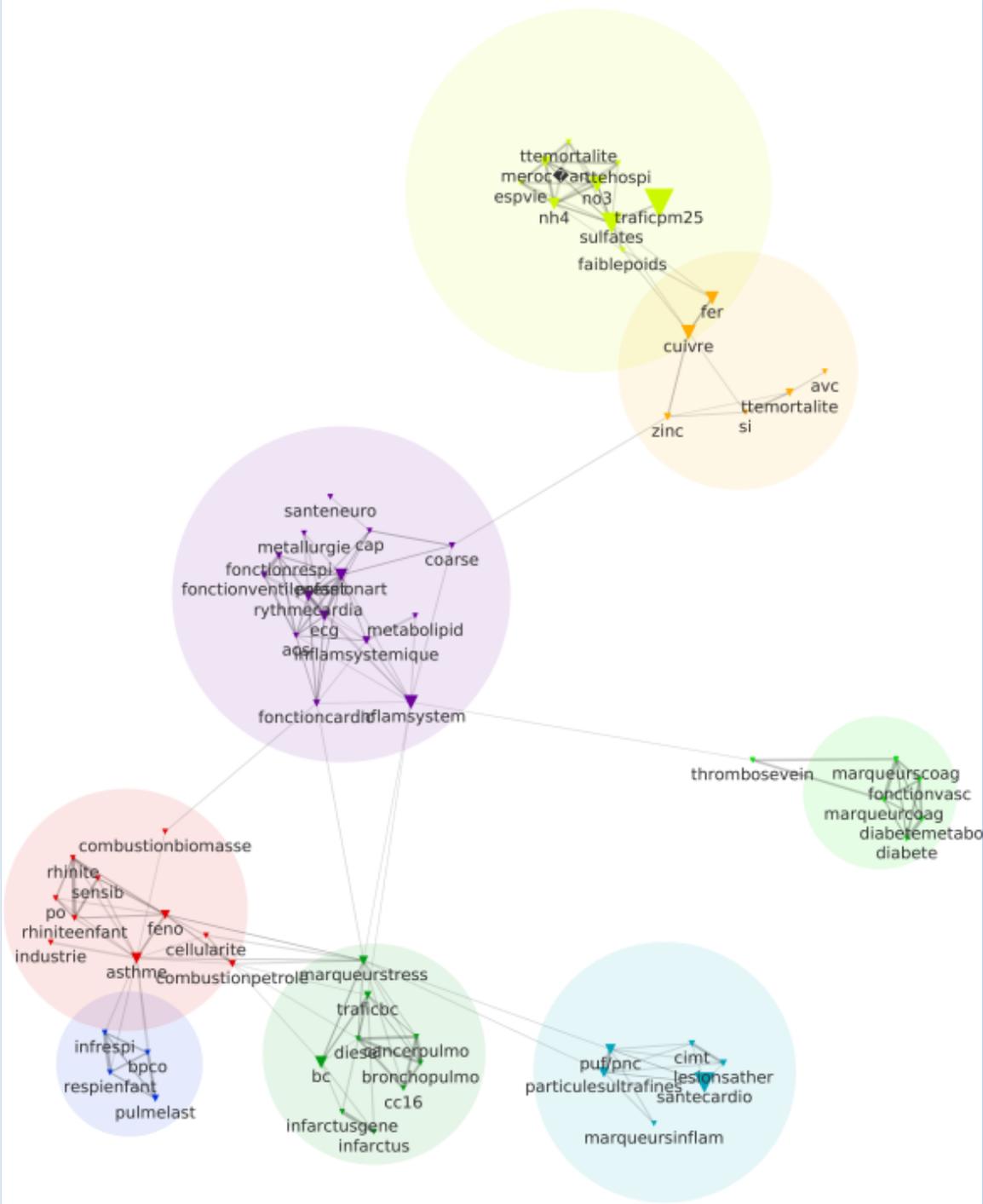


## ▫ Carte de cooccurrence des termes dans les études épidémiologiques

Les termes définis par l'ANSES, lorsqu'on les regroupe en fonction de leur fréquence de cooccurrence dans les abstracts des études épidémiologiques, s'organisent comme sur la carte ci-après. Ils dessinent 8 clusters, dont la teneur montre que la recherche s'organise par composés (et de là par les effets qui sont liés à ces composés), plus que par source des polluants. Les deux clusters (vert et orange) en haut de la carte lient un ensemble de composés, dominé par les particules fines 2.5 issues du trafic routier, à des *endpoints* généraux (toute mortalités, hospitalisation)<sup>2</sup>. Le cluster mauve du milieu traite des particules épaisses (« coarse »), et montre que celles-ci sont sémantiquement liées à la discussion d'effets respiratoires, cardiaques (pression artérielle est la catégorie la plus importante ici) et en partie neurologiques. La métallurgie est la seule source apparente. Deux clusters

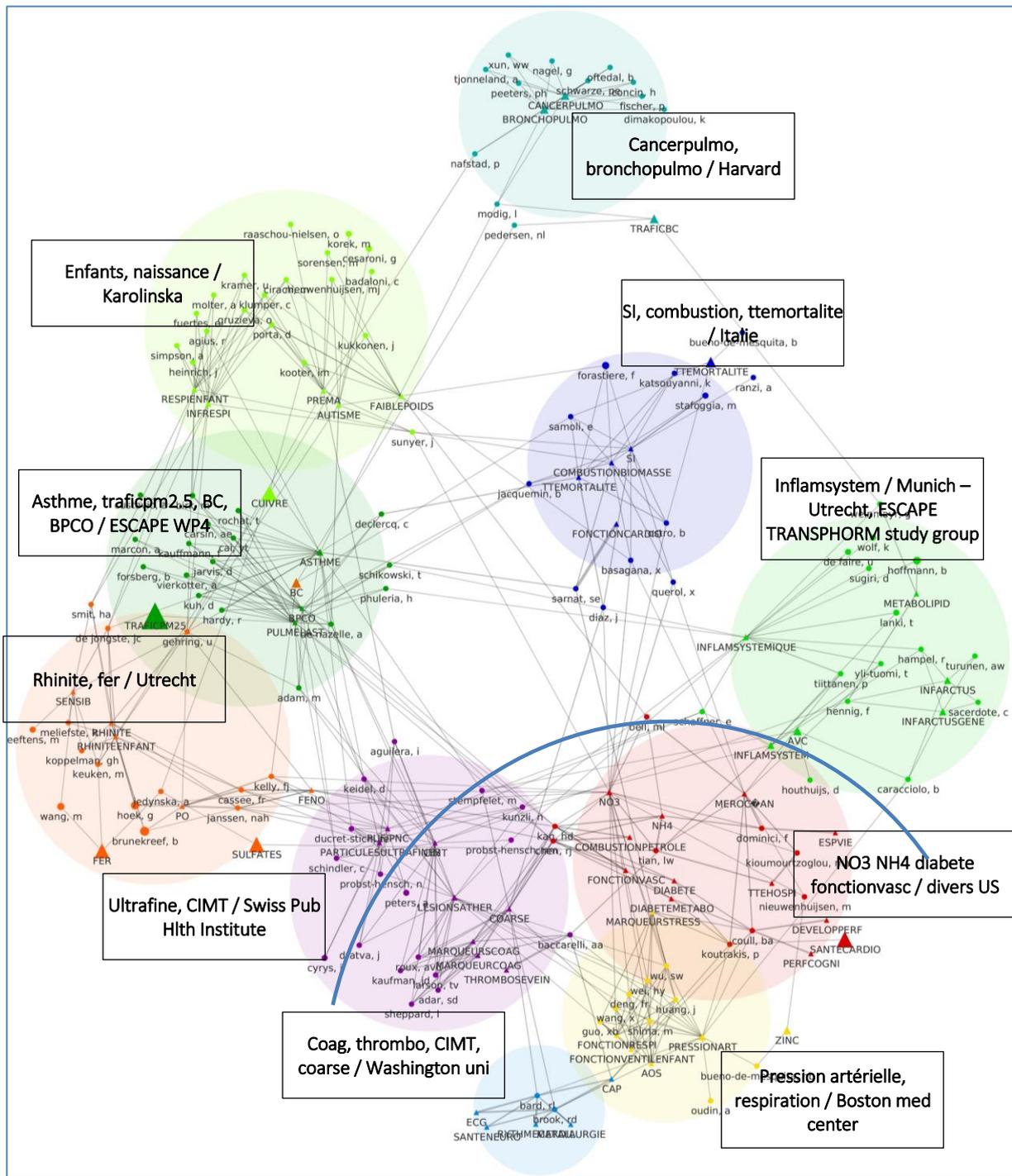
<sup>2</sup> Le fait qu'un composé et un effet soient mentionnés conjointement dans un cluster ne signifie pas, ou pas nécessairement, que la littérature établit un lien de cause à effet entre les deux. Ce pourrait éventuellement être le cas sur une population d'études très large, en considérant que les termes utilisés par les auteurs pour composer un résumé sont les termes entre lesquels l'article établit une association positive. La population de 130 études considérées ici paraît trop restreinte.

emboîtés, rouge et bleu, en bas à gauche, traite des affections touchant particulièrement les enfants (asthme, rhinite), en lien avec des substances comme po et feno. Le cluster vert en bas au milieu est organisé autour du composé black carbon et le cc16, liés à l'infarctus, au cancer des poumons et à la bronchite pulmonaire. En bas à droite, le cluster des particules ultrafines et de la santé cardiologique. Au milieu de la carte à droit, le cluster du diabète.



▫ **Réseau des auteurs en fonction de leur proximité avec les termes définis par l'ANSES**

La carte suivante est un réseau hétérogène, composé des relations entre les auteurs du corpus d'études épidémiologiques et les termes définis par l'ANSES. Les auteurs sont rapprochés de ces termes s'ils sont présents dans les résumés des articles qu'ils ont publiés. Des clusters se dégagent, mais les interrelations entre eux sont nombreuses. Sur la partie gauche de la carte (clusters vert, vert foncé et orange), des auteurs qui travaillent sur les composés ou substances comme fer, sulfates, particules fines, cuivre, en lien, notamment, avec les effets observables chez les enfants (prématurité, faible poids de naissance, rhinite, autisme...). En bas, et en bas à droite de la carte, les clusters font davantage place aux particules plus grosses, au NO<sub>3</sub> et NH<sub>4</sub>, et respectivement, en partant du bas au milieu (cluster mauve), aux problèmes artériels, aux problèmes de diabète et problèmes inflammatoires, ainsi qu'aux infarctus. Ces clusters sont associés à des équipes. Là encore, on peut remarquer que la recherche s'organise principalement par composés, ou autour d'associations composés-effets. La carte montre que, sur une période courte de 4 ans couverte par le corpus, les équipes sont relativement fixées sur de telles associations, plus qu'elles n'explorent des associations diverses. Des auteurs se retrouvent fréquemment en situation de faire des ponts entre les clusters, montrant que le domaine est structuré par des réseaux de publication assez larges, qui permettent des mobilités individuelles vers d'autres sujets (par ex. Nicole Probst-Hensch du Swiss Tropical and Public Health Institute, Basle, professeur en épidémiologie génétique). Même si les équipes sont fixées sur une association ou un composé, le réseau d'équipes, et les recombinaisons des collaborations entre les individus, semblent permettre de couvrir l'ensemble des composés et associations composés-effets notés dans la liste de termes fournis par l'agence. Les équipes américaines et leurs centres d'intérêt sont positionnés en bas, le reste de la carte étant largement structuré par les clusters européens, notamment issus de la coordination ESCAPE.

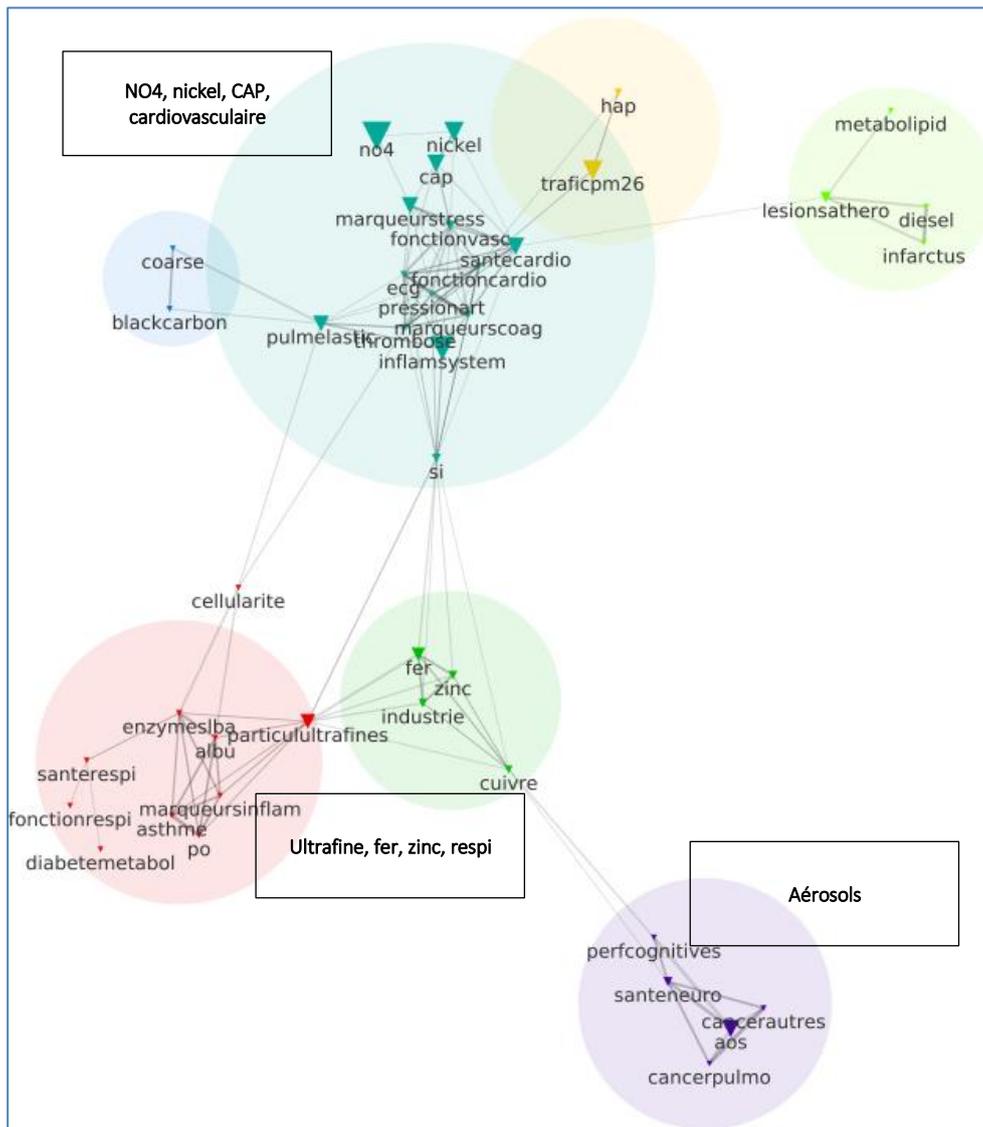


▫ **Réseau des relations entre institutions et termes ANSES**

Des associations entre universités et centres d'intérêt en termes de source, composés ou effets apparaissent. Des formes de spécialisation sont à noter, mais aussi des relations de fait, entre universités ou institutions de recherche, avec une forte imbrication des recherches produites dans les universités américaines et les universités chinoises, mais aussi une association assez forte entre agences nord-américaines (EPA, Health Canada) et la New York University, où est installé le Health Effects Institute, souvent financé par les agences (l'EPA notamment) mais aussi l'industrie automobile.

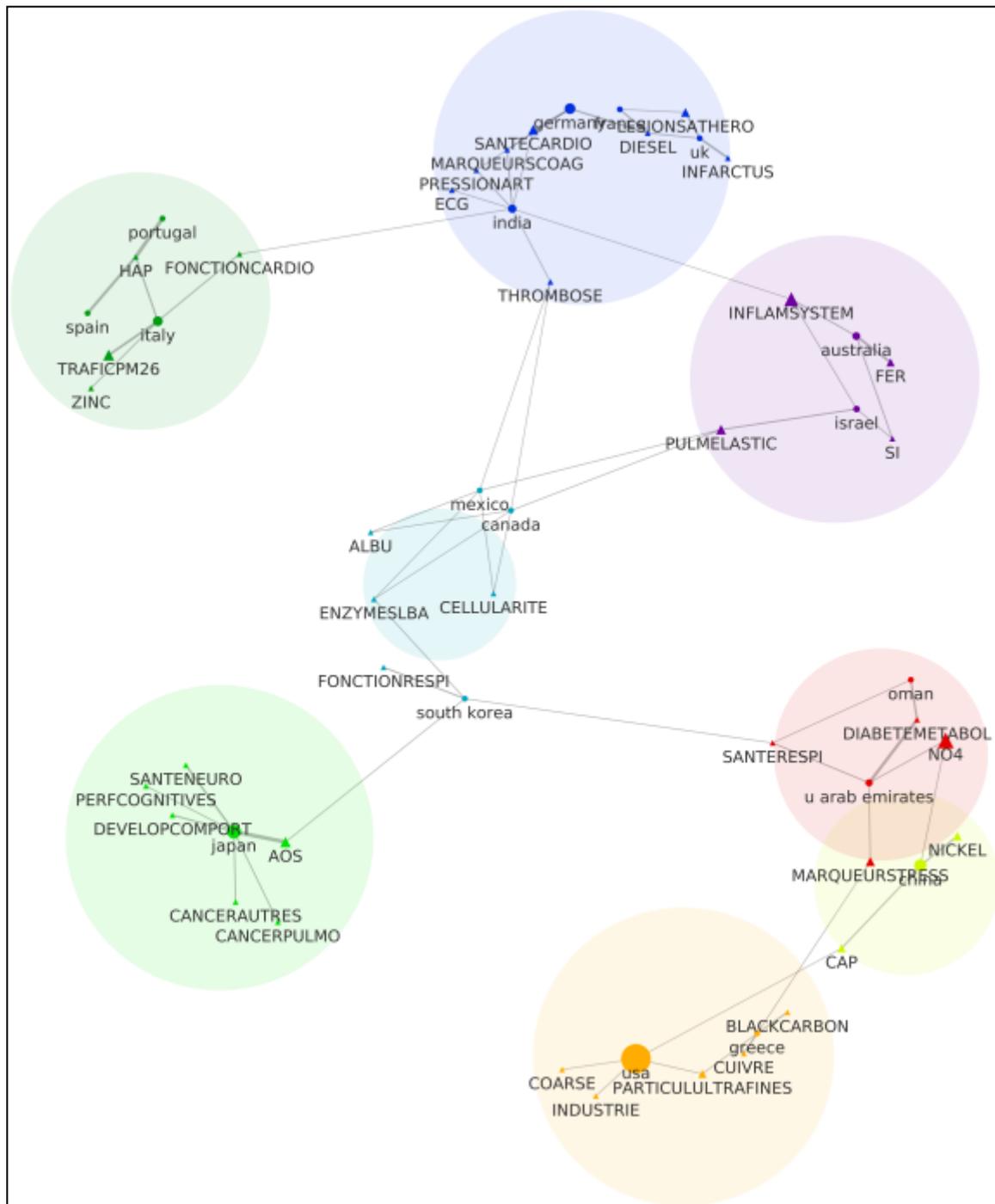






▫ **Réseau des relations entre pays d'origine des auteurs et termes ANSES**

La carte pays/termes ANSES n'est pas très significative, puisque les équipes d'auteurs sont généralement internationales, on aboutit donc à une distribution des termes ANSES entre les pays quelque peu aléatoire, et non représentative du domaine des études toxicologiques des effets des particules dans son entièreté.



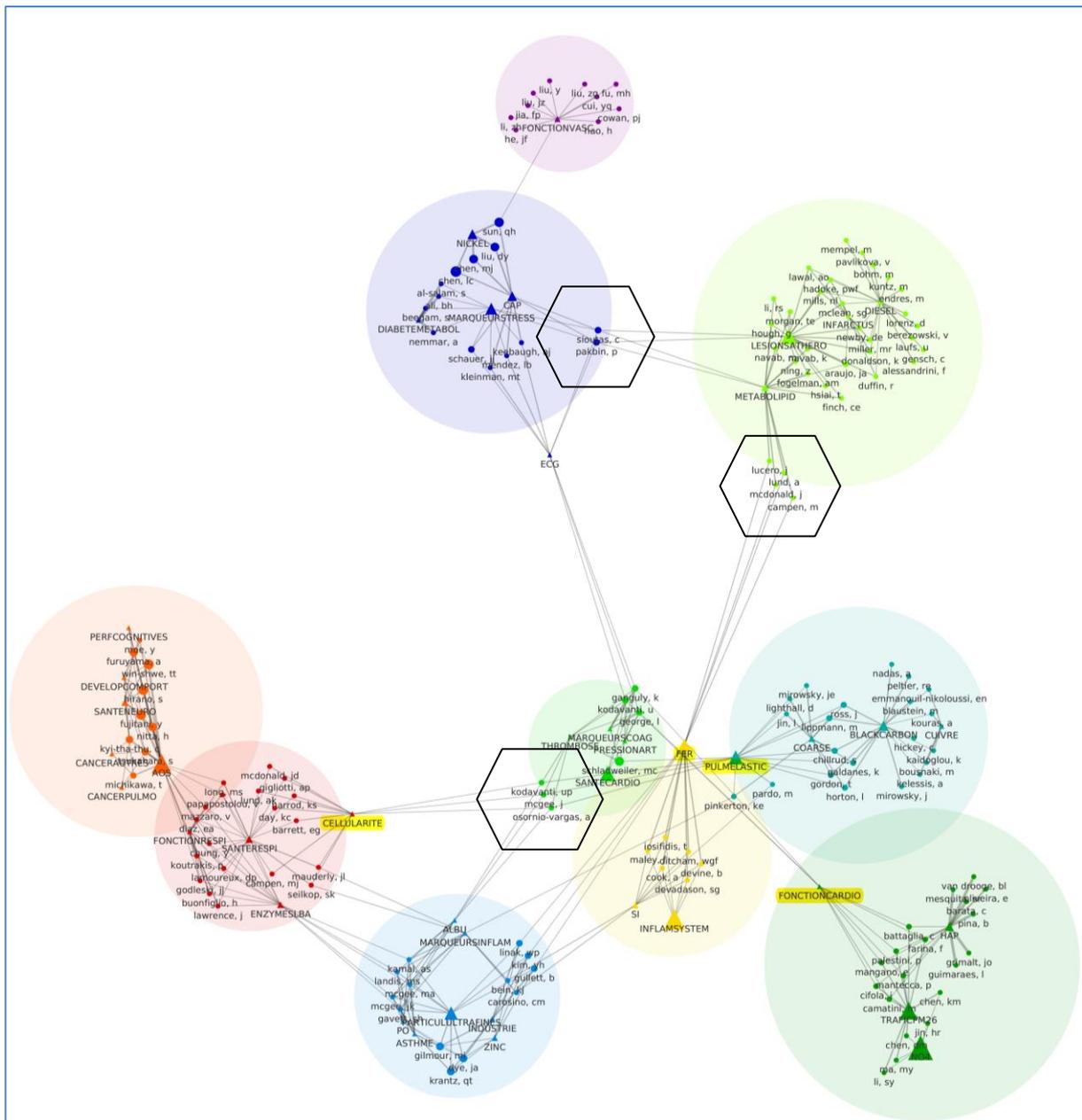
▫ **Réseau des relations entre institutions et termes ANSES**

La carte institutions/termes ANSES conduit à faire des hypothèses sur l'existence de collaborations spécialisées entre institutions. Un double cluster situé à gauche de la carte laisse apparaître une forme de spécialisation des institutions situées en caroline du Nord (Université Caroline du Nord, les laboratoires de l'EPA situés dans le Research Triangle Park, et les laboratoires nationaux Oak Ridge, non loin), et la participation à cet ensemble d'une institution privée (Arcadis). Les institutions japonaises, au bas de la carte, sur les quelques études du corpus, semblent se spécialiser dans l'étude des effets de gaz aérosols. Les



▫ Réseau des auteurs en fonction de leur proximité avec les termes définis par l'ANSES

Logiquement, les clusters sont bien dessinés, puisqu'ils correspondent à une ou deux études produites par un même groupe d'auteurs, et que l'échantillon d'ensemble est limité. Il est à noter que l'on voit apparaître quelques termes structurants, reliant les grappes d'auteurs : fer, élasticité pulmonaire, cellularité, fonctioncardio (surlignés en jaune). Quelques collectifs font aussi le lien entre équipes (entourés d'un losange) : Kodavanti et al., de l'EPA ; Lucero/Luce/Campen/McDonald, du Lovelace Respiratory Research Institute de Albuquerque ; Sitouas et Pakbin de l'université de Californie sud.



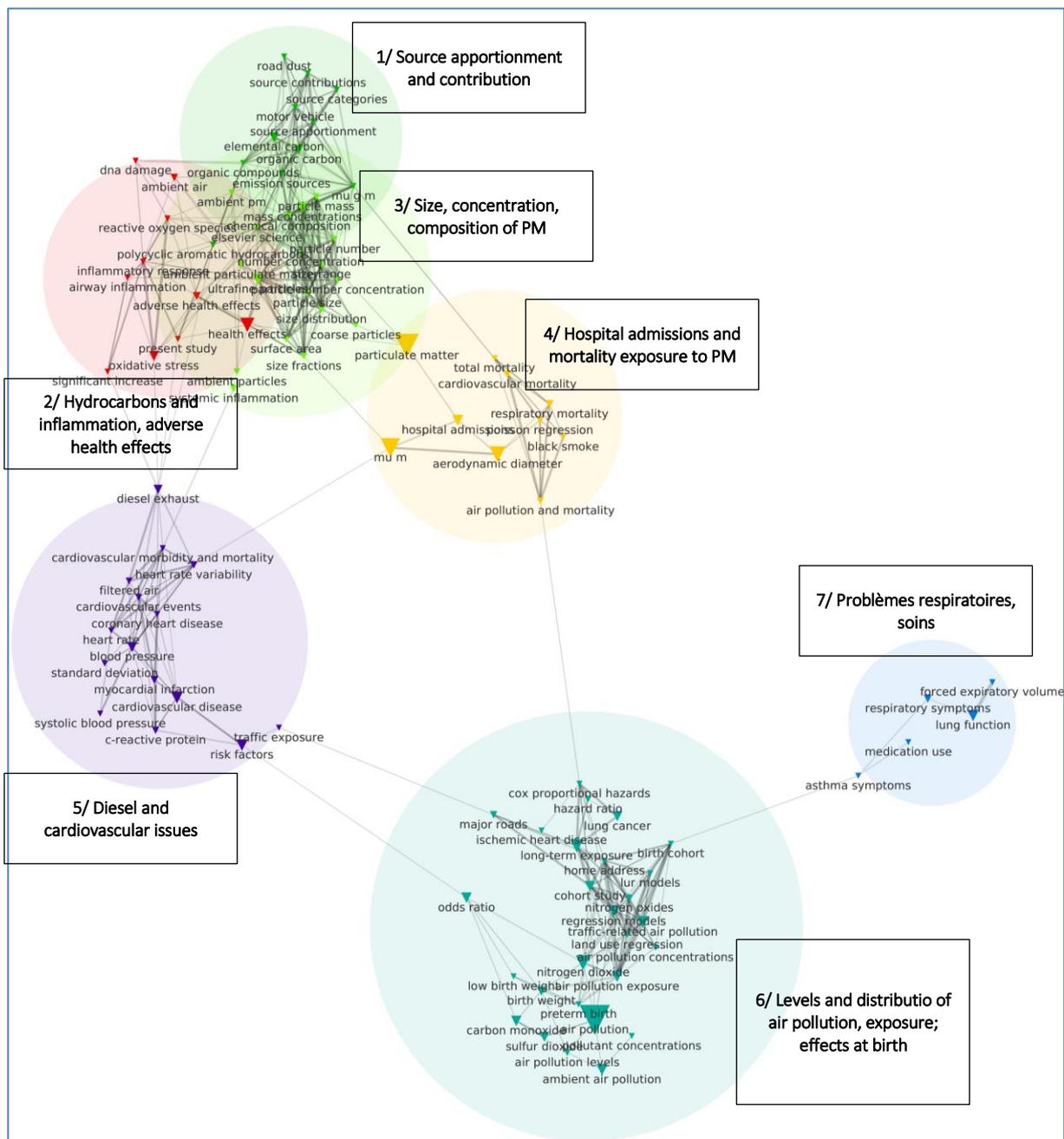
## ▫ **Positionnement des termes ANSES dans le corpus étendu**

La liste des termes fournie par l'ANSES compose un champ sémantique, contenant les thèmes d'intérêt pour la saisine. Il est apparu pertinent de positionner ces termes dans le paysage qui résulte de l'analyse du corpus, pour examiner dans quelle partie du champ de recherche sur la pollution de l'air (tel que représenté à travers ce corpus élargi) est-ce que ces thématiques sont travaillées, et quelles autres thématiques, structurent ce champ. Les observations faites sont à prendre avec précaution, puisque le corpus d'études utilisées pour l'avis n'a pas été construit en partant du corpus élargi. Ce dernier n'est qu'un des environnements dans lequel le corpus utilisé pour l'avis peut être replacé. Il n'a d'intérêt que dans le cadre d'une étude exploratoire, montrant le type de questionnement que l'on peut appliquer à un corpus construit pour répondre à une saisine.

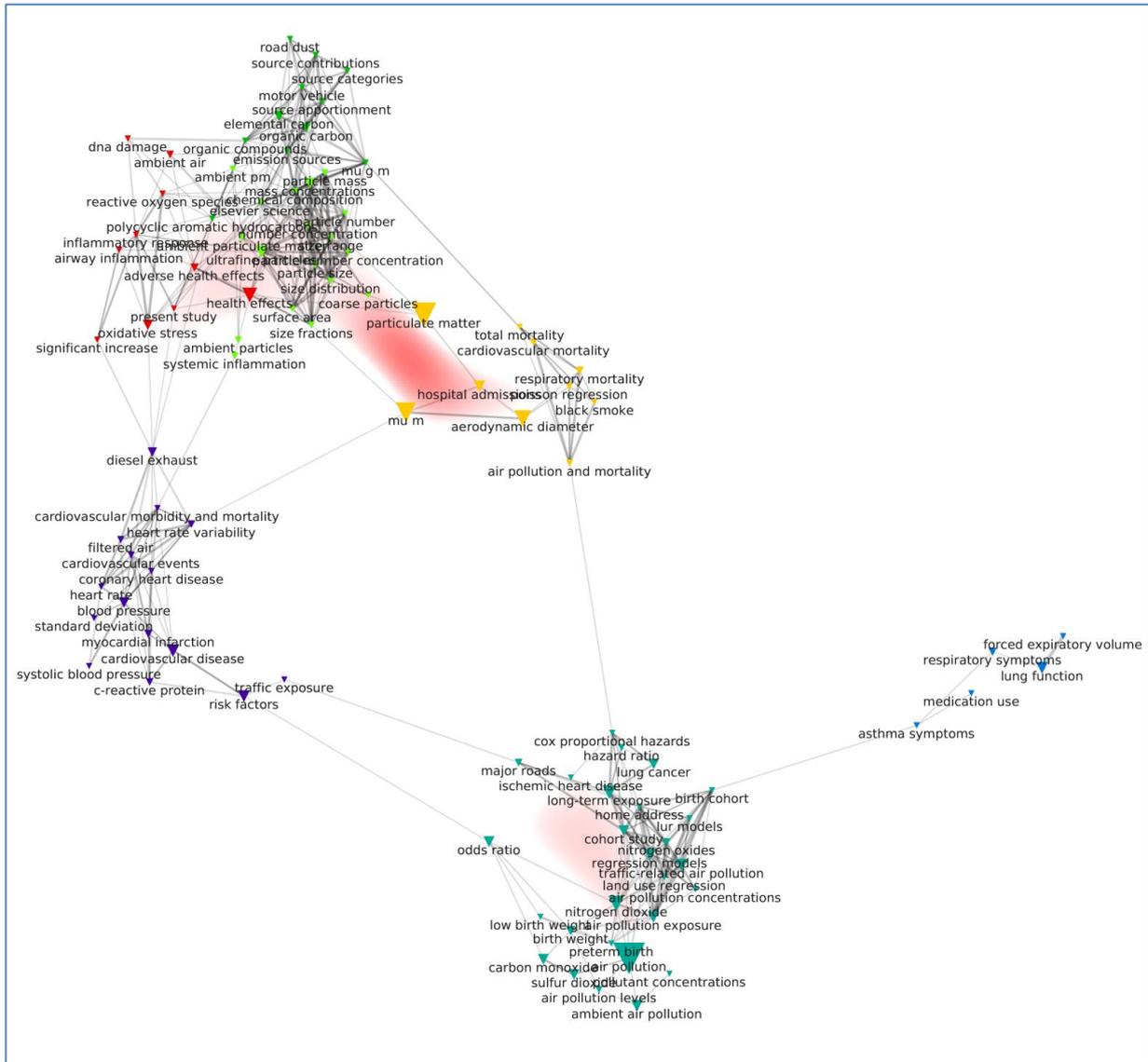
La méthode employée pour produire des indications sur ces questions consiste à représenter, sous forme de « zones de chaleur », les points de la carte des associations entre les termes les plus fréquemment utilisés dans les résumés des articles du champ de recherche, et qui correspondent également à des termes sélectionnés par l'ANSES. Le champ de recherche est représenté à partir d'un corpus élargi de 2151 références, comprenant l'ensemble des études épidémiologiques et toxicologiques examinées par l'ANSES (le « corpus-mère » étudié jusque ici) ainsi que l'ensemble des références citées par ces études. Ce corpus élargi contient donc, en plus du corpus-mère, tout l'« historique » des recherches par rapport auxquelles les auteurs d'études du corpus-mère se situent.

Le champ de recherche apparaît organisé par six clusters principaux :

- 1 - Source apportionment and contributions
- 2 - Hydrocarbons and inflammation, adverse health effects
- 3 - Size, concentration and composition of PM
- 4 - Hospital admission and mortality related to exposure to PM
- 5 - Cardiovascular problems and diesel, exposure to traffic
- 6 - Levels and distribution of air pollution, exposure; effects at birth
- 7 - Asthma, respiratory symptoms, medication



La superposition des termes ANSES sur cette carte active trois zones en particulier. La zone la plus « chaude », celle où la correspondance entre les termes ANSES et les termes utilisés par les auteurs est la plus forte, est à la frontière des clusters 3 et 4. Cette zone concentre des termes comme « particulate matter », « coarse particles », « hospital admission ». La deuxième zone, proche de la première, est celle des termes comme « health effects », « adverse health effects », « particle ». La troisième s’inscrit dans le cluster 6, et couvre le domaine des « cohort study », des calculs d’exposition de long terme à base de « land use regression models », en lien notamment avec la circulation routière.



La superposition des cartes permet de situer le champ sémantique composé par les termes ANSES dans le champ de recherche plus large, et de mesurer sa distance avec les différents clusters qui le composent. Dans le cas présent, le champ sémantique que compose la liste des termes ANSES est relativement distant de la question source apportionment and contribution (cluster 1), de la question de la inflammatory response/hydrocarbons (cluster 2), mais aussi de l'ensemble du cluster 5 sur les questions cardiovasculaires. La partie du cluster 6 qui concerne le lien entre carbon monoxide/sulfur dioxide et birth weight, est à l'écart également de la zone chaude touchant ce cluster.

Cette analyse permet de voir dans quelle partie du champ de recherche global est-ce que la saisine, et la recherche de littérature, va être amené à puiser, en fonction des termes dans lesquels elles sont construites. Elle peut également permettre, lorsqu'on cible la liste de termes avec laquelle on croise la carte de départ, dans quel cluster est traité tel ou tel objet de recherche, que l'on parle d'un effet, d'un composé ou d'une source.

Certaines zones semblent être peu marquées par la présence de termes faisant partie de la liste de référence ANSES. Cela est surprenant, dans la mesure où les termes composant le

cluster, sur la carte, correspondent bien à des termes de la liste de référence ANSES. Le cluster de points violets, sur la gauche, correspond à la recherche sur les problèmes cardiovasculaires lié à l'exposition aux émissions du trafic automobile – autant de sujets ou termes qui sont effectivement inclus dans la liste de termes de référence. Cela pourrait traduire le fait que les termes et expressions utilisées dans ces études, tout en prenant le même type de problèmes sanitaires comme objet (les maladies cardiovasculaires), varient légèrement par rapport à la terminologie de la liste de référence. Ainsi, les expressions « cardiovascular events » et « cardiovascular disease » sont très fréquents, et apparaissent centraux dans ce cluster violet, mais sont plus génériques que ceux qui sont utilisés dans la liste de référence de l'agence pour décrire ce type d'effets. Cela confirme la grande sensibilité de recherches de littérature scientifique à partir de requêtes décrivant le sujet des études. Cela peut aboutir à extraire des études dans un gisement donné, adjacent à un autre gisement, qui serait moins exploité. De ce point de vue, la méthode adoptée ici confirme l'intérêt de la constitution d'un corpus à partir d'une liste de références présentes dans une revue donnée, exploitée exhaustivement.

# Conclusion

A travers les différentes analyses présentées ici, quelques éléments de conclusion concernant la structuration du corpus semblent émerger. Ces éléments montrent qu'une telle étude permet de générer des éléments d'appréciation sur la pertinence de la littérature existante pour les questions posés dans une saisine. En l'occurrence, quelques observations peuvent être formulées, de manière tentative:

- Le corpus porte prioritairement sur les particules fines, bien plus que sur les ultrafines, par exemple. Les termes décrivant des sources ne semblent pas organiser fortement les clusters et le corpus dans son ensemble. La recherche est assez distribuée à travers les différents effets sur la santé.
- Le champ de recherche, analysé par exemple à travers la carte de cooccurrence de termes dans les titres et abstracts des études épidémiologiques (p. 23), sans être fragmenté, laisse apparaître des clusters disjoints. Au vu du nombre d'études et du travail de sélection qui a été fait pour les intégrer dans le corpus, on aurait pu s'attendre à ce que la carte présente des clusters plus proches les uns des autres, voire superposés. Il semble exister des ensembles composés/effets, qui rendent chaque cluster assez spécifique. En tout état de cause, il peut s'avérer utile de se pencher plus avant sur les raisons de cette différenciation entre clusters, et donc de structuration de ce champs de recherche, car cela éclaire sur la manière dont les recherches sont conçues, et donc leur adéquation pour répondre aux questions posées par la saisine.
- On peut examiner l'hypothèse d'une spécialisation géographique des recherches, puisque l'on trouve des équipes relativement localisées (en termes de ville dans laquelle elles sont situées), notamment du côté d'Amérique du nord ou d'Asie (tandis que les équipes européennes travaillant en épidémiologie collaborent fortement, et examinent conjointement les données concernant des cohortes constituées dans différents pays européens). Il est possible que les équipes localisées s'intéressent plus directement à des composés ou des sources pertinentes pour le territoire où elles sont localisées (ou pour le financeur de leurs études).
- L'analyse permet d'objectiver le fait que le corpus est influencé par les études produites dans le cadre du projet européen ESCAPE, donc par les réseaux de collaboration qui sont formés dans le cadre de ce projet, et les cohortes qui sont utilisées. Le projet européen promeut une forme de coordination transeuropéenne de la recherche dans le domaine. L'analyse du corpus est donc aussi une analyse des choix faits dans la conception et la conduite de ce projet européen, et de la manière dont il permet, ou non, de répondre aux questions posées dans le cadre de la saisine que traite l'ANSES.
- Des clusters dédiés à la question des sources et des composés sont apparus lorsque l'on a élargi le corpus. Il est intéressant que ces questions soient constitutives d'un cluster à part, et non reliés à ou intégrés dans d'autres clusters, comme si l'étude des effets des particules n'étaient pas nécessairement reliés à l'étude de la distribution des particules ou autres composés, par sources.

Dans l'ensemble, et même si le corpus étudié est de petite taille (puisque conçu pour traiter d'une saisine spécifique, et non pour être représentatif de l'ensemble du champ de recherche international sur les effets sanitaires de l'exposition aux particules présentes dans l'air ambiant), on peut gagner par l'étude de ce corpus une compréhension de la manière dont les recherches sur ce sujet se structurent, et notamment sur les formes de spécialisation interne à ce champ. Les analyses et les cartes présentées ici contribuent à la vision synoptique de la sélection réalisée par les experts, à l'explicitation des choix faits. Elles peuvent guider la recherche d'études supplémentaires, le cas échéant.

Sans supposer à aucun moment que le travail de délimitation et d'analyse peut remplacer le travail de sélection des connaissances pertinentes pour une saisine, on peut avancer que des approches scientométrique de la littérature scientifique peuvent avoir un intérêt d'accompagnement de la sélection et d'animation du travail fait avec les experts du GT. Il s'agirait de pouvoir présenter la façon dont des paysages et des analyses de la littérature peuvent être réalisés, et d'irriguer le travail initial des experts pour les aider à constituer les équations de recherche et à contrôler leur propre sélection en discutant des effets du périmétrage. D'un point de vue sociologique, les différences entre corpus renseignent sur les pratiques et critères qui guident le travail de constitution de bases de connaissances pour l'évaluation et la gestion des risques, et les différences, similitudes et frontières entre ces deux régimes de connaissance qu'on appelle traditionnellement — mais probablement de manière insatisfaisante — recherche et expertise.

# Annexes

## Annexe 1 : liste de termes décrivant sources, composés et effets couverts par la saisine

### Endpoints sanitaires (études animales)

lésions atherosclérotiques	atherosclerotic / atherosclerotic / atherosclerosis
infarctus et évènements coronariens	infarct / myocardial infarct / heart attack / coronary events / coronary heart disease
pression artérielle	arterial pressure / blood pressure
rythme cardiaque (ECG)	heart rythm / ECG / heart frequency / heart rate / electrocardiogram
fonction cardiaque	heart function / cardiac / heart failure
fonction vasculaire	vascular function / vasodilatation / vasoconstriction
thrombose veineuse et vasculaire périphérique	thrombosis / thrombotic / phlebitis
dépôt lipidique	lipid deposit / fat deposit / plaque deposit / atherom
métabolisme lipidique	lipid metabolism / cholesterol
marqueurs inflammation systémique	systemic inflammation / inflammation / inflammatory
marqueurs coagulation	coagulation / blood clotting
marqueurs de stress oxydant systémique	systemic oxidative stress / oxidative stress
hématocrite/hémoglobine	Hematocrit / Hemoglobin
fonction respiratoire	respiratory function / pulmonary function / breathing pattern / spirometry / plethysmography
élasticité pulmonaire/thoracique et résistance des voies aériennes	pulmonary / thoracic elasticity and airway resistance
histologie pulmonaire (ex : remodelages, infiltrats inflammatoires)	pulmonary histology / lung histology / remodeling / inflammatory infiltrates
cellularité du LBA (neutrophiles, macrophages)	broncho-alveolar lavage cellularity /neutrophil / macrophage
protéines du LBA (albumine)	broncho-alveolar lavage proteins / albumin
enzymes du LBA (LDH, $\gamma$ -GT, NAG)	broncho-alveolar lavage enzymes / LDH / lactate dehydrogenase / $\gamma$ -GT / NAG / N-acetyl glucosamine
mortalité périnatale	perinatal mortality / neonatal mortality
malformations	Defect / birth defect / birth outcome
développement embryonnaire	Embryonic development / embryo development
développement du comportement (sociabilité, exploration, anxiété, locomotion, etc. chez la descendance des mères exposées)	behavioral development / sociability / exploration / anxiety / locomotion
développement des performances cognitives (mémorisation, apprentissage, etc. chez la descendance des mères exposées)	cognitive performances / memorization / learning
comportement (sociabilité, exploration, anxiété, locomotion, etc.)	
comportement (sociabilité, exploration, anxiété, locomotion, etc.) _juvéniles	
développement du comportement (sociabilité,	

exploration, anxiété, locomotion, etc. chez la descendance des mères exposées)	
performances cognitives (mémorisation, apprentissage, etc.)	
performances cognitives (mémorisation, apprentissage, etc.) _juvéniles	
développement des performances cognitives (mémorisation, apprentissage, etc. chez la descendance des mères exposées)	
incidence de tumeurs broncho-pulmonaires	bronchopulmonary carcinoma / lung cancer
histopathologie de cancer broncho-pulmonaire	
incidence de tumeurs (cancers autres que broncho-pulmonaire)	
histopathologie de cancer (cancers autres que broncho-pulmonaire)	
diabète (avec caractéristiques pathologie humaine diabétique type I)	type I diabetes / diabetes
diabète (avec caractéristiques pathologie humaine diabétique type II)	type 2 diabetes / diabetes

### Endpoints sanitaires (études humaines)

mortalité toutes causes	all-cause mortality / natural mortality
hospitalisation toute cause	hospitalization / hospital admission
espérance de vie	life expectancy
mortalité (toute cause cardiovasculaire)	cardiovascular mortality
hospitalisations (toute cause cardiovasculaire)	hospitalization / hospital admission
pression artérielle	arterial pressure / blood pressure
CIMT et CAC	CIMT / CAC / coronary artery calcification / carotid intima-media thickness
infarctus et évènements coronariens	infarct / myocardial infarct / heart attack / coronary events / coronary heart disease
AVC	Stroke / cerebrovascular
insuffisance et congestion cardiaque	Heart failure / congestive heart failure
altérations du rythme cardiaque (ECG)	heart rythm / ECG / heart frequency / heart rate / electrocardiogram
fonction vasculaire	vascular function / vasodilatation / vasoconstriction
marqueurs inflammation systémique	systemic inflammation / inflammation
marqueurs coagulation	coagulation
marqueurs de stress oxydant systémique	systemic oxidative stress / oxidative stress
marqueurs altération de l'endothélium vasculaire	vascular endothelium
naissance prématurée	premature birth / gestational age
faible poids de naissance	birth weight
circonférence crânienne	cranial circumference
pré-éclampsie	Preeclampsia
mortalité (toute cause cardiovasculaire) _enfant	
hospitalisations (toute cause cardiovasculaire) _enfant	
pression artérielle _enfant	
CIMT et CAC _enfant	
altérations du rythme cardiaque (ECG) _enfant	
mortalité (toute cause respiratoire) _enfant	
hospitalisations (toute cause respiratoire) _enfant	
asthme (incidence, prévalence et phénotype)/sifflements _enfant	asthma / wheezing
sensibilisation allergique _enfant	sensitization / allergy

rhinite _enfant	rhinitis
infections respiratoires et ORL _enfant	respiratory infection /ENT infections / ENT / pneumonia
fonction ventilatoire _enfant	respiratory function / pulmonary function / breathing pattern / spirometry / plethysmography
FeNO _enfant	FeNO / exhaled nitric oxide / exhaled NO
marqueurs inflammation dans le condensat de l'air exhalé _enfant	
autisme _enfant	autism
performances cognitives (mémorisation, apprentissage, etc.) _enfant	cognitive performances / memorization / learning
hyperactivité	hyperactivity
mortalité (toute cause respiratoire)	
hospitalisations (toute cause respiratoire)	
asthme (incidence, prévalence et phénotype)/sifflements	asthma / wheezing
BPCO et bronchite chronique	COPD / chronic bronchitis / chronic pulmonary disease
rhinite	rhinitis
infections respiratoires et ORL	pneumonia
fonction ventilatoire	
FeNO	
Autres marqueurs inflammation dans le condensat de l'air exhalé	
marqueurs d'inflammation dans le lavage bronchoalvéolaire	bronchoalveolar lavage
CC16	CC16
performances cognitives (mémorisation, apprentissage, etc.)	
cancer broncho-pulmonaire	bronchopulmonary carcinoma / lung cancer
cancer colon/rectum	colon cancer / rectum cancer / colorectal cancer
cancer autres digestifs	Digestive cancer
cancer hématologique	hematological cancer / blood cancer
cancer vessie	bladder cancer
cancer sein	breast cancer
cancer cerveau	brain cancer
cancer indéterminé	
cancer indéterminé _enfant	
dépression	Depression / depressive

### Catégorie d'effets sanitaires

mortalité toutes causes	all-cause mortality
santé cardiovasculaire	cardiovascular
santé respiratoire	respiratory
cancer bronchopulmonaire	lung cancer
cancer autres	cancer
santé neurologique	Neurologic* / nervous
fertilité	Fertility / reproductive
santé périnatale	perinatal
diabète/troubles du métabolisme	diabetes / metabolism disorder

## Composition et sources

BC/EC/PMabs	Black Carbon / BC / Elementary Carbon / EC / soot / PM Absorbance / black smoke
OC	Organic Carbon / OC
HAP	PAH / polycyclic aromatic hydrocarbon / hydrocarbon
PUF/PNC	PUF / PNC / ultrafine / particle number
Coarse	Coarse / PM10-2.5
CAP	CAP / concentrated ambient particle
particules ultrafines de carbone	ultrafine carbon particle PUF / PNC / ultrafine / particle number ET spark discharge
NH <sub>4</sub>	NH <sub>4</sub> / ammonium
NO <sub>3</sub> /N	NO <sub>3</sub> / N / Ammonia / nitrate
S/Sulfates	S / sulphate / sulfate / sulfur / sulphur
AOS (incluant indicateur S)	SOA / secondary organic aerosol / organic aerosol / secondary particle / organic particle
AIS (incluant NH <sub>4</sub> NO <sub>3</sub> , (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> , ...)	SIA / secondary inorganic aerosol / inorganic aerosol / secondary particle / inorganic particle
Ni	Ni / Nickel
Zn	Zn / Zinc
Cu	Cu / copper
V	V / Vanadium
Si/SiO <sub>2</sub>	Si / SiO <sub>2</sub> / silica / silicium
Fe	Fe / iron
Autres oligo-éléments	
Endotoxines	Endotoxin / beta-glucan
PO	Oxidative potential / OP
Trafic BC	Traffic / Black Carbon / BC / Elementary Carbon / EC / soot / PM Absorbance / black smoke
Trafic PM2,5	Traffic / PM2.5 / fine particulate matter / fine particle
Diesel	Diesel / Diesel Engine Exhausts / diesel emission
Nouveau Diesel (ex : FAP)	
Essence	Gasoline / Gasoline Engine Exhausts / gasoline emission
Brake dust	Brake dust
Road dust / Crustal material	Road dust / Crustal material
Combustion charbon (incluant indicateur Se)	Coal combustion
Combustion pétrole	Oil combustion / rofa / Residual oil fly ash / petrol combustion / fossil fuel
Emission centrale électrique à charbon	power station
Industrie	Industry / facility / facilities / manufacture / manufacturing
Combustion biomasse	Biomass combustion / woodsmoke / vegetative combustion / vegetative burning / biomass burning / wildfire / prescribed fire / wood-burning
Desert dust	Desert dust
Mer et océan	sea salt / sea spray / ocean spray / sodium
Site de déchet dangereux	Hazardous waste site
Agriculture / Rural	Agriculture / Rural
Métallurgie	Metallurgy / metalworking / metal industry / steel industry / steel manufacture / steel manufacturing / smelter

## Annexe 2

Mise en forme de la liste de termes fournie par l'ANSES pour le traitement par la plateforme CorText. La colonne de gauche contient les termes fournis par l'ANSES, la colonne de droite l'intitulé de catégorie qui est utilisé par le logiciel sur les cartes pour signaler la présence d'un terme de la colonne de gauche.

to be merged	into
Agriculture	Agri
Rural	Agri
SIA	SIA
secondary inorganic aerosol	SIA
inorganic aerosol	SIA
secondary particle	SIA
inorganic particle	SIA
broncho-alveolar lavage proteins	Albu
albumin	Albu
SOA	AOS
secondary organic aerosol	AOS
organic aerosol	AOS
secondary particle	AOS
organic particle	AOS
asthma	Asthme
wheezing	Asthme
autism	autisme
Stroke	AVC
cerebrovascular	AVC
Black Carbon	BC
BC	BC
Elementary Carbon	BC
EC	BC
soot	BC
PM Absorbance	BC
black smoke	BC
COPD	BPCO
chronic bronchitis	BPCO
chronic pulmonary disease	BPCO
Brake dust	Brakedust
bronchopulmonary carcinoma	bronchopulmo
lung cancer	bronchopulmo
Digestive cancer	Cancerautredig
cancer	Cancerautres
brain cancer	Cancercerveau
colon cancer	Cancercol
rectum cancer	Cancercol
colorectal cancer	Cancercol

hematological cancer	Cancerhemato
blood cancer	Cancerhemato
bronchopulmonary carcinoma	Cancerpulmo
lung cancer	Cancerpulmo
lung cancer	Cancerpulmo
breast cancer	Cancersein
bladder cancer	Cancervessie
CAP	CAP
concentrated ambient particle	CAP
Heart failure	Cardio
congestive heart failure	Cardio
CC16	CC16
broncho-alveolar lavage cellularity /neutrophil	Cellularite
macrophage	Cellularite
CIMT	CIMT
CAC	CIMT
coronary artery calcification	CIMT
carotid intima-media thickness	CIMT
cranial circumference	Circrane
Coarse	Coarse
PM10-2.5	Coarse
Biomass combustion	Combustionbiomasse
woodsmoke	Combustionbiomasse
vegetative combustion	Combustionbiomasse
vegetative burning	Combustionbiomasse
biomass burning	Combustionbiomasse
wildfire	Combustionbiomasse
prescribed fire	Combustionbiomasse
wood-burning	Combustionbiomasse
Coal combustion	Combustioncharbon
Oil combustion	Combustionpetrole
rofa	Combustionpetrole
Residual oil fly ash	Combustionpetrole
petrol combustion	Combustionpetrole
fossil fuel	Combustionpetrole
Cu	Cuivre
copper	Cuivre
Hazardous waste site	Dechetsdangereux
Depression	Depression
depressive	Depression
Desert dust	Desertdust
behavioral development	Developbehav
sociability	Developbehav
exploration	Developbehav
anxiety	Developbehav

locomotion	Developbehav
cognitive performances	Developperf
memorization	Developperf
learning	Developperf
Embryonic development	Developtembryo
embryo development	Developtembryo
type 1 diabetes	Diabete
diabetes	Diabete
type 2 diabetes	Diabete
diabetes	Diabete
diabetes	Diabetemetabo
metabolism disorder	Diabetemetabo
Diesel	Diesel
Diesel Engine Exhausts	Diesel
diesel emission	Diesel
heart rythm	ECG
ECG	ECG
heart frequency	ECG
heart rate	ECG
electrocardiogram	ECG
power station	Emissioncharbon
Endotoxin	Endotox
beta-glucan	Endotox
broncho-alveolar lavage enzymes	enzymesLBA
LDH	enzymesLBA
lactate dehydrogenase	enzymesLBA
?-GT	enzymesLBA
NAG	enzymesLBA
N-acetyl glucosamine	enzymesLBA
life expectancy	espvie
Gasoline	Essence
Gasoline Engine Exhausts	Essence
gasoline emission	Essence
birth weight	Faiblepoids
FAP	FAP
FeNO	FENO
exhaled nitric oxide	FENO
exhaled NO	FENO
Fe	Fer
iron	Fer
Fertility	Fertilite
reproductive	Fertilite
heart function	Fonctioncardio
cardiac	Fonctioncardio
heart failure	Fonctioncardio

respiratory function	fonctionrespi
pulmonary function	fonctionrespi
breathing pattern	fonctionrespi
spirometry	fonctionrespi
plethysmography	fonctionrespi
vascular function	Fonctionvasc
vasodilatation	Fonctionvasc
vasoconstriction	Fonctionvasc
vascular function	Fonctionvasc
vasodilatation	Fonctionvasc
vasoconstriction	Fonctionvasc
respiratory function	Fonctionventilenfant
pulmonary function	Fonctionventilenfant
breathing pattern	Fonctionventilenfant
spirometry	Fonctionventilenfant
plethysmography	Fonctionventilenfant
PAH	HAP
polycyclic aromatic hydrocarbon	HAP
hydrocarbon	HAP
Hematocrit	Hemato
Hemoglobin	Hemato
hyperactivity	Hyperactivite
Industry	Industrie
facility	Industrie
facilities	Industrie
manufacture	Industrie
manufacturing	Industrie
infarct	Infarctus
myocardial infarct	Infarctus
heart attack	Infarctus
coronary events	Infarctus
coronary heart disease	Infarctus
infarct	Infarctusgene
myocardial infarct	Infarctusgene
heart attack	Infarctusgene
coronary events	Infarctusgene
coronary heart disease	Infarctusgene
systemic inflammation	Inflamsystem
inflammation	Inflamsystem
inflammatory	Inflamsystem
systemic inflammation	Inflamsystemique
inflammation	Inflamsystemique
pneumonia	Infrespi
atherosclerotic	lesionsather
atherosclerotic	lesionsather

atherosclerosis	lesionsather
lipid deposit	Lipid
fat deposit	Lipid
plaque deposit	Lipid
atherom	Lipid
Defect	malformations
birth defect	malformations
birth outcome	malformations
coagulation	Marqueurcoag
coagulation	Marqueurscoag
blood clotting	Marqueurscoag
bronchoalveolar lavage	Marqueursinflam
systemic oxidative stress	Marqueurstress
oxidative stress	Marqueurstress
systemic oxidative stress	Marqueurstress
oxidative stress	Marqueurstress
vascular endothelium	Marqueurvasc
sea salt	Merocan
sea spray	Merocan
ocean spray	Merocan
sodium	Merocan
lipid metabolism	Metabolipid
cholesterol	Metabolipid
Metallurgy	Metallurgie
metalworking	Metallurgie
metal industry	Metallurgie
steel industry	Metallurgie
steel manufacture	Metallurgie
steel manufacturing	Metallurgie
smelter	Metallurgie
perinatal mortality	Mortaliteperinat
neonatal mortality	Mortaliteperinat
NH4	NH4
ammonium	NH4
Ni	Nickel
Nickel	Nickel
NO3	NO3
N	NO3
Ammonia	NO3
nitrate	NO3
Organic Carbon	OC
OC	OC
PUF	Particulesultrafines
PNC	Particulesultrafines
ultrafine	Particulesultrafines

ultrafine carbon particle	Particulesultrafines
cognitive performances	Perfcogni
memorization	Perfcogni
learning	Perfcogni
Oxidative potential	PO
OP	PO
Preeclampsia	Preecl
premature birth	Prema
gestational age	Prema
arterial pressure	Pressionart
blood pressure	Pressionart
arterial pressure	Pressionart
blood pressure	Pressionart
PUF	PUF/PNC
PNC	PUF/PNC
ultrafine	PUF/PNC
particle number	PUF/PNC
pulmonary	Pulmelast
thoracic elasticity and airway resistance	Pulmelast
pulmonary histology	Pulmhisto
lung histology	Pulmhisto
remodeling	Pulmhisto
inflammatory infiltrates	Pulmhisto
respiratory infection /ENT infections	Respienfant
ENT	Respienfant
pneumonia	Respienfant
rhinitis	rhinite
rhinitis	Rhiniteenfant
Road dust	Roaddust
Crustal material	Roaddust
heart rythm	rythmecardia
ECG	rythmecardia
heart frequency	rythmecardia
heart rate	rythmecardia
electrocardiogram	rythmecardia
cardiovascular	Santecardio
Neurologic	Santeneuro
nervous	Santeneuro
perinatal	Santeperinatale
respiratory	Santerespi
sensitization	Sensib
allergy	Sensib
Si	SI
SiO2	SI
silica	SI

silicium	SI
S	Sulfates
sulphate	Sulfates
sulfate	Sulfates
sulfur	Sulfates
sulphur	Sulfates
thrombosis	Thrombosevein
thrombotic	Thrombosevein
phlebitis	Thrombosevein
Traffic	TraficBC
Black Carbon	TraficBC
BC	TraficBC
Elementary Carbon	TraficBC
EC	TraficBC
soot	TraficBC
PM Absorbance	TraficBC
black smoke	TraficBC
Traffic	TraficPM25
PM2.5	TraficPM25
fine particulate matter	TraficPM25
fine particle	TraficPM25
hospitalization	Ttehospi
hospital admission	Ttehospi
all-cause mortality	Ttemortalite
natural mortality	Ttemortalite
cardiovascular mortality	ttemortalite
V	Vanadium
Vanadium	Vanadium
Zn	Zinc
Zinc	Zinc