

EHESP



Santé Publique et Risques

Environnementaux

Promotion : **2011-2012**

Date du Jury : **Juillet 2012**

**Influence de l'environnement sur la Mortalité
Infantile dans l'Agglomération de Marseille**

Etude épidémiologique à fine échelle géographique

Hermine Mével

Remerciements

Je remercie tout d'abord Denis Zmirou de m'avoir offert l'opportunité de travailler sur ce sujet aussi passionnant.

Je veux également remercier Séverine Deguen, mon maître de stage, de m'avoir encadrée, et épaulée tout au long de ce stage. Je tiens à lui faire part de toute ma reconnaissance pour son soutien et son accompagnement à chaque étape de ce travail, ainsi qu'à la rédaction du mémoire. Je lui dis aussi un grand merci pour ses encouragements et conseils qui m'ont enrichie aussi bien sur le plan professionnel que personnel.

Merci à toute l'équipe d'Equit'area, et en particulier à Claire Petit, Wahida Kihal, Cindy Padilla et Benoît Lalloué, qui m'ont aidée et ont toujours pris du temps pour répondre à mes questions.

Merci à Fei Gao, non seulement de sa bonne humeur et de son soutien, mais aussi de m'avoir aidée dans l'apprentissage de logiciels qui étaient nouveaux pour moi.

Merci à Fabienne Jutel qui a pris de son temps pour m'aider dans toutes les démarches administratives.

Je tiens enfin à remercier tout le personnel de l'EHESP, en particulier des départements Epidémiologie et Biostatistiques et Santé-Environnement-Travail (DSET) pour m'avoir accueillie et accompagnée pendant ces quelques mois.

Sommaire

Introduction	2
1 Cadre et Objectifs de l'étude	3
1.1 Le Projet Equit'Area	3
1.1.1 Contexte	3
1.1.2 Caractéristiques du projet	4
1.1.3 Les différentes variables	5
1.2 Présentation de l'étude	6
1.2.1 Objectifs	6
1.2.2 Déroulement	6
2 Méthodologie	8
2.1 Recueil des données sanitaires	8
2.1.1 Naissances	8
2.1.2 Décès	8
2.2 Agrégation des cas à l'échelle de l'IRIS	10
2.2.1 Géocodage	10
2.2.2 Cartographie	12
2.3 Bilan des données disponibles et modification de la zone d'étude	12
2.4 Analyses statistiques	14
2.4.1 Statistiques descriptives	14
2.4.2 L'Analyse bayésienne	16
3 Résultats	20

4 Discussion.....	23
Conclusion	27
Bibliographie	28
Liste des annexes	I

Liste des sigles utilisés

API : Application Programming Interface (Interface de programmation)

CépiDc : Centre épidémiologique sur les causes médicales de Décès

CNIL : Commission Nationale de l'Informatique et des Libertés

DSET : Département Santé-Environnement-Travail

EHESP : Ecole des Hautes Etudes en Santé Publique

IGN : Institut national de l'informatique géographique et forestière

INSEE : Institut National de la Statistique et des Etudes Economiques

INSERM : Institut National de la Santé et de la Recherche Médicale

IRIS : Ilots regroupés pour l'Information Statistique

ISS : Inégalités Sociales de Santé

OMS : Organisation Mondiale de la Santé

SES : Statut Socio-Économique

SIG : Système d'Information Géographique

Introduction

Des Inégalités Sociales de santé (ISS) sont présentes dans la plupart des pays industrialisés. En plus des inégalités socio-économiques, l'environnement a récemment été rapporté comme pouvant lui aussi jouer un rôle dans les inégalité sociales de santé. Déterminer comment les expositions environnementales peuvent expliquer en partie les inégalités sociales de santé représente aujourd'hui un axe majeur de la Recherche en santé publique. Par ailleurs, la mortalité infantile est un excellent indicateur de l'état de santé d'une population [1,2]. C'est pour cette raison que l'on s'y intéresse ici.

Les décisions politiques jouent un rôle dans les ISS, en particulier celles modifiant les conditions environnementales (système de transport, génie urbain, etc.). Une littérature scientifique abondante documente le lien entre expositions environnementales et effets sanitaires. Une des vocations de la Recherche est donc d'aider les décideurs à mettre en place des politiques visant à réduire ces inégalités en identifiant les quartiers qui cumuleraient une vulnérabilité sociale et environnementale. En France, la diminution des inégalités sanitaires et environnementales constitue un des objectifs du deuxième Plan National Santé Environnement (PNSE 2).

C'est dans ce contexte qu'un groupe de chercheurs de l'EHESP a développé le projet Equit'area. Celui-ci, débuté en 2009, s'intéresse aux interactions entre inégalités sociales de santé et inégalités environnementales dans les quatre plus grandes agglomérations de France. Il s'intègre au groupe de travail « Inégalités environnementales » du PNSE 2.

Le présent mémoire a pour vocation de décrire, dans le cadre d'Equit'Area, le lien entre expositions environnementales, mortalité infantile et statut socio-économique dans l'agglomération de Marseille.

1 Cadre et Objectifs de l'étude

1.1 Le Projet Equit'Area

1.1.1 Contexte

L'étude de la distribution des expositions environnementales entre les populations de différents niveaux socio-économiques et démographiques a été initiée aux Etats-Unis et au Canada, suivis par le Royaume-Uni et la Suède. En France, peu d'études se sont penchées sur la question. *Laurian et al* [3] ont montré que les zones les plus pauvres étaient en moyenne exposées à de plus hauts niveaux de pollution. Par la suite, *Harvard et al* [4] ont mis en évidence que les quartiers de Strasbourg avec un niveau moyen de défaveur étaient plus exposés à la pollution de l'air que les quartiers plus privilégiés.

Une combinaison de deux types de « différentiels » est envisagée pour expliquer l'interaction entre les inégalités environnementales et les ISS :

- Le différentiel d'exposition : la différence est au niveau de la quantité d'exposition, celle-ci étant inégalement répartie parmi les groupes de population. L'hypothèse est que des groupes d'individus sont plus ou moins exposés aux nuisances environnementales selon leur niveau d'éducation, leur comportement, ou leur exposition dans certains « microenvironnements » (sur le lieu de travail, chez soi...). Les individus les plus défavorisés seraient donc plus exposés aux nuisances environnementales ;
- Le différentiel de vulnérabilité : il indique la variabilité de la réponse de l'organisme à l'exposition et de la vulnérabilité des individus. Ainsi, une même exposition à une nuisance environnementale pourrait avoir des effets sanitaires plus graves chez certains groupes socialement plus défavorisés.

Le fait que les groupes de population les plus défavorisés seraient plus exposés aux nuisances environnementales, résultant à des effets sanitaires plus importants, a fait l'objet de plusieurs études. Une majorité d'entre elles, mais pas toutes, a conclu que les concentrations de polluants avaient des niveaux plus élevés dans les quartiers où résident les populations les plus défavorisées que dans les quartiers des populations plus privilégiées.

Cependant, le différentiel de vulnérabilité des populations les plus défavorisées aux expositions environnementales est moins documenté.

Par ailleurs, de plus en plus d'études mettent en lumière un lien entre le taux de mortalité infantile et les expositions environnementales[5], d'une part, et le statut socio-économique d'autre part [6].

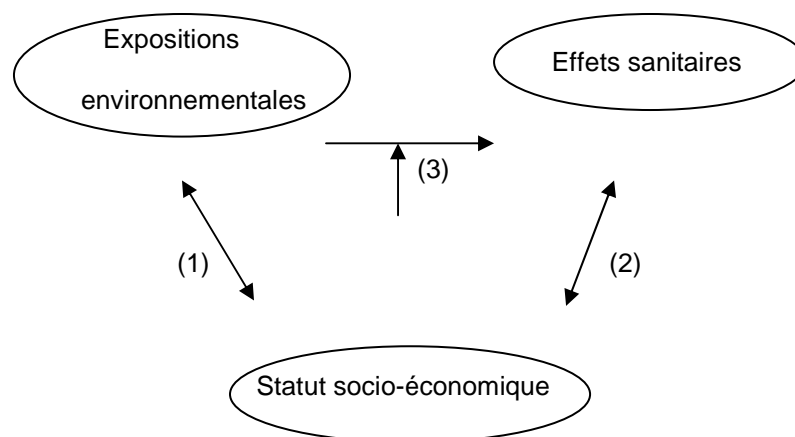
Ainsi, le projet Equit'Area vise à explorer les interactions entre trois familles de variables (environnementales, socio-économiques et événements sanitaires), qui ne sont encore que très rarement considérées conjointement dans la littérature. Aussi, l'évaluation de l'effet de l'interaction entre les ISS et les inégalités environnementales pourrait aider à mettre en place des politiques publiques visant à diminuer les risques et les inégalités.

1.1.2 Caractéristiques du projet

A) Structure du projet

Le projet est structuré en trois composantes (cf schéma ci dessous) :

- La caractérisation des inégalités environnementale (1),
- La caractérisation des inégalités sociales de santé (2),
- La contribution des expositions environnementales aux inégalités sociales de santé (3).



B) Schéma d'étude d'Equit'Area

Equit'Area est une étude écologique multicentre, réalisée à l'échelle de l'IRIS (Ilôts Regroupés pour l'Information Statistique). Il s'agit de la plus petite unité géographique française pour laquelle des données démographiques et socio-économiques sont disponibles. L'étude s'intéresse aux quatre plus grandes agglomérations de France : Paris, Marseille, Lyon et Lille. Un des intérêts de choisir plusieurs agglomérations est de pouvoir distinguer les caractéristiques socio-économiques retrouvées dans les quatre agglomérations et celles spécifiques à chaque métropole.

1.1.3 Les différentes variables

A) Les évènements sanitaires

La variable sanitaire choisie pour l'étude est la mortalité infantile, reconnue par l'Organisation Mondiale de la Santé comme un indicateur de l'état de santé d'une population et de l'efficacité du système de soins [1,2]. En effet, d'importantes disparités sont visibles, non seulement entre différents pays, mais également à des échelles plus fines, par exemple entre les régions. Par ailleurs, une récente hausse du taux de mortalité infantile a été remarquée. Depuis une cinquantaine d'années, ce taux avait diminué malgré la persistance des inégalités socio-économiques. Ce projet pourrait donc éclairer le rôle possible des expositions environnementales dans ce phénomène.

B) Les données environnementales

Les données environnementales prises en compte sont :

- la pollution atmosphérique (dioxyde d'azote NO₂). Le NO₂ est sélectionné car il constitue un bon traceur de la pollution atmosphérique générée par le trafic et présente une hétérogénéité spatiale supérieure à celles des autres polluants atmosphériques (les données sont recueillies auprès d'Atmo PACA),
- la présence et le nombre d'industries polluantes par IRIS et donc leur proximité par rapport aux cas,
- le bruit.

C) Les données socio-économiques

Les variables utilisées pour décrire le statut socio-économique sont extraites du recensement de l'Insee de 1999. L'indice de défaveur a été construit à l'aide d'une analyse en composantes principales (ACP) conduite sur les 52 variables initiales. La stratégie retenue pour restreindre la mesure du niveau socio-économique à un seul indicateur a été de maximiser l'inertie de la première composante principale. Ainsi, plusieurs ACP consécutives ont été conduites pour ne conserver à chaque fois que les variables les plus fortement corrélées à la première composante et contribuant le plus à sa construction. Finalement, pour chaque IRIS une valeur d'indice de défaveur est calculée à partir de la combinaison linéaire des 19 variables finales (présentées dans le Tableau 1) retenues sur la première composante principale.

1.2 Présentation de l'étude

Le présent mémoire s'intéresse à l'une des quatre zones du projet Equit'Area, à savoir l'agglomération de Marseille.

1.2.1 Objectifs

Les objectifs principaux de ce mémoire sont, d'une part, d'étudier la distribution spatiale de la mortalité infantile dans l'agglomération de Marseille à l'échelle de l'IRIS (plus petite unité géographique utilisée par l'INSEE) et, d'autre part, d'analyser la relation de l'événement sanitaire avec le niveau socio-économique, ainsi que l'impact de la présence ou non d'industries polluantes sur les ISS.

1.2.2 Déroulement

L'étude se structure en trois étapes principales.

La première étape correspond au recueil de données sanitaires. Il s'agit alors d'obtenir, auprès des services d'état civil des mairies de l'agglomération, les données de mortalité infantiles qui seront exploitées.

La deuxième étape consiste à géocoder les données sanitaires, c'est-à-dire leur attribuer un jeu de coordonnées géographiques. Ceci permet de déterminer dans quel IRIS ces données se situent, les agrégeant ainsi à la même échelle que les autres données à traiter (pollution atmosphérique, indice socio-économique).

Enfin, la dernière étape est l'analyse épidémiologique des données sanitaires, recueillies lors de la première étape, et des données environnementales et socio-économiques disponibles pour l'agglomération de Marseille.

2 Méthodologie

2.1 Recueil des données sanitaires

Les données sanitaires exploitées dans la présente étude sont les taux de mortalité infantile de chaque IRIS de l'unité urbaine Marseille Aix-en-Provence. Il s'agit du « rapport entre le nombre d'enfants décédés à moins d'un an et l'ensemble des enfants nés vivants » (Insee). Cette partie de l'étude consiste donc à recueillir les données de naissances et les données de décès pour chaque IRIS de l'unité urbaine sur la période 2000-2009.

2.1.1 Naissances

Les données de naissances par IRIS ont été partiellement fournies par l'Insee. Ainsi, les bases de données fournies par l'institut comprennent le nombre de naissances domiciliées par IRIS pour les années 1999, 2005, 2006, 2007, 2008 et 2009. Une autre base de données contient le nombre de naissances domiciliées à l'échelle de la commune pour chaque année sur la période 1999-2000.

Une estimation des naissances par IRIS est donc nécessaire pour les années 2000, 2001, 2002, 2003 et 2004. Elle est réalisée de la manière suivante : à partir de la base de données des naissances de 1999, une proportion est calculée pour chaque IRIS. On émet l'hypothèse que la proportion est sensiblement la même d'une année à l'autre. Ainsi, pour les années qui s'étendent sur la période 2000-2004, le nombre de naissances par commune est, pour chaque IRIS, multiplié par la proportion calculée à partir des naissances de 1999.

2.1.2 Décès

Le CépiDC de l'Inserm permet de disposer du nombre de décès à moins d'un an pour chaque commune de la zone d'intérêt.

L'étude étant réalisée au niveau de l'IRIS, un recueil à échelle plus fine est nécessaire. Le seul moyen d'obtenir l'information à cette échelle est de recueillir l'adresse de domicile de chaque cas figurant sur l'acte de décès, puis agréger les cas par IRIS. Les actes de décès sont émis et conservés au service d'état civil dans la commune de décès, et une retranscription de l'acte est également réalisée dans la commune de domicile du défunt.

Dans un premier temps, les services d'état civil des municipalités de l'unité urbaine Marseille-Aix-en-Provence sont contactés par téléphone, afin de leur présenter le projet, l'intérêt de leur participation, et de connaître la forme sous laquelle sont conservées les données. Cette dernière information est utile pour organiser les déplacements sur le terrain. En effet, des données conservées sous base de données Excel ont l'avantage de pouvoir être directement envoyées par courriel, tandis que les données présentes uniquement sur les actes de décès papier impliquent non seulement un déplacement dans la commune concernée pour avoir accès aux registres d'état civil, mais aussi plusieurs heures (voire jours) de présence physique au sein du service pour recueillir les données manuellement.

Une fois cette prise de contact téléphonique réalisée, est envoyé un courrier signé de Denis Zmirou, directeur du Département Santé-Environnement-Travail de l'EHESP, présentant les objectifs du projet ainsi que la liste des données nécessaires, à savoir :

- La date de naissance,
- La date de décès,
- Le sexe,
- La profession du père et de la mère,
- L'adresse de domicile du défunt.

Ce courrier est accompagné d'une copie de l'accord de la CNIL pour l'accès aux informations.

L'étape suivante consiste à collecter les données proprement dites. Que ce soit par courriel ou sur le terrain, une phase de relance téléphonique est nécessaire. Il s'agit de s'assurer que le courrier est bien reçu et que le responsable a donné son accord. Les données sont rarement informatisées. Sur le terrain, les données sont alors collectées de la manière suivante : Les services des municipalités comptant au moins un hôpital sont visités en premier. Est alors obtenu, théoriquement, la totalité des données pour la commune visitée ainsi qu'un certain nombre de cas domiciliés dans d'autres communes de la zone d'étude. Grâce au CépiDC, le nombre total de cas à recueillir par commune est connu. Il est donc possible de savoir dans quelles autres communes il est le plus judicieux de se rendre, selon le nombre de cas déjà collectés et le nombre de cas restant à obtenir pour atteindre le nombre de cas total recensé par le CépiDC. Cette méthode évite de perdre du temps à se

rendre dans des communes pour lesquelles la totalité des données est déjà obtenue grâce au service d'état civil d'une municipalité voisine.

En pratique, l'accès aux données est très délicat à obtenir par le responsable du service d'état civil, même en possédant l'accord de la CNIL, et certaines données sont manquantes pour cette raison. De plus, les retranscriptions d'actes dans les communes de domicile sont en réalité très rarement faites. Ainsi, dans les cas où l'enfant n'est décédé ni dans sa ville de résidence, ni dans une autre ville de la zone d'étude, les données ne peuvent pas être récupérées et sont alors considérées comme manquantes.

Par ailleurs, les actes sont souvent incomplets, et la profession des parents est rarement précisée. Il est à noter que ces difficultés administratives sont rencontrées uniquement pour l'unité urbaine de Marseille-Aix-en-Provence, les données pour le Grand Lyon et Lille Métropole étant plus complètes et accessibles.

Pour finaliser l'étape de recueil des données, ces dernières sont rassemblées en tableau de données sous *Excel*.

2.2 Agrégation des cas à l'échelle de l'IRIS

Il s'agit, à ce stade, de déterminer le nombre d'évènements sanitaires survenus dans chaque IRIS de l'unité urbaine Marseille-Aix-en-Provence sur la période 2000-2009. La démarche se décompose en 2 principales étapes : le géocodage de chaque cas et le placement de ces cas sur un fond de carte.

2.2.1 Géocodage

Le géocodage consiste à retrouver la localisation géographique, en longitude (coordonnée X) et latitude (coordonnée Y), d'une adresse donnée.

L'étape du géocodage est réalisée avec la page web *Batch géocodeur*, qui utilise l'interface de programmation *Google Maps API*. Les adresses sont tout d'abord extraites de leur fichier *Excel*, qui est compatible avec le géocodeur. Ce dernier permet ensuite d'obtenir les données géocodées pour les cartographier et réaliser des analyses spatiales avec le logiciel *Arcview*.

Les données de géocodage comportent différentes informations, dont :

- la latitude et la longitude du point trouvé, en degrés décimaux (système de coordonnées WGS84 – EPSG 4326),
- la pertinence du positionnement trouvé :
 - 1 si seul le pays est localisé,
 - 2 si la région est localisée,
 - 3 si le département est localisé,
 - 4 si la ville est localisée,
 - 5 si le code postal est localisé,
 - 6 si la rue est localisée,
 - 7 si l'intersection est localisée,
 - 8 si l'adresse complète est localisée.

La précision du positionnement dépend de la manière dont est renseignée l'adresse à géocoder, ainsi que l'exhaustivité des données géographiques exploitées (dans ce cas celles de *Google Maps*). L'étude étant réalisée à l'échelle de l'IRIS, les données ne sont retenues que lorsque la pertinence est au moins égale à 6. Or, en utilisant les adresses « brutes » collectées à partir des registres d'actes de décès, moins du tiers des positionnements trouvés sont suffisamment pertinents. Un travail de standardisation des adresses est donc indispensable.

En effet, les fautes d'orthographe pour le nom des rues ou résidences (e.g « Les Mordelles » au lieu des « Mauredelles ») sont relativement fréquentes et empêchent le géocodeur de trouver l'adresse. Par ailleurs, de nombreux cas avaient pour adresse uniquement un nom de résidence ou de lotissement. Il faut donc retrouver le nom de la rue où se situe cette résidence ou ce lotissement. Parfois, le nom de la rue utilisé par Google Maps est quelque peu différent du nom sous lequel la rue est connue par les habitants de la commune. Ainsi, la « route Nationale de la Viste » devient dans Google Maps la « Rue Nationale ». Les adresses non trouvées ou dont la pertinence du positionnement est insuffisante sont donc retravaillées au cas par cas, jusqu'à obtenir un score minimal de 6.

2.2.2 Cartographie

Le but de cette étape est de faire apparaître les différents cas précédemment dotés de coordonnées (X,Y) sur des fonds de carte représentant les contours d'IRIS de la zone d'étude.

La cartographie est réalisée à l'aide du logiciel *Arcview*, qui est l'une des composantes du pack *ArcGIS* de la société Esri (Environmental Systems Research Institute). Il s'agit d'un logiciel de Système d'Information Géographique (SIG) permettant de visualiser, de gérer, de créer et d'analyser des données géographiques.

Les *contours...Iris* ® sont le fruit d'une coédition entre l'Insee et l'IGN. Il s'agit d'un fond numérisé des Iris définis par l'Insee.

Ces contours sont importés dans *Arcview*, de même que toutes les coordonnées obtenues grâce au géocodage. Sont alors obtenus deux plans indépendants superposés, qu'il suffit de joindre l'un à l'autre dans le menu « outils d'analyse », pour que le logiciel puisse attribuer à chaque cas un IRIS. A l'inverse, et c'est le but de la tâche ici, une autre manipulation simple permet d'obtenir un tableau récapitulatif le nombre de cas reportés dans chaque IRIS.

2.3 Bilan des données disponibles et modification de la zone d'étude

Au moment de la réalisation de ce travail, certaines données ne sont pas disponibles pour toutes les communes de la zone d'étude, ni pour les dix ans de la période considérée. Ainsi, la zone d'étude est réduite à l'ensemble des communes pour lesquelles toutes les données ont pu être recueillies. Il est donc nécessaire de faire un bilan des données disponibles :

- L'indice socio-économique est construit pour toutes les communes de l'unité urbaine de Marseille Aix-en-Provence,
- Les données de pollution ne sont disponibles que pour la commune de Marseille et quelques communes de la Communauté du Pays d'Aix, pour la période s'étendant de 2002 à 2009,
- Les données sanitaires sont disponibles pour la période 2000-2009 pour l'unité urbaine Marseille-Aix-en-Provence, à l'exception de la ville d'Aix-en-Provence. Les seuls cas obtenus pour cette ville proviennent des services d'état civil des

communes alentour. Les données manquantes étant trop importantes, Aix en Provence est exclue de la zone d'étude.

Certains IRIS pour lesquels les données sanitaires et socio-économiques sont disponibles ne disposent pas de données de pollution. Après renseignement pris auprès d'Atmo PACA, ces IRIS ne font pas partie du réseau de mesure de pollution atmosphérique. Or parmi ceux-ci, certains forment un îlot en périphérie Nord-Est de la commune de Marseille et appartiennent au 15^e arrondissement. Ce dernier est l'arrondissement qui comporte le plus de cas de mortalité infantile relevés par le CépiDC, et aussi le plus de données sanitaires manquantes (50 cas recueillis auprès du service d'état civil, alors que le CépiDC en relève 70). Ainsi, plutôt que de supprimer ces IRIS et perdre encore quelques cas pour cet arrondissement, il est choisi d'estimer leur concentration de NO₂ en calculant la moyenne des concentrations de NO₂ des IRIS voisins. Une vérification de la répartition des concentrations de NO₂ sur les modélisations permet d'évaluer grossièrement la qualité de l'estimation. Ainsi, les concentrations de NO₂ estimées semblent bien être du même ordre de grandeur que les concentrations des IRIS voisins, sauf pour un IRIS du quinzième arrondissement au milieu duquel passe une autoroute. Dans ce cas, la concentration est probablement sous-estimée.

Par ailleurs, 17 IRIS sont supprimés par manque de renseignements concernant la situation socio-économique.

Seulement trois zones d'industries sont recensées dans la zone sélectionnée, ce qui n'est pas suffisant pour réaliser des analyses. Elles ne sont donc pas prises en compte dans la suite de ce mémoire.

Au final, 26 communes (chaque arrondissement de Marseille étant considéré comme une commune) et 423 IRIS constituent la zone d'étude pour le travail présenté ici. La période d'étude est quant à elle réduite à la période 2002-2009, puisque les données de NO₂ ne sont pas disponibles pour les années précédentes.

2.4 Analyses statistiques

2.4.1 Statistiques descriptives

La première étape de l'analyse statistique consiste à réaliser l'analyse descriptive des variables. Il s'agit donc de décrire les différentes familles de variables dont on dispose :

- NO₂,
- Mortalité infantile,
- Caractéristiques socio-économiques.

A) Résumé numérique

Le résumé numérique des variables de mortalité infantile et socio-économique est réalisé à l'aide du logiciel R.

B) Représentations graphiques

Deux types de graphiques sont effectués pour décrire les variables.

Tout d'abord, il est choisi de représenter l'évolution de la moyenne des concentrations de NO₂ sur la période de 2002 à 2009. En considérant que les concentrations évoluent de façon similaire dans les différents IRIS, l'évolution globale sur huit ans des concentrations de NO₂ couvrant la zone d'étude est représentée. Pour ce faire, un graphe avec une « boîte à moustaches » pour chaque année est réalisé, à l'aide du logiciel R, sous la formule suivante.

L'évolution est représentée sous forme de boîtes à moustaches, puisqu'elles ont l'avantage de résumer simplement les principales caractéristiques de position (médiane, quartiles, extrêmes).

C) Représentations cartographiques

Pour un meilleur aperçu de la répartition spatiale des données, il est choisi de représenter chacune des trois variables à l'aide de cartes. Pour ce faire, le logiciel ArcGIS est utilisé. Les cartes sont réalisées de la manière suivante. ArciGIS fonctionne avec un système de « couches ». Dans le cas présenté ici, la première « couche » correspond au fond de carte avec les contours IRIS. La deuxième couche correspond à la base de données préalablement importée au format Excel (les dernières versions d'ArcGIS sont capables

d'importer ce format, ce qui évite les problèmes de conversions de fichiers). La couche d'IRIS contient une table attributaire avec, entre autres les colonnes suivantes :

- « FID », qui correspond à un numéro d'identifiant,
- « Shape », qui correspond à la forme de l'IRIS (il s'agit ici pour tous les IRIS de polygones),
- DCOMIRIS, qui correspond au numéro complet de l'IRIS (code de la commune selon l'Insee + code de l'IRIS lui-même),
- « NOM COM », qui correspond au nom de la commune.

La deuxième couche, la base de données réalisée dans Excel, contient également une colonne équivalente à « DCOMIRIS », c'est-à-dire les numéros d'IRIS, ainsi que les colonnes correspondant aux valeurs de chaque variable pour chaque IRIS.

Les deux couches sont, à ce stade, indépendantes. L'une est une carte, l'autre une base de données. On cherche à représenter le contenu de la base de données sur la carte. Il est donc nécessaire de joindre les deux couches, la clé de jointure étant le numéro d'IRIS (« DCOMIRIS »). Cette manipulation se réalise grâce à la fonction « joindre » du logiciel. En pratique, ArcGIS étant très instable, en particulier lorsqu'il s'agit de lire le format Excel, le format des données peut poser problème. En effet, pour pouvoir joindre les deux couches en fonction de la colonne DCOMIRIS, celle-ci doit être présentée sous le même format dans la table attributaire et dans la base de données, à savoir le format texte. Cependant, dans le cas présent, les valeurs de cette colonne sont bien au format texte lorsqu'elles sont lues sous Excel, mais sont reconnues comme étant au format numérique sous ArcGIS, empêchant de réaliser la jointure entre les deux couches. La solution finalement trouvée est d'importer le fichier Excel vers le logiciel de gestion de bases de données Access, puis d'y modifier le format de la colonne et de sauvegarder le fichier modifié sous Excel. Les données sont, cette fois-ci, correctement reconnues par ArcGIS comme étant du texte, et la jointure des deux couches est alors possible.

L'onglet « symbologie » du menu « propriétés » de la couche « IRIS » permet ensuite de choisir de représenter, dans le cas présent, des quantités, en utilisant un dégradé de gris. Des menus déroulants permettent de sélectionner la variable visible souhaitée : taux de mortalité infantile, moyenne NO₂, Indice SES. Il est possible de choisir le nombre de classes représentées, ici cinq, pour les 3 variables.

2.4.2 L'Analyse bayésienne

L'approche bayésienne est un moyen de tenir compte de l'autocorrélation spatiale lors de l'analyse statistique des données.

A) Le concept d'autocorrélation spatiale

C'est un phénomène fréquent dans les études de type écologique. Il s'agit de l'estimation de la corrélation d'une variable en référence à sa localisation dans l'espace. L'autocorrélation spatiale mesure le degré d'interaction et d'interdépendance entre les observations spatialement localisées. En effet, la représentation géographique d'observations n'est généralement pas aléatoire et répond à des schémas de répartition définis. Dans le cas présent, ce qui est observable dans un IRIS particulier dépend de ce qui est observable dans les IRIS voisins. Ainsi, si l'autocorrélation spatiale est positive, les valeurs similaires d'une variable considérée sont regroupées, de sorte que les IRIS voisins se ressemblent davantage que les IRIS éloignés. Plusieurs outils statistiques existent pour mesurer l'autocorrélation spatiale. L'indice de Moran est celui considéré comme le plus robuste et le plus performant pour estimer le degré de dépendance entre des observations localisées spatialement.

B) L'analyse bayésienne

a) *La matrice de contiguïté*

Cette matrice définit la dépendance spatiale entre plusieurs IRIS selon la notion de contiguïté, c'est-à-dire sur le fait de partager un point commun. Dans le cas présent, il est considéré que les IRIS partagent au moins une frontière en commun. On définit les voisins d'ordre 1 comme les IRIS partageant une frontière en commun avec l'IRIS de référence. Les voisins d'ordre 2 représentent quant à eux les voisins d'ordre 1, auxquels s'ajoutent leurs propres voisins d'ordre 1, et ainsi de suite. Ce type de matrice permet de sélectionner des voisins directement en rapport avec l'unité de référence. Elle ne tient cependant pas compte des distances dans la détermination du voisinage.

La réalisation de la matrice nécessite la création d'un fichier .shp (« shapefile ») comportant uniquement les IRIS sélectionnés pour l'analyse. Ce fichier est créé à l'aide du

logiciel Arcview. Le fichier .shp des contours IRIS du département des Bouches du Rhône dont nous disposons y est donc transformé en utilisant une requête en langage SQL. Une nouvelle couche « sur-mesure » est ainsi créée en ne sélectionnant que les IRIS d'intérêt. Ce nouveau fichier .shp sert ensuite de base pour créer la matrice de contiguïté réalisée grâce au logiciel R. Cette matrice est réalisée au format .txt afin d'être utilisée ensuite par le logiciel WinBUGS.

b) *Le Modèle bayésien utilisé*

L'association entre le niveau de pollution, le niveau socio-économique et la mortalité infantile est estimée à l'aide d'un modèle hiérarchique bayésien proposé par *Besag et al* en 1991 (connu comme le BYM model) [7,8]. Ce modèle permet de contrôler à la fois l'hétérogénéité (sans structure spatiale) et l'autocorrélation spatiale des données en introduisant des composantes aléatoires.

Le modèle BYM est un modèle gaussien autorégressif conditionnel structuré selon trois niveaux hiérarchiques.

Le premier niveau suppose que le nombre d'événements sanitaires observés O_i dans l'IRIS i (dans le cas présent, i est compris entre 1 et 430, c'est-à-dire le nombre d'IRIS dans la zone d'étude), conditionnellement au risque relatif de survenue de l'événement θ_i , suit une loi de Poisson d'espérance $E_i\theta_i$, où E_i est le nombre d'événements attendus dans l'IRIS i (équation 1). La loi de Poisson est pertinente pour décrire des événements, en particulier rares (tels que la mortalité infantile) sur des intervalles de type spatial.

$$O_i \sim \text{Poisson}(E_i\theta_i) \quad (\text{Equation 1})$$

Le deuxième niveau associe au logarithme du risque relatif θ_i une distribution de probabilité définie comme la somme d'un paramètre constant α , de la ou des variable(s) explicative(s) X_k et de deux composantes aléatoires indépendantes u_i et v_i . Ces deux composantes tiennent compte respectivement de la variabilité structurée spatialement (composante de clustering) et de la variabilité non structurée (composante d'hétérogénéité) (équation 2).

$$\text{Log}(\theta_i) = \alpha + X_{ik}\beta_k + u_i + v_i \quad (\text{Equation 2})$$

où α est donc une constante, X_k sont les k variables auxquelles sont associés les coefficients de régression β_k , et u_i et v_i correspondent aux composantes de clustering et d'hétérogénéité respectivement.

On considère que u_i est distribuée selon une loi normale conditionnelle autorégressive et que v_i suit une loi normale (équation 3).

$$\begin{aligned} u_i / u_j &\sim N(\bar{u}_i, \sigma_u^2 / m_i) \\ v_i &\sim N(0, \sigma_v^2) \end{aligned} \quad (\text{Equation 3})$$

où u_j correspond aux effets spatiaux pour tous les IRIS autres que l'IRIS i , u_i la moyenne de pour tous les IRIS voisins de l'IRIS i , m_i le nombre de voisins de l'IRIS i , et les variances respectives des distributions des composantes u_i et v_i . La définition du voisinage utilisé pour ce travail repose sur une matrice de contiguïté nodale du 1er ordre comme celle présentée dans le paragraphe précédent.

Le troisième niveau associe à chacun des paramètres des distributions appelées « distributions hyper a priori ». Ainsi, on considère qu' α est distribué selon une loi normale avec une variance très importante, les coefficients de régression β_k sont considérés comme distribués selon une loi normale de variance très importante, et les paramètres de précision (qui sont les inverses des variances) $1 / \sigma_u^2$ et $1 / \sigma_v^2$ sont distribués selon des distributions Gamma (équation 4).

$$\begin{aligned} \alpha &\sim N(0, 10000) \\ \beta_k &\sim N(0, 10000) \\ 1 / \sigma_u^2 &\sim Ga(0.01, 0.01) \\ 1 / \sigma_v^2 &\sim Ga(0.01, 0.01) \end{aligned} \quad (\text{Equation 4})$$

Les paramètres des distributions gamma ont été sélectionnés après analyse de sensibilité utilisant le DIC ((Deviance Information Criterion) développé par Spiegelhalter [9]).

Ce modèle est décliné en trois modèles différents réalisés sous le logiciel WinBUGS. Le premier modèle permet de calculer l'effet du niveau de défaveur sur la mortalité infantile. Pour ce faire, la variable socio-économique est divisée, à l'aide du logiciel R, en cinq classes de défaveur, correspondant aux quintiles. La classe 1 (niveau de défaveur le plus

faible) est choisie comme classe de référence, et chaque paramètre s'interprète par rapport à cette classe de référence.

Le premier modèle permet donc d'obtenir le risque relatif de survenue d'événement sanitaire pour chaque IRIS ainsi que pour chaque classe de défaveur par rapport à la classe de référence.

Le deuxième modèle évalue l'effet du NO₂ sur la mortalité infantile. De la même manière que pour le premier modèle la variable de NO₂ est divisée en 5 classes, la classe 1 étant celle où l'exposition est la plus faible. Elle est choisie comme classe de référence.

Enfin, le troisième modèle calcule le risque relatif de mortalité infantile selon la classe de NO₂, ajusté sur le niveau de défaveur.

Chacun de ces modèles est mis en oeuvre avec et sans contrôle de l'autocorrélation spatiale. Pour ne pas prendre en compte l'autocorrélation spatiale, il suffit de lancer le modèle après avoir supprimé u_i et v_i , qui caractérisent ce phénomène.

c) *Application avec WinBUGS*

Le modèle compare le nombre de cas de mortalité infantile observés avec le nombre de cas attendus dans chaque IRIS. Le nombre de cas attendus correspond au nombre de cas présent dans chaque IRIS si le taux de mortalité infantile était identique dans tous les quartiers. Pour pouvoir appliquer le modèle, ce nombre de cas attendus est donc obtenu en multipliant le taux de mortalité infantile (2,6 pour 1000 naissances vivantes) par le nombre de naissances (ou l'estimation du nombre de naissances) dans l'IRIS.

L'estimation des paramètres du modèle bayésien utilisé fait appel à des intégrales qui ne peuvent pas être calculées via des méthodes analytiques. Des méthodes de simulation (algorithme de Monte Carlo par chaînes de Markov) sont donc utilisées. Ce travail est effectué à l'aide de deux chaînes parallèles indépendantes de 30000 itérations chacune. En réalité, 60000 itérations sont lancées, et seules les 30000 dernières sont retenues, les 30000 premières correspondant au « burn-in ». Cette étape de « préchauffe » est nécessaire pour éviter les problèmes de convergence.

La qualité des modèles est évaluée par le DIC (plus le DIC est faible, meilleur est le modèle).

3 Résultats

Pour les 26 communes retenues, sur la période s'étendant de 2002 à 2009, le nombre de naissances est estimé à 102 756. Sur cette même période, le nombre de décès d'enfants de moins d'un an est de 365 d'après le CépiDC. Parmi ces 365 cas, 308 sont collectés auprès du service d'Etat civil des communes concernées. Le taux d'exhaustivité atteint ainsi 84,4%. Les données manquantes concernent plusieurs communes, notamment le 15^e arrondissement de la ville de Marseille, qui compte 19,3% des données manquantes.

Sur la période 2002-2009, la moyenne du taux de mortalité infantile est de 2,6 pour 1000 naissances vivantes, avec un écart-type de 0,0039 et atteint un maximum de 24 pour 1000 naissances vivantes.

La Figure 1 montre les variations spatiales des taux de mortalité infantile des différents IRIS de la zone d'étude. Les taux de mortalité infantile les plus élevés correspondent aux IRIS occupant une large partie Nord-Ouest de la zone d'étude.

La Figure 2 est un graphe montrant les variations des moyennes des concentrations de NO₂ des différents IRIS sur la période allant de 2002 à 2009. Globalement, les valeurs n'ont pas beaucoup évolué sur 8 ans, la majorité des concentrations se situant entre 25 et 42 µg/m³. Des augmentations globales sont cependant remarquables en 2003 et en 2005, atteignant des valeurs maximales de 67,1 µg/m³ et 63,8 µg/m³, respectivement.

Des concentrations globalement plus élevées concernent les IRIS de la moitié Nord-Ouest de la ville de Marseille ainsi que les IRIS situés plus haut au Nord-Ouest à l'extérieur de Marseille, englobant la commune de Vitrolles. Un noyau de concentrations très élevées est remarquable dans les premiers arrondissements de Marseille, en particulier à l'approche du port et des docks. Plus les IRIS sont éloignés de ce noyau, plus les concentrations diminuent. Cependant, une ligne de concentrations plus élevées est également notable sur l'axe Est-Ouest de Marseille, qui correspond au passage de l'autoroute A50. En effet, le NO₂ a pour sources principales les trafics routier, maritime et aérien. Ainsi, les concentrations élevées au niveau de la commune de Vitrolles peuvent s'expliquer par le passage de l'autoroute A7 et de la présence de l'aéroport Marseille-Marignane à l'Ouest.

Les concentrations élevées au niveau de la partie Nord Ouest de Marseille peuvent, quant à elles, être expliquées par la présence des autoroutes A7 et A55, d'une part, et par le trafic maritime côtier à l'approche du port de Marseille, d'autre part.

Le Tableau 1 résume les principales caractéristiques socio-économiques de la population étudiée. De larges écarts sont observables. La Figure 3 montre la distribution spatiale de l'indice de défaveur, et d'importants contrastes sont visibles. Ainsi, les IRIS pour lesquels l'indice de défaveur est le plus élevé occupent majoritairement une grande part Nord-Ouest de la ville de Marseille, avec un noyau de valeurs très élevées au niveau des 1^{er}, 2^e, 3^e, 14^e, 15^e et 16^e arrondissements. La commune de Vitrolles (au Nord-Ouest de Marseille) présente également des valeurs moyennes voire élevées. Les valeurs les plus faibles se retrouvent en périphérie côté Est et côté Nord-Est de Marseille, ainsi qu'au Sud, où les valeurs sont faibles à moyennes. Des valeurs très faibles sont également observées dans les communes voisines au Nord-Est de Marseille (Simiane-Collongue, Fuveau, Meyreuil, Mimet, Venelles, Le Tholonet, etc.). Les quartiers les plus défavorisés présentent des taux élevés de chômage, de population étrangère, de familles monoparentales, de personnes n'ayant pas de diplôme. Les IRIS les plus privilégiés, quant à eux, présentent un nombre plus important de personnes résidant en maisons, étant propriétaires, possédant deux voitures.

Le Tableau 2 présente les risques relatifs de mortalité infantile par classe de défaveur ainsi que leur intervalle de confiance à 95%, et la figure 5 montre la distribution spatiale des risques relatifs de chaque IRIS de la zone d'étude retenue. Pour chaque risque relatif présenté, la classe de référence est la classe pour laquelle le niveau de défaveur est le plus faible (C1). Les résultats montrent qu'il existe un gradient de mortalité infantile selon le niveau socio-économique. Pour toutes les classes et quelque soit le modèle utilisé (avec ou sans prise en compte de l'autocorrélation spatiale), le risque de mortalité infantile est significatif pour toutes les classes de défaveur et dès la deuxième classe. Il augmente avec le niveau de défaveur jusqu'à obtenir un risque près de trois fois plus important pour les quartiers les plus défavorisés comparativement aux quartiers les plus favorisés. Le contrôle de l'autocorrélation spatiale diminue sensiblement l'association entre le taux de mortalité infantile et le niveau de défaveur, mais améliore la qualité du modèle. De plus, même si le gradient est diminué, les risques gardent la même tendance et restent significatifs.

Le tableau 4 présente les risques relatifs de mortalité infantile par classe de NO₂, en référence à la classe pour laquelle les concentrations de NO₂ sont les plus faibles. La figure 6 présente la distribution des risques relatifs de tous les IRIS de la zone d'étude. Les risques relatifs sont statistiquement significatifs pour toutes les classes lorsqu'il n'y a pas de contrôle de l'autocorrélation spatiale. Lorsque celle-ci est contrôlée, les risques relatifs diminuent légèrement et celui de la classe 4 par rapport à la classe 1 n'est plus significatif, mais présente un intervalle de confiance « borderline » (RR = 1,78 [IC 95% = 0,98 ; 3,31]). La relation entre la mortalité infantile et les concentrations de NO₂, bien que significative, n'est pas monotone et linéaire. Les risques relatifs ont tendance à diminuer quand le niveau de NO₂ augmente, sauf pour la classe la plus exposée, où le risque de mortalité infantile est beaucoup plus important (RR = 3,96 [IC 95% = 1,85 ; 8,50] dans les zones où les concentrations de NO₂ sont les plus élevées comparativement aux quartiers où les concentrations sont les plus faibles). Cette fois encore, le contrôle de l'autocorrélation spatiale diminue l'association entre les deux variables mais améliore la qualité du modèle.

Le tableau 5 présente les risques relatifs de mortalité infantile par classe de NO₂, après ajustement sur le niveau de défaveur, avec et sans contrôle de l'autocorrélation spatiale. Il existe toujours un gradient d'exposition. Les risques relatifs sont diminués après ajustement, mais gardent la même tendance et restent significatifs pour les classes 2, 3 et 5. Le contrôle de l'autocorrélation diminue dans ce cas encore l'intensité du gradient, et les risques relatifs restent significatifs, sauf pour la classe quatre. Même après ajustement sur l'indice de défaveur, les quartiers les plus exposés à la pollution au NO₂ ont encore un risque de mortalité infantile jusqu'à trois fois plus important que dans les quartiers les moins exposés.

4 Discussion

En utilisant des données agrégées à l'échelle de l'IRIS et une analyse statistique appropriée, cette étude met en évidence l'existence d'un gradient socio-économique significatif. Ce résultat est cohérent avec d'autres publications. En effet, différentes études se sont penchées sur les inégalités de santé socio-spatiales et les issues de grossesses. Guildea, au Pays de Galles, a montré que la mortalité post-néonatale (décès survenant entre 24 semaines de gestation et un an après la naissance) est fortement associée à l'indice de défaveur de Townsend avec un risque 1,53 fois plus important dans les quartiers les plus pauvres par rapport aux quartiers les plus privilégiés [10]. Une étude américaine conduite sur une longue période (1969 – 2001) a montré un gradient significatif de mortalité infantile selon le niveau de défaveur à l'échelle du quartier [11]. Les comportements individuels et l'éducation sanitaire jouent un rôle important dans les inégalités sociales de santé, et en particulier lorsqu'il s'agit de mortalité infantile. Un régime alimentaire insuffisamment varié pendant la grossesse, le fait de fumer ou de boire de l'alcool, et le manque de soins pendant la période prénatale sont des facteurs fortement associés à la mortalité infantile [10,11] et varient significativement selon le niveau socio-économique du quartier de résidence et le niveau d'éducation de la mère [12]. Par exemple, en France, 21% des femmes fument et ce pourcentage tend à augmenter avec le niveau de défaveur du quartier de résidence [13]. La quantité et la qualité de soins pendant la période prénatale sont aussi associées au statut socio-économique. L'association entre le commencement tardif des soins pendant la période prénatale et naissance prématurée est bien documentée [14]. Le statut socio-économique peut également être perçu comme un déterminant indirect de mortalité infantile. En générant des situations de stress, le niveau de défaveur peut diminuer le temps de gestation et constituer un facteur de risque de mort prématurée. Une étude américaine montre une forte association entre le stress de la mère pendant la grossesse et le faible poids de naissance ainsi que la mortalité. Cette étude prend en compte des variables socio-économiques telles que les revenus, le nombre de personnes habitant dans le même logement et le niveau d'éducation du quartier de résidence [15].

Etant donné les données disponibles au moment de la réalisation de ce travail, l'indice de défaveur est construit à partir des données provenant du recensement de la population française en 1999.

Cette étude étant de type écologique, les résultats obtenus à l'échelle de l'IRIS ne peuvent pas être extrapolés à l'échelle de l'individu. Il n'y a donc pas d'information concernant les comportements individuels (mère qui fume pendant la grossesse, consommation de fruits et légumes, activité physique, etc.) souvent utilisés dans la littérature. Cependant, l'indice de défaveur est robuste et validé. Il est en effet constitué de variables prenant en compte tous les aspects de la notion de défaveur (incluant l'activité professionnelle, le niveau d'éducation, etc.) généralement cités dans la littérature. L'IRIS est la plus fine échelle géographique disponible pour la France (regroupant environ 2000 habitants). Ainsi, l'indice de défaveur de chaque IRIS peut donner une bonne idée des caractéristiques socio-économiques des individus résidant dans ces quartiers.

L'étude montre également l'existence d'un gradient de mortalité selon l'exposition au NO₂ qui est significatif. Ce résultat concorde avec ce qui est publié dans la littérature. De plus en plus d'études démontrent l'existence d'un lien entre le taux de mortalité infantile et les expositions environnementales, en particulier l'exposition à la pollution de l'air. Une synthèse de la littérature relative à l'exposition aux particules dans l'air et à la mortalité infantile publiée en 2004 souligne un accroissement significatif du risque de mortalité infantile et de mortalité infantile tardive à des niveaux élevés de particules dans l'air ambiant [16].

Ce travail montre également que le gradient d'exposition au NO₂ existe toujours après ajustement sur le niveau de défaveur, et que la relation entre la mortalité infantile et l'exposition au NO₂ garde la même tendance et la même significativité. D'autres publications ont montré un lien entre la pollution de l'air et la mortalité infantile. Une récente étude conduite à Séoul a étudié la modification de l'effet des PM₁₀ sur les accouchements prématurés en fonction du niveau socio-économique, particulièrement sur les mères résidant dans des zones géographiques à faible SES [17].

Dans le cas présent, les risques relatifs de mortalité infantile ont tendance à diminuer quand l'exposition au NO₂ augmente, sauf dans la classe pour laquelle l'exposition est la plus importante. La quatrième classe d'exposition présente un risque relatif non significatif. Une des hypothèses avancées pour expliquer cette tendance pourrait être l'existence d'une interaction entre le niveau de défaveur et l'exposition au NO₂. Une prochaine étape peut donc être de réaliser un modèle comprenant un terme d'interaction entre le NO₂ et l'indice de défaveur. Le choix de créer cinq classes d'exposition en utilisant les quintiles est une manière simple de transformer une variable quantitative en variable qualitative tout en mettant en évidence un gradient. Au vu des résultats, il peut

être intéressant de créer les classes d'une autre manière (par exemple créer plus de classes, ou utiliser une autre méthode que celle des quintiles. Un autre travail est actuellement en cours pour créer des classes en réalisant une Analyse en Composantes Principales), pour avoir une idée plus précise de la tendance, en particulier entre les classes 4 et 5.

Ce travail est basé sur l'hypothèse que tous les individus vivant dans un IRIS donné sont exposés au même niveau de NO₂. Ainsi, il ne prend pas en compte les caractéristiques individuelles, comme les expositions multiples et la notion de « temps-activité », qui peut varier sensiblement selon les individus et le statut socio-économique. Cependant, la littérature récente montre que l'environnement de résidence de la mère et de l'enfant a une influence sur les issues de grossesse, indépendamment des risques individuels [18,19,20]. Le statut socio-économique du quartier de résidence a été montré comme le principal facteur de risque des issues de grossesse [21]. Mais il est possible que les populations les plus défavorisées soient plus susceptibles à la pollution de l'air que les populations vivant dans des quartiers privilégiés. En effet, plusieurs facteurs ayant un impact plus important dans les quartiers défavorisés peuvent modifier la relation pollution-mortalité[17].

Par ailleurs, modéliser les concentrations de NO₂ à l'échelle de l'IRIS donne de bonnes estimations de l'exposition à la pollution de l'air ambiant dans ces zones. Une analyse à l'échelle du quartier permet de révéler les inégalités au niveau local, qui sont souvent masquées lors d'étude à plus large échelle (commune, région, pays).

Le nombre d'IRIS présentant la classe d'exposition la plus importante est faible, ce qui peut expliquer l'intervalle relativement large. Malgré ce faible nombre d'IRIS présentant cette classe, et donc la faible puissance statistique, le risque relatif est très grand et très significatif.

Le recueil des données sanitaires n'a pas permis d'obtenir un taux d'exhaustivité supérieur à 85%, et les données manquantes ne sont pas équi-réparties. En effet, le quinzième arrondissement de Marseille présente près de 20% de données manquantes (50 cas recueillis au lieu des 70 cas recensés par le CépiDC). Cependant, le nombre de cas est de toute façon tellement élevé par rapport aux autres arrondissements et communes de la zone d'étude, que l'obtention de ces données manquantes ne ferait qu'accentuer encore plus les gradients déjà présentés dans les résultats.

La zone d'étude finale s'est trouvée extrêmement réduite par rapport à la zone d'étude initiale, réduisant fortement la puissance statistique et donnant des résultats difficilement extrapolables à l'extérieur de la zone. Mais les résultats relativement significatifs malgré

cette puissance statistique très réduite laissent à penser qu'élargir la zone d'étude amplifierait encore plus les gradients déjà montrés dans ces résultats.

Un des points forts de cette étude est qu'elle prend en compte l'autocorrélation spatiale dans la mesure de la relation entre les différentes variables. L'importance de prendre en compte ce phénomène a déjà été décrite dans d'autres publications [4,22] qui montrent de quelle manière l'autocorrélation spatiale peut influencer les résultats. Dans le travail présenté ici, la majorité des paramètres restent significatifs avec le contrôle de l'autocorrélation spatiale, et la qualité du modèle n'en est qu'améliorée. Le risque de mortalité infantile pour les IRIS appartenant à la classe 4 d'exposition par rapport aux IRIS les moins exposés est significatif quand le phénomène d'autocorrélation spatiale n'est pas pris en compte mais diminue et devient non significatif après contrôle du phénomène. Ceci montre que si l'autocorrélation spatiale n'était pas prise en compte, on pourrait prendre le risque de conclure à tort qu'une association positive et significative existe entre cette classe d'exposition et la mortalité infantile. Ceci confirme la nécessité de ne pas ignorer ce phénomène lors de la mise en place d'études écologiques afin d'éviter les résultats présentant de fausses associations [4,22,23,24].

Conclusion

Cette étude confirme, dans le cas d'une grande agglomération européenne, que le risque de mortalité infantile augmente avec l'indice de défaveur utilisé. Ce dernier est constitué de différentes variables sociales, dont certaines peuvent être plus facilement influencées que d'autres par la mise en place de politiques à l'échelle locale.

Ce travail met également en évidence l'existence d'un gradient de mortalité infantile selon l'exposition des populations à la pollution atmosphérique liée au trafic. Le risque de mortalité infantile peut être jusqu'à trois fois plus important pour les populations vivant dans les quartiers les plus exposés par rapport aux populations vivant dans les secteurs les moins exposés, et ceci après ajustement sur le niveau de défaveur. La tendance est différente pour les classes d'exposition intermédiaires.

L'importance de la prise en compte de l'autocorrélation dans une étude écologique est également montrée dans ce mémoire. Ne pas contrôler ce phénomène peut aboutir à la conclusion de l'existence de fausses associations.

De plus fines analyses pour caractériser la relation entre les trois familles de variables, et notamment l'étude de l'interaction entre les deux variables explicatives sont nécessaires pour guider la prise de décisions visant à réduire les inégalités sociales de santé.

Bibliographie

1. Reidpath DD, Allotey PJ. Infant mortality rate as an indicator of population health. *Epidemiology and Community Health* 2003;57(5):344-46.
2. Shi L, Starfield B, Kennedy B, Kawachi I. Income Inequality, primary care, and health indicators. *J Fam Pract* 1999;48(4):275-84.
3. Laurian L. Environmental Injustice in France. *Journal of Environmental Planning and Management* 2008;51(1):55-79.
4. Havard S, Deguen S, Zmirou-Navier D, Schillinger C, Bard D. Traffic-related air pollution and socioeconomic status: a spatial autocorrelation study to assess environmental equity on a small-area scale. *Epidemiology (Cambridge, Mass.)* 2009;20(2):223-30.
5. Glinianaia SV, Rankin J, Bell R, Pless-Mullooli T, Howel D. Does particulate air pollution contribute to infant death? A systematic review. *Environ Health Perspect.* 2004;112(14):1365-71.
6. American Academy of Pediatrics CoEH. Noise: a Hazard for the Fetus and Newborn. *Pediatrics* 1997;100(4):724-27.
7. Besag J, York J, Mollie A. Bayesian image restoration with two applications in spatial statistics. *Ann Inst Stat Math* 1991;43:1-21.
8. Mollie A. Bayesian mapping of Hodgkin's disease in France. In: Elliott P, Wakefield J, Best N, Briggs D, eds. *Spatial epidemiology. Methods and applications*. Oxford: Oxford University Press; 2000:267-285.

9. Spiegelhalter D, Best N, Carlin B, et al. Bayesian deviance, the effective number of parameters and the comparison of arbitrarily complex models. *J Roy Stat Soc B* 2002;64:583-640.
10. Guildea ZE, Fone DL, Dunstan FD, Sibert JR, Cartlidge PH. Social deprivation and the causes of stillbirth and infant mortality. *Arch Dis Child* 2001 Apr;84(4):307-10.
11. Singh GK, Kogan MD. Persistent socioeconomic disparities in infant, neonatal, and postneonatal mortality rates in the United States, 1969-2001. *Pediatrics* 2007 Apr;119(4):e928-e939.
12. Mathews TJ, MacDorman MF. Infant mortality statistics from the 2005 period linked birth/infant death data set. *Natl Vital Stat Rep* 2008 Jul 30;57(2):1-32.
13. Lelong N, Blondel B, Kaminski M. Smoking during pregnancy in France between 1972 to 2003: Results from the national perinatal surveys. *J Gynecol Obstet Biol Reprod (Paris)* 2011 Feb;40(1):42-9.
14. Luo ZC, Wilkins R, Kramer MS. Effect of neighbourhood income and maternal education on birth outcomes: a population-based study. *CMAJ* 2006 May 9;174(10):1415-20.
15. Nkansah-Amankra S, Dhawain A, Hussey JR, Luchok KJ. Maternal social support and neighborhood income inequality as predictors of low birth weight and preterm birth outcome disparities: analysis of South Carolina Pregnancy Risk Assessment and Monitoring System survey, 2000-2003. *Matern Child Health J* 2010 Sep;14(5):774-85.
16. Glinianaia S. V., Rankin J., Bell R., Pless-Mullooli T. & Howel D. (2004). Does particulate air pollution contribute to infant death? A systematic review. *Environ Health Perspect.*, 112 (14), 1365-71.
17. Yi O., Kim H. & Ha E. (2010). Does area level socioeconomic status modify the effect of PM10 on preterm delivery? *Environmental research* , 110 (1), 55-61.

18. O'Campo P, Xue X, Wang MC, Caughy M. Neighborhood risk factors for low birthweight in Baltimore: a multilevel analysis. *Am.J.Public Health*. 1997;87(7):1113-1
19. Luo ZC, Wilkins R, Kramer MS. Effect of neighbourhood income and maternal education on birth outcomes: a population-based study. *CMAJ*. 2006;174(10):1415-20.
20. Génèreux M, Auger N, Goneau M, Daniel M. Neighbourhood socioeconomic status, maternal education and adverse birth outcomes among mothers living near highways. *J Epidemiol Community Health* 2008a;62:695–700.
21. Carbajal-Arroyo L, Miranda-Soberanis V, Medina-Ramón M, Rojas-Bracho L, Tzintzun G, Solís-Gutiérrez P, Méndez-Ramírez I, Hurtado-Díaz M, Schwartz J, Romieu I. Effect of PM(10) and O(3) on infant mortality among residents in the Mexico City Metropolitan Area: a case-crossover analysis, 1997-2005. *J Epidemiol Community Health*. 2011 Aug;65(8):715-21.
22. Havard S, Deguen S, Bodin J, Louis K, Laurent O, Bard D. A small-area index of socioeconomic deprivation to capture health inequalities in France. *Soc Sci Med* 2008 Dec;67(12):2007-16.
23. Deguen S, Lalloue B, Bard D, Havard S, Arveiler D, Zmirou-Navier D. A small-area ecologic study of myocardial infarction, neighborhood deprivation, and sex: a Bayesian modeling approach. *Epidemiology* 2010 Jul;21(4):459-66.
24. Deguen S, Paulus H, Zmirou-Navier D. Indicators of residential proximity to industrial facilities, a literature review dealing with proximity index. *Environment International* 2011.

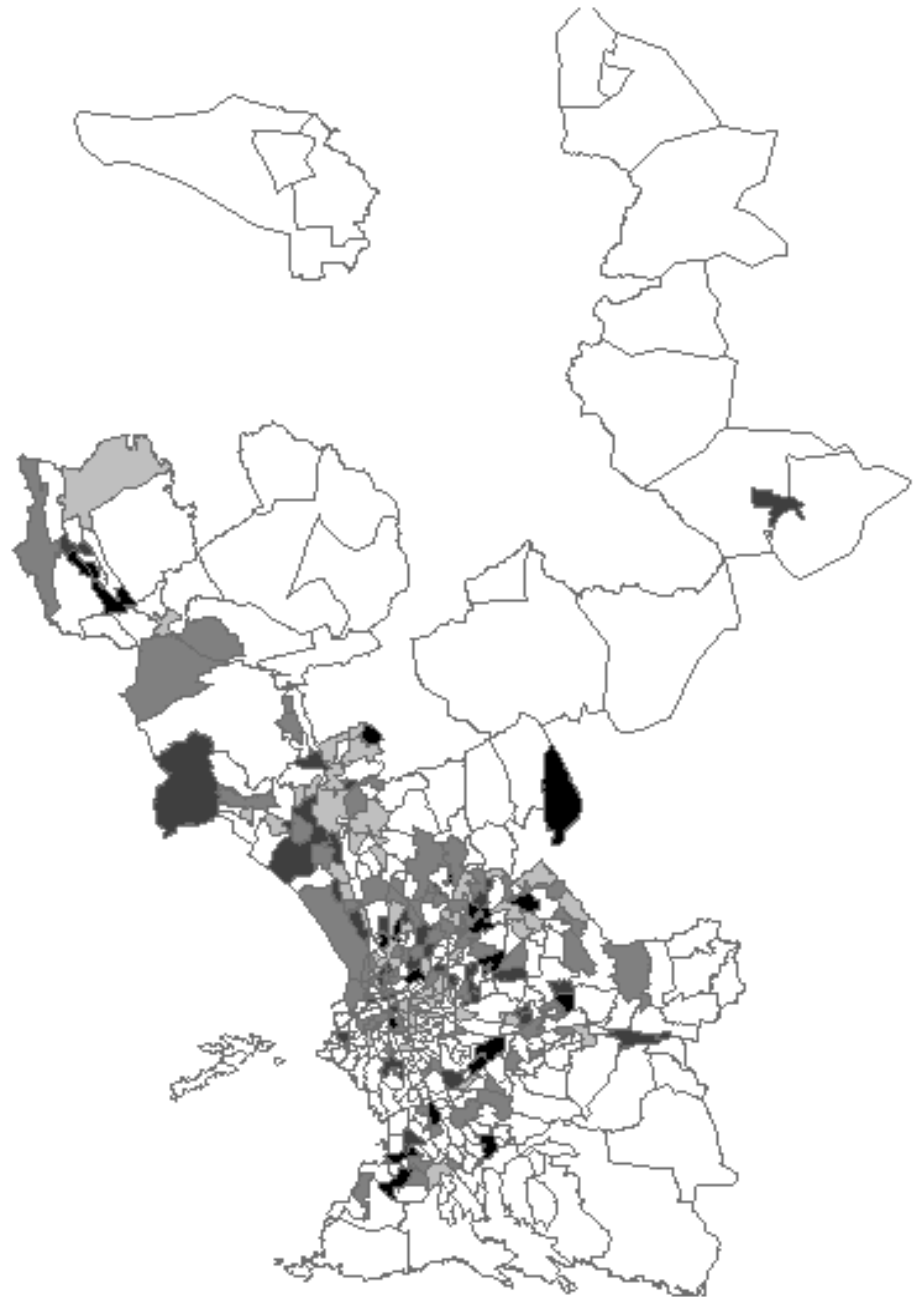
Liste des annexes

Liste des Figures

Figure 1 : Distribution spatiale des taux de mortalité infantile	II
Figure 2 : Evolution des concentrations de NO ₂ (µg/m ³) de 2002 à 2009	III
Figure 3 : Distribution spatiale des moyennes des concentrations de NO ₂ (µg/m ³).....	IV
Figure 4 : Distribution de l'indice de défaveur	VIII
Figure 5 : Distribution des risques relatifs de mortalité infantile en prenant en compte le niveau de défaveur.....	X
Figure 6 : Distribution des risques relatifs de mortalité infantile en tenant compte du niveau d'exposition au NO ₂	XII

Liste des Tableaux

Tableau 1 : Caractéristiques socio-économiques	VI
Tableau 2 : Risques relatifs de mortalité infantile selon la classe de défaveur.....	IX
Tableau 3 : Risques relatifs de mortalité infantile selon la classe de NO ₂	XI
Tableau 4 : Risques relatifs de mortalité infantile selon la classe de NO ₂ après ajustement sur l'indice de défaveur	XIII



Taux de Mortalité Infantile

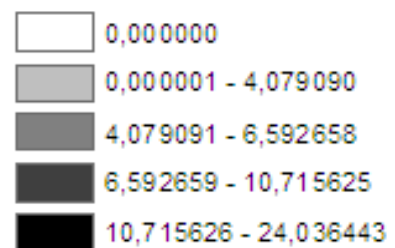


Figure 1 : Distribution spatiale des taux de mortalité infantile

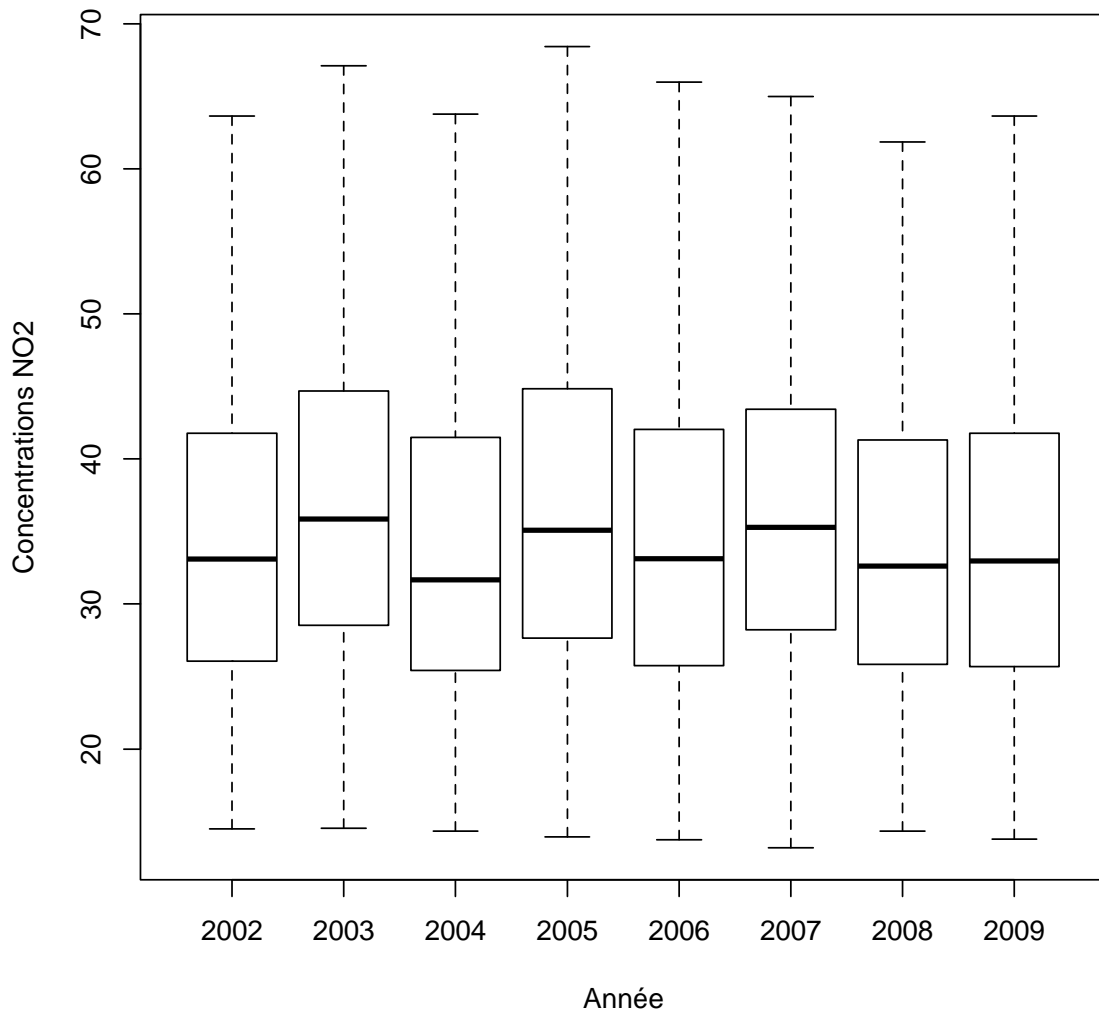
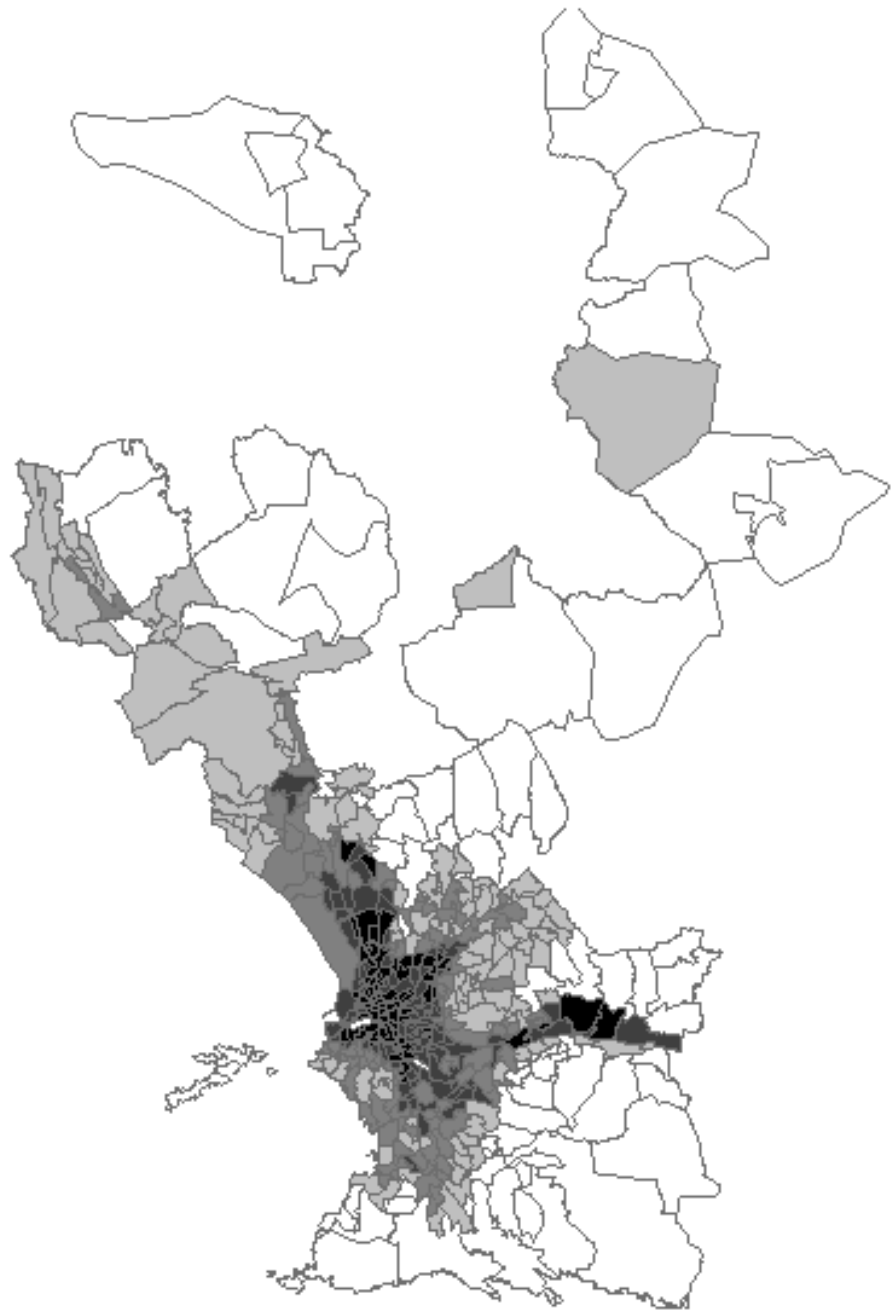


Figure 2 : Evolution des concentrations de NO₂ (µg/m³) de 2002 à 2009



Moyennes NO2

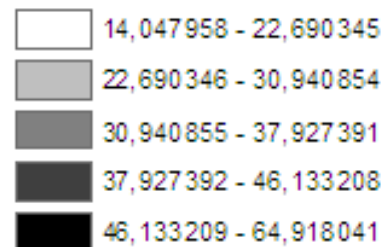
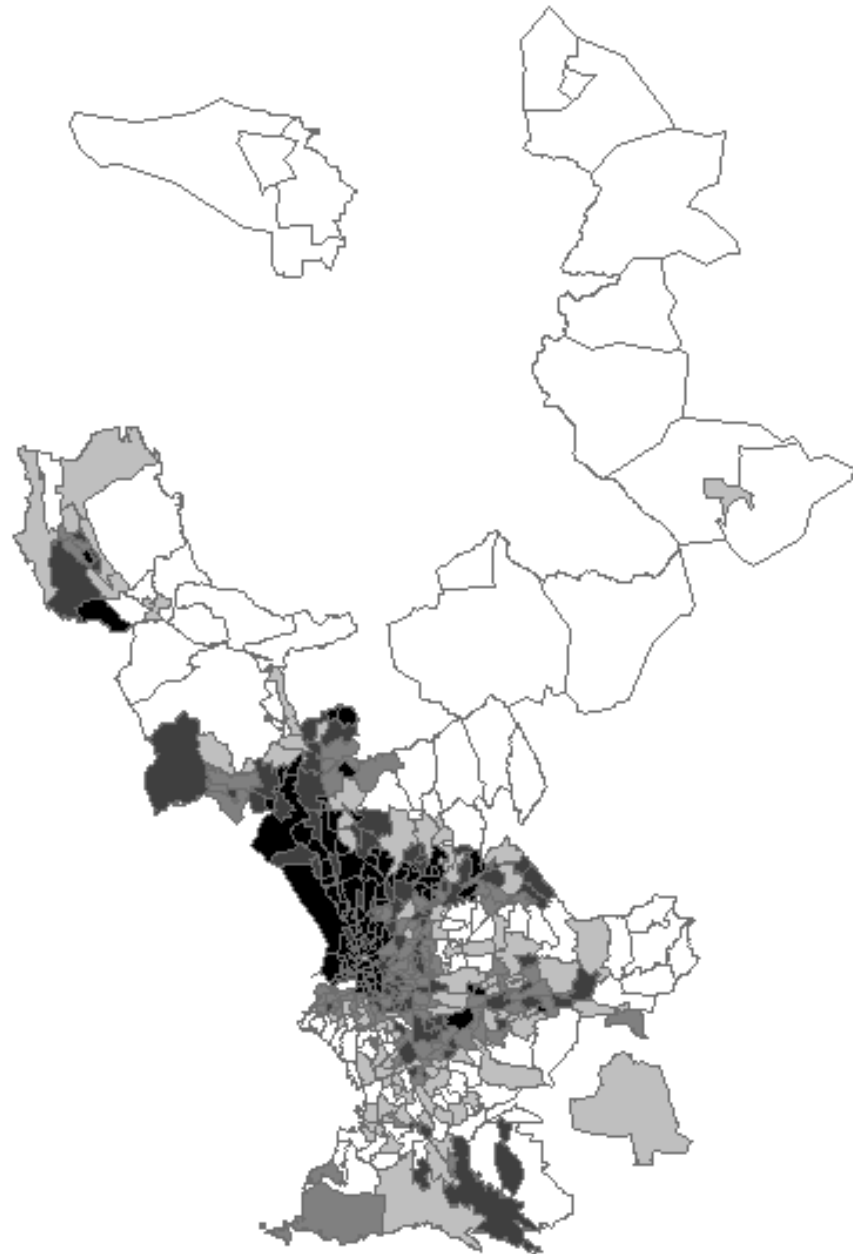


Figure 3 : Distribution spatiale des moyennes des concentrations de NO₂ (µg/m³)

Variables	Moyenne ± Ecart-type	Centile 5%	Centile 50%	Centile 95%
Etrangers dans la population (%)	6,30 ± 6,86	0,88	3,78	20,50
Immigrants étrangers (%)	2,65 ± 3,22	0,15	1,66	7,85
Part de la population au chômage depuis plus d'un an (%)	22,63 ± 12,69	9,38	19,19	49,84
Part de la population non salariée (%)	8,88 ± 7,10	2,33	7,58	18,96
Part de la population ayant un emploi précaire (%)	10,54 ± 3,55	5,97	10,43	60,36
Part de la population ayant un emploi stable (%)	57,71 ± 12,98	32,16	60,36	72,38
Cadres (%)	12,53 ± 12,34	0,00	9,72	31,82
Familles monoparentales (%)	19,56 ± 9,07	7,74	18,84	33,33
Part de la population sans diplôme (%)	20,66 ± 14,04	8,07	17,65	42,42
Part de la population ayant obtenu le bac/brevet (%)	10,91 ± 4,02	4,86	10,97	15,99
Part de la population ayant obtenu un diplôme Bac + 2 (%)	7,71 ± 4,32	1,84	7,82	13,10
Part de la population résidant dans une maison (%)	24,95 ± 29,81	0,09	9,81	90,37
Part de la population résidant en immeuble (%)	72,20 ± 30,62	5,88	86,34	98,62

Part de la population disposant d'un garage (%)	37,47 ± 24,88	2,62	34,21	82,63
Part de la population non propriétaire (%)	55,05 ± 22,47	20,49	54,23	98,76
Part des foyers sans voiture (%)	28,90 ± 17,13	4,57	27,45	18,55
Part des foyers possédant au moins deux voitures (%)	22,30 ± 15,66	4,64	18,55	54,03
Part des foyers avec plus d'une personne par pièce (%)	10,67 ± 8,07	4,23	7,93	27,55
Nombre moyen de personnes par pièce	0,74 ± 0,11	0,64	0,71	0,98
Revenu médian (Euros/an)	20 733,81 ± 7224,14	9827,65	19525	33427,95

Tableau 1 : Caractéristiques socio-économiques



Indice de défaveur






	-1,794635 -- -0,815941
	-0,815634 -- -0,194744
	-0,174502 - 0,297347
	0,309885 - 0,943055
	0,970821 - 3,261319

Figure 4 : Distribution de l'indice de défaveur

Classe de défaveur*	RR (IC 95%)	
	Avec autocorrélation spatiale	Sans autocorrélation spatiale
C1	1 -	1 -
C2	1,70 (1,05 - 2,82)	1,83 (1,17 - 2,98)
C3	1,86 (1,13– 3,13)	2,10 (1,35 - 3,41)
C4	2,02 (1,16– 3,53)	2,39 (1,50- 3,96)
C5	2,88 (1,52– 5,42)	3,39 (1,98 - 5,91)
DIC**	873,5	886,8

* C1 est la classe la moins défavorisée, C5 la plus défavorisée.

**Deviance Information Criterion

Tableau 2 : Risques relatifs de mortalité infantile selon la classe de défaveur

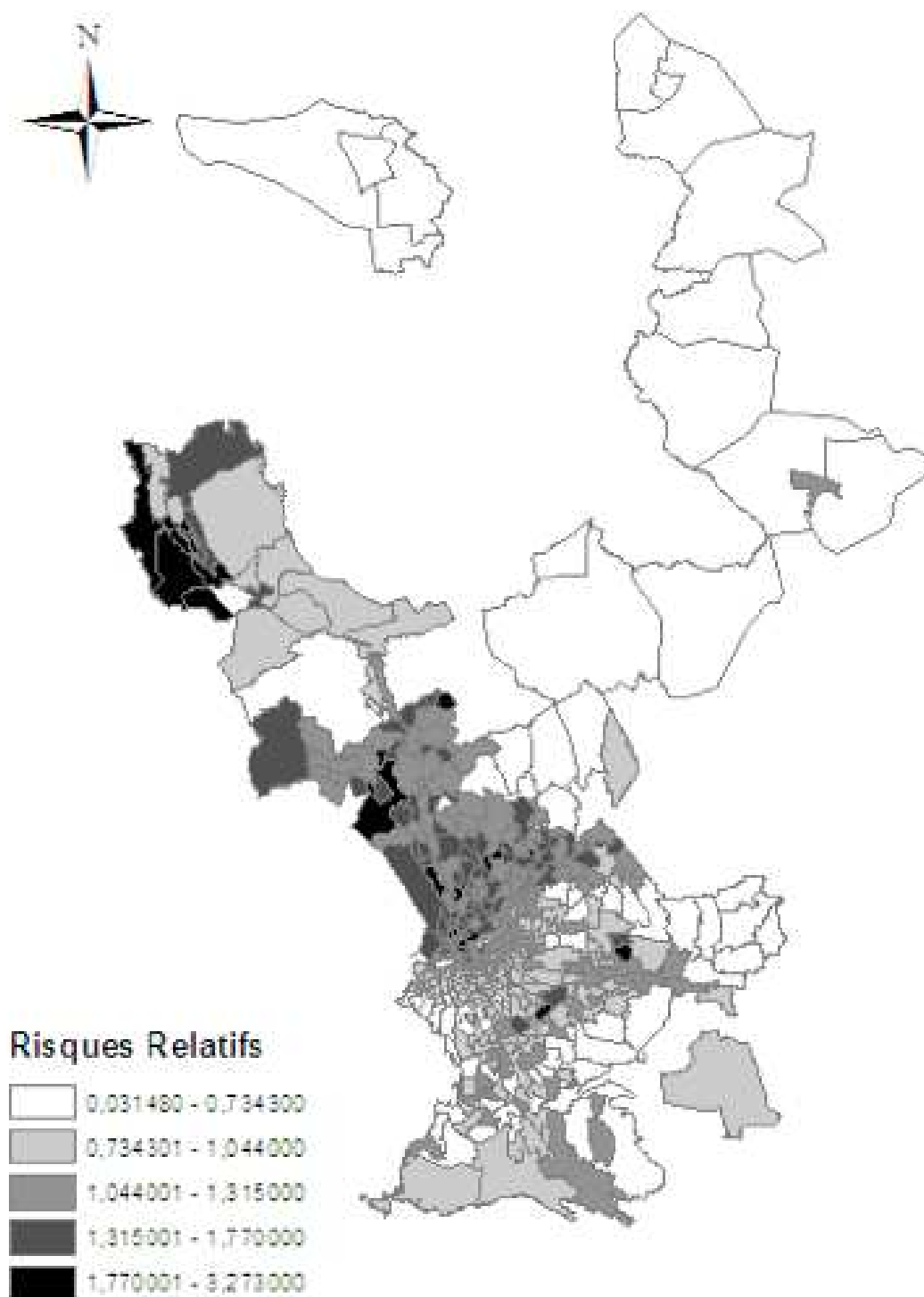


Figure 5 : Distribution des risques relatifs de mortalité infantile en prenant en compte le niveau de défaveur.

Classe de NO2*	RR (IC 95%)	
	Avec autocorrélation spatiale	Sans autocorrélation spatiale
C1	1 -	1 -
C2	2,10 (1,26 - 3,66)	2,44 (1,52 - 4,08)
C3	2,07 (1,19- 3,72)	2,17 (1,33 - 3,67)
C4	1,78 (0,98- 3,31)	1,94 (1,16- 3,33)
C5	3,96 (1,85- 8,50)	4,70 (2,47 - 8,89)
DIC**	869,1	885,3

* C1 est la classe d'exposition la plus faible

**Deviance Information Criterion

Tableau 3 : Risques relatifs de mortalité infantile selon la classe de NO2

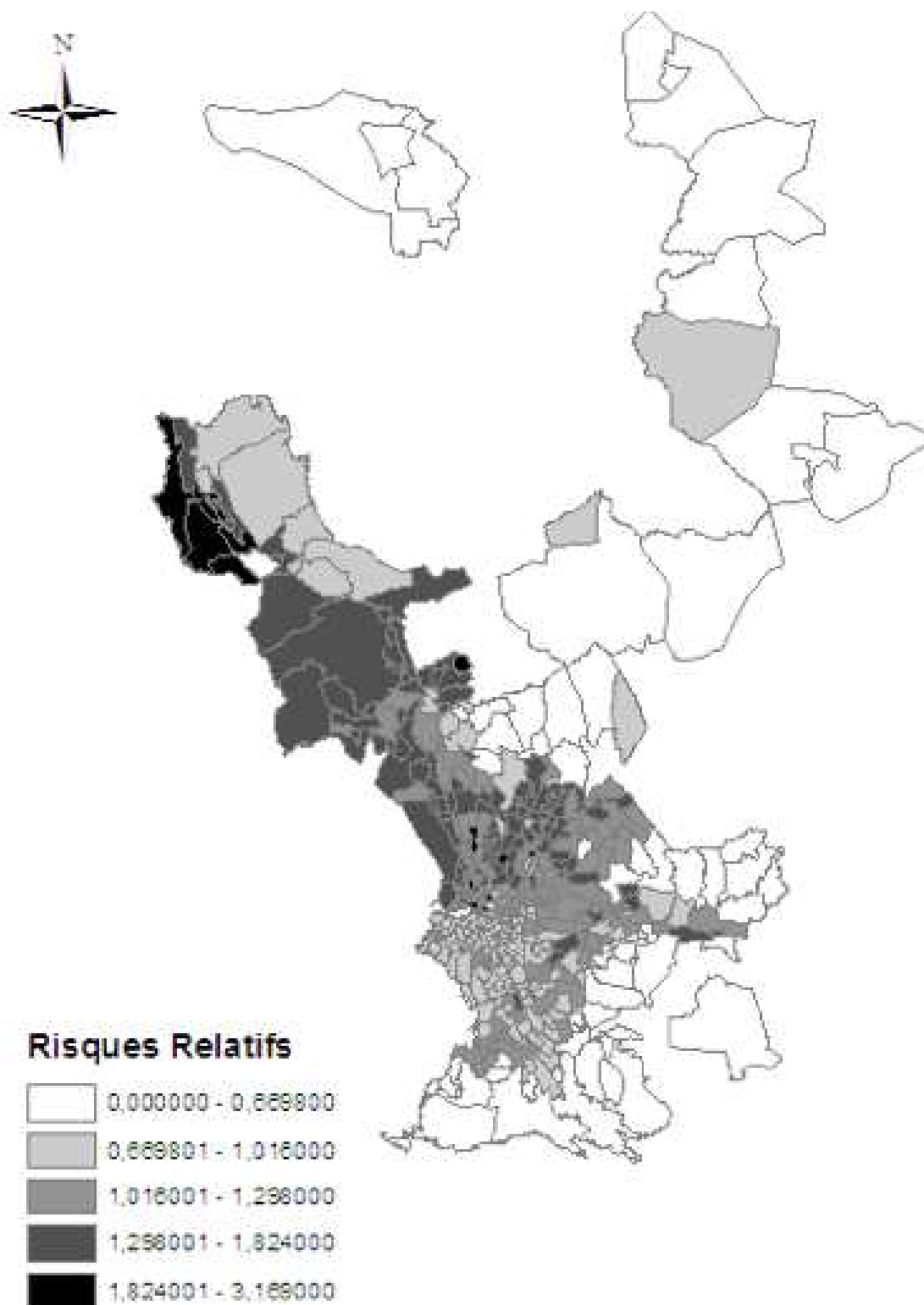


Figure 6 : Distribution des risques relatifs de mortalité infantile en tenant compte du niveau d'exposition au NO₂

Classe de NO2*	RR (IC 95%)		RR (IC 95%)	
	Avec autocorrélation spatiale		Sans autocorrélation spatiale	
C1	1	-	1	-
C2	1,95 (1,17 – 3,39)		2,11 (1,30 - 3,56)	
C3	1,81 (1,06 – 3,23)		1,83 (1,1 – 3,13)	
C4	1,47 (0,82 – 2,73)		1,47 (0,86 – 2,59)	
C5	3,10 (1,46 – 6,64)		3,21 (1,63 – 6,28)	
DIC**	867,0		874,2	

* C1 est la classe d'exposition la plus faible

**Deviance Information Criterion

Tableau 4 : Risques relatifs de mortalité infantile selon la classe de NO2 après ajustement sur l'indice de défaveur

Environment and Infant mortality in Marseille Metropolitan Area

Background- Evidence of social health inequalities is well established: deprived populations are more strongly affected by various health problems. Environmental nuisances are suspected to be part of these inequalities: socio-economically disadvantaged groups may be both more exposed to pollution sources and more sensitive to their health effects, because of a degraded health status and a lack of early health care. Assessing how environmental exposures may partly explain social inequalities in health is today a major public health research issue. This issue is part of the second national environmental health plan (PNSE2).

Objective- The Equit'Area project, conducted by a research team at the french national school of public health, aims to project aims to explore in France the interaction between three families of variables: environmental, socioeconomic and health events (infant mortality), which are not often considered jointly in epidemiological studies. The study carried out for this thesis consists in assessing the relationship between traffic-related air pollution and infant mortality, taking into account the socio-economic status of the population.

Methods- An ecological study design is used. The geographical unit used is the French census residential block. This thesis is focused on Marseille metropolitan area. The socio-economic variables are available from the national census. From these was created a deprivation index, which is the final variable used to characterize the socio-economic level in each census block. Atmospheric nitrogen dioxide mean concentrations are the environmental exposure considered here. After a few descriptive statistics, a Bayesian model was implemented to investigate the link between infant mortality and, on the one hand, the deprivation level and, on the other hand, exposure to nitrogen dioxide, after adjustment for the deprivation level. The model was run with and without taking into account spatial autocorrelation.

Results / Conclusion- A significant association between nitrogen dioxide and infant mortality was found. The risk of infant mortality is multiplied by almost 4 times in the most exposed census blocks compared to the less exposed ones. This risk, still significant, is actually multiplied by up to 3 times after adjustment for the deprivation level. This study shows how not considering spatial autocorrelation can strongly influence the results, leading to spurious associations.

Environnement et mortalité infantile dans l'agglomération de Marseille

Contexte- Le constat des inégalités sociales de santé est aujourd'hui solidement établi: les populations les plus défavorisées sont plus affectées par différents problèmes de santé. Les nuisances environnementales pourraient être à l'origine d'une partie de ces inégalités: les groupes socio-économiquement défavorisés seraient à la fois souvent plus exposés à des sources polluantes et plus sensibles aux effets sanitaires en résultant, en raison d'un état sanitaire comparativement dégradé et d'un moindre accès aux soins précoces. L'estimation de la contribution des expositions environnementales à l'explication des ISS est aujourd'hui une question majeure en Santé publique, constituant un axe fort du second plan national santé environnement (PNSE2).

Objectif- Le projet Equit'area, réalisé par une équipe de chercheurs de l'Ecole des hautes études en santé publique (EHESP), a pour but d'explorer l'interaction pouvant exister entre trois familles de variables: nuisances environnementales, niveau socio-économique et les événements sanitaires, qui ne sont que rarement considérés conjointement dans les études épidémiologiques. L'étude menée pour la réalisation de ce mémoire consiste en l'évaluation de la relation entre la pollution de l'air liée au trafic et la mortalité infantile, en prenant en compte le statut socio-économique de la population.

Méthodes- L'étude réalisée est de type écologique. L'unité géographique utilisée est l'IRIS (Ilôts Regroupés pour l'Information Statistique). Ce mémoire est concentré sur l'agglomération de Marseille. Les variables socio-économiques sont disponibles à partir du recensement de 1999, et à partir de celles-ci est créé un indice socio-économique, qui est la variable finale utilisée pour caractériser le niveau de défaveur de chaque IRIS. Les concentrations moyennes de NO₂ atmosphérique représentent l'exposition environnementale qui est considérée ici. Après quelques statistiques descriptives, un modèle bayésien est appliqué pour évaluer le lien entre la mortalité infantile et le niveau de défaveur, puis entre la mortalité infantile et l'exposition au NO₂, d'abord considérée seule puis en ajustant sur le niveau socio-économique. Le modèle est réalisé avec et sans contrôle de l'autocorrélation spatiale.

Résultats / Conclusion- Cette étude montre une association statistiquement significative entre l'exposition au NO₂ et Le risque de mortalité infantile est pratiquement multiplié par quatre dans les quartiers les plus exposés par rapport aux quartiers les moins exposés. Ce risque, toujours significatif, est multiplié par trois lorsque le niveau de défaveur est pris en compte. Cette étude montre également comment ne pas prendre en compte l'autocorrélation spatiale peut influencer les résultats et mener à de fausses associations.