

Equit'Area

Agrégation socio-spatiale de l'exposition à des nuisances environnementales à risque sanitaire

Exposition au dioxyde d'azote
Note technique

Octobre 2011

Rédacteurs :

Charles Beugard – Atmo Nord – Pas de Calais
Géraldine Le Nir – AIRPARIF
Philippe Rossello – Atmo PACA
Florence Troude – Atmo Rhône Alpes

Equit'Area – Agrégation socio-spatiale de l'exposition à des nuisances environnementales à risque sanitaire

1. Contexte et objectifs du projet

Ce projet s'inscrit dans le cadre du programme de recherche de l'EHESP sur les inégalités socio-spatiales de la santé et la contribution de l'environnement à ces inégalités, programme financé par l'Agence Nationale de la Recherche (ANR), nommé Equit'Area – Agrégation socio-spatiale de l'exposition à des nuisances environnementales à risque sanitaire.

L'objectif général de ce programme de recherche consiste à étudier la contribution des expositions environnementales aux inégalités sociales de santé sur le territoire national et plus particulièrement sur 4 agglomérations pilotes représentant une fraction importante de la population nationale : Lille, Paris, Lyon et Marseille-Communauté du Pays d'Aix (CPA). Par le choix de zones d'observation variées quant à leurs caractéristiques socio-démographiques, économiques et leur tissu industriel, ce projet permet de tester la robustesse de l'approche et d'identifier les conditions de son extension au plan national.

Le travail a été réalisé conjointement par les 4 associations agréées pour la surveillance de la qualité de l'air (AASQA) des territoires concernés : Atmo Nord Pas-de-Calais, AirParif, Atmo Rhône-Alpes et Atmo PACA. Ces associations disposent en effet non seulement de mesures de la qualité de l'air depuis les années 80, mais aussi d'un savoir-faire en termes de modélisation permettant d'élaborer des cartographies de polluants à une résolution spatiale fine compatible avec les unités de statistiques socio-démographiques de l'INSEE (IRIS).

De manière plus technique, le travail réalisé par les AASQA des 4 agglomérations dans le cadre de ce projet et présenté dans ce document a consisté à élaborer et tester une démarche méthodologique permettant de décrire et de caractériser les disparités d'expositions environnementales à une échelle territoriale fine, pour la période allant de 2002 à 2009 pour chaque domaine :

1. mise à disposition de l'ensemble des mesures disponibles issues des réseaux de mesures (stations fixes) pour l'EHESP ;
2. réalisation d'une « image » des concentrations annuelles de NO₂ par année ;
3. élaboration d'une méthode d'agrégation des résultats par iris ;
4. réalisation d'indicateurs par iris : NO₂ moyen par iris, écart-type des valeurs de NO₂ par iris, estimation de la population¹ exposée à la valeur limite (>40 µg/m³) par bâtiment et par iris, classification ascendante hiérarchique (CAH) des concentrations moyennes de NO₂ et représentativité des stations par iris.

¹ Population estimée par habitation en fonction de la population par iris et du volume (ou de sa superficie totale = superficie x nombre d'étages) du bâti

2. Données d'entrées aux modèles

Selon les AASQA, les cartes de concentrations à l'échelle locale sont obtenues par différents modèles numériques : ADMS Urban, SIRANE ou encore STREET, avec ou sans adaptation géostatistique. Ces modèles, bien qu'ayant chacun leurs spécificités numériques et phénoménologiques, permettent d'obtenir une cartographie fine des concentrations de polluants, en l'occurrence, ici, le dioxyde d'azote (NO₂).

Bien que s'agissant de modèles différents, les données d'entrées sont quant à elles de même type. Il s'agit en effet de caractériser la météorologie (température, vent, humidité, pression), des émissions de polluants (notamment celles du trafic routier) et de la pollution de fond pour les années représentées. Les caractéristiques de ces données d'entrées utilisées pour les différentes agglomérations sont synthétisées dans les tableaux ci-dessous (Tableau 1 et Tableau 2). La dispersion des polluants est alors calculée par le modèle et restituée via une cartographie.

Tableau 1

Agglomération	Modèle	Caractéristiques météo	Pollution de fond	Résolution spatiale de base
Lille	ADMS Urban	Mesure Météo-France à Lesquin pour vitesse et direction du vent, température, précipitations. Modèle MM5 pour les paramètres hauteur de couche limite et flux de chaleur sensible.	Mesure de la station de Salomé (périurbaine)	50 m
Lyon	SIRANE	Mesure Météo-France à Bron (St Exupéry pour 2007), horaire de chaque année mesurée (vitesse et direction du vent, température, précipitations, pression). Autres grandeurs calculées directement par le modèle.	Mesure de la station de St-Exupéry	10 m
Paris	STREET pour la pollution de proximité, ESMERALDA en fond + traitement géostatistique avec les concentrations des stations de proximité et utilisation d'une fonction de décroissance des niveaux entre la proximité et le fond (prise en compte de l'urbanisme).	Modèle MM5 Pour STREET : données de météo moyennes annuelles en Ile-de-France (VV=3m/s et DV = Sud-Ouest)	2002 à 2006 : cartographie ISATIS (géostatistique) à partir des données stations de fond et du cadastre année 2000. 2007 à 2009 : résultats du modèle interrégional ESMERALDA	25 m
Marseille	STREET + traitement géostatistique, utilisation d'une fonction de décroissance des niveaux entre la proximité et le fond (prise en compte de la densité bâti).	Mesure à la station Hippodrome	2001 à 2005 : champs 2D constitués à partir de 9 stations au maximum de mesures fixes (Plombières, Rabatau, Saint-Louis, Prado, Cinq avenues, Sainte-Marguerite, Thiers/Noailles, Vallée Huveaune) + 214 points de mesure par tubes, spatialisation de la pollution de fond avec logiciel ISATIS	20 m
Communauté du Pays d'Aix (CPA)	ADMS-Urban	Mesure Météo-France d'Aix en Provence	Mesure station Aix en Provence	50 m

Tableau 2

Agglomération	Emissions trafic	Emissions Surfaciennes		Emissions ponctuelles
		année	résolution	
Lille	Modèle de trafic Communauté Urbaine de Lille	2006	communale	31 grandes sources ponctuelles
Lyon	Cadastre annuel intégrant des comptages journaliers, COPERT IV	annuel calculé à partir du cadastre 2008	1 x 1 km ²	91 grandes sources ponctuelles
Paris	2002 à 2006 : utilisation d'un réseau trafic régional réactualisé fourni par DRIEA (année 2000), avec calage. Calcul avec le modèle HEAVEN, COPERT III 2007 à 2009 : utilisation d'un réseau trafic régional réactualisé fourni par DRIEA (année 2007), avec calage. Calcul avec le modèle HEAVEN, COPERT IV	2000	1 x 1 km ²	200 grandes sources ponctuelles
Marseille	Logiciel Street : 650 brins, données trafic fournies par CETE (EMME2 2001) et Marseille Provence Métropole (Opéra 1997), facteurs d'émission fournis par le logiciel IMPACT de l'ADEME	1999	3 x 3 km ²	non
Communauté du Pays d'Aix (CPA)	Comptages et estimations réalisés par la Communauté du Pays d'Aix (CPA)	2004	1 x 1 km ²	une dizaine de grandes sources ponctuelles

Pour certains domaines, des années ont été reconstituées via des adaptations géostatistiques (Cf. partie 5.). Les données utilisées sont les suivantes (Tableau 3) :

Tableau 3

Agglomération	Carte de référence	Méthode	Données d'entrée	Années concernées
Lille	2009	krigeage avec dérive externe	valeurs de NO ₂ annuelles aux stations et la carte de référence en dérive externe	2002 à 2008
Marseille	2005			2002 à 2004 et 2006 à 2009
Communauté du Pays d'Aix (CPA)	2007			

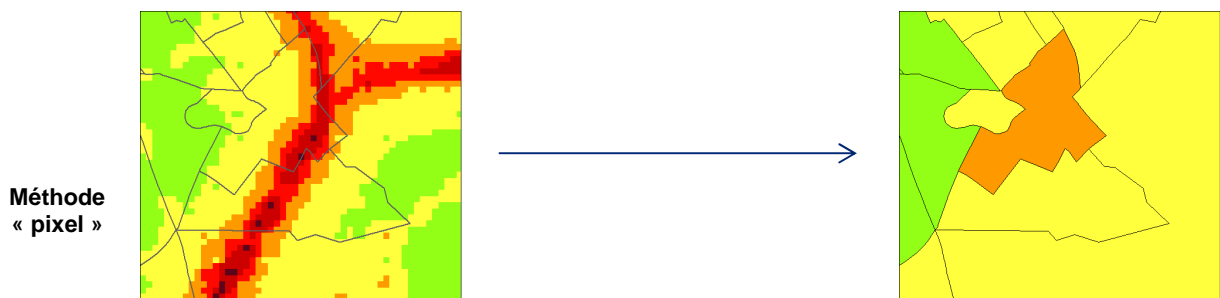
L'unité géographique utilisée est l'IRIS (Ilots Regroupés pour l'Information Statistique), découpage infra communal mis en œuvre par l'Institut National de la Statistique et des Études Économiques au cours du recensement de 1999. L'IRIS correspond en moyenne à un quartier de résidence de 2000 habitants et constitue la plus petite unité géographique française pour laquelle l'intégralité des données démographiques et socio-économiques du recensement est disponible. Le découpage des quartiers de résidence en IRIS tient compte des obstacles physiques introduisant des ruptures dans le paysage urbain (infrastructures autoroutières, cours d'eau, espaces verts...) et recherche le plus d'homogénéité en termes de taille de population, de caractéristiques socio-économiques et d'utilisation de l'espace urbain. La géométrie des IRIS est éditée par l'IGN et l'INSEE. Les données de bâti ont également été utilisées. Elles proviennent de la « BD TOPO » de l'IGN (couche « Bâti indifférencié »).

3. Méthode d'agrégation à l'IRIS des concentrations spatialisées

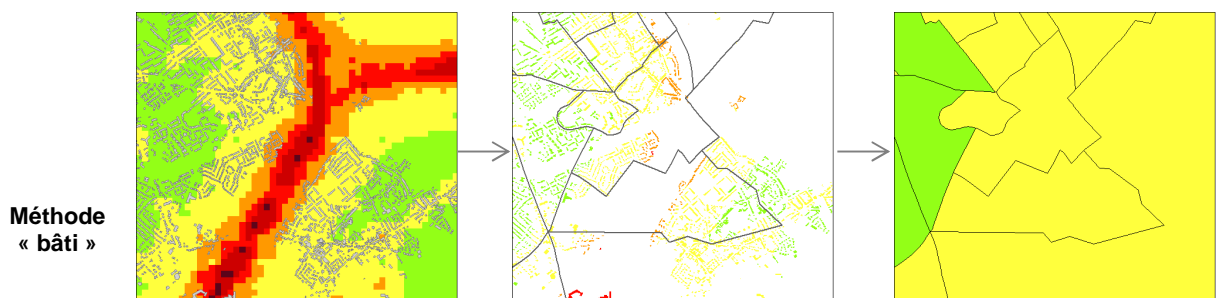
3.1. Description des méthodes testées

L'agrégation des concentrations spatialisées à l'IRIS met en œuvre les différentes couches d'informations (données de concentrations, base IRIS, couche bâti, population) au sein d'une application SIG, en l'occurrence les logiciels MapInfo ou ArcView. Les requêtes spatiales prédéfinies pour l'agrégation de variables par croisement de deux couches d'informations ont été mises en œuvre, et en particulier la fonction de moyenne proportionnelle. Dans le cas présent, la variable attribut de l'objet cartographique A est agrégée ou désagrégée proportionnellement à la surface de cet objet qui entre en intersection, dans un plan de projection cartographique, avec celle l'objet B. Dans le cas présent, un nombre important d'objets issus de la modélisation devaient entrer en intersection avec des objets plus grands – les IRIS, d'où le recours à la fonction de moyenne proportionnelle.

Ainsi deux méthodes ont été testées, l'une basée sur le simple croisement des couches de concentrations avec les contours IRIS (méthode « pixel »), l'autre faisant intervenir la couche bâti de la BD TOPO lors d'une étape intermédiaire.



Croisement de la couche de concentrations spatialisées sous forme de pixels (cellules de grille de résolution native du modèle) avec la couche des contours IRIS, avec calcul d'une moyenne proportionnelle par IRIS.



1. Croisement de la couche des concentrations spatialisées avec la couche BD TOPO (Bâti indifférencié), calcul d'une moyenne proportionnelle par bâtiment
2. Croisement de la couche bâti avec la couche des contours IRIS, calcul d'une moyenne proportionnelle par IRIS.

Les deux méthodes ont été testées sur les agglomérations de Lille, Lyon, Marseille, et la Communauté du Pays d'Aix. Sur l'agglomération parisienne, seule la méthode pixel a été testée en raison du grand nombre d'objets dans la couche bâti qui rendait le temps de calcul excessif.

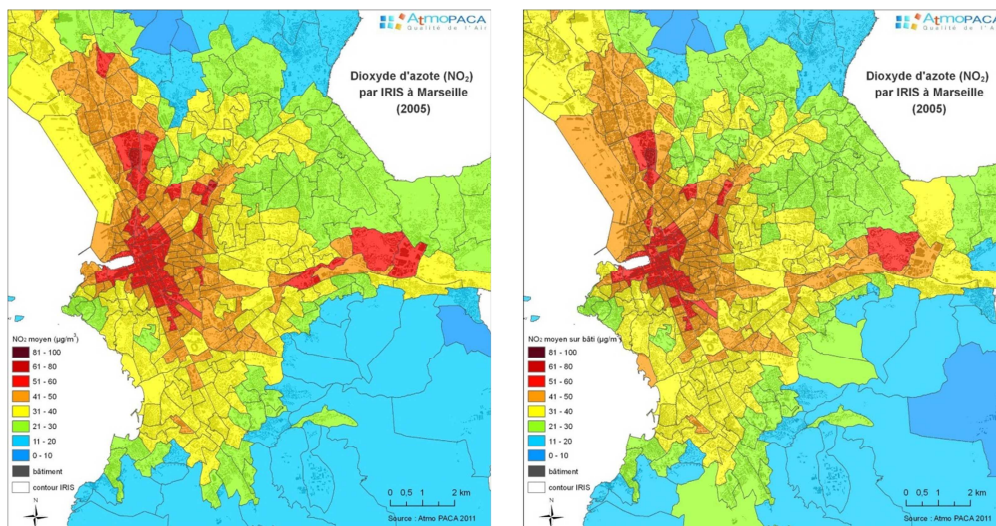
3.2. Résultats et choix de la méthode

Pour chaque méthode, différents paramètres ont été calculés au niveau des concentrations agrégées à l'IRIS avec, en commun sur toutes les agglomérations, la moyenne en NO₂ par IRIS, et l'écart-type par IRIS. Cet écart-type est calculé dans l'ensemble de cellules de grille (ou pixels) qui entrent en intersection avec l'objet IRIS. Le Tableau 4 suivant reprend des résultats des tests obtenus sur Lille et Marseille :

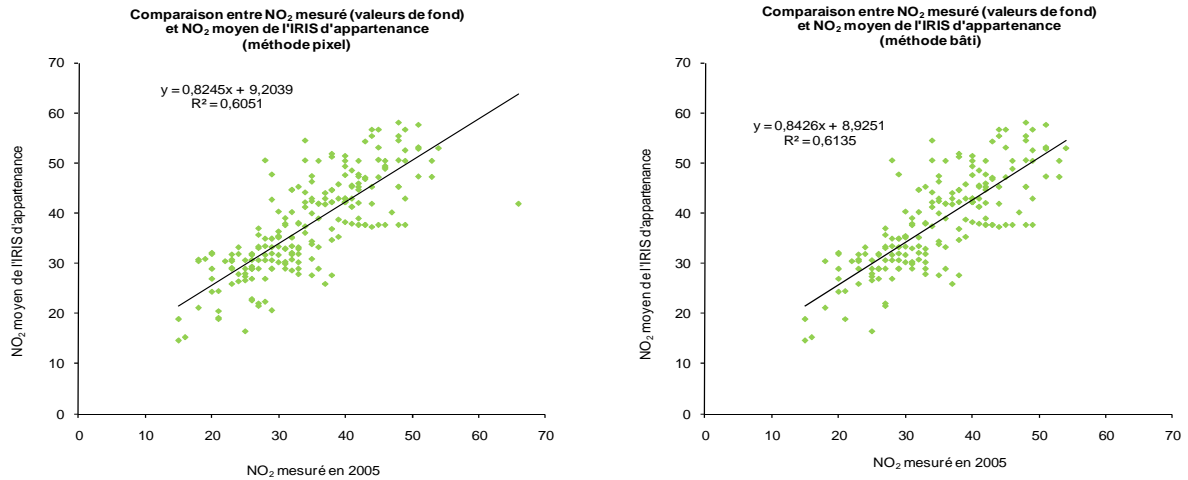
Tableau 4

		Moyenne µg/m ³	Ecart-type µg/m ³
Marseille	NO ₂ méthode pixel	36.1	12.7
	NO ₂ méthode bâti	35.8	11.6
	écart (pixel - bâti)	0.3	2.3
	écart-type bâti dans l'IRIS	4.7	2.6
Lille	NO ₂ méthode pixel	31.68	5.42
	NO ₂ méthode bâti	31.31	4.73
	écart (pixel - bâti)	0.62	2.87
	écart-type bâti dans l'IRIS	2.22	1.32

Exemples de cartes de concentrations agrégées sur l'agglomération de Marseille selon la méthode pixel (à gauche) et la méthode bâti (à droite) :



Comparaison mesures/concentrations agrégées à l'IRIS selon les deux méthodes sur Marseille :



Les conclusions tirées de tests sont les suivantes :

1. il n'existe pas d'écart sensible entre les deux méthodes (pixel et bâti) ; les écarts les plus importants sont observés dans les IRIS influencés par les grands axes ;
2. l'hétérogénéité des concentrations dans les IRIS tend à diminuer avec la méthode bâti (diminution des écarts-type) ;
3. les écarts-types les plus élevés mettent en évidence les IRIS traversés par les axes routiers majeurs ;
4. on retrouve une corrélation entre les valeurs agrégées à l'IRIS et les valeurs sur les stations de fond incluses dans l'IRIS (Cf. résultats sur Marseille).

Compte tenu des faibles écarts entre les deux méthodes obtenus dans les résultats et de la difficulté à réaliser l'étape nécessaire à la méthode bâti sur l'agglomération parisienne, il a été décidé d'exploiter les résultats de l'agrégation par la méthode pixel qui est plus aisée à mettre en œuvre, année par année entre 2002 et 2009.

4. Classification Ascendante Hiérarchique

Les résultats en termes de concentrations obtenues à l'échelle des IRIS fournissent une information fine à l'échelle géographique, mais l'échelle temporelle (annuelle) reste trop grossière au vu des données socio-démographiques disponibles.

Par ailleurs, les données à l'échelle horaire des stations de mesures réparties sur chacun des territoires étudiés fournissent des informations temporelles nécessaires.

Par conséquent, une méthode permettant de reproduire les variations journalières observées aux stations de mesures au niveau de l'IRIS a été recherchée. Plus précisément, l'objectif été d'associer à chaque IRIS une station de mesure permanente, pour ensuite lui attribuer les variations journalières. C'est la Classification Ascendante Hiérarchique (CAH) qui a été retenue.

4.1. Classification Ascendante Hiérarchique

La classification ascendante hiérarchique (CAH) est l'une des méthodes de classification qui a pour objectif de regrouper des individus similaires caractérisés par différentes variables. Les classes (groupes d'individus) ainsi formées doivent être les plus homogènes possibles en fonction des critères choisis. La classification ascendante hiérarchique associée à une distance et à un critère d'agrégation considère chaque individu séparément avant de les regrouper pas à pas jusqu'à n'obtenir qu'une seule classe comprenant l'ensemble des individus. Pour évaluer la qualité des groupes obtenus, l'inertie inter-classe est généralement utilisée. Il faut ainsi arbitrer entre le nombre de groupe interprétable (assez restreint) et la qualité des classifications, c'est-à-dire entre interprétabilité et fiabilité de l'information.

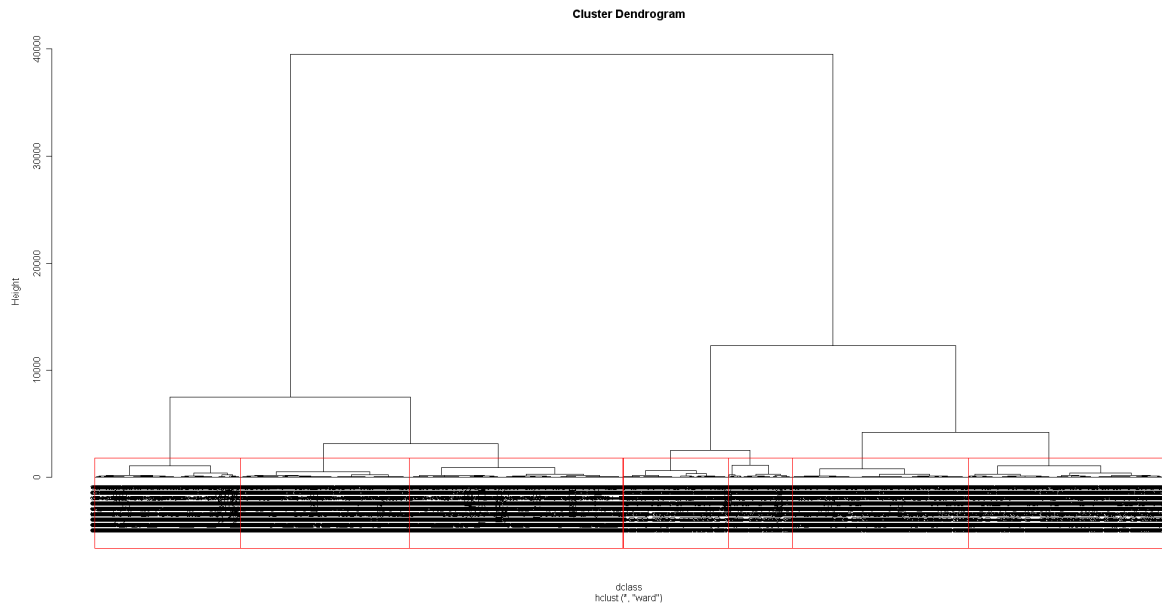
Cette méthode a été retenue pour sa facilité d'utilisation et sa maîtrise (cette technique statistique est notamment utilisée pour l'exploitation de données de campagne de mesures, par exemple, afin de connaître les sites de mesure temporaires ayant une variabilité dans ses résultats comparables à ceux de certaines stations permanentes).

4.2. Résultats

Les CAH ont été réalisées à l'échelle des agglomérations. Les données d'entrée de la CAH sont les résultats à l'IRIS (moyennes annuelles en NO_2) pour les années disponibles, ainsi que les mêmes résultats aux stations permanentes de l'agglomération choisie. Concrètement, sur l'agglomération parisienne, les moyennes annuelles pour les années 2002 à 2009 ont été utilisées. Pour les autres agglomérations, les années disponibles concernent également les 8 années.

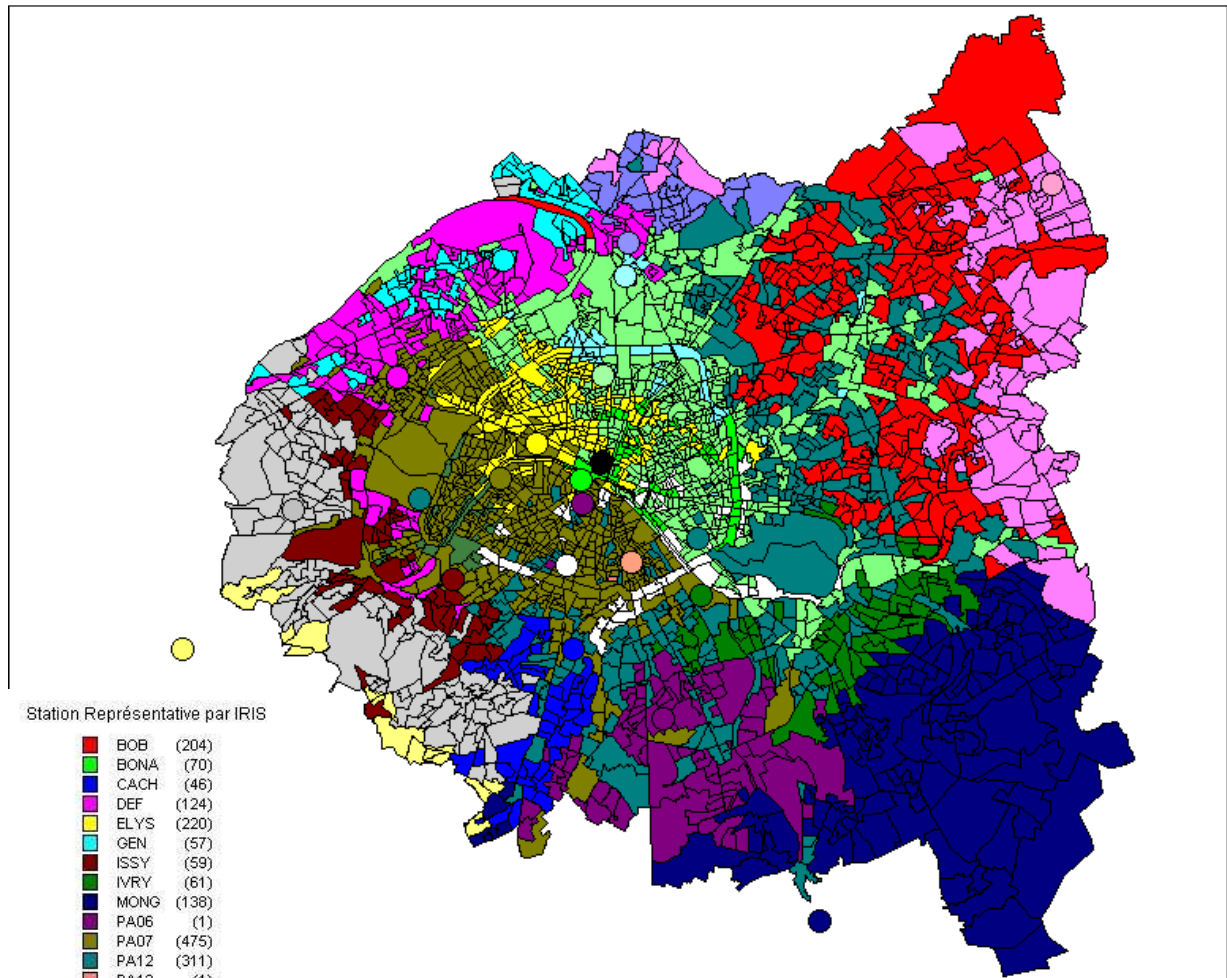
À partir de ces informations, la CAH fournit un groupement d'individus (Cf. illustration ci-dessous). En abscisse figure le nom des IRIS et des stations permanentes utilisées, en ordonnée la distance. Il faut ensuite manuellement choisir le nombre de groupes (traits en rouge), l'important étant que l'inertie inter-classe soit la plus grande possible, mais en même temps que l'intra-classe, soit la plus petite possible, cela correspond à des groupes assez bien séparés tout en restant stables à l'intérieur. Vis-à-vis de notre problématique, il faut également s'assurer que dans chaque groupe figure une station permanente, dont les données seront utilisées pour reconstituer les variations temporelles.

Résultat de la CAH sur Paris et sa Petite couronne : en rouge choix des 7 groupes :



Au final, on obtient des groupes d'IRIS et de stations permanentes. Sur Paris et la Petite Couronne, 7 groupes ont été définis. Ils sont en lien avec la typologie des stations permanentes (stations trafics, de fond)

Une sélection finale est réalisée afin d'associer une seule station permanente à chaque IRIS. En effet, les groupes définis par CAH peuvent contenir plusieurs stations permanentes. La station représentative retenue est celle la plus proche géographiquement. Le résultat final de la station représentative est présenté à la figure suivante.



Station Représentative par IRIS

■	BOB	(204)
■	BONA	(70)
■	CACH	(46)
■	DEF	(124)
■	ELYS	(220)
■	GEN	(57)
■	ISSY	(59)
■	IVRY	(61)
■	MONG	(138)
■	PA06	(1)
■	PA07	(475)
■	PA12	(311)
■	PA13	(1)
■	PA18	(465)
■	STDE	(32)
■	TREM	(94)
■	VERS	(15)
■	AutA1	(45)
■	BASCH	(69)
■	BPAut	(28)
■	GARCH	(120)
■	VITRY	(81)

5. Adaptation géostatistique des concentrations spatialisées aux années non modélisées

5.3. Démarche générale

Pour reconstituer les concentrations moyennes annuelles de NO₂ de 2002 à 2009, la démarche consiste à spatialiser les mesures des stations annuelles en utilisant une méthode d'interpolation spatiale. L'objectif est de kriger les données locales en fonction de la différenciation spatiale du modèle de référence. Après des tests (krigeage simple, krigeage des innovations...), le krigeage avec dérive externe a été retenu.

5.4. Description simplifiée du krigeage avec dérive externe

Les relations entre le modèle et la mesure diffèrent selon les méthodes. Pour le krigeage avec dérive externe, la relation linéaire entre la mesure et le modèle est ajustable localement par voisinage. Il s'agit d'une fonction linéaire locale (dans un voisinage de x) entre Z et S avec une absence de corrélation spatiale entre S et le résidu R :

$$Z(x) = a_x \cdot S(x) + b_x + R(x)$$

avec :

- Z la variable de concentration à estimer sur le domaine d'étude à un instant t_0 donné. Elle est mesurée sur un ensemble de points d'observation notés x_α ;
- S la simulation issue du modèle pour cet instant t_0 . Elle est définie sur une grille de points notés x_β ;
- R le résidu ;
- les concentrations sont estimées aux points d'une grille notés x_j qui coïncident ici avec les x_β .

À noter que la modélisation du variogramme est non stationnaire et indirecte (par validation croisée) et qu'en première approximation, le modèle de variogramme est ajusté sur le variogramme expérimental.

Pour la cartographie, considérant la relation $Z(x) = a \cdot \text{MODELE}(x) + b + R(x)$ en un point de grille x_j à un instant t_0 , la concentration est estimée par la formule :

$$Z^{KDE}(x_j) = \sum_{x_\alpha \in V_j} \lambda_\alpha^{KDE} \cdot Z(x_\alpha)$$

Le poids du krigeage correspond à :

$$\begin{cases} \sum_{V_j} \lambda_\alpha^{KDE} = 1 \\ S(x_j) = \sum_{V_j} \lambda_\alpha^{KDE} \cdot S(x_\alpha) \end{cases}$$

avec :

- λ_α^{KDE} poids calculés de façon à minimiser la variance d'erreur d'estimation
- V_j comme voisinage d'estimation associé à un point x_j .

5.5. Résultats et limites

De par le faible nombre de valeurs disponibles certaines années (seulement deux stations en 2009 sur la CPA par exemple), la détermination d'un modèle de variogramme pour la structure des résidus est très difficile, car la nuée variographique n'est composée que de quelques points. Dès lors, à la suite de tests par validation croisée et sur les coefficients de corrélation (Tableau 5), les meilleurs résultats ont été obtenus avec le krigeage avec dérive

externe, avec des modèles de résidus qui sont linéaires, exponentiels, sphériques ou encore purement pépitiqes. Dans le cas d'une structure pépitiqie, une fois les concentrations issues des modèles « adaptées » avec les données aux stations des années antérieures ou postérieures, les écarts modèle/mesures ne sont plus corrélés spatialement. Ce résidu pourrait donc être considéré comme aléatoire et la covariance entre ce résidu et les concentrations spatialisées pourrait être considérée comme un bruit autour d'une dérive structurellement bien identifiée constituée par les données modèles.

Tableau 5 : coefficients de corrélation obtenus sur le territoire de Lille par validation croisée avec krigeage avec dérive externe et krigeage des innovations :

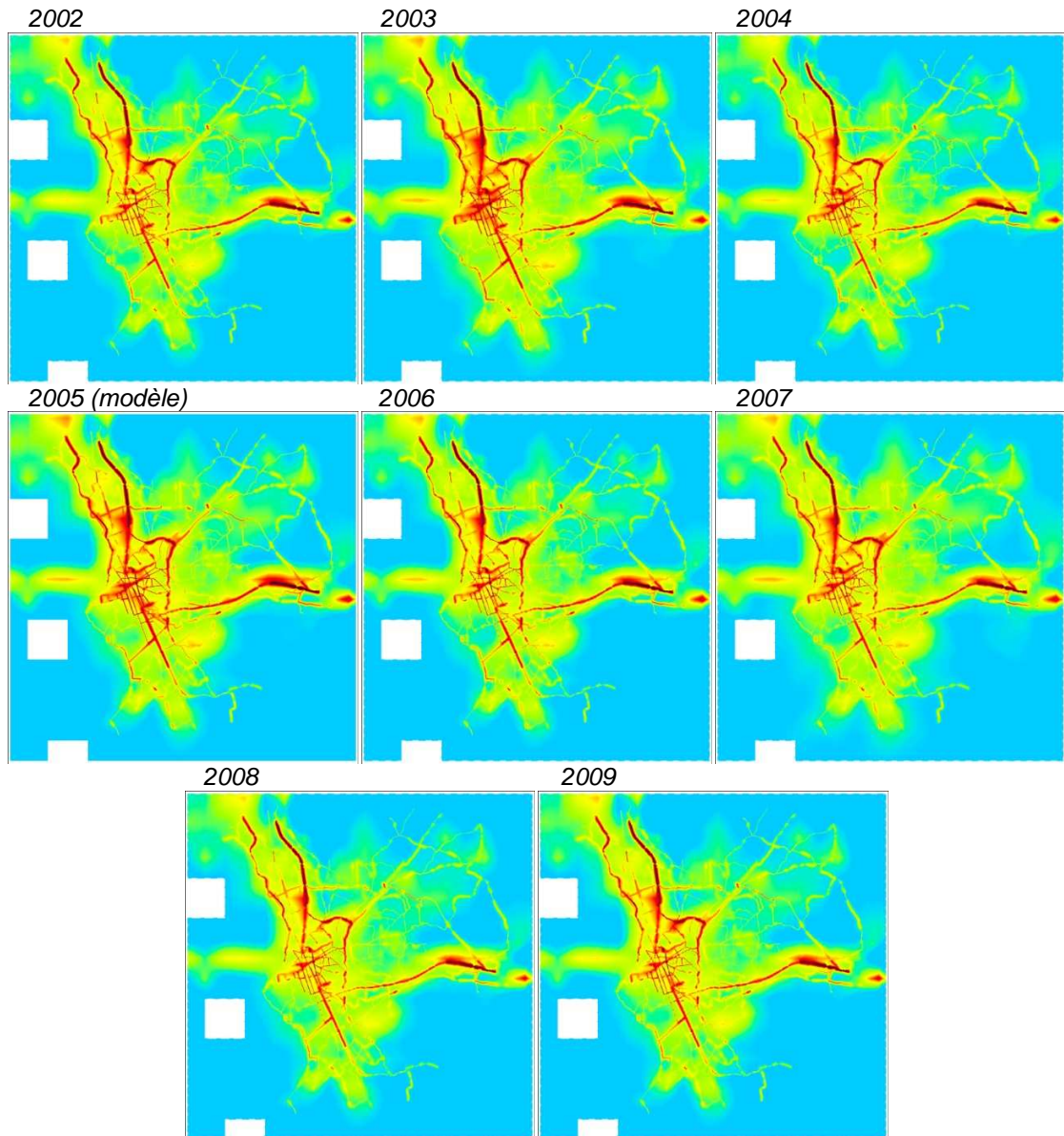
	2002	2003	2004	2005	2006	2007	2008	Moyenne
Krigeage avec dérive externe	0,75	0,55	0,57	0,76	0,50	0,45	0,64	0,60
Krigeage des innovations	0,83	0,62	0,51	0,67	0,51	0,37	0,44	0,56

De plus, une tendance à la sous-estimation des concentrations en limite de domaine est observée, ce qui peut être expliquée par le comportement du krigeage. C'est un interpolateur exact au niveau des données expérimentales, mais qui aura naturellement tendance à extrapoler en bordure de domaine, si aucune autre donnée n'existe aux limites de celui-ci pour quantifier le gradient décroissant des concentrations du centre de l'agglomération vers la périphérie. La structure spatiale de certains domaines, présentant de forts contrastes de concentrations sur de courtes distances, a aussi favorisé la sous-estimation des valeurs élevées sur et à proximité des principaux axes de communication. C'est notamment le cas sur la Communauté du Pays d'Aix. Des stations virtuelles ont été créées pour limiter cette sous-estimation sur les zones les plus sensibles, mais le biais n'a pas été totalement corrigé.

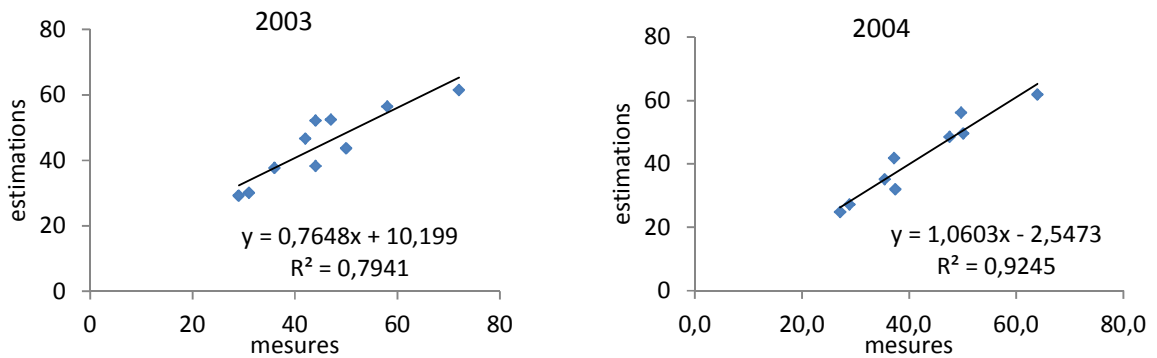
Enfin, l'adaptation des concentrations spatialisées issues de la modélisation sur une année de référence implique une l'hypothèse de rejets de polluants constants, hypothèse qui est naturellement de moins en moins robuste à mesure que l'on s'éloigne dans le temps de l'année de référence. Les évolutions des rejets liées au renouvellement du parc de véhicules, la construction de nouvelles infrastructures, etc. nécessitent en pratique des actualisations des inventaires d'émissions de polluants qui servent à la modélisation.

Néanmoins, la spatialisation des données de 2002 à 2009 a été réalisée sur l'ensemble des domaines avec des résultats de validation croisée satisfaisants :

Cas de Marseille :



Résultats de la validation croisée à Marseille (exemples : cas extrêmes) :



Pour information, le krigeage avec dérive externe a été réalisé avec le logiciel ISATIS ou le logiciel R.

6. Type de données livrées

■ Atmo PACA

Marseille et Communauté du Pays d'Aix (CPA) : données fournies sous forme de tableaux Excel avec comme identifiant commun le numéro DCOMIRIS (département, commune et iris) et sous forme de fichiers shape (*.shp en Lambert II étendu).

■ Air Nord Pas-de-Calais

Lille : données fournies sous forme de fichier shape (*.shp en Lambert II étendu) avec comme identifiant commun le numéro DCOMIRIS (département, commune et iris).

■ AirParif

Paris et Petite Couronne : données fournies sous forme d'un tableau Excel, l'identifiant IRIS étant le DCOMIRIS (département, commune et iris). Un onglet de ce tableau donne d'autres éléments de chaque IRIS (nom, superficie, etc.).

■ Atmo Rhône-Alpes

Grand-Lyon : données fournies sous forme d'un tableau Excel, l'identifiant IRIS étant le DCOMIRIS (département, commune et iris) et sous forme de fichiers MapInfo (.tab en UTM 31 North).