



**HAL**  
open science

# Modélisation des pathologies respiratoires des écoliers niçois par les Réseaux Bayésiens

Sandra Perez

► **To cite this version:**

Sandra Perez. Modélisation des pathologies respiratoires des écoliers niçois par les Réseaux Bayésiens. First edition of International Colloquium Health and Space, Frédéric AUDARD, Aix-Marseille University, Marion BORDERON, Aix-Marseille University, Vincent LAPERRIERE, Aix-Marseille University, Sébastien OLIVEAU, Aix-Marseille University, Sandra PEREZ, University of Nice Sophia Antipolis, Sep 2010, Marseille, France. hal-01283742

**HAL Id: hal-01283742**

**<https://hal.science/hal-01283742>**

Submitted on 6 Mar 2016

**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.

# Modélisation des pathologies respiratoires des écoliers niçois par les Réseaux Bayésiens

Sandra PEREZ, UMR Espace 6012 - Université de Nice-Sophia-Antipolis

## 1- Contexte

Nous respirons quotidiennement  $15 \text{ m}^3$  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 que nous ingérons. Les pathologies en lien avec la pollution de l'air sont variées, elles vont des allergies à des sensibilisations cutanées, en passant par des bronchites à répétition qui peuvent déboucher sur de l'asthme. Certaines catégories de la population sont plus vulnérables que d'autres, notamment les enfants, dont le système respiratoire immature est particulièrement fragile. En effet, 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 petites<sup>1</sup>, 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 d'eczéma, ou bien encore de dermatite. Ces allergies risquent 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 (de 5 %, à 10-12 % actuellement). 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 des symptômes suivants : difficultés à respirer, toux nocturne, bronchites à répétition sans fièvre, sentiment d'oppression. Les facteurs favorisant son expression sont complexes et interagissent entre eux, pour induire une susceptibilité à l'asthme. Cette complexité est liée à l'interaction entre des facteurs de natures différentes liés à l'hérédité, mais aussi aux conditions socio-économiques, et environnementales, autant de facteurs qui de plus, peuvent se potentialiser. Ainsi, selon une étude menée par l'IRDES (2008) 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 5 (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 » (IRDES, 2008).

Nous proposons au sein de cette étude de prendre simultanément en compte une série de variables de nature socio-économique et environnementale, dont on suppose avoir un lien avec les pathologies respiratoires (asthme et allergies) observées lors des bilans de santé scolaires, auxquels sont soumis les écoliers niçois, au cours de leur scolarité dans le premier degré. Nous cherchons à identifier d'éventuelles relations entre l'état de santé des écoliers niçois, leur situation socio-économique (évaluée d'après celle de leurs parents), et les caractéristiques environnementales de leur IRIS d'habitation. Nous partons du postulat que la sectorisation imposée par la carte scolaire est respectée, et que les élèves sont scolarisés dans l'établissement scolaire le plus proche de leur lieu de résidence (dans le même IRIS). En effet, l'affectation d'un élève dans une école maternelle ou une école élémentaire relevant de l'enseignement public obéit à la sectorisation scolaire telle qu'elle est définie par décision du conseil municipal : les élèves sont scolarisés dans l'établissement correspondant à leur lieu de résidence. Une dérogation peut être accordée, par le maire de la commune. De plus, même si la carte scolaire est assouplie depuis maintenant 3 ans, c'est surtout lors du passage dans le secondaire que les demandes de dérogation affluent, car pour les parents

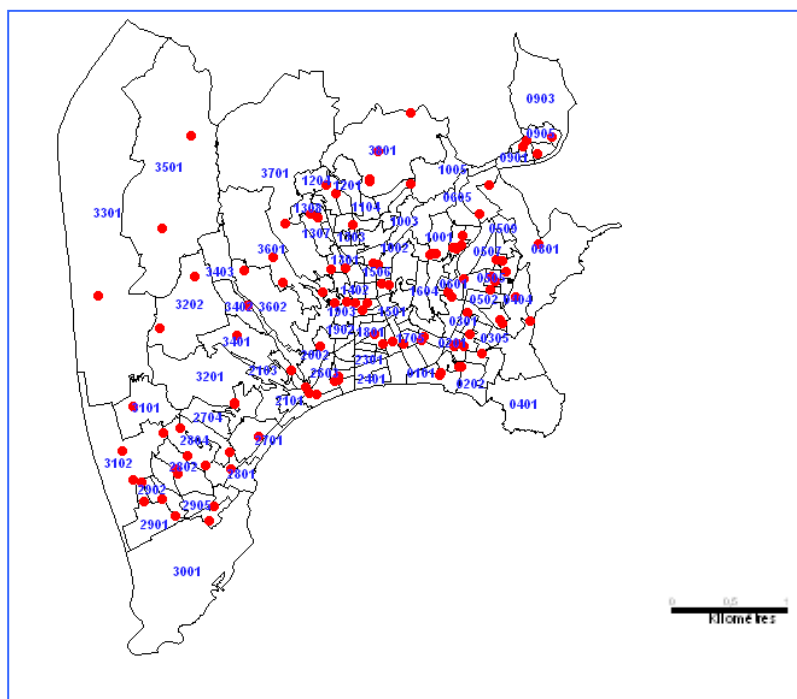
<sup>1</sup> 4 mm chez l'enfant, contre 8 mm pour l'adulte

c'est à ce moment là que les choses véritablement sérieuses commencent. D'après le Service Education de la mairie de Nice les demandes de dérogation étaient de l'ordre de 6.5%<sup>2</sup> pour la rentrée scolaire 2009-2010. Autre postulat : les élèves sont pour l'année scolaire 2008-2009<sup>3</sup> dans l'école qui est la leur depuis le début de leur scolarité, sachant qu'il est rare de quitter un établissement en cours de cycle.

En santé environnementale, les variables estimées exercer un rôle fondamental sur une pathologie donnée sont soumises à une batterie de tests, et ce, dans une perspective essentiellement éco-toxicologique qu'elle soit prédictive ou rétrospective. Or, souvent les pathologies, et notamment celles dont un environnement dégradé peut en être l'origine, sont multifactorielles. Nous proposons ici de prendre en compte non pas 2 ou 3 variables mais un faisceau de variables de natures différentes, et d'analyser la part respective de chacune dans l'apparition des pathologies respiratoires infantiles.

## 2- Travail préparatoire : terrain d'étude, échelle et données

Dès lors que l'on cherche à mettre en relation les états de santé des individus avec leur localisation dans l'espace géographique, et notamment avec leur exposition à des substances potentiellement néfastes pour leur santé, les chercheurs doivent disposer de données les plus fines possibles, afin notamment d'évaluer de la manière la plus juste cette exposition, et le risque sanitaire associé. En milieu urbain, l'ilôt pourrait représenter dans l'idéal cette échelle, sauf que les données socio-économiques ne sont pas disponibles à ce niveau, puisque pour l'INSEE « la brique de base en matière de diffusion de données infra-communales » est l'IRIS (Ilots Regroupés pour des Indicateurs Statistiques). De ce compromis entre l'exigence de précision et la disponibilité des données, le choix de travailler au niveau des IRIS<sup>4</sup> s'est imposé. La ville de Nice est décomposée en 146 IRIS et compte 191 écoles fréquentées par 27 767 élèves. Les écoles sont réparties comme suit : 106 écoles maternelles et 85 écoles élémentaires, dont 22 écoles privées. Ces écoles ne sont pas bien entendu uniformément réparties dans les 146 IRIS qui découpent la ville, certains ne possèdent pas d'écoles, et d'autres plusieurs. Les données sanitaires ne sont pas à l'échelle des IRIS, mais sont réparties par écoles. Les écoles ont donc été dans un premier temps géocodées, puis affectées à l'IRIS correspondant, certains IRIS n'ayant pas d'écoles, ils n'ont donc pas de données, et ne sont donc pas pris en compte dans cette étude. Au total, seuls 60 IRIS sur les 146, sont pris en considération dans l'analyse.



Carte 1. En rouge les écoles géocodées, en bleu le code INSEE des IRIS  
Avec l'aide de Pierre-Alain Mannoni

<sup>2</sup> 1806 dérogations demandées pour les écoles maternelles et primaires sur un total de 27 767 élèves

<sup>3</sup> Année de référence pour les données sanitaires

<sup>4</sup> IRIS-géographie 2008

Les données sanitaires ont été collectées auprès du service de santé scolaire de la ville de Nice. Le problème de l'anonymat ne se pose pas ici étant donné que ces données sont recueillies par école, et que nous répartissons ensuite ces écoles dans leur IRIS de localisation. Le service des infirmières scolaires de la mairie de Nice nous a communiqué les statistiques sanitaires concernant les élèves dépistés dans les écoles niçoises. En fait, seuls les élèves de Grande Section de maternelle et de CM2 font l'objet d'un dépistage systématique, ce qui représente un total de 3798 enfants. Dans le cadre de cette étude nous ne retenons que le nombre d'enfants asthmatiques ou allergiques présents dans les écoles maternelles et primaires pour l'année scolaire 2008-2009. Les données sont actualisées tous les ans, mais la dernière collecte de données disponible concerne 2008-2009. On compte à Nice 11.24% enfants asthmatiques, soit 437 enfants sur les 3798 qui ont bénéficié d'un dépistage, c'est plus que la moyenne nationale (estimée autour de 10%), pour les enfants allergiques (toutes allergies confondues) le pourcentage est de 6.45 %, soit 245 enfants sur 3798.

Nous avons choisi de travailler sur la ville de Nice, car au-delà de la vitrine touristique, Nice présente des inégalités socio-économiques assez marquées entre ses habitants. Ainsi, la moyenne du revenu par ménage pour l'année 2007 atteignait 16 376 euro pour l'IRIS du Mont Gros, contre 56 608 euro pour l'IRIS du Mont Boron (Ecart type de la série 9432 euro). Les données socio-économiques issues de la base de données locales de l'INSEE ont été collectées entre 2006 et 2009, elles sont axées sur la précarité, et concernent : le nombre de chômeurs, le nombre de chômeurs depuis 2 ans, le nombre de personnes travaillant à temps partiel, ayant un niveau d'étude inférieur au Bac, percevant le RMI<sup>5</sup>, le nombre d'étrangers, le revenu moyen des habitants de l'IRIS, le premier quartile de leurs revenus, ainsi que le premier et dernier décile, le nombre de personnes non imposables, percevant la CMUC<sup>6</sup>, en situation de monoparentalité ou bien de parent isolé, le nombre de personnes dont 50% ou 100% du revenu est composé par les allocations familiales, et enfin le nombre de personnes habitant dans un HLM, un habitat de type collectif ou un logement ancien<sup>7</sup>.

Les données environnementales proviennent d'AtmoPACA, plus précisément de l'"Inventaire AtmoPACA 2004". Elles représentent les émissions des principaux polluants primaires exprimés en nombre de kg/an à l'échelle des IRIS. La ville de Nice, hormis l'incinérateur, est exempte d'industries lourdes, potentiellement « impactantes » sur la santé des riverains. La pollution atmosphérique est due essentiellement au transport routier. Le centre urbain est en effet entouré par une série d'axes routiers avec l'autoroute A8 au nord, d'orientation ouest-est (70 000 véhicules/jour), la pénétrante (Voie Mathis) légèrement décalée par rapport au littoral et qui constitue un autre transect ouest-est, la Promenade des anglais en bordure du littoral (65 000 véhicules/jour), et la Route Nationale 202 à l'ouest de la ville, d'orientation cette fois sud-nord (50 000 véhicules/jour). Les transports routiers sont émetteurs d'environ 80 % des oxydes d'azote et du monoxyde de carbone, ils sont également à l'origine pour 64% des particules fines, et pour 50 % des COVNM (Composés Organiques Volatils Non Méthaniques) émis. Le monoxyde de carbone est lié à la combustion incomplète lors des embouteillages en centre ville, c'est la raison pour laquelle ses « émissions sont plus marquées en centre ville par rapport aux axes plus circulants de type autoroute ou pénétrante » (source : AtmoPACA). Le transport, cette fois non routier, lié à la présence du port, à l'est, et de l'aéroport à l'ouest<sup>8</sup> est également à prendre en considération, surtout pour les émissions de dioxyde de soufre SO<sub>2</sub> (56%). Le CO<sub>2</sub> est quant à lui émis majoritairement dans la zone urbaine où se concentrent « plusieurs types d'activités (combustions dues aux installations de chauffage, aux chaufferies industrielles, aux transports) » (source : AtmoPACA).

Les cartes ci-après ont été réalisées en 2007<sup>9</sup> par AtmoPACA dans le cadre d'une étude d'impact environnementale, et ce, afin de connaître l'état initial de la qualité de l'air pour le futur tracé du tramway. Elles ont été réalisées en prenant appui sur une maille qui recouvre l'agglomération Nice Côte d'Azur, nous avons ajouté un cercle noir qui circonscrit la commune de Nice. Ces polluants ont été choisis au regard de leur effet sur la santé, et plus particulièrement sur le système respiratoire. En effet, les asthmatiques y sont particulièrement sensibles. Le CO<sub>2</sub> accroît la fréquence respiratoire, le SO<sub>2</sub> est irritant pour les bronches, et peut provoquer une inflammation de ces dernières, le NO<sub>2</sub> a non seulement des propriétés irritantes, mais également oxydantes sur les bronchioles et les alvéoles, il atteint profondément l'arbre respiratoire, car c'est un gaz peu soluble, et augmente la fréquence et la gravité des crises d'asthme. Les particules fines

<sup>5</sup> Les données datent de 2007 avant donc la mise en place du RSA

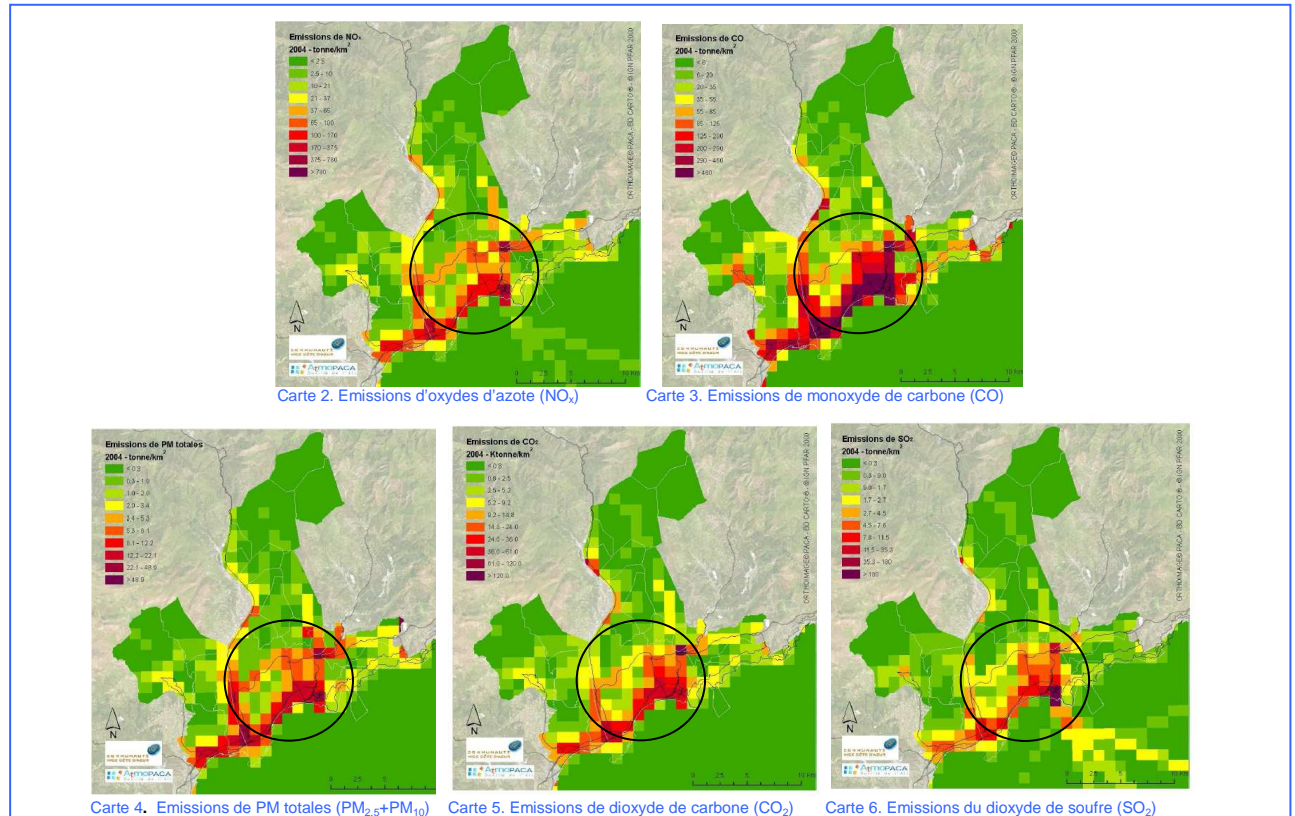
<sup>6</sup> Couverture Maladie Universelle Complémentaire

<sup>7</sup> Année de construction antérieure à 1974

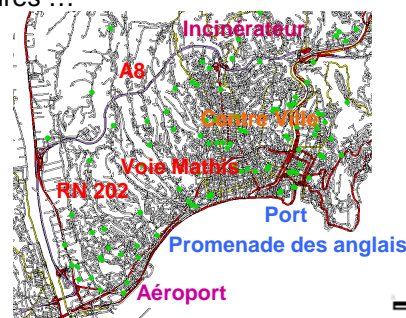
<sup>8</sup> 2<sup>ème</sup> aéroport national après les ADP (10 millions de passagers/an)

<sup>9</sup> Avec comme année de référence : 2004

pénètrent d'autant plus loin dans l'arbre respiratoire que leur taille est petite, elles peuvent de plus adhérer à elles d'autres polluants (hydrocarbures, métaux lourds), elles accentuent elles aussi l'asthme, et de courtes expositions peuvent provoquer de la toux, des irritations, ainsi que des inflammations des bronches. Les particules fines interagissent également avec les pollens pour accroître la sensibilité aux allergies. L'ozone ( $O_3$ ) est fréquent à Nice durant la saison estivale, bien qu'étant connu pour réduire la fonction respiratoire des enfants il n'a pu être intégré dans notre étude, car nous ne disposons tout simplement pas de données d'ozone à l'échelle des IRIS<sup>10</sup>. Toutefois, nous pouvons nous référer au travail de thèse de Nicolas MARTIN (2008) qui indique que « l'espace le plus touché par l'ozone [à Nice] est le littoral ».



A la vue de ces cartes nous observons que certains espaces sont concernés par l'émission simultanée de plusieurs polluants, c'est notamment le cas du centre-ville, du port, de la bordure littorale, et de la plaine du Var à l'ouest. Les émissions d'oxyde d'azote et de dioxyde de soufre en mer sont dues aux rejets des cheminées des bateaux (activité portuaire, ferries avec la Corse). Notons également qu'au niveau de la plateforme aéroportuaire de nombreux polluants sont émis : oxydes d'azote ( $NO_x$ ), dioxyde de soufre ( $SO_2$ ), monoxyde de carbone (CO), Composés Organiques Volatils (COV), et particules. Ces composés chimiques proviennent « de toutes les réactions de combustion présentes sur l'aéroport (avions, véhicules spéciaux, centrale d'énergie), et de l'évaporation de certains polluants comme les COV (avitaillement, stockage de carburant) » source : AtmoPACA. La Carte ci-dessous qui replace les écoles (en vert) dans leur contexte « environnemental » (proximité avec les principaux axes de communication, avec le port, l'aéroport, ou bien en centre ville) se passe de commentaires ...



<sup>10</sup> En effet, l'étude d'AtmoPACA ne concernait que les polluants primaires, or, l'ozone est un polluant secondaire produit par la réaction des hydrocarbures imbrulés et des oxydes d'azote avec l'oxygène de l'air et sous l'influence de la lumière solaire

Au point où nous en sommes de l'analyse, il est temps d'examiner la nature et l'intensité des liens présumés entre les données de pollution, les variables socio-économiques et les pathologies respiratoires observées au sein des écoles niçoises, et ce, à l'aide des réseaux bayésiens.

### 3- Méthodologie et résultats

Les réseaux bayésiens doivent leur nom au mathématicien britannique Thomas BAYES (1702-1761) qui travailla sur la théorie des probabilités, « initiés par Judea PEARL dans les années 1980 [ils] 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. Les réseaux bayésiens permettent d'établir des relations de dépendances ou d'indépendances entre différentes variables, aussi bien quantitatives que qualitatives, 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.

Les réseaux bayésiens permettent également d'acquérir, rassembler, fusionner des connaissances de diverses natures dans un même modèle, de créer de la connaissance manquante, voire d'extraire de la connaissance contenue dans une base de données, et ce, en utilisant le principe du *data mining*. 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. Ils s'apparentent à un graphe dont les nœuds définissent les variables du système, et dont les arcs définissent l'existence de relations entre ces variables. Plus précisément, ils décrivent de manière qualitative 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 sont associées à une probabilité. Ainsi, l'observation d'une cause ou de plusieurs causes n'entraîne pas systématiquement l'effet ou les effets qui en dépendent, mais modifie leur probabilité de survenue.

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 » (NOVACK, 2008). 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. Dans notre cas, nous avons choisi d'extraire de la connaissance contenue dans la base de données<sup>11</sup>.

La structure du réseau est visuellement représentée par les arcs orientés figurant l'ensemble des relations causales reliant les variables. La mise en place proprement dite du réseau se fait par la suite en associant au formalisme graphique de la structure du réseau, un contenu probabiliste qui constitue les paramètres du réseau. Les probabilités étant calculées à partir de variables discrètes, si elles sont continues, il convient au préalable de les discrétiser, ce que nous avons fait en utilisant la méthode du K-means qui permet d'obtenir des intervalles par la recherche de valeurs moyennes en minimisant la variance intra-classe.

Comme nous sommes dans le cadre d'un apprentissage non supervisé puisque nous cherchons à extraire de la connaissance à partir de la base de données, nous faisons appel à un algorithme de découverte d'associations (apprentissage structurel), afin d'identifier l'ensemble des relations probabilistes présentes au sein de nos variables. Il existe plusieurs méthodes d'apprentissage structurel, après les avoir toutes testées notre choix s'est porté sur celle appelée « arbre de recouvrement maximal », où l'intensité des relations entre les variables est calculée par le biais du coefficient de corrélation de Pearson et la divergence de Kullback-Leibler qui mesure la différence entre les distributions de probabilités représentées par le réseau avec l'arc de relation et sans cet arc. L'épaisseur de l'arc est proportionnelle à l'importance de la relation probabiliste. Plus l'arc est épais, plus son importance est grande pour la structure globale du réseau. Une couleur est associée à la nature des données : rouges pour les données sanitaires, violet pour celles relatives à la pollution, vert pour les données socio-économiques, et jaune pour l'habitat.

---

<sup>11</sup> Tous les traitements ont été réalisés avec le logiciel BayesiaLab 4.6.8

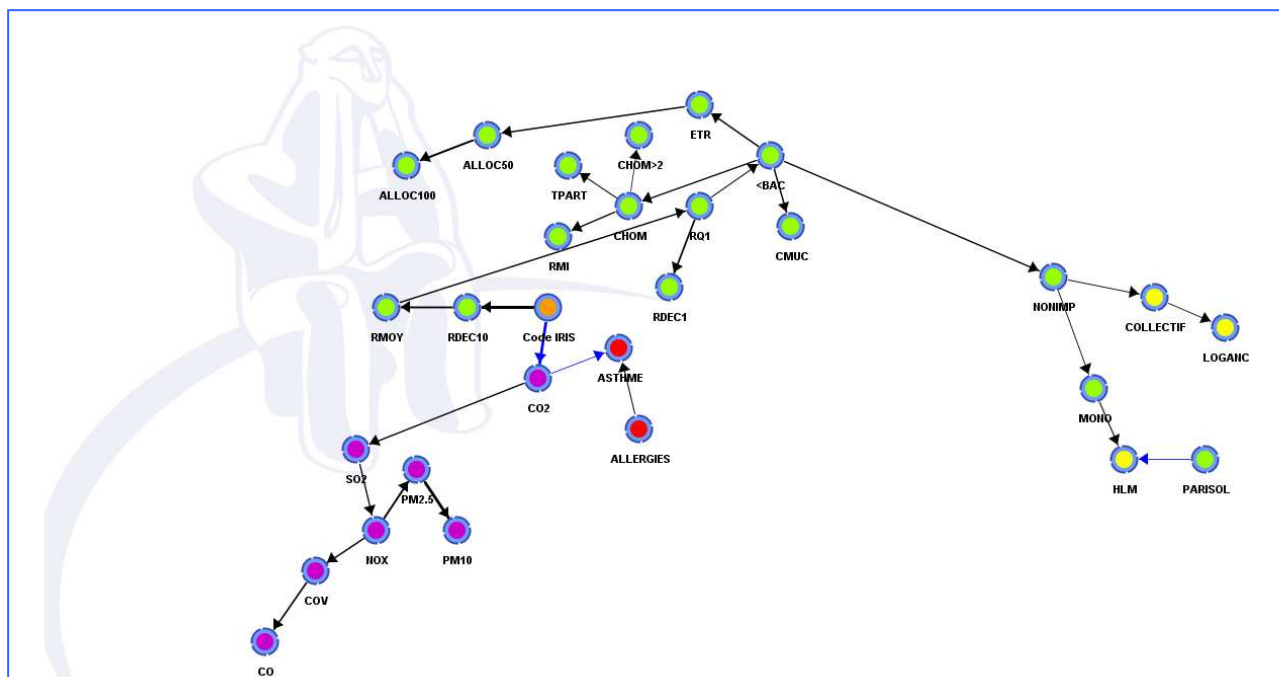


Figure 1. Le réseau bayésien extrait de la base de données

L'analyse du réseau dans son ensemble révèle que les données ne sont pas indépendantes les unes des autres puisqu'il existe plusieurs relations entre les nœuds. Ainsi, il existe bien une relation entre les allergies et l'asthme, mais, nous ne savons pas en regardant le graphique quelle est l'intensité de cette relation, ni si elle est positive ou négative. Pour cela, nous devons nous pencher sur le tableau ci-dessous afin de connaître le coefficient de corrélation de Pearson. Cette relation est positive, elle est de l'ordre de 0.36, mais elle est faible, par conséquent si les allergies jouent bien un rôle dans l'apparition de l'asthme, elles ne sont pas *a priori* le seul facteur explicatif. Le CO<sub>2</sub> joue lui aussi un rôle à hauteur de 0.25. Les coefficients de corrélation les plus élevés s'observent entre les différents polluants, ce qui est logique puisqu'une émission importante de CO<sub>2</sub> présuppose l'émission des autres polluants, il en est de même entre les variables de nature socio-économiques, avec par exemples de fortes relations entre les personnes qui ont un niveau d'étude inférieur au Bac, et le fait qu'elles soient au chômage, bénéficiant de la CMUC, ou bien sont non imposables. Les variables chômage, chômage de plus de 2 ans, RMI Allocation 50 et 100 sont également très liées, suivant une logique de dégradation économique bien connue.

Parent	Enfant	Divergence de Kullback-Leibler	Contribution globale	Information mutuelle	Corrélation de Pearson
Code IRIS	RDEC10	2,537761	7,32%	2,537761	0,3944
PM2.5	PM10	2,405215	6,94%	2,405215	0,9999
Code IRIS	CO2	2,289961	6,61%	2,289961	0,4083
ALLOC50	ALLOC100	1,668728	4,81%	1,668728	0,9575
NOX	PM2.5	1,644635	4,75%	1,644635	0,9545
RDEC10	RMOY	1,583111	4,57%	1,583111	0,9071
RQ1	RDEC1	1,499816	4,33%	1,499816	0,9231
COV	CO	1,416349	4,09%	1,416349	0,8066
<BAC	ETR	1,261682	3,64%	1,261682	0,8885
ETR	ALLOC50	1,245486	3,59%	1,245486	0,9335
NOX	COV	1,241476	3,58%	1,241476	0,8524
<BAC	CMUC	1,228735	3,55%	1,228735	0,8918
RMOY	RQ1	1,207456	3,48%	1,207456	0,8685
CO2	SO2	1,180227	3,41%	1,180227	0,8430
<BAC	NONIMP	1,159198	3,34%	1,159198	0,8693
CHOM	RMI	1,152652	3,33%	1,152652	0,8581
<BAC	CHOM	1,132105	3,27%	1,132105	0,8756
SO2	NOX	1,116051	3,22%	1,116051	0,8562

MONO	HLM	0,915526	2,64%	0,490277	0,4238
CHOM	TPART	0,874595	2,52%	0,874595	0,7417
RQ1	<BAC	0,867449	2,50%	0,867449	-0,7610
ALLERGIES	ASTHME	0,841711	2,43%	0,375246	0,3679
NONIMP	COLLECTIF	0,773368	2,23%	0,773368	0,5416
NONIMP	MONO	0,739269	2,13%	0,739269	0,6814
COLLECTIF	LOGANC	0,689278	1,99%	0,689278	0,6533
PARISOL	HLM	0,679516	1,96%	0,254267	0,3028
CO2	ASTHME	0,672117	1,94%	0,205651	0,2517
CHOM	CHOM>2	0,636622	1,84%	0,636622	0,6527

Tableau 1. Les premiers indicateurs statistiques et les relations entre les variables

Si nous analysons non plus le réseau dans sa globalité, mais un nœud en particulier comme l'ASTHME, ce nœud est placé au cœur du réseau bayésien, et devient un nœud cible<sup>12</sup> (l'apprentissage est cette fois supervisé). La clarté des carrés inscrits dans les nœuds est proportionnelle à la quantité d'information apportée par le nœud sur la variable cible, autrement-dit, plus les carrés sont clairs, plus le nœud correspondant est corrélé à la cible. Le nœud PARISOL (parent isolé) n'ayant pas de carré, cela signifie qu'il est marginalement indépendant de la cible, tandis que ALLERGIES est le nœud le plus corrélé, suivi des IRIS, des émissions en CO<sub>2</sub>, puis des autres polluants.

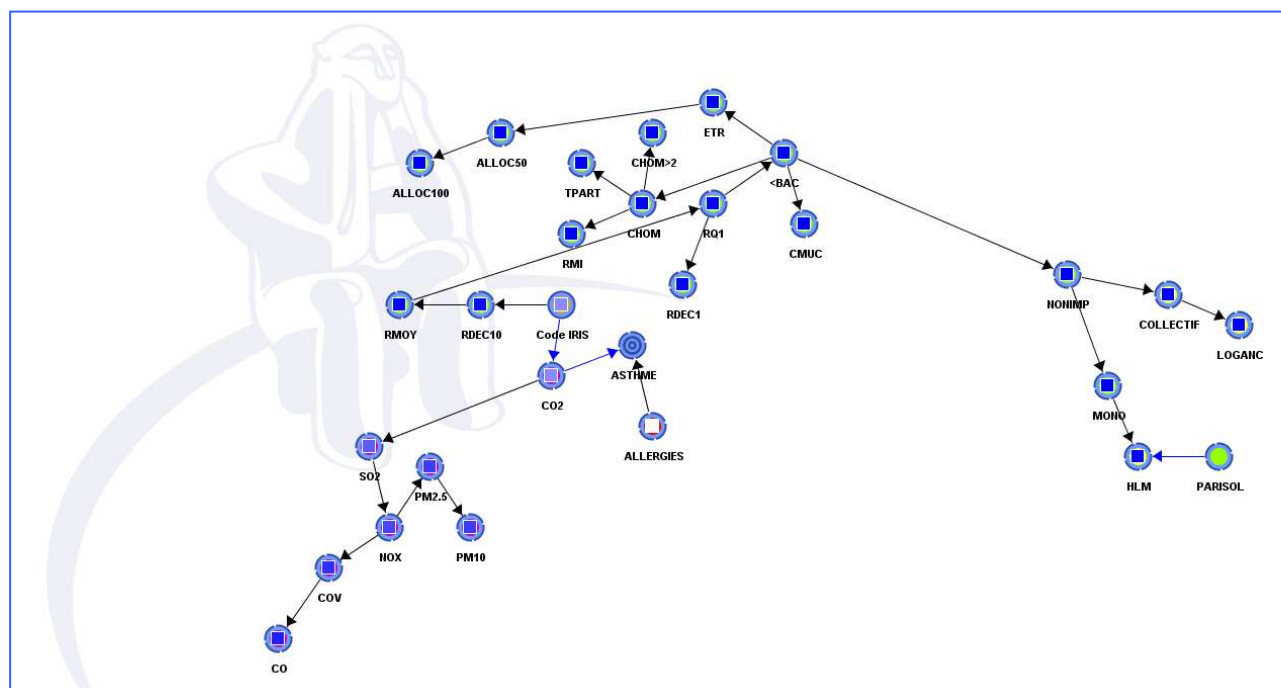


Figure 2. L'apport d'information des variables sur la variable cible ASTHME

<sup>12</sup> Pour la modalité la plus élevée des asthmatiques : 9.836



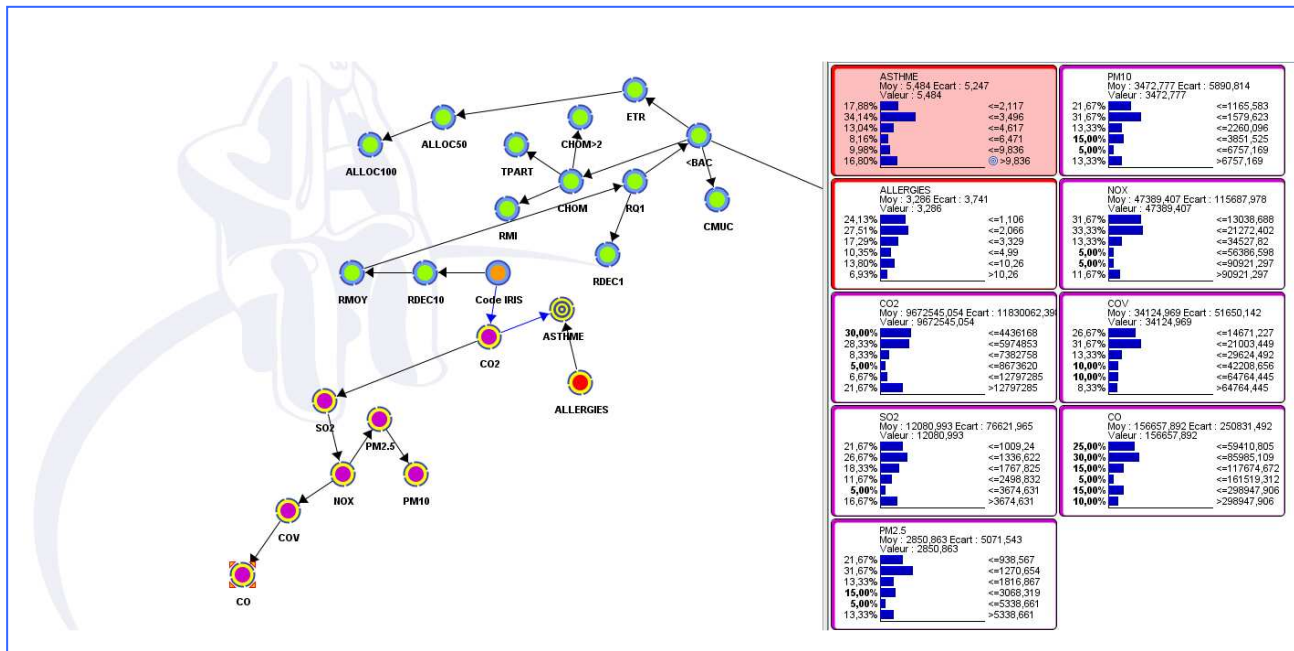


Figure 3. Monitoring entre les variables les plus corrélées avec la variable cible ASTHME

Un monitoring entre le nœud ASTHME et les variables avec lesquelles il est le plus corrélé donne les résultats suivants (Figure 3) : 34% des IRIS analysés comptent plus de 3.5 enfants asthmatiques dans leurs écoles, et pour 16.8% d'entre eux le taux grimpe à plus de 9.8 enfants. Concernant les enfants allergiques la moitié des IRIS se situe dans les deux premières classes, 7% des IRIS comptent cependant plus de 10 enfants allergiques. Relativement aux principaux polluants en lien avec l'asthme, là encore, plus de la moitié des IRIS se trouvent dans les deux premiers intervalles. Mais, nous avons vu sur les cartes que les plus forts taux de pollution étaient très conditionnés à la proximité des axes routiers, ce résultat n'est donc pas contradictoire avec les observations faites d'après les cartes. Examinons à présent ce qui se passe si nous prenons comme référence la classe la plus élevée d'enfants asthmatiques (9.836) (Figure 4). Les probabilités associées sont alors mises à jour, et des flèches indiquent le sens de la modification (augmentation ou diminution). Cette classe extrême concentre plus de 48% des enfants allergiques (lien connu entre les 2 variables), et comme on pouvait s'y attendre elle est en lien avec les taux les plus élevés des différents polluants (25% des NO<sub>x</sub>, 35% du SO<sub>2</sub>, 27% des PM<sub>2.5</sub>) et notamment pour plus de 45% du taux le plus élevé de CO<sub>2</sub> !

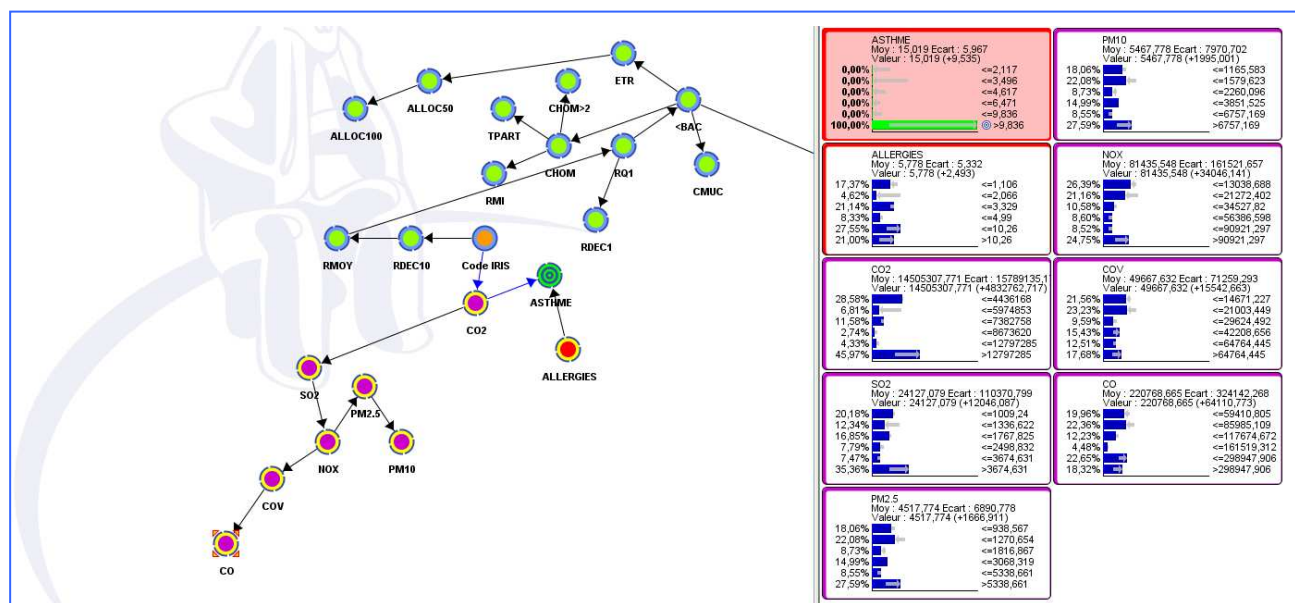


Figure 4. Monitoring sur la classe la plus élevée de la variable cible ASTHME

Examinons à présent ce qui se passe pour la modalité 3.496 qui correspond à la situation non plus extrême, mais la plus fréquente (Figure 5). La probabilité d'avoir plus de 10 enfants allergiques tombe à 1.7%, la

majeure partie des IRIS compte seulement 1 ou 2 enfants allergiques. Néanmoins, les probabilités associées aux dernières classes des polluants sont augmentées<sup>13</sup> (comparées à l'analyse globale de la Figure 3) de 9% pour le CO<sub>2</sub>, 7% pour le SO<sub>2</sub>, les autres polluants variant quant à eux entre 5 et 3%. Les écoliers habitant dans ces IRIS seraient ainsi soumis à une pollution de fond, cette situation étant la plus fréquemment rencontrée.

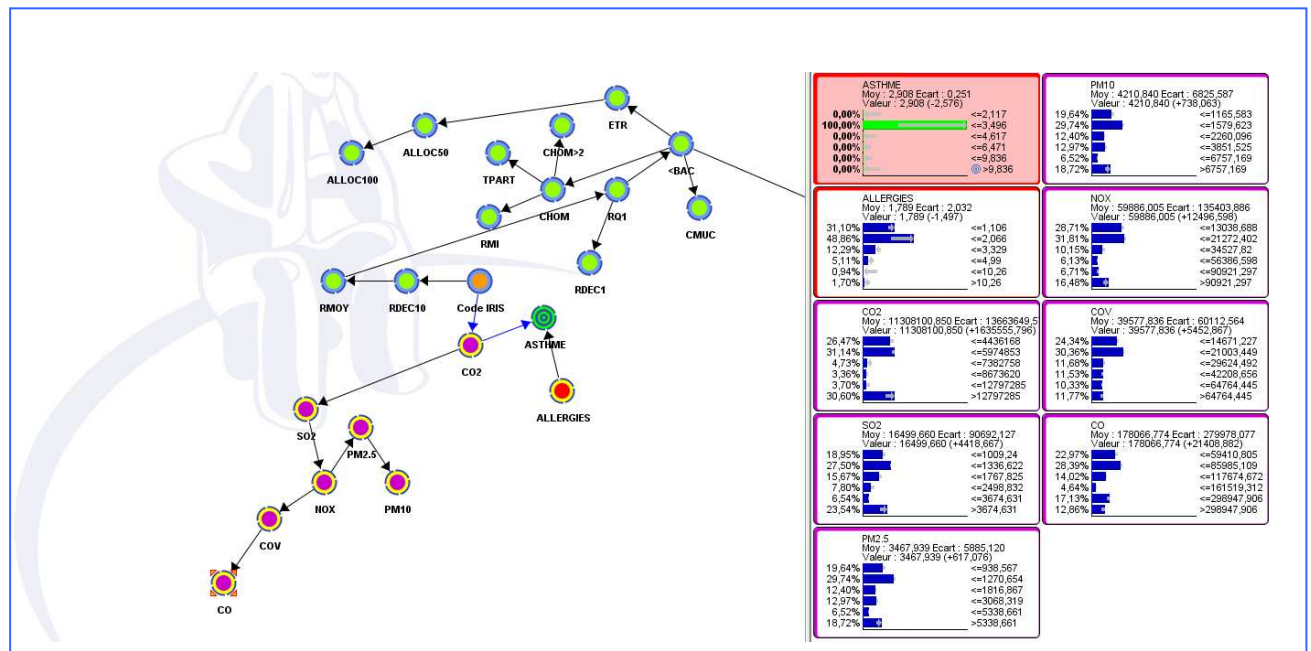


Figure 5. Monitoring sur la classe la plus fréquente de la variable cible ASTHME

L'analyse de sensibilité de la cible ASTHME nous permet à présent de voir à quelles variables les différentes modalités de l'asthme sont sensibles. Ainsi, la modalité la plus fréquemment rencontrée (3.496) est la plus liée aux allergies (la probabilité d'être allergique lorsque nous avons plus de 3 enfants asthmatiques dans une école est de 0.6), ainsi qu'au CO<sub>2</sub> et au SO<sub>2</sub> comme nous l'avons vu précédemment. Les variables de nature socio-économiques apparaissent au deuxième plan, après les variables de pollution, et ce, quelque soient les modalités considérées. Cependant, si nous réitérons cette analyse de sensibilité non plus sur la variable cible ASTHME, mais sur la variable cible IRIS (Figure 7), et que nous considérons les 9 IRIS où le nombre d'écoliers asthmatiques est nettement supérieur à la moyenne, cette séparation entre les facteurs écologiques et socio-économiques n'est plus aussi marquée, et les caractéristiques socio-économiques de ces IRIS émergent à nouveau.

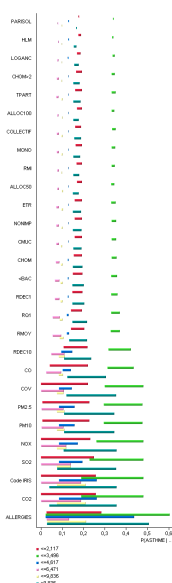
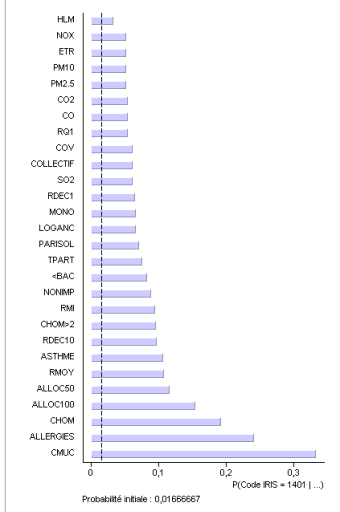
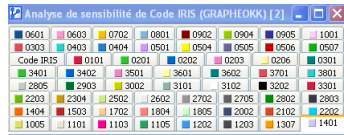
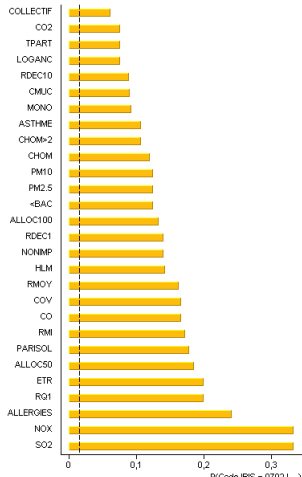


Figure 6. Analyse de sensibilité des différentes modalités de la variable cible ASTHME

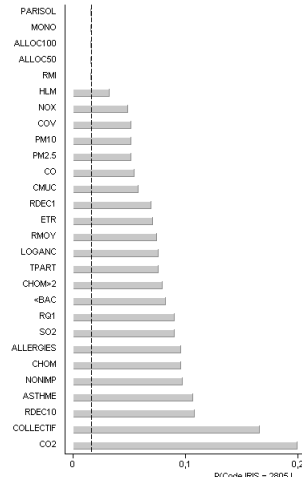
<sup>13</sup> Comme indiqué par le sens des flèches



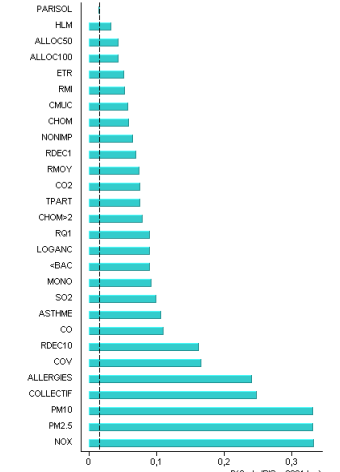
**IRIS Saint Barthélémy [1401]**



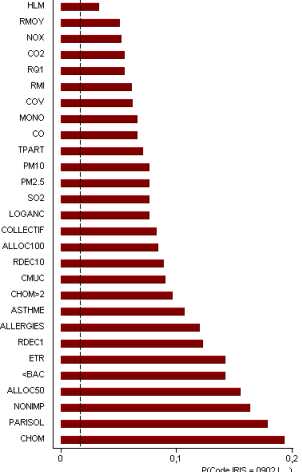
**IRIS Bon voyage [0702]**



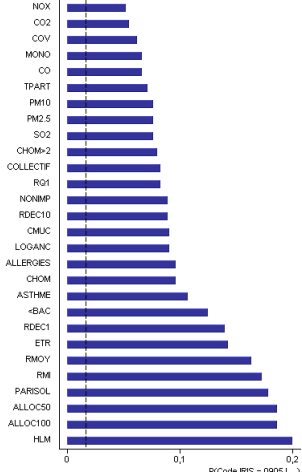
**IRIS Napoléon III [2805]**



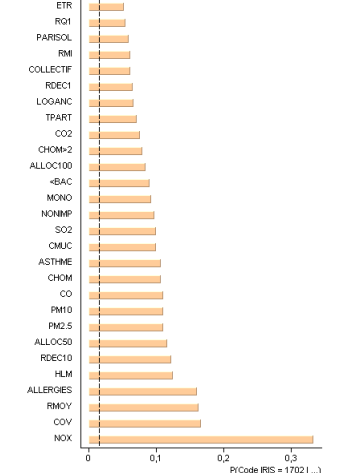
**IRIS Gairaut [3801]**



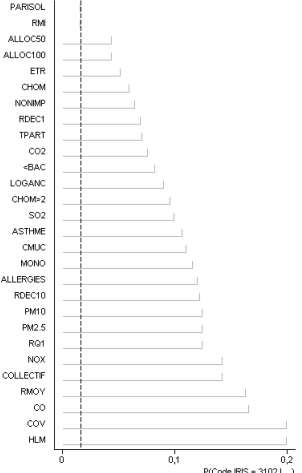
**IRIS Ariane Monzie [0902]**



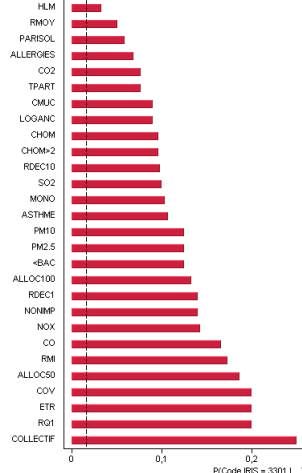
**IRIS Ariane Saramito [0905]**



**IRIS Saint Roch [1702]**



**IRIS Nikaia [3102]**



**IRIS Lingostière-Saint Isidore [3301]**

Figure 7. Analyse de sensibilité de la variable cible IRIS

En fait, les probabilités liées aux variables de nature socio-économiques sont pour ces IRIS mêlées à celles des polluants. Elles dessinent alors une configuration particulièrement néfaste à une bonne santé respiratoire 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, RMI etc..) qui habitent dans des HLM ou de l'habitat collectif, et où la qualité de l'air intérieur peut là encore aggraver l'asthme.

## Discussion

Certains facteurs n'ont pas été pris en compte faute de moyens. Notamment, toute la partie liée non plus à l'air extérieur, mais à l'air intérieur, que ce soit dans les lieux d'habitation des enfants, ou bien dans leur lieu de scolarisation, où ils passent en moyenne 8 heures par jour. Il en est de même pour les prédispositions génétiques (atopie) dont le rôle dans l'apparition de l'asthme infantile est reconnu, et de l'impact des pollens dans la potentialisation des polluants. L'exposition est ici évaluée par rapport aux émissions et non aux concentrations des polluants dans l'air, plus difficile à cerner à l'échelle d'une ville, en raison de sa grande viabilité spatio-temporelle (parfois d'une rue à l'autre). Malgré cela, la performance du réseau bayésien pour la variable cible ASTHME est estimée par un coefficient de détermination de 0.59. Par leur simplicité d'utilisation, les réseaux bayésiens s'avèrent être des outils particulièrement efficaces dans le domaine de l'étiologie. Le recours aux réseaux bayésiens pour l'étude des risques sanitaires, ou en amont d'études écotoxicologiques, leur assurent un avenir prometteur.

## Bibliographie Sélective

- ASCHAN-LEYGONIE Ch., BAUDET-MICHEL S., 2008, *La santé respiratoire dans les villes françaises : une caractérisation par les hospitalisations*, Environnement, Risques et Santé n°2, pp.130-137.
- AtmoPACA., 2009, *Etat initial de la qualité de l'air projet NCA-Tramway*, 164p.
- BAILLY A., RACINE J-B., 1988, *Qualité de la vie, bien être, indicateurs sociaux territoriaux : l'homo géographique entre choix et contraintes*, *L'espace géographique*, 3, pp. 232-240.
- BASLY L., 2000 *Téledétection pour la qualité de l'air en milieu urbain*, Thèse de doctorat de l'Université de Nice - Sophia Antipolis, 168 p.
- BERCHE A., 2006, *L'impact de l'environnement sur la santé et le bien-être en Picardie*, 117p.
- BLEY D., CHAMPAUD J., BAUDOT P., BRUN B., PAGEZY H., VERNAZZA-LICHT N.,(eds), 1997, *Villes du Sud et environnement*, Editions SEH/de Bergier, 241 p.
- BLEY D., VERNAZZA-LICHT N., 1997, "Qualité de vie : santé, écologie, environnement", *Prévenir* n°33, 224 p.
- BLEY D., CHAMPAUD J., BAUDOT., BRUN B., PAGEZY H., VERNAZZA-LICHT N., 1997, *L'environnement urbain est-il une idée neuve ? "Ville du Sud et environnement"*, SEH/Edition de Bergier eds, pp.7-17.
- BLEY D., 2001 *La santé, entre biologie et culture : un domaine majeur d'une écologie humaine*, *NSS*, 4, p. 3
- BLEY D., 2001 *Une charte de la qualité de vie. Un outil pour la prise de décision en santé publique*, entretien publié dans "Le Panorama du Médecin", jeudi 17 mai 2001, 10p.
- BLEY D., novembre 2000 *Population, environnement, qualité de vie, Habilitation à Diriger les Recherches*, document multigraphié, 78 p.
- BRAINARD J-S., JONES A-P., BATEMAN I-J., LOVETT A-A., FALLON P-J., 2002, *Modelling environmental equity : access to air quality in Birmingham UK*, *Environment and Planning A*, (4), pp. 695-716
- BUSSON O., LUCAS-GABRIELLI V., ROCHEREAU T., SERMET C., TONNELIER F., 1999, *Environnement local, santé, soins et protection sociale*, CREDES, 126p.
- CHARPIN D., 2004, *L'air et la santé* Flammarion, Medecine-Sciences, 305p.
- CHAUVIN P., 2006, *Territoires, vulnérabilités et santé*, Conseil économique et social régional, Paris.
- CORNUT P., BAULER T., ZACCAI E., 2007, *Environnement et inégalités sociales*, Editions de l'Université Libre de Bruxelles
- 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.
- DJEBBAR A., MEROUANI H-F., 2006, *Vers une modélisation de la base de cas par un réseau bayésien : application d'aide au diagnostic de pathologies hépatiques*, 8p., 6e

Conférence Francophone de MOdélisation et SIMulation - MOSIM'06 - du 3 au 5 avril 2006 – Rabat- Maroc « Modélisation, Optimisation et Simulation des Systèmes : Défis et Opportunités ».

- FROMENT VEDRINE M., 2008, *Médecins, chercheurs et risques sanitaires*. Journées scientifiques de l'AFSSET, LE CONCOURS MEDICAL, n°1 30, pp. 75-93.
- FUSCO G., 2008, *Spatial dynamics in France* pp.87-109 in *Bayesian Networks a practical guide to application*, Wiley, 428p.
- Global INitiative for Asthma (GINA)., 2009 *Global strategy for asthma management and prevention*, Rapport, 92p.
- GRANGE D., HOST S., GREMY I., 2007, *Les composés organiques volatils (COV) : État des lieux*, 6p.
- IRDES., 2008, *L'asthme en France en 2006 : prévalence et contrôle des symptômes*, Questions d'économie de la Santé, n°138, décembre 2 008, 8p.
- JEDOR B., RIBERON J., MANDIN C., et al, 2006, *Lieux de vie fréquentés par les enfants : typologie et qualité de l'air intérieur*, Centre Scientifique et Technique du Bâtiment. (C.S.T.B.). Marne-la-Vallée, Institut National de l'Environnement Industriel et des Risques. (I.N.E.R.I.S.) Verneuil-en-Halatte, 95p.
- KAWACHI I., BERKMAN L., 2003 *Neighborhoods and Health*, Oxford University Press, 347p.
- KIRCHNER S., 2007, *État de la pollution dans les logements français*, Congrès Qualité de l'air dans les logements, La Rochelle, POLLUTION ATMOSPHERIQUE n°194, pp. 139-142.
- KOSTRZEWA A., 2007, *Impacts économiques des pathologies liées à la pollution - Étude d'impact sur les coûts que représentent pour l'Assurance maladie certaines pathologies liées à la pollution : Illustration avec l'asthme et le cancer*, 139p, Université de Bordeaux 2.
- KOSTRZEWA A., 2007, *Pollution atmosphérique particulaire et santé respiratoire des enfants à Bordeaux*, Thèse pour le diplôme d'état de docteur en médecine, Université de Bordeaux 2, 114p.
- LAIGLE L., OEHLER V., 2004, *Les enjeux sociaux et environnementaux du développement urbain : la question des inégalités écologiques*, Recherche exploratoire pour le PUCA / MELT, Centre Scientifique et Technique du Bâtiment.
- LANOE J-L., et MAKDESSI-RAYNAUD Y., 2005, *L'état de santé en France en 2003, Santé perçue, morbidité déclarée et recours aux soins à travers l'enquête décennale santé*, Etudes et Résultats, DREES.
- LERAY Ph., 2008, *Réseaux bayésiens - introduction et apprentissage, modélisation et découverte de connaissances*, EGC 2008
- MAES S., MEGANK S., LERAY P., 2007, *An integral approach to causal inference with latent variables*, London College Publications pp.17-41.
- MAIGNANT G., 2002, *Pollution et développement durable des villes françaises, étude de cas : Nice, Marseille, Lyon et Paris*. Thèse de Doctorat de l'Université de Nice Sophia-Antipolis, 3 volumes 460 p + 1 atlas de 135 pages.
- MAIGNANT G., 2005 *Compacité et forme urbaine, une analyse environnementale dans la perspective d'un développement urbain durable*, in *Incertitude et environnement*, Edisud, 17p.
- MARTIN N., 2008 *La pollution par l'ozone et la climatologie dans un espace méditerranéen : les Alpes-Maritimes*, Thèse de Doctorat, soutenue à Nice le 8 décembre 2008, 290p.
- MISC, 2007 *Modèles graphiques probabilistes*, Revue d'Intelligence Artificielle, n°21, Hermès.
- NAIM P., WUILLEMIN P-H., LERAY P., POURRET, O., and BECKER A., 2004, *Réseaux bayésiens*, Eyrolles.
- NOVAK, J. D. *La théorie qui sous-tend les cartes conceptuelles et la façon de les construire*, Université de Cornell In <http://sites.estvideo.net/gfritsch/doc/rezo-cfa-410.htm> (juillet 2010)
- PARIZOT I., RENAHY E., PECHOUX S., QUAGLIA M., VIVIER G., CHAUVIN P., 2006, *Santé, inégalités et ruptures sociales (SIRS) : une cohorte francilienne d'épidémiologie sociale*. Colloque ADELFF « épidémiologie sociale et inégalités de santé », Toulouse, 18-19 mai 2006.
- PAUGAM S., 2007, *La santé pour tous : un enjeu territorial*, Colloque « Les inégalités sociales de santé en France aujourd'hui », Centre National de la Fonction Publique Territoriale, Nancy, 25 juin 2007.

- PEARL J., 1988, *Probabilistic Reasoning in Intelligent Systems*, San Mateo, CA, Morgan Kaufman
- PEARL J., 2000, *Causality : Models, Reasoning, and Inference*, Cambridge University Press.
- PEWSNER D., BLEUER J-P., BUCHER H-C., BATTAGLIE M., 2001, *Sur la voie de l'intuition : Théorème de Bayes et Diagnostic en médecine générale*, p. 41-45.
- PICHERAL H., 1999, *Risques et inégalités de santé : de la salubrité à l'équité*, Hérodote, n°92, pp.50-68.
- Qualitair., 2004 *Etude de la qualité de l'air sur l'aéroport « Nice Cote d'Azur »* 22p.
- RICAN S., 2002, *Géographie des affections respiratoires en France*, Doctorat, Nanterre, Université Paris X, 283 p.
- RICAN S., SALEM G., JOUGLA E ; « *Villes et santé respiratoire en France* », *Géocarrefour*, vol. 78/3, 2003, [En ligne], mis en ligne le 12 mars 2008. URL : <http://geocarrefour.revues.org/index2048.html>. Consulté juillet 2010
- RIZK C., 2003, *Le cadre de vie des ménages les plus pauvres*, INSEE n°926.
- SALEM G., 1998, *La santé dans la ville. Géographie d'un petit espace dense : Pikine (Sénégal)*, Khartala-Orstom, 360p.
- SALEM G., 1999, *Les problèmes de santé dans les villes : une urgence négligée*, *Médecine Tropicale*, vol. LIX, n°2, pp. 31-33.
- SEGALA Cl., POIZEAU D., MESBAH M., WILLEMS S., MAIDENBERG M., 2008, *Winter air pollution and infant bronchiolitis in Paris*, *ENVIRONMENTAL RESEARCH*, n°106, pp. 96-100.
- SOTTAS P-E., BRUZZI R., VERNEZ D., MANN S., GUILLEMIN M., DROZ P-O., 2005, *A Bayesian framework for the combination of different occupational exposure assessment methods*, IOHA.
- TASCAN-MENNETRIER Cl., 2007, *Villes, santé et développement durable*, La Documentation française, 550p.
- THEYS J., EMELIANOFF C., 2001, *Les contradictions de la ville durable*, *Le débat* 113, pp.122-135.
- VERNAZZA-LICHT N., BLEY D., 2003, *La qualité de vie in Manuel d'Anthropologie* (CHIARELLI B., REBATO E., SUSANNE S. eds), De Boeck Université.
- VILLALBA B., ZACCAI E., 2007, *Inégalités écologiques, inégalités sociales : interfaces, interactions, discontinuités ?*, Développement durable et territoire, Dossier 9 : Inégalités écologiques, inégalités sociales.