Février 2018

CIRCÉ : CANCER, INÉGALITÉS RÉGIONALES, CANTONALES ET ENVIRONNEMENT

Étude des associations spatiales entre expositions environnementales, socio-économiques et sanitaires en Île-de-France

RAPPORT & ENQUÊTE

OBSERVATOIRE
RÉGIONAL DE SANTÉ
ÎLE-DE-FRANCE

ORS
Île-de-France

CIRCÉ : CANCER, INÉGALITÉS RÉGIONALES, CANTONALES ET ENVIRONNEMENT

Étude des associations spatiales entre expositions environnementales, socio-économiques et sanitaires en Île-de-France

Février 2018

Auteurs

Maylis TELLE-LAMBERTON (ORS Île-de-France)

Valérie FÉRON (ORS Île-de-France)

Adrien SAUNAL (ORS Île-de-France)

Sous la direction d'Isabelle GRÉMY (ORS Île-de-France)

Étude collaborative

Étude multicentrique : Bretagne, Île-de-France, Lorraine, Picardie, Rhône-Alpes.

Ce rapport ne concerne que la région Île-de-France

Couverture : crédits photos Fred Romero/Flickr

Remerciements

Nous adressons nos plus vifs remerciements à nos collègues de l'Ineris et des ORS partenaires de ce projet sans qui celui-ci n'aurait pu être mené à bien ; notamment : Julien Caudeville et Mahdi Saïb (Ineris), Emilie Gardeur (ORS Lorraine), Léna Pennognon et Philippe Clappier (ORS Bretagne), Abdoul Sonko (ORS Rhône-Alpes).

Nous remercions également Pierre Goovaerts (BioMedware, université de Floride) réalisateur du logiciel d'analyse spatiale utilisée dans ce projet, qui a introduit la géostatistique dans la thématique et développé les méthodes mobilisées dans cette étude.

Nous remercions Mahdi Saïb, Julien Caudeville et Emilie Gardeur, pour leur relecture attentive et leurs remarques constructives.

Nous remercions Christine Canet, documentaliste à l'ORS Île-de-France, pour son aide précieuse dans l'élaboration de ce document.

Nous remercions Isabelle Grémy, médecin de santé publique et directrice de l'ORS Île-de-France, pour son soutien dans la conduite de ce projet, sa relecture attentive et ses remarques constructives.

Enfin, notre reconnaissance va à l'Agence régionale de santé (ARS) Île-de-France et à la Région Île-de-France dont les financements ont contribué à ce travail.

Relecteurs

- Christine Canet, documentaliste, ORS Île-de-France
- Julien Caudeville, ingénieur, Ineris
- Emilie, Gardeur, Coordinatrice du pôle Études, ORSAS Lorraine
- Mahdi Saïb, statisticien, Ineris, actuellement à l'ARS Île-de-France
- Isabelle Grémy, directrice de l'ORS Île-de-France

Suggestion de citation

Telle-Lamberton M, Féron V, Saunal A, Grémy I. CIRCÉ : Cancer, inégalités régionales, cantonales et environnement. Étude des associations spatiales entre expositions environnementales, socio-économiques et sanitaires en Île-de-France. Paris : Observatoire régional de santé Île-de-France, 2018.

Il peut être reproduit ou diffusé librement pour un usage personnel et non destiné à des fins commerciales ou pour des courtes citations.

Pour tout autre usage, il convient de demander l'autorisation auprès de l'ORS.

RÉSUMÉ

Contexte et objectifs

Le projet CIRCÉ, Cancer, inégalités régionales, cantonales et environnement, conduit dans cinq régions, s'inscrit dans l'objectif suivant du 3e Plan national santé environnement (2015-2019) : « caractériser les inégalités environnementales et territoriales de santé ». Il vise à analyser les inégalités géographiques à partir de croisements écologiques entre données sanitaires, environnementales et sociales et plus particulièrement les zones associant des valeurs de taux élevées de mortalité et des corrélations locales avec les indicateurs environnementaux et sociaux. Ce rapport présente les résultats pour la région Île-de-France.

Méthodes

Les indicateurs utilisés pour les 258 cantons d'Île-de-France et utilisés pour les études d'association sont les suivants : taux de mortalité par cancer pour 17 localisations chez les hommes, les femmes et les deux sexes confondus, standardisés sur la population française de 2006 et calculés pour la période 2002-2011 ; indicateur de défaveur sociale FDep08 ; expositions aux particules PM10 et expositions au cadmium, au plomb, au nickel et au chrome (concentration atmosphérique et dose journalière d'exposition cumulant les doses par ingestion des différentes voies de transfert : eau, sol, aliments).

La méthode comprend les étapes suivantes, communes aux cinq régions : (i) exploration des corrélations entre les taux standardisés de mortalité non lissés, puis lissés par krigeage de Poisson, les indicateurs environnementaux et l'indice de défaveur sociale, pour identifier les indicateurs à retenir ; (ii) régression linéaire non spatiale du risque de mortalité à partir des indicateurs environnementaux et de l'indice de défaveur sociale pour identifier la combinaison des variables d'intérêt pour l'analyse spatiale ; (iii) régression géographiquement pondérée (GWR) prenant en compte les variations spatiales au niveau loco-régional ; (iv) analyse des résultats obtenus pour détecter au sein des zones où les indicateurs sont corrélés les niveaux de mortalité et d'exposition les plus élevés et les plus faibles.

Résultats

La régression aspatiale a identifié une association entre la mortalité par cancer du poumon chez les hommes et la défaveur sociale. La régression géographiquement pondérée entre ces deux variables a mis en évidence deux zones où le modèle était le plus performant : une zone nord-ouest sud-est centrée sur Paris et deux cantons au sud de la région, zones à inégalités territoriales sociales et de santé caractérisées. Huit zones de points « chauds » ont été identifiées au nord de Paris, à l'est et dans le Val-de-Marne, et un au sud de la région ainsi que 29 zones de points « froids » figurant dans Paris ouest, dans l'ouest et le sud des Hauts-de-Seine et trois dans le Val-de-Marne.

Conclusion

Cette démarche exploratoire de recherche de zones d'inégalités a produit des résultats cohérents avec la littérature, à savoir une association entre cancer du poumon chez les hommes et défaveur sociale. Elle a permis d'identifier des zones où la corrélation est particulièrement forte.

SOMMAIRE

1 | INTRODUCTION 9
2 | CONTEXTE ET OBJECTIFS 11
2.1 Contexte 11
2.2 Objectif du présent rapport 11
3 | MÉTHODOLOGIE 13
3.1 Préambule : les études de corrélation écologique 13
3.2 Matériel 14
3.2.1 Échelle géographique d'analyse : le canton 14
3.2.2 Indicateurs utilisés 15
3.3 Méthode 19
3.3.1 Lissage des indicateurs sanitaires 19
3.3.2 Analyse des corrélations écologiques 19
3.3.3 Le déroulé du traitement géostatistique 20
3.3.4 Logiciel utilisé 21
4 | RÉSULTATS 23
4.1 Taux standardisés de mortalité 23
4.2 Liens entre mortalité par canton et expositions environnementales et sociales 26
4.2.1 Première étape : liens entre taux et expositions sans prise en compte des effets de voisinage entre cantons 26
4.2.2 Deuxième étape : prise en compte des effets de voisinage entre cantons par lissage spatial 27
4.2.3 Détermination des associations (expositions, mortalité) à retenir pour l'analyse par regroupement de cantons 32
4.3 Identification des regroupements de cantons où l'association entre mortalité et exposition est forte 34
4.3.1 Préliminaires 34
4.3.2 Résultats 36
4.4 Identification des points « chauds » (mortalité et défaveur élevées) et « froids » au sein des regroupements de cantons. 42
5 | DISCUSSION-CONCLUSION 45
5.1 Synthèse des résultats 45
5.2 Limites 46
5.3 Perspectives 47
LISTE DES SIGLES ET ACRONYMES 49
BIBLIOGRAPHIE 51
ANNEXES 55
Annexe 1 : Biais écologique, données fictives illustratives, radon domestique et cancer du poumon 55
Annexe 2 : les 258 cantons d'Île-de-France 57
Annexe 3 : Compléments méthodologiques 60
Annexe 4 : Corrélations entre les expositions et les taux de mortalité des cantons d'Île-de-France 66
Annexe 5 : Régression géographiquement pondérée entre la défaveur sociale et le taux de mortalité par cancer du poumon chez les hommes d'Île-de-France. Résultats détaillés. 73

LISTE DES TABLEAUX

Tableau 1. Pathologies retenues dans le projet CIRCÉ 15
Tableau 2. Taux standardisés de mortalité en Île-de-France - période 2002-2011..... 25
Tableau 3. Coefficients de corrélations linéaires entre les taux de mortalité standardisés des cantons d'Île-de-France (hommes) et les facteurs sociaux et environnementaux supérieurs à 0,35..... 26
Tableau 4. Coefficients de corrélations linéaires entre les taux de mortalité standardisés des cantons d'Île-de-France lissés (hommes) et les facteurs sociaux et environnementaux supérieurs à 0,35..... 30
Tableau 5. Coefficients de corrélations linéaires entre les taux de mortalité standardisés des cantons d'Île-de-France lissés (femmes) et les facteurs sociaux et environnementaux 30
Tableau 6. Coefficients de corrélations linéaires entre les taux de mortalité standardisés des cantons d'Île-de-France lissés (hommes et femmes) et les facteurs sociaux et environnementaux..... 31
Tableau 7. Coefficients de détermination des modèles linéaires expliquant les taux de mortalité standardisés et lissés par l'exposition, pour les associations avec un coefficient de corrélation supérieur à 0,35 (hommes)..... 32
Tableau 8. Coefficients de détermination des modèles linéaires expliquant les taux de mortalité standardisés et lissés par l'exposition, pour les associations avec un coefficient de corrélation supérieur à 0,35 (femmes)..... 33
Tableau 9. Coefficients de détermination des modèles linéaires taux de mortalité standardisé lissé~ exposition pour les associations avec un coefficient de corrélation supérieur à 0,35 (hommes et femmes) 33
Tableau 10. Zones de points « chauds » associant des taux de mortalité par cancer du poumon chez les hommes élevés, un indice de défaveur élevé (FDep08) et une association spatiale entre les deux dans un rayon de 30 km 42
Tableau 11. Zones de points « froids » associant des taux de mortalité par cancer du poumon chez les hommes faibles, un indice de défaveur faible (FDep08) et une association spatiale entre les deux dans un rayon de 30 km 44
Tableau 12. Coefficients de corrélations linéaires entre les taux de mortalité standardisés des cantons d'Île-de-France (hommes) et les facteurs sociaux et environnementaux... 67
Tableau 13. Coefficients de corrélations linéaires entre les taux de mortalité standardisés des cantons d'Île-de-France (femmes) et les facteurs sociaux et environnementaux.... 68
Tableau 14.Coefficients de corrélations linéaires entre les taux de mortalité standardisés des cantons d'Île-de-France (hommes et femmes) et les facteurs sociaux et environnementaux..... 69
Tableau 15. Coefficients de corrélations linéaires entre les taux de mortalité standardisés des cantons d'Île-de-France lissés (hommes) et les facteurs sociaux et environnementaux 70
Tableau 16. Coefficients de corrélations linéaires entre les taux de mortalité standardisés des cantons d'Île-de-France lissés (femmes) et les facteurs sociaux et environnementaux 71
Tableau 17. Coefficients de corrélations linéaires entre les taux de mortalité standardisés des cantons d'Île-de-France lissés (hommes et femmes) et les facteurs sociaux et environnementaux..... 72

LISTE DES CARTES

Carte 1. Dose journalière d'exposition au chrome dans les cantons d'Île-de-France..... 17
Carte 2. Indice de défaveur sociale (FDep08) dans les cantons d'Île-de-France..... 18
Carte 3. Taux de mortalité standardisé par cancer du poumon chez les hommes, période 2002-2011 avant lissage..... 27
Carte 4. Taux de mortalité standardisés par cancer du poumon chez les hommes, période 2002-2011, lissés par krigeage de Poisson..... 28
Carte 5. Variance associée au krigeage des taux de mortalité standardisés par cancer du poumon chez les hommes, période 2002-2011..... 28

Carte 6. Indice de défaveur sociale (FDep08) dans les cantons d'Île-de-France 35
Carte 7. Taux de mortalité lissés par cancer du poumon chez les hommes entre 2002 et 2011 35
Carte 8. Coefficient de détermination associé au paramètre du modèle de régression géographiquement pondérée entre le taux de mortalité lissé par cancer du poumon chez les hommes entre 2002 et 2011 et la défaveur sociale (FDep08) (R2).....37
Carte 9. Corrélations entre le taux de mortalité lissé par cancer du poumon chez les hommes entre 2002 et 2011 et la défaveur sociale (FDep08).....38
Carte 10. Significativité du coefficient associé à l'exposition dans le modèle de régression géographiquement pondérée entre le taux de mortalité lissé par cancer du poumon chez les hommes entre 2002 et 2011 et la défaveur sociale (FDep08) 38
Carte 11. Défaveur sociale en 2008 (Fdep08) et zones où le modèle explique plus de 70% de la variance (en vert foncé) et 50% de la variance (en vert clair) 39
Carte 12. Taux de mortalité lissé par cancer du poumon chez les hommes entre 2002 et 2011 et zones où le modèle explique plus de 70% de la variance (en vert foncé) et 50% de la variance (en vert clair).....40
Carte 13. Carte des cantons d'Île-de-France pour lesquels une association robuste entre défaveur sociale et cancer du poumon chez les hommes est observée dans un rayon de 30 km 41
Carte 14. Zones de points « chauds » et froids associant des taux de mortalité par cancer du poumon chez les hommes élevés, un indice de défaveur élevé (FDep08) et une association spatiale entre les deux dans un rayon de 30 km 43
Carte 15. Les cantons d'Île-de-France de Paris et sa proche couronne 57
Carte 16. Les cantons d'Île-de-France en grande couronne 58

LISTE DES FIGURES

Figure 1. Démarche générale adoptée pour l'étude CIRCÉ..... 8
Figure 2. Critère d'information d'Akaïke (AIC) dans la régression géographiquement pondérée, en fonction de la distance.....36
Figure 3. Biais écologique : données fictives mais plausibles. Radon domestique et cancer du poumon dans quatre régions fictives. Données groupées.....55
Figure 4. Biais écologique : données fictives mais plausibles. Radon domestique et cancer du poumon dans quatre régions fictives. Données individuelles. 56
Figure 5. Distribution de la population des cantons d'Île-de-France en 2006.....59
Figure 6. Scattergrammes entre les valeurs de pH du sol séparées par une distance de 1 m, 2 m, 3 m et 4 m (58) 61
Figure 7. Semi-variogramme des taux de mortalité par cancer du cerveau entre 50 et 69 ans calculés à l'aide de l'estimateur pondéré en fonction de la population, femmes blanches, États-Unis, Nouvelle-Angleterre, 295 comtés, période 1950-1969 (64). 63
Figure 8. Semivariogrammes des taux de mortalité par cancer du sein en Nouvelle-Angleterre et modèle ajusté associé 64

Figure 1. Démarche générale adoptée pour l'étude CIRCÉ

Adaptée de (1)

1 | INTRODUCTION

L'un des objectifs du premier Plan national santé environnement – PNSE - (2004-2008) est de prévenir les pathologies d'origine environnementale et notamment les cancers. Dans la continuité, le deuxième PNSE (2009-2013) propose des mesures visant à réduire les expositions responsables de pathologies à fort impact sur la santé (cancers, maladies cardiovasculaires, pathologies respiratoires, neurologiques…). Un objectif affiché du troisième PNSE enfin (2015-2019) est de réduire autant que possible et de la façon la plus efficace les impacts des facteurs environnementaux sur la santé afin de permettre à chacun de vivre dans un environnement favorable à la santé.

Le projet CIRCÉ vise à caractériser les inégalités territoriales environnementales, sociales et sanitaires de cinq régions françaises.

C'est ainsi qu'a été initié en 2004, le projet CIRCÉ (Cancer inégalités régionales cantonales et environnement). Il regroupe actuellement l'unité Impact sanitaire et expositions (ISAE) de l'Institut national de l'environnement industriel et des risques (Ineris) et les Observatoires régionaux de santé de cinq régions françaises[1] (Bretagne, Île-de-France, Lorraine, Picardie et Rhône-Alpes). Les ORS apportent leur expérience dans l'exploitation des bases de données sanitaires et leur connaissance des territoires régionaux. L'Ineris apporte son expertise sur la caractérisation des inégalités environnementales et le traitement des données spatialisées.

L'objectif général du projet CIRCÉ est d'identifier les inégalités géographiques du cancer, d'évaluer leurs relations spatiales avec les facteurs sociaux et environnementaux et de développer un SIG (Système d'informations géographiques) sur les cancers et les expositions environnementales.

Le projet CIRCÉ s'organisait en trois phases :

- La phase 1 était composée de deux étapes, l'une réalisée par les ORS : produire des indicateurs sanitaires (de mortalité et de morbidité) et socio-économiques dans chacune des régions étudiées, et l'autre réalisée par l'Ineris : produire des indicateurs spatialisés à partir des principales bases de données environnementales sur l'ensemble des territoires étudiés (2) (cf Figure 1).

- La phase 2 consistait en une analyse des inégalités géographiques à partir des croisements écologiques de données sanitaires, environnementales et sociales.

- La phase 3 enfin visait à mettre en évidence les zones contrastées associant des valeurs de taux élevées et des corrélations locales statistiquement significatives.

La première phase a donné lieu en Île-de-France à deux publications portant sur la mortalité par cancer, l'une pour la période 1990-1999, l'autre pour la période 2000-2007 et à la réalisation d'une cartographie de l'exposition aux métaux (3-5).

Le présent rapport porte sur les phases 2 et 3.

Il s'agit d'une démarche collective, innovante par les approches géostatistiques qu'elle propose. Toutefois, la méthode écologique demande de très grandes précautions en raison de ses limites précisées plus loin.

---

[1] Selon la définition des régions au moment du début du projet, soit en 2004.

2 | CONTEXTE ET OBJECTIFS

2.1 Contexte

À la suite de la publication des atlas sur la mortalité par cancer par les ORS et des atlas sur les expositions aux éléments traces métalliques (ETM) par l'Ineris, s'engagent les phases 2 et 3 du projet relatives aux croisements des données sanitaires, environnementales et sociales et à la mise en évidence de zones contrastées. Les ORS de quatre régions, Bretagne, Île-de-France, Lorraine et Rhône-Alpes se sont associés pour ce travail en collaboration avec l'Ineris.

Outre une analyse « tous cancers », les localisations retenues sont celles pour lesquelles on soupçonne des facteurs de risques environnementaux : lèvres, bouches et pharynx (LBP) ; œsophage ; côlon-rectum ; estomac ; foie ; pancréas ; larynx ; trachée, bronches et poumon [1]; plèvre ; prostate ; sein chez la femme ; ovaire ; vessie ; système nerveux central ; lymphome malin non hodgkinien et leucémies (toutes leucémies confondues).

2.2 Objectif du présent rapport

Le présent rapport a pour objectif d'une part d'analyser les associations spatiales entre expositions et santé par une étude de corrélation écologique au niveau cantonal et d'autre part de mettre en évidence des zones contrastées où le canton central cumule des valeurs d'exposition élevées et des indicateurs sanitaires défavorables.

Ce rapport présente une analyse des associations spatiales entre les expositions environnementales et sociales et le risque de mortalité par cancer en Île-de-France.

Ce travail se base sur ceux effectués dans la région Picardie dans le cadre d'une thèse de recherche « Construction d'une méthode de caractérisation des inégalités territoriales, environnementales et sociales de santé » (1) menée en collaboration avec l'Ineris, l'OR2S (Picardie) et l'INSERM et pour laquelle la région Picardie était la zone d'étude pilote. Ainsi dans le cadre de ce travail, les indicateurs et les méthodes géostatistiques utilisées sont déclinées dans les quatre régions impliquées.

---

[1] résumé par poumon par la suite

3 | MÉTHODOLOGIE

3.1 Préambule : les études de corrélation écologique

Les études de corrélation écologique sont de plus en plus utilisées pour analyser le lien entre indicateurs environnementaux, sociaux et sanitaires. Elles peuvent aider à générer des hypothèses quant aux effets d'une exposition sur la santé à des niveaux géographiques agrégés qui peuvent dans un second temps amorcer une réflexion sur la causalité de ces effets. Elles présentent à la fois des forces et des limites.

Forces des méthodes de corrélation écologique :

- **l'utilisation de données déjà disponibles** grâce aux nombreuses bases régionales ou nationales géoréférencées et collectées en routine qui existent en France : au niveau environnemental, différentes agences, instituts ou observatoires assurent par exemple la surveillance de la qualité des eaux, des sols ou encore de l'air. Pour l'étude de l'état de santé des populations, il existe de nombreuses bases de données : causes de décès (centralisées par le Cépi-DC de l'INSERM[1]), morbidité hospitalière (PMSI[2]), morbidité médico-économique (ALD[3]), registres de maladies (cancers, malformations congénitales, cardiopathies ischémiques,…). Pour caractériser les aspects sociodémographiques, les données de population sont fournies par les recensements de la population et sont disponibles auprès de l'Institut national de la statistique et des études économiques (Insee[4]).

Les études de corrélation écologique, permettent de détecter des zones de cumul, c'est-à-dire où les indicateurs environnementaux, sociaux et sanitaires sont tous défavorables (points chauds) ou au contraire favorables (points froids).

- l'utilisation, à l'échelle d'unités administratives, de données agrégées et d'estimations de facteurs de risque pour l'ensemble de la population résidant dans une zone géographique, ce qui permet de bénéficier d'une puissance statistique importante (du fait de la taille conséquente des populations prises en compte).

- **l'objectivation de contrastes d'exposition plus élevés** que ceux qui auraient été mesurés a priori au niveau individuel.

- **l'aide à la génération d'hypothèses,** qui pourront être testées par des études étiologiques adaptées.

Limites des méthodes de corrélation écologique :

- **le biais écologique :** des relations existant au sein des unités géographiques étudiées peuvent disparaître lorsque les données sont agrégées dans ces unités du fait de la présence de facteurs de confusion (illustration en annexe 1).
- **la prise en compte de la latence entre expositions et pathologies :** les expositions environnementales historiques sont difficiles à reconstituer.
- **la mobilité des populations :** en considérant que l'on parvient à constituer une bonne approximation de l'exposition passée sur le territoire, les populations

Les études de corrélation écologique comportent de nombreux biais qui imposent une grande prudence dans l'interprétation des résultats.

---

[1] Cépi-DC : Centre **d'**épidémiologie sur les causes médicales décès
[2] PMSI : Programme de médicalisation des systèmes d'information
[3] ALD : Affections de longue durée
[4] Insee : Institut national de statistique et des études économiques. https://www.Insee.fr/fr/information/2008354, accédé le 02/11/2017

peuvent avoir changé de résidence entre leur période d'exposition et l'apparition de leur pathologie.

Du fait de ces limites, nous ne nous situons pas dans la recherche d'une relation causale dans cette étude mais dans la démarche qui suit.

Les études de corrélation écologique, permettent d'identifier des zones où les indicateurs environnementaux, sociaux et sanitaires sont tous défavorables (points chauds) ou au contraire favorables (points froids). L'analyse permet alors de mettre en évidence des zones de cumuls d'inégalité.

La recherche d'éventuelles explications causales demanderait des études approfondies des biais, de leur importance et de leur direction. En l'absence, la plus grande prudence est requise dans l'interprétation des résultats.

3.2 Matériel

3.2.1 Échelle géographique d'analyse : le canton

Dans la continuité des atlas régionaux de la mortalité par cancer produits dans la phase 1 du projet CIRCÉ, l'analyse des relations spatiales entre les différents indicateurs a été réalisée à l'échelle cantonale, en utilisant le découpage canton-ou-ville[1] (ou pseudo-cantons dans sa définition avant le redécoupage cantonal intervenu en 2014) de l'Insee. La carte des 258 cantons-villes d'Île-de-France et la distribution de leur population en 2006, année médiane de la période, figurent en annexe 2. Dans la suite du corps du document nous simplifierons en utilisant le terme canton pour désigner cette unité géographique.

La majorité des cantons d'Île-de-France compte entre 20 000 et 40 000 habitants. Un seul canton fait moins de 10 000 habitants, Villiers-Saint-Georges en Seine-et-Marne, qui compte 8 300 habitants et 15 en comptent plus de 100 000, il s'agit des arrondissements périphériques de Paris et, par ordre de population décroissante de Saint-Denis, Boulogne-Billancourt, Versailles, Argenteuil, Montreuil.

---

[1] « À la différence du canton, le « canton-ou-ville » (ou pseudo-canton) est un regroupement d'une ou plusieurs communes entières. Dans les agglomérations urbaines, chaque canton comprend en général une partie de la commune principale et éventuellement une ou plusieurs communes périphériques. Dans ce cas, l'Insee considère la commune principale, entière, comme un pseudo-canton unique et distinct. Pour la ou les communes périphériques, le pseudo-canton considéré est alors identique au vrai canton amputé de la fraction de la commune principale que comprend le vrai canton. » (Insee : https://www.insee.fr/fr/metadonnees/definition/c1725, accédé le 02/11/2017

3.2.2 Indicateurs utilisés

Indicateurs sanitaires

L'indicateur le plus couramment utilisé pour décrire la mortalité par cancer est le taux de mortalité sur une période donnée, défini par le nombre de décès survenu dans cette période rapporté à la population à risque sur l'ensemble de la période (personnes-années). La mortalité par cancer est très dépendante de l'âge et les populations des différents cantons peuvent avoir des structures d'âge différentes. Pour tenir compte de cette différence de structure, on utilise des taux standardisés, calculés à partir des taux spécifiques à chaque classe d'âge et égaux à la somme pondérée de ces taux. Les poids utilisés correspondent à ceux d'une population de référence, ici la population française de 2006, les deux sexes confondus, avec 21 tranches d'âges (< 1 an, 1-4 ans, 5-9 ans …, 90-94 ans, ≥ 95 ans).

Les localisations retenues sont décrites dans le Tableau 1.

Tableau 1. Pathologies retenues dans le projet CIRCÉ

Localisations cancéreuses	Code CIM10
Tous cancers	C00-C97
Lèvre cavité buccale pharynx	C00-C14
Œsophage	C15
Estomac	C16
Colon-rectum	C18-C21
Foie	C22
Pancréas	C25
Larynx	C32
Poumon	C33-C34
Plèvre	C384, C450, C459
Sein	C50
Ovaire	C56, C570-C574, C578
Prostate	C61
Vessie	C67
Rein	C64, C66-68
Système nerveux central	C70-C72
Lymphome non Hodgkinien	C82-C85, C96
Toutes leucémies	C91 -C95

Indicateurs d'exposition environnementale

Les données utilisées reposent sur les travaux menés par l'Ineris concernant l'exposition aux polluants de la famille des éléments traces métalliques. Quatre polluants ont été retenus en raison de leurs effets sanitaires et du fait de l'accessibilité de leurs mesures dans les différents compartiments environnementaux[1] (6). Il s'agit du plomb, du nickel, du cadmium et du chrome (chrome total) : « une des particularités de ces éléments est d'entraîner des effets majeurs sur la santé humaine (caractère toxique des substances) et/ou sur les écosystèmes (caractère écotoxique des substances) en cas d'exposition » (7, 8).

---

[1] Le Centre international de recherche sur le cancer (CIRC) a classé le cadmium, le nickel et ses composés, et le chrome VI parmi le groupe 1 (cancérigène certain pour l'homme), le plomb et ses dérivés inorganiques dans le groupe 2A (potentiellement cancérigène pour l'homme).

La cartographie de ces quatre éléments traces métalliques a été réalisée dans le cadre du développement de Plaine (Plateforme intégrée pour l'analyse des inégalités environnementales) (9-11) ; la résolution spatiale initiale de cartographie de ces indicateurs est une géographie carroyée par maille de 1km². Dans la présente étude, les indicateurs utilisés ont été agrégés par ratio surfacique à l'échelle des cantons et les données ont été transmises aux ORS à l'échelle des cantons.

L'exposition par inhalation est caractérisée par la concentration atmosphérique (CA) et l'exposition par ingestion est caractérisée par la Dose journalière d'exposition[1] (DJE), combinaison des expositions relatives à l'ensemble des milieux d'exposition par ingestion (eau, sol, aliments) et construite à partir du modèle d'exposition multimedia MODUL'ERS (11).

À titre d'exemple la Carte 1 présente la moyenne des DJE au chrome pour la population de 0 à 70 ans.

Pour le cadmium, le nickel et le plomb, pour lesquels on dispose de valeurs toxicologiques de référence (VTR), les quotients de danger par inhalation (CA/VTR) et par ingestion (DJE/VTR) ont été calculés ainsi qu'un indice spatialisé relatif (ISR), sommant les deux quotients de danger. Un indice composite a également été calculé : l'ISR total, sommant les trois ISR : cadmium, nickel et plomb.

Pour le chrome, seul le chrome hexavalent (chrome VI) est classé par le CIRC[2] (cancérogène certain) et celui-ci ne représente qu'une proportion réduite de la teneur en chrome total que l'on ne connait pas. Il a donc été décidé de ne pas calculer de quotient de danger et de ne pas inclure la composante chrome dans l'indice composite précité.

Encadré 1.La plate-forme Plaine de l'Ineris

Afin de permettre aux pouvoirs publics de réduire les inégalités face à la pollution environnementale, l'Ineris cartographie le degré d'exposition des populations aux contaminations de l'air, de l'eau, des sols… et ce, sur tout le territoire. L'institut a ainsi développé la Plate-forme d'analyse des inégalités environnementales (Plaine). Elle intègre, sur une trame géographique, des données de nature diverse sur les différentes contaminations de l'air, de l'eau, des sols et de l'alimentation, ceci pour différents polluants. Elle a d'ores et déjà permis d'élaborer un premier atlas régional des inégalités environnementales intégrant l'exposition à quatre métaux (cadmium, plomb, nickel et chrome). Des travaux plus approfondis dans certaines régions ont aussi permis aux pouvoirs publics de sélectionner des zones d'action immédiate et des zones d'investigations sanitaires et sociales. Enfin, Plaine est utilisée dans des programmes de recherche, et de nouvelles méthodes d'agrégation des voies d'exposition devraient améliorer les indicateurs intégrés afin qu'ils rendent mieux compte de la réalité biologique des expositions environnementales suspectées d'avoir un lien avec la santé humaine : c'est le concept d'exposome. L'harmonisation des méthodes et outils ainsi que le partage des données et des indicateurs au niveau national sont des pré-requis pour développer des démarches opérationnelles cohérentes sur l'ensemble du territoire. Des recommandations sont formulées en ce sens dans le troisième Plan national santé environnement.[3]

---

[1] Dose journalière d'exposition (DJE) : Dose de substance reçue par l'organisme rapportée au poids de l'individu et au nombre de jours d'exposition (dans le cas d'une substance non cancérigène) ou au nombre de jours de la vie entière (dans le cas d'une substance cancérigène) : définition Ineris. Unité : µg/kg/j

[2] http://monographs.iarc.fr/ENG/Classification/latest_classif.php, accédé le 27/10/2017

[3] https://www.ineris.fr/fr/les-risques/dossiers-thematiques/inegalites-environnementales accédée le 19/12/2017

Carte 1. Dose journalière d'exposition au chrome dans les cantons d'Île-de-France

Indice de défaveur sociale

Les indices de « désavantage ou défaveur social(e) » (« deprivation index » en anglais) sont couramment utilisés pour analyser les inégalités sociales de santé. Les deux indicateurs les plus utilisés et connus sont ceux de Townsend (12) et Carstairs (13). Ceux-ci ont cependant été construits dans des contextes majoritairement urbains et anglo-saxons et sont peu adaptés au mode d'organisation socio-spatiale du territoire français du fait des variables qui le composent (pourcentage des foyers sans voiture, chômage, classe sociale et/ou ménages non propriétaires de leur logement, logement surpeuplé).

En 2011, Rey *et al.* ont proposé une méthode de calcul d'un indice sur l'ensemble de la France métropolitaine : le FDep 99. Construit à l'échelle des communes, des cantons et des régions, le FDep 99 est basé sur quatre variables issues du recensement de la population de 1999 et des données de déclaration d'impôt de 2001 (14) : la part des ouvriers dans la population active, la part des diplômés de niveau baccalauréat chez les 15 ans et plus, la part des chômeurs dans la population active et le revenu médian par foyer. L'indice FDep99 a été défini comme la première composante de l'analyse en composante principale (ACP) de ces quatre variables, pondérée par la taille de la population de chaque commune. Ces quatre variables représentent toutes une dimension fondamentale du niveau socio-économique, de signification homogène dans l'ensemble du pays. Afin de tenir compte des modalités variées des structurations socio-spatiales entre zones rurales et zones urbaines, l'ACP a été stratifiée en quatre groupes selon le nombre d'habitants par commune :

- rural (population de la commune inférieure à 2 000 habitants),
- quasi-rural (population supérieure ou égale à 2 000 et inférieure à 10 000 habitants),
- quasi-urbain (population supérieure ou égale à 10 000 et inférieure à 100 000 habitants),
- urbain (population supérieure ou égale à 100 000 et inférieure à 2 000 000 habitants).

Par ailleurs les auteurs ont montré une homogénéité de l'association entre le FDep99 et la mortalité toutes causes, quel que soit le contexte rural/urbain ou la région, illustrant le caractère discriminant de cet indice sur l'ensemble du territoire.

Enfin, le Haut Conseil de santé publique (HCSP) préconise l'utilisation du FDep afin de caractériser les déterminants socioéconomiques pour l'étude des inégalités sociales de santé (15). Dans le cadre du travail de la thèse précitée et en collaboration avec l'université de Nanterre, l'indicateur de défaveur a été construit sur la base des quatre variables utilisées pour la construction du FDep99 mais il a été décliné à l'échelle des IRIS[1] puis agrégés aux cantons à partir des données de recensement de l'année 2008 (16). Cet indicateur actualisé est dénommé FDep08 par analogie avec le FDep99.

La Carte 2 représente la défaveur sociale dans les cantons d'Île-de-France et fait apparaître les contrastes habituellement constatés : le département de Seine-Saint-Denis et les communes limitrophes apparaissent comme les plus défavorisés, suivis de près par l'est de la Seine-et-Marne ; à l'inverse l'ouest parisien et l'est des Yvelines sont les zones les plus favorisées (17).

Dans la suite du document nous utiliserons le terme environnemental pour désigner à la fois les expositions physicochimiques, issues de l'environnement au sens strict et la défaveur sociale, composante de l'environnement envisagé dans une définition plus large.

Carte 2. Indice de défaveur sociale (FDep08) dans les cantons d'Île-de-France

Sources : Références (1, 14, 16)

---

[1] Les IRIS ou îlots regroupés pour l'information statistique constituent la plus petite unité géographique en France pour laquelle les données démographiques et socio-économiques du recensement national sont disponibles.

3.3 Méthode

3.3.1 Lissage des indicateurs sanitaires

Un des enjeux pour obtenir une représentation cartographique optimale d'un indicateur sanitaire est le choix de la taille des unités géographiques. Celle-ci doit combiner les deux contraintes suivantes : les zones doivent être suffisamment petites pour permettre une vue assez précise des caractéristiques géographiques de l'incidence ou de la mortalité de la pathologie, mais de taille suffisante pour limiter l'incertitude causée par des valeurs extrêmes non fiables sur des zones peu peuplées (18, 19). Cet effet est connu sous le nom de "problème du petit nombre". En effet, dans les unités géographiques les moins peuplées, le nombre de cas ou de décès observé peut être faible, ce qui induit une importante instabilité du taux calculé, les valeurs les plus élevées ou les plus faibles étant plus fréquemment observées dans les zones géographiques les moins peuplées (20). Ce problème est d'autant plus marqué que l'événement étudié est rare et génère de plus fortes variations locales potentiellement dues au hasard. Une stratégie consiste à lisser spatialement les indicateurs calculés par la prise en compte de l'influence des unités géographiques avoisinantes (21).

Plusieurs techniques de lissage spatial d'indicateurs sanitaires ont été utilisées dans la littérature (cf annexe 3). La méthode du **krigeage de Poisson** a été retenue car elle offre un cadre flexible pour modéliser cette variabilité spatiale et estimer le risque sous-jacent en tenant compte à la fois de la forme et la taille des unités géographiques (22). Elle fournit également une variance associée aux estimations obtenues.

On suppose que le nombre de décès durant la période d'observation suit une loi de Poisson, de paramètre égal au produit de la taille de la population par le risque local. Ce risque local peut être conçu comme un taux de mortalité pour une aire, affranchi du « bruit statistique ».

Le krigeage ATA (Area-To-Area) utilise le barycentre de chaque canton pour le calcul des distances estimées à partir de la matrice des contiguïtés cantonales d'ordre 1[1] (deux cantons sont considérés comme voisins s'ils ont au moins une portion de frontière commune). Ces distances sont ensuite utilisées pour calculer la portée des points de mesure et identifier les cantons corrélés qui seront utilisés pour calculer les valeurs des taux lissés.

Les taux lissés sont cartographiés en y adjoignant la carte de la variance des estimations. La précision de l'estimation peut en effet être évaluée par le calcul d'une variance d'estimation (appelée variance de krigeage). La variance de krigeage, par définition, quantifie la dispersion possible de la valeur vraie autour de la valeur estimée. Moins le phénomène est structuré et plus sa variabilité est grande, plus la variance de krigeage augmente. Ainsi, les cantons qui présentent une variance très élevée pèsent peu dans l'analyse.

3.3.2 Analyse des corrélations écologiques

Les études de corrélation écologique nécessitent l'utilisation de modèles statistiques ou géostatistiques adaptés parmi lesquels les modèles de type bayésien (23) ou les modèles de régression géographiquement pondérée (Geographically weighted regression, GWR) (24). Ces derniers permettent d'évaluer la variation spatiale des coefficients de régression qui estiment les associations entre les différentes variables et ils fournissent des statistiques permettant de visualiser la répartition spatiale de ces associations (24-26). Ces modèles supposent que différents processus d'association sont à l'œuvre dans la région d'étude et que ces variations peuvent être dues aux hétérogénéités spatiales de

---

[1] La matrice des contigüités, symétrique, est définie de la façon suivante : la cellule (i,j) est égale à 1 si les deux cantons partagent une portion de frontière et à 0 sinon (les cellules (i,i) valent 0 également).

cofacteurs potentiels (comportement, autres facteurs environnementaux, vulnérabilité, mobilité et migration des populations). Ils ont donc été identifiés comme l'une des méthodes les plus adaptées dans le contexte des études de santé (27, 28).

3.3.3 Le déroulé du traitement géostatistique

La méthode résulte d'un travail conjoint entre les experts de l'Ineris et des ORS participant à l'étude, avec l'appui scientifique de Pierre Goovaerts, université de Floride. Elle a été appliquée de façon identique sur les cinq régions. Le croisement des données utilisées est réalisé sans a priori, c'est-à-dire sans tenir compte, a priori, dans le choix des croisements à faire, des facteurs de risques environnementaux évoqués dans les études toxicologiques et épidémiologiques et dans la littérature qui en rendent compte. Elle permet ainsi une analyse spatiale spécifique à chaque région.

La démarche proposée se décompose en six grandes étapes permettant d'une part de réaliser des cartes d'indicateurs sanitaires tenant compte de l'instabilité liée aux petits effectifs (taux lissés) et, d'autre part, de détecter des zones où le risque de mortalité est le plus élevé.

Les grands principes sont décrits ci-dessous et les détails statistiques sont précisés en annexe 3.

1. Exploration des corrélations avec les taux standardisés de mortalité non lissés
   Cette première étape donne une première image des corrélations entre les indicateurs sanitaires et environnementaux. **Le seuil de 0,35, pragmatique, est adopté pour considérer qu'une corrélation a un niveau suffisant**.[1]

2. Lissage des taux standardisés de mortalité
   Ici, le modèle de Poisson (krigeage de Poisson) est utilisé pour corriger l'effet du problème des petits nombres.

3. Examen des corrélations avec les taux standardisés lissés de mortalité
   On considère également que la **corrélation entre deux variables est suffisante si elle est d'au moins 0,35**. Les variables correspondantes seront retenues pour l'étape suivante.

4. Régression non spatiale ou aspatiale
   Chaque association (pathologie, exposition) donne lieu à une régression linéaire pondérée par l'inverse de la variance. En effet, la variance du taux de mortalité d'un canton donné est inversement proportionnelle à la taille de sa population. La régression pondérée permet de tenir compte de ces variations cantonales (29).
   Pour chaque association, la régression cherche à expliquer le taux de mortalité lissé de la pathologie considérée par l'exposition et le Fdep08, celui-ci étant inclus systématiquement du fait de son rôle dans les inégalités de survenue des cancers (30).[2]
   Également à titre pragmatique, les combinaisons de variables qui seront considérées comme « intéressantes » sont celles pour lesquelles la

---

[1] Dans la méthode d'origine, les corrélations entre expositions sont également examinées, pour tenir compte dans les modèles explicatifs des colinéarités entre variables explicatives. Ainsi, des variables très corrélées (coefficient d'au moins 0,8 par exemple) seront « regroupées » et remplacées par leur somme ou toute autre combinaison linéaire pertinente (choisie, par exemple, par analyse en composantes principales) (1).

[2] Dans le cas où le taux considéré est corrélé à plusieurs variables d'exposition physico-chimiques, la méthode d'origine propose d'examiner également les modèles incluant ces variables simultanément, en tenant compte des colinéarités éventuelles (cf note précédente). Le modèle retenu est alors celui pour lequel le critère d'AIC (Akaïke Information Criterion) est le plus faible. Étant donné la complexité d'interprétation des études écologiques mentionnée plus haut, nous avons choisi d'intégrer le nombre minimum de variables dans les modèles, avant d'aller, le cas échéant, vers des analyses plus poussées.

régression aspatiale explique plus de 50% de la variance. Il s'agit d'un seuil pragmatique à partir duquel on considère qu'il peut y avoir une association « réelle » entre les variables environnementales et sanitaires.

5. Régressions géographiques pondérées (GWR)
   Les modèles retenus à l'étape précédente donnent lieu à des régressions géographiques pondérées pour chaque canton. Dans un premier temps, la bande passante est calculée. Ce terme utilisé en géostatistique désigne la distance optimale au-delà de laquelle il n'est plus utile d'intégrer les cantons adjacents pour améliorer la performance du modèle. Les régressions sont alors réalisées en tenant compte des cantons inclus dans le voisinage de rayon égal à cette bande passante.

6. Détermination des zones « froides » et « chaudes »
   Les zones « chaudes » répondent aux quatre critères suivants :
   - le modèle de régression local permet d'expliquer plus de 50% de la variance ;
   - il correspond à une corrélation linéaire d'au moins 0,5 ;
   - les paramètres associés aux variables explicatives sont significatifs au seuil de 5% ;
   - les indicateurs correspondants sont supérieurs au dernier quintile de chacune des distributions.

   On obtient ainsi des zones cumulant exposition et mortalité élevées, et pour lesquelles dans un rayon de X km, X étant la bande passante, les deux indicateurs sont fortement corrélés.
   Les zones « froides » sont définies de façon analogue, avec à l'inverse les deux indicateurs inférieurs aux deux premiers quintiles des distributions.

3.3.4 Logiciel utilisé

SpaceStat 3.0.64

SpaceStat© est un logiciel de système d'information géographique développé par la société BioMedware permettant la visualisation et l'analyse de données espace-temps.

Pour plus d'informations consulter le site de Biomedware : https://www.biomedware.com/?module=Home

GWR4

GWR4 est un logiciel développé par le NCG (National centre for geocomputation) de l'université nationale d'Irlande Maynooth et le département de géographie de l'université de Ritsumeikan au Japon[1]. Ce logiciel intégrant les modèles de régression géographiquement pondérée permet d'ajuster le paramétrage de la régression géographiquement pondérée sous SpaceStat et notamment de déterminer la bande passante.

---

[1] http://www.ritsumei.ac.jp/~nakaya

4 | RÉSULTATS

LES POINTS ESSENTIELS

> La corrélation linéaire entre la défaveur sociale et le risque de décès tous cancers est de 0,6 chez les hommes et dans l'ensemble de la population, elle est de 0,7 pour le cancer du poumon ;

> Les variables éligibles pour la régression géographiquement pondérée sont la mortalité par cancer du poumon chez les hommes et la défaveur sociale ;

> L'association entre les deux variables est robuste au sein de deux zones de la région : une zone nord-ouest sud-est centrée sur le sud de Paris et deux cantons au sud de la région ;

> Après standardisation, lissage et application du modèle de régression, aucune association statistiquement significative n'a été observée entre les expositions chimiques et les taux de mortalité.

4.1 Taux standardisés de mortalité

Le Tableau 2 décrit les taux de mortalité standardisés à l'échelle de la région, leurs intervalles de confiance à 95% ainsi que les valeurs minimales et maximales obtenues pour les différents cantons. On constate que l'amplitude entre ces dernières valeurs est importante y compris pour les localisations fréquentes. Ainsi, le taux de mortalité tous cancers varie chez les hommes de 228 pour 100 000 à 676 pour 100 000 et de 116,1 à 264,6 pour 100 000 chez les femmes.

Tableau 2. Taux standardisés de mortalité en Île-de-France - période 2002-2011

Localisations cancéreuses	Code CIM10	Hommes				Femmes				Total			
		Taux et IC95% en Île-de-France		Taux min-max dans les cantons	Effectif de décès en Île-de-France	Taux et IC95% en Île-de-France		Taux min-max dans les cantons	Effectif de décès en Île-de-France	Taux et IC95% en Île-de-France		Taux min-max dans les cantons	Effectif de décès en Île-de-France
Tous cancers	C00-C97	319,8	[318-321,7]	[228,5-675,9]	119 450	169,1	[168-170,2]	[116,1-264,6]	93 312	228,4	[227,5-229,4]	[179-301,9]	212 762
Lèvre cavité buccale pharynx	C00-C14	11,2	[10,8-11,5]	[2,3-67,5]	4 935	2,4	[2,2-2,5]	[0-17,1]	1 298	6,3	[6,1-6,5]	[2-15,1]	6 233
Œsophage	C15	9,4	[9,1-9,7]	[1,6-49,9]	3 844	2,3	[2,2-2,4]	[0-9,5]	1 256	5,4	[5,2-5,5]	[0,8-11,3]	5 100
Estomac	C16	10,8	[10,5-11,2]	[3,1-74,7]	4 108	4,4	[4,2-4,6]	[0-11,8]	2 465	7,0	[6,9-7,2]	[2,7-15,2]	6 573
Colon-rectum	C18-C21	31,5	[30,9-32,1]	[10,8-216,5]	11 143	19,9	[19,6-20,3]	[10,1-43,3]	11 136	24,5	[24,2-24,8]	[11,9-35,6]	22 279
Foie	C22	18,6	[18,2-19,1]	[5,8-47,2]	7 293	5,4	[5,2-5,6]	[0-17,9]	2 934	11,0	[10,8-11,2]	[4,8-20,2]	10 227
Pancréas	C25	16,5	[16,1-16,9]	[5,3-46,6]	6 270	10,9	[10,7-11,2]	[2,4-23,1]	5 976	13,3	[13,1-13,6]	[4,8-28,9]	12 246
Larynx	C32	3,9	[3,7-4,1]	[0-99]	1 602	0,5	[0,4-0,6]	[0-5,4]	267	1,9	[1,9-2]	[0-8,3]	1 869
Poumon	C33-C34	77,2	[76,4-78,1]	[47,5-123,1]	30 838	21,9	[21,5-22,2]	[9,7-37,1]	11 852	45,0	[44,6-45,5]	[26,5-70,1]	42 690
Plèvre	C384, C450, C459	3,8	[3,6-4]	[0-14,7]	1 352	1,2	[1,1-1,3]	[0-7,8]	658	2,2	[2,1-2,3]	[0-9,4]	2 010
Sein	C50	0,8	[0,7-0,9]	[0-9,5]	294	33,7	[33,2-34,2]	[22,1-59,2]	18 621	19,9	[19,6-20,1]	[12,4-33,2]	18 915
Ovaire	C56, C570-C574, C578					9,1	[8,9-9,4]	[0-18,1]	4 944				
Prostate	C61	37,7	[37-38,4]	[10,9-121,1]	11 443								
Vessie	C67	15,4	[15-15,9]	[5,8-137,9]	5 175	3,1	[2,9-3,2]	[0-9,9]	1 732	7,7	[7,5-7,9]	[3,5-17]	6 907
Rein	C64, C66-68	24,9	[24,4-25,5]	[12,5-238,3]	8 616	6,2	[6-6,4]	[1,8-15,3]	3 495	13,4	[13,2-13,6]	[8,3-24,4]	12 111
Système nerveux central	C70-C72	6,2	[6-6,5]	[0,7-42]	2 764	3,9	[3,8-4,1]	[0-11,8]	2 137	4,9	[4,8-5,1]	[2,1-9,6]	4 901
Lymphome non Hodgkinien	C82-C85, C96	9,0	[8,7-9,3]	[0-20,1]	3 255	5,3	[5,1-5,5]	[0-18,1]	2 954	6,8	[6,6-7]	[2,1-13,7]	6 209
Toutes leucémies	C91 -C95	11,8	[11,5-12,2]	[1,6-46,1]	4 188	6,7	[6,5-6,9]	[0,5-17,1]	3 733	8,6	[8,4-8,8]	[4,3-14,5]	7 921

Taux standardisé pour 100 000 habitants, population de référence : population française de 2006 tous sexes confondus. 21 classes d'âge : [0-1[, [1-5[, [5-10[, [10-15[, …[90-95[, ≥95 ans ; IC95% : intervalle de confiance à 95%, Min-Max valeurs minimum et maximum. N décès
Sources : expositions : Ineris. Décès : CépiDC. Population : Insee. Exploitation ORS Île-de-France

4.2 Liens entre mortalité par canton et expositions environnementales et sociales

4.2.1 Première étape : liens entre taux et expositions sans prise en compte des effets de voisinage entre cantons

Le Tableau 3 ci-dessous synthétise les associations (exposition, pathologie) pour lesquelles une corrélation linéaire supérieure à 0,35 a été observée, seuil choisi à titre pragmatique par les experts. Nous les qualifions de corrélation suffisante par la suite. L'intégralité des résultats est présentée en annexe 4.

Chez les femmes, aucune corrélation suffisante n'est observée entre les taux de mortalité et les expositions.

Chez les hommes, les corrélations suffisantes portent toutes sur **l'indicateur de défaveur sociale**. Les localisations concernées sont par ordre de corrélation décroissante : le cancer du poumon, le cancer du foie et l'ensemble des cancers.

Dans l'ensemble de la population, les mêmes localisations donnent lieu à des corrélations suffisantes avec l'indicateur de défaveur sociale, auxquelles sont ajoutées le cancer LBP et le cancer de l'estomac. On observe également des corrélations suffisantes entre la mortalité tous cancers et les DJE cadmium et chrome.

Tableau 3. Coefficients de corrélations linéaires entre les taux de mortalité standardisés des cantons d'Île-de-France (hommes) et les facteurs sociaux et environnementaux supérieurs à 0,35

Localisations cancéreuses	Code CIM10	Hommes	Tous		
		Défaveur sociale (FDep08)	Défaveur sociale (FDep08)	Doses journalières d'exposition au cadmium	Doses journalières d'exposition au chrome
Tous cancers	C00-C97	0,56	0,60	0,38	0,41
Lèvre cavité buccale pharynx	C00-C14	/	0,49	/	/
Estomac	C16	/	0,47	/	/
Foie	C22	0,47	0,54	/	/
Poumon	C33-C34	0,65	0,61	/	/

Sources : expositions : Ineris. Décès : CépiDC. Population : Insee. Exploitation ORS Île-de-France

4.2.2 Deuxième étape : prise en compte des effets de voisinage entre cantons par lissage spatial

- Lissage spatial des taux

Les cartes ci-après (Carte 3 etCarte 4) permettent d'apprécier l'impact du lissage par krigeage de Poisson.

Carte 3. Taux de mortalité standardisé par cancer du poumon chez les hommes, période 2002-2011 avant lissage

Population de référence : population française de 2006 tous sexes confondus. 21 classes d'âge : [0-1[, [1-5[, [5-10[, [10-15[, …[90-95[, ≥95 ans

Carte 4. Taux de mortalité standardisés par cancer du poumon chez les hommes, période 2002-2011, lissés par krigeage de Poisson

Population de référence : population française de 2006 tous sexes confondus. 21 classes d'âge : [0-1[, [1-5[, [5-10[, [10-15[, …[90-95[, ≥95 ans

Carte 5. Variance associée au krigeage des taux de mortalité standardisés par cancer du poumon chez les hommes, période 2002-2011

La variance de krigeage est à la fois fonction de la taille de la population du canton considéré et de ses voisins, et de la variabilité locale (17). On observe ici qu'elle est élevée sur les frontières est et sud de la région. Pour ces cantons, l'effet du lissage est important. Ainsi le taux de Donnemarie-Dontilly passe de 70 à 85 pour 100 000 après lissage. À l'inverse, la variance est moins élevée au centre de la Seine-et-Marne où la densité n'est

pas élevée mais où la variabilité locale est faible, ainsi les taux passent à Rozay-en-Brie 92 à 95 pour 100 000 après lissage et ceux de Coulommiers ne changent presque pas.

- Corrélations entre taux lissés et expositions

Les tableaux suivants (Tableau 4,Tableau 5 et 6) synthétisent les associations (exposition, pathologie) pour lesquelles une corrélation linéaire supérieure à 0,35 a été observée entre les taux lissés et les expositions, corrélation dite suffisante. L'intégralité des résultats est présentée en annexe 4.

Les associations pour lesquelles des corrélations suffisantes ont été observées à partir des taux non lissés donnent lieu à des corrélations avec les taux lissés également suffisantes et un peu plus élevées.

Chez les **hommes**, la corrélation maximale observée est de 0,70 et porte sur l'association (défaveur, mortalité par cancer du poumon). On observe par ailleurs une corrélation de 0,60 entre la défaveur et la mortalité tous cancers et entre la défaveur et la mortalité par cancer du foie. Les corrélations avec les expositions environnementales sont au maximum de 0,42 (chrome et cancer de l'œsophage) et de 0,39 (cadmium et cancer de l'œsophage).

Chez les **femmes**, alors qu'aucune corrélation supérieure à 0,35 n'était repérée à partir des taux non lissés, on observe 16 associations vérifiant ce critère à partir des taux lissés. Quatre des cinq localisations corrélées chez les hommes à plus de 0,35 avec la défaveur le sont également chez les femmes. Il s'agit de l'ensemble des cancers, des cancers LBP, de l'œsophage, de l'estomac et du foie. Par ailleurs, on observe des corrélations respectives de 0,43 et 0,45 entre les doses journalières d'exposition au cadmium et au chrome et la mortalité par cancer. Ces corrélations sont respectivement de 0,38 et 0,49 pour la mortalité par cancer du côlon et du rectum. La DJE au chrome est également corrélée à la mortalité par cancer du sein (0,50) et à la mortalité par leucémie (0,47). Les corrélations entre expositions et mortalité par cancer du poumon chez les femmes ont un profil particulier : elles dépassent 0,35 pour l'exposition aux PM10, au cadmium, au plomb et au nickel par voie atmosphérique. La corrélation maximale observée est de 0,50 et porte sur la DJE chrome et la mortalité par cancer du sein. Elle est suivie de près par la corrélation entre la DJE chrome et la mortalité par cancer colorectal (0,49).

Lorsque l'on examine la mortalité **tous sexes confondus**, on observe une corrélation suffisante avec l'indicateur de défaveur pour neuf des localisations examinées ainsi que pour la mortalité tous cancers, par ordre décroissant : cancers du foie et du poumon, estomac, LBP, larynx, vessie, œsophage, colon-rectum et rein. Les DJE au cadmium et au chrome sont corrélées à plus de 0,35 pour les localisations suivantes : œsophage, colon-rectum, vessie et rein et pour l'ensemble des cancers. On observe également une corrélation de 0,41 entre la DJE au plomb et la mortalité par cancer du rein et entre de 0,50 entre la DJE au chrome et la mortalité par leucémie.

Tableau 4. Coefficients de corrélations linéaires entre les taux de mortalité standardisés des cantons d'Île-de-France lissés (hommes) et les facteurs sociaux et environnementaux supérieurs à 0,35

Localisations cancéreuses	Code CIM10	Défaveur	Doses journalières d'exposition (DJE)	
Hommes		FDep08	Cd	Cr
Tous cancers	C00-C97	0,60	/	/
Lèvre cavité buccale pharynx	C00-C14	0,39	/	/
Œsophage	C15	/	0,39	0,42
Estomac	C16	0,38	/	/
Colon-rectum	C18-C21	/	/	/
Foie	C22	0,60	/	/
Poumon	C33-C34	0,71	/	/

Taux standardisé pour 100 000 habitants, population de référence : population française de 2006 tous sexes confondus.
21 classes d'âge : [0-1[, [1-5[, [5-10[, [10-15[, ...[90-95[, ≥95 ans
FDep : indice de défaveur sociale (1, 14, 16), PM10 : particules de diamètre inférieur au millième de millimètre,
Cd :cadmium, Pb : plomb, Ni : nickel, Cr : chrome
Sources : expositions : Ineris. Décès : CépiDC. Population : Insee. Exploitation ORS Île-de-France

Tableau 5. Coefficients de corrélations linéaires entre les taux de mortalité standardisés des cantons d'Île-de-France lissés (femmes) et les facteurs sociaux et environnementaux

Localisations cancéreuses	Code CIM10	Défaveur	Particules	Concentrations atmosphériques				Doses journalières d'exposition (DJE)		ISR
Femmes		FDep08	PM10	Cd	Pb	Ni	Cr	Cd	Cr	Ni
Tous cancers	C00-C97	0,45	/	/	/	/	/	0,43	0,45	/
Lèvre cavite buccale pharynx	C00-C14	0,39	/	/	/	/	0,41	/	/	/
Estomac	C16	0,51	/	/	/	/	/	/	/	/
Colon-rectum	C18-C21	/	/	/	/	/	/	0,38	0,49	/
Foie	C22	0,44	/	/	/	/	/	/	/	/
Poumon	C33-C34	/	0,38	0,42	0,41	0,41	/	/	/	0,37
Sein	C50	/	/	/	/	/	/	/	0,50	/
Toutes leucémies	C91-C95	/	/	/	/	/	/	/	0,47	/

Taux standardisé pour 100 000 habitants, population de référence : population française de 2006 tous sexes confondus.
21 classes d'âge : [0-1[, [1-5[, [5-10[, [10-15[, ...[90-95[, ≥95 ans
ISR : indice spatialisé relatif
FDep : indice de défaveur sociale (1, 14, 16), PM10 : particules de diamètre inférieur au millième de millimètre,
Cd :cadmium, Pb : plomb, Ni : nickel, Cr : chrome
Sources : expositions : Ineris. Décès : CépiDC. Population : Insee. Exploitation ORS Île-de-France

Tableau 6. Coefficients de corrélations linéaires entre les taux de mortalité standardisés des cantons d'Île-de-France lissés (hommes et femmes) et les facteurs sociaux et environnementaux

Localisations cancéreuses	Code CIM10	Défaveur	Doses journalières d'exposition (DJE)		
Tous		FDep08	Cd	Pb	Cr
Tous cancers	C00-C97	0,65	0,44	/	0,48
Lèvre cavite buccale pharynx	C00-C14	0,59	/	/	/
Œsophage	C15	0,42	0,59	/	0,54
Estomac	C16	0,63	/	/	/
Colon-rectum	C18-C21	0,42	0,38	/	0,50
Foie	C22	0,70	/	/	/
Larynx	C32	0,49	0,44	/	/
Poumon	C33-C34	0,70	/	/	/
Vessie	C67	0,44	0,39	/	0,36
Rein	C64, C66-68	0,39	0,38	0,41	0,54
Toutes leucémies	C91 -C95	/	/	/	0,50

Taux standardisé pour 100 000 habitants, population de référence : population française de 2006 tous sexes confondus. 21 classes d'âge : [0-1[, [1-5[, [5-10[, [10-15[, …[90-95[, ≥95 ans
FDep : indice de défaveur sociale (1, 14, 16), PM10 : particules de diamètre inférieur au millième de millimètre, Cd :cadmium, Pb : plomb, Ni : nickel, Cr : chrome
Sources : expositions : Ineris. Décès : CépiDC. Population : Insee. Exploitation ORS Île-de-France

4.2.3 Détermination des associations (expositions, mortalité) à retenir pour l'analyse par regroupement de cantons

Les associations pour lesquelles des corrélations suffisantes ont été observées sont au nombre de 46 : 7 pour la mortalité chez les hommes, 16 pour la mortalité chez les femmes et 23 pour les deux sexes confondus.

Les tableaux ci-dessous (Tableau 7,Tableau 8,Tableau 9) présentent les coefficients de détermination (pourcentage de variance expliquée) obtenus à partir des régressions aspatiales. On observe que seul le modèle incluant la défaveur et la mortalité par cancer du poumon chez les hommes permet d'expliquer plus de 50% de la variance. On note également deux modèles proches de ce critère de 50% : celui incluant défaveur et mortalité par cancer du poumon pour les deux sexes confondus et celui incluant défaveur et mortalité par cancer du foie pour les deux sexes confondus. Pour les femmes, le pourcentage de variance expliqué ne dépasse pas 28%.

La régression géographiquement pondérée va donc porter sur le modèle cherchant à « expliquer » localement la mortalité par cancer du poumon chez les hommes par la défaveur sociale.

Tableau 7. Coefficients de détermination des modèles linéaires expliquant les taux de mortalité standardisés et lissés par l'exposition, pour les associations avec un coefficient de corrélation supérieur à 0,35 (hommes)

Localisations cancéreuses	Code CIM10	Défaveur	Doses journalières d'exposition (DJE)	
Hommes		FDep08	Cd	Cr
Tous cancers	C00-C97	0,33	/	/
Lèvre cavité buccale pharynx	C00-C14	0,10	/	/
Œsophage	C15	/	0,14	0,18
Estomac	C16	0,08	/	/
Colon-rectum	C18-C21	/	/	/
Foie	C22	0,34	/	/
Poumon	C33-C34	0,51	/	/

Taux standardisé pour 100 000 habitants, population de référence : population française de 2006 tous sexes confondus. 21 classes d'âge : [0-1[, [1-5[, [5-10[, [10-15[,…[90-95[, ≥95 ans
FDep : indice de défaveur sociale (1, 14, 16), PM10 : particules de diamètre inférieur au millième de millimètre, Cd : cadmium, Pb : plomb, Ni : nickel, Cr : chrome
En rouge : coefficients de détermination > 0,50
Sources : expositions : Ineris. Décès : CépiDC. Population : Insee. Exploitation ORS Île-de-France
Note de lecture : La défaveur sociale « explique » 51% de la variance du taux de mortalité par cancer du poumon observée chez les hommes dans les cantons.

Tableau 8. Coefficients de détermination des modèles linéaires expliquant les taux de mortalité standardisés et lissés par l'exposition, pour les associations avec un coefficient de corrélation supérieur à 0,35 (femmes)

Localisations cancéreuses	Code CIM10	Défaveur	Particules	Concentrations atmosphériques				Doses journalières d'exposition (DJE)		ISR
Femmes		FDep08	PM10	Cd	Pb	Ni	Cr	Cd	Cr	Ni
Tous cancers	C00-C97	0,20	/	/	/	/	/	0,22	0,23	/
Lèvre cavité buccale pharynx	C00-C14	0,20	/	/	/	/	0,28	/	/	/
Estomac	C16	0,20	/	/	/	/	/	/	/	/
Colon-rectum	C18-C21	/	/	/	/	/	/	0,14	0,27	/
Foie	C22	0,16	/	/	/	/	/	/	/	/
Poumon	C33-C34	/	0,07	0,10	0,09	0,10	/	/	/	0,08
Sein	C50	/	/	/	/	/	/	/	0,26	/
Toutes leucémies	C91 -C95	/	/	/	/	/	/	/	0,19	/

Taux standardisé pour 100 000 habitants, population de référence : population française de 2006 tous sexes confondus. 21 classes d'âge : [0-1[, [1-5[, [5-10[, [10-15[, ...[90-95[, ≥95 ans
ISR : indice spatialisé relatif
FDep : indice de défaveur sociale (1, 14, 16), PM10 : particules de diamètre inférieur au millième de millimètre, Cd : cadmium, Pb : plomb, Ni : nickel, Cr : chrome
Note de lecture : La défaveur sociale « explique » 20% de la variance du taux de mortalité tous cancers observée chez les femmes dans les cantons.
Sources : expositions : Ineris. Décès : CépiDC. Population : Insee. Exploitation ORS Île-de-France

Tableau 9. Coefficients de détermination des modèles linéaires taux de mortalité standardisé lissé~ exposition pour les associations avec un coefficient de corrélation supérieur à 0,35 (hommes et femmes)

Localisations cancéreuses	Code CIM10	Défaveur	Doses journalières d'exposition (DJE)		
Tous		FDep08	Cd	Pb	Cr
Tous cancers	C00-C97	0,39	0,18	/	0,24
Lèvre cavite buccale pharynx	C00-C14	0,34	/	/	/
Œsophage	C15	0,17	0,34	/	0,32
Estomac	C16	0,37	/	/	/
Colon-rectum	C18-C21	0,16	0,13	/	0,26
Foie	C22	0,45	/	/	/
Larynx	C32	0,24	0,19	/	/
Poumon	C33-C34	0,49	/	/	/
Vessie	C67	0,16	0,15	/	0,13
Rein	C64, C66-68	0,15	0,12	0,21	0,31
Toutes leucémies	C91 -C95	/	/	/	0,25

Taux standardisé pour 100 000 habitants, population de référence : population française de 2006 tous sexes confondus. 21 classes d'âge : [0-1[, [1-5[, [5-10[, [10-15[, ...[90-95[, ≥95 ans
FDep : indice de défaveur sociale (1, 14, 16), PM10 : particules de diamètre inférieur au millième de millimètre, Cd : cadmium, Pb : plomb, Ni : nickel, Cr : chrome
En gras : coefficients de corrélation linéaire proches de 0,50
Note de lecture : La défaveur sociale « explique » 49% de la variance du taux de mortalité par cancer du poumon observée dans les cantons.
Sources : expositions : Ineris. Décès : CépiDC. Population : Insee. Exploitation ORS Île-de-France

4.3 Identification des regroupements de cantons où l'association entre mortalité et exposition est forte

4.3.1 Préliminaires

Le modèle retenu pour la régression géographiquement pondérée porte sur le taux de mortalité par cancer du poumon chez les hommes, retenu comme variable à expliquer et la défaveur sociale, caractérisée par l'indicateur FDep08, retenu comme variable explicative.

Ces variables vérifient en effet les critères adoptés dans la démarche, plus précisément :

- Leur corrélation linéaire est supérieure à 0,35. On observe :
  - une corrélation entre le taux de mortalité non lissé et l'indicateur de défaveur de 0,65
  - une corrélation entre le taux de mortalité lissé et l'indicateur de défaveur de 0,71

- La régression aspatiale pondérée permet « d'expliquer » plus de 50% de la variance, (51%).

On considère donc qu'un phénomène d'association globale, indépendant de la composante spatiale, peut exister entre les deux variables. La régression géographiquement pondérée va permettre d'examiner si cette association est plus forte dans certaines zones de la région d'étude. L'hypothèse sous-jacente est que des facteurs de confusion peuvent amoindrir l'association sur l'ensemble de la zone et que, sur des zones où ceux-ci sont constants ou peu variables, l'association « réelle » peut être mise en évidence.

Les cartesCarte 6 etCarte 7 illustrent visuellement les corrélations observées.

Carte 6. Indice de défaveur sociale (FDep08) dans les cantons d'Île-de-France

Carte 7. Taux de mortalité lissés par cancer du poumon chez les hommes entre 2002 et 2011

4.3.2 Résultats

Rappelons que la régression comprend deux étapes :

- **Détermination de la bande passante**

Lorsque l'on choisit de réaliser des régressions locales, il est nécessaire de fixer la distance entre le canton considéré et ceux inclus dans la régression[1], dite bande passante. Au-delà d'une certaine distance, tous les cantons seront inclus et le modèle sera alors égal au modèle global. Le critère d'information d'Akaïké (AICs[2]) associé au modèle de régression géographique pondéré permet de déterminer la distance pour laquelle les associations locales sont les plus élevées (31). La distance choisie sera celle correspondant à l'AICs le plus faible. La Figure 2 illustre les AICs obtenus pour les différentes bandes passantes et la distance optimale égale à 30 km. Les résultats détaillés du calcul de la bande passante figurent en annexe 5.

Figure 2. Critère d'information d'Akaïke (AIC) dans la régression géographiquement pondérée, en fonction de la distance

La bande passante obtenue est de 30 km.

- **Régressions locales**

À cette étape, des régressions pondérées sont réalisées pour chaque canton. On obtient donc pour chaque canton :

  - Un coefficient de détermination R^2 donnant le pourcentage de variance expliquée,
  - Un paramètre associé à chaque variable explicative, donc ici au FDep08,
  - La p-value correspondante, permettant de juger de la significativité statistique du paramètre,
  - Le coefficient de corrélation linéaire entre les deux variables.

---

[1] Plus exactement entre le centroïde du canton, ici défini par le barycentre du polygone définissant le canton.

[2] L'AIC « Akaïke Information Criteria » d'origine n'est pas valable lorsque le nombre d'observations est faible. Il est alors nécessaire d'utiliser le critère d'Akaïke corrigé (AICs), qui corrige les erreurs dues au nombre d'observations.

Les cartes qui suivent présentent respectivement :

- Les coefficients de détermination associés aux modèles de régression géographique locaux. Il s'agit, lorsqu'ils sont exprimés en pourcentage, de la part de variance expliquée par les modèles (Carte 7).

- Les corrélations entre le taux de mortalité lissé par cancer du poumon chez les hommes entre 2002 et 2011 et la défaveur sociale (FDep08) dans une zone de 30 km autour des cantons d'Île-de-France (Carte 8).

- Les valeurs de significativité statistique (p-value) du coefficient associant la défaveur sociale et la mortalité par cancer du poumon dans chaque modèle local (Carte 9).

Deux zones combinent à la fois des valeurs élevées pour le pourcentage de variance expliquée, la significativité statistique du coefficient associé au modèle et les corrélations entre taux et défaveur. Il s'agit d'une part d'une zone nord-ouest sud-est centrée sur le sud de Paris et d'autre part de deux cantons au sud de la région : La Chapelle-la-Reine et Nemours.

À l'inverse, deux autres zones combinent des valeurs faibles pour ces trois indicateurs. Il s'agit du croissant nord et est de la région et du sud de l'Essonne autour du canton de La Ferté-Alais, zones essentiellement rurales.

Carte 8. Coefficient de détermination associé au paramètre du modèle de régression géographiquement pondérée entre le taux de mortalité lissé par cancer du poumon chez les hommes entre 2002 et 2011 et la défaveur sociale (FDep08) (R2)

Note de lecture : À Ivry, le canton le plus foncé sur la carte, la défaveur sociale «explique» entre 80% et 82% de la variance du taux de mortalité par cancer du poumon observée dans les cantons situés dans une zone de moins de 30 km. Les classes ont été choisies a priori. Ivry est également la valeur maximale.

Carte 9. Corrélations entre le taux de mortalité lissé par cancer du poumon chez les hommes entre 2002 et 2011 et la défaveur sociale (FDep08)

Note de lecture : Pour un canton en rouge, la corrélation entre défaveur sociale et cancer du poumon chez les hommes, dans une zone de moins de 30 km autour du canton considéré, est supérieure à 0,80.

Carte 10. Significativité du coefficient associé à l'exposition dans le modèle de régression géographiquement pondérée entre le taux de mortalité lissé par cancer du poumon chez les hommes entre 2002 et 2011 et la défaveur sociale (FDep08)

Dans les Carte 11 etCarte 12, nous avons représenté la zone à l'intérieur de laquelle le modèle local explique plus de 70% de la variance (trait vert foncé). Nous avons délimité cette zone en prenant les cantons pour lesquels la variance expliquée est localement de plus 70% et ceux distant de ceux-ci d'au plus 30 km. Le trait vert clair représente, dans la même logique, la zone où le modèle explique plus de 50% de la variance. La Carte 11 représente le taux de mortalité par cancer du poumon chez les hommes et la Carte 12 la défaveur sociale. On observe sur ces territoires des valeurs contrastées tant de la défaveur que du taux de mortalité par cancer du poumon chez les hommes. Ces contrastes sont notamment marqués pour la zone autour de Paris.

Carte 11. Défaveur sociale en 2008 (Fdep08) et zones où le modèle explique plus de 70% de la variance (en vert foncé) et 50% de la variance (en vert clair)

Note de lecture : À l'intérieur de la zone en vert foncé, la défaveur sociale explique plus de 70% de la variance du taux de mortalité par cancer du poumon chez les hommes. Les couleurs représentent le niveau de défaveur des cantons.

Carte 12. Taux de mortalité lissé par cancer du poumon chez les hommes entre 2002 et 2011 et zones où le modèle explique plus de 70% de la variance (en vert foncé) et 50% de la variance (en vert clair)

Note de lecture : À l'intérieur de la zone en vert foncé, la défaveur sociale explique plus de 70% de la variance du taux de mortalité par cancer du poumon chez les hommes. Les couleurs représentent le taux de mortalité des cantons.

- **Choix des modèles locaux satisfaisants**

Il est logique dans le choix de ces modèles de tenir compte à la fois d'un pourcentage de variance expliquée minimum, d'une significativité statistique minimum pour les paramètres estimés et d'un niveau de corrélation minimum observé entre les variables explicatives et à expliquer. Rappelons que les seuils retenus dans cette analyse, communs aux ORS participant à l'étude, sont respectivement de 50% pour la variance expliquée par le modèle, de 0,05 pour la significativité statistique du coefficient associé à la variable explicative et de 0,5 pour la corrélation linéaire entre les deux variables.

Les cantons pour lesquels le modèle local permet d'expliquer plus de 50% de la variance sont représentés en bleu dans la Carte 13. On retrouve les zones décrites précédemment élargies à l'ouest pour la zone parisienne ainsi que les quatre cantons isolés de Bonnières-sur-Seine, Maurepas, Limours et les Ulis. Notons aussi que pour deux cantons de la zone concentrique de Paris la variance expliquée est inférieure à 50% : Paris 11e et Choisy-le-Roi.

Pour caractériser la prise en compte du second critère seul, celui de la significativité statistique du paramètre associé à la variable explicative du modèle, il est plus facile de désigner les zones ne le satisfaisant pas, c'est-à-dire celles pour lesquelles p-Value>0,05, zones non hachurées horizontalement sur la carte : un croissant nord-ouest sud-est et une zone plein sud (Carte 13).

Enfin, deux zones satisfont le troisième critère (corrélation linéaire supérieure à 0,5) hachurées verticalement : un croissant ouest-sud et deux cantons plein est.

La combinaison de ces trois critères fait ressortir les 114 cantons figurant en bleu et doublement hachurés sur la carte.

Carte 13. Carte des cantons d'Île-de-France pour lesquels une association robuste entre défaveur sociale et cancer du poumon chez les hommes est observée dans un rayon de 30 km

4.4 Identification des points « chauds » (mortalité et défaveur élevées) et « froids » au sein des regroupements de cantons

Les zones de cumul caractérisées ici sont à la fois des zones où les deux indicateurs étudiés, la défaveur sociale et la mortalité par cancer du poumon chez les hommes sont élevés (resp. faibles) et où une association statistique suffisante est observée entre les deux. Le seuil retenu pour caractériser les valeurs élevées (respectivement faibles) est le quatrième quintile ou 80e percentile (respectivement le premier quintile ou 20e percentile).

L'application de ces critères fait apparaître huit zones de points « chauds » et 29 zones de points « froids ». Les caractéristiques de ces zones sont décrites dans les Tableau 10 etTableau 11.

Tableau 10. Zones de points « chauds » associant des taux de mortalité par cancer du poumon chez les hommes élevés, un indice de défaveur élevé (FDep08) et une association spatiale entre les deux dans un rayon de 30 km

cv	Cantons	FDep08	Taux lissé de mortalité par cancer du poumon	Corrélation	Coefficient de détermination	Pvalue
7724	Nemours	1,9	90,5	0,83	69%	<0,0001
9240	Villeneuve-la-Garenne	2,9	103,7	0,83	68%	<0,0001
9325	Romainville	3,2	99,3	0,74	54%	0,0001
9398	Saint-Denis	4,0	101,4	0,79	62%	<0,0001
9403	Bonneuil-sur-Marne	2,5	100,8	0,83	69%	<0,0001
9420	Orly	3,2	94,3	0,78	61%	<0,0001
9488	Villeneuve-Saint-Georges	3,1	90,1	0,78	61%	<0,0001
9499	Vitry-sur-Seine	2,3	108,3	0,84	70%	<0,0001
	Minimum	-4,3	50,6			
	Médiane	-0,0	81,6			
	Q80 (4ème quintile)	1,7	89,9			
	Maximum	5,5	116,6			

Note de lecture : (i) les cantons de ce tableau ont une défaveur sociale et un taux de mortalité parmi les 20% les plus élevés de l'ensemble de la région, (ii) à 30 km alentour de ces cantons, la corrélation spatiale est robuste c'est-à-dire qu'elle respecte les critères retenus par les experts

Les zones de points « chauds » figurent au nord de Paris, à l'est et dans le Val-de-Marne, à l'exception de Nemours, au sud de la région (Carte 14).

Les zones de points « froids » figurent dans Paris ouest, dans l'ouest et le sud des Hauts-de-Seine auquel il faut ajouter trois cantons du Val-de-Marne (Carte 14).

Carte 14. Zones de points « chauds » et froids associant des taux de mortalité par cancer du poumon chez les hommes élevés, un indice de défaveur élevé (FDep08) et une association spatiale entre les deux dans un rayon de 30 km

Tableau 11. Zones de points « froids » associant des taux de mortalité par cancer du poumon chez les hommes faibles, un indice de défaveur faible (FDep08) et une association spatiale entre les deux dans un rayon de 30 km

cv	Cantons	FDep08	Taux lissé de mortalité par cancer du poumon	Corrélation	Coefficient de détermination	Pvalue
7515	PARIS 1er Arrondissement	-2,6	68,7	0,78	61%	<0,0001
7518	PARIS 4e Arrondissement	-2,2	68,1	0,80	64%	<0,0001
7519	PARIS 5e Arrondissement	-2,3	57,2	0,86	75%	<0,0001
7520	PARIS 6e Arrondissement	-3,0	51,0	0,86	75%	<0,0001
7521	PARIS 7e Arrondissement	-3,5	51,4	0,83	69%	<0,0001
7522	PARIS 8e Arrondissement	-3,3	51,7	0,75	57%	<0,0001
7523	PARIS 9e Arrondissement	-2,0	67,4	0,85	72%	<0,0001
7528	PARIS 14e Arrondissement	-1,8	69,8	0,83	70%	<0,0001
7529	PARIS 15e Arrondissement	-2,4	59,7	0,85	73%	<0,0001
7530	PARIS 16e Arrondissement	-3,4	51,4	0,84	71%	<0,0001
7531	PARIS 17e Arrondissement	-1,8	59,6	0,84	71%	<0,0001
7803	La Celle-Saint-Cloud	-2,9	67,1	0,82	68%	<0,0001
7811	Maisons-Laffitte	-3,1	67,0	0,83	68%	<0,0001
7813	Marly-le-Roi	-2,9	62,7	0,82	68%	<0,0001
7822	Saint-Nom-la-Bretèche	-4,2	64,0	0,72	52%	<0,0001
7830	Le Vésinet	-3,5	67,4	0,86	73%	<0,0001
7835	Le Pecq	-2,8	59,4	0,81	66%	<0,0001
9211	Chaville	-3,4	56,2	0,83	69%	<0,0001
9235	Saint-Cloud	-3,6	59,0	0,83	70%	<0,0001
9237	Sèvres	-2,7	65,4	0,78	61%	<0,0001
9285	Châtenay-Malabry	-1,9	63,2	0,77	60%	<0,0001
9287	Antony	-2,2	61,4	0,78	62%	<0,0001
9289	Boulogne-Billancourt	-2,2	65,2	0,86	74%	<0,0001
9293	Courbevoie	-2,1	61,3	0,84	71%	<0,0001
9298	Neuilly-sur-Seine	-4,3	56,0	0,79	63%	<0,0001
9299	Rueil-Malmaison	-2,6	68,0	0,82	68%	<0,0001
9419	Nogent-sur-Marne	-2,2	65,6	0,81	65%	<0,0001
9422	Saint-Mandé	-2,6	61,2	0,75	57%	<0,0001
9426	Villecresnes	-2,0	70,6	0,77	60%	<0,0001
	Minimum	-4,3	50,6			
	Q20 (1er quintile)	-1,7	71,3			
	Médiane	-0,0	81,6			
	Maximum	5,5	116,6			

Note de lecture : (i) les cantons de ce tableau ont une défaveur sociale et un taux de mortalité parmi les 20% les moins élevés de l'ensemble de la région, (ii) à 30 km alentour de ces cantons, la corrélation spatiale est robuste c'est-à-dire qu'elle respecte les critères retenus par les experts

5 | DISCUSSION-CONCLUSION

5.1 Synthèse des résultats

La première étape de l'étude CIRCÉ avait permis de caractériser les inégalités de mortalité par cancer à l'échelle cantonale en Île-de-France (ORS, 2012). La seconde étape, objet de ce rapport, a permis d'examiner les associations entre les caractéristiques environnementales au sens large et la mortalité par cancer de façon approfondie. La démarche adoptée a consisté à aller au-delà des simples corrélations linéaires entre indicateurs d'environnement et de santé en imposant des exigences supplémentaires : examen des corrélations à partir des taux lissés pour tenir compte des unités géographiques de faible population, régression globale aspatiale satisfaisante et enfin régressions géographiques locales pour mettre en évidence des zones où les associations étaient robustes. Ces zones pouvaient ainsi être considérées comme zones d'inégalités territoriales environnementales et de santé (ITESS) caractérisées.

L'application à l'Île-de-France de la démarche a fourni les résultats suivants :

L'analyse globale des corrélations entre taux non lissés et expositions n'a fait apparaître que très peu de corrélations supérieures à 0,35 : aucune chez les femmes, trois chez les hommes portant toutes sur l'indicateur de défaveur sociale et sept pour l'ensemble de la population dont cinq portant également sur l'indicateur de défaveur. Les deux corrélations restantes portaient sur les DJE au cadmium et au chrome corrélées à la mortalité tous cancers, les niveaux peu au-dessus de 0,35 (resp. 0,38 et 0,41).

La même analyse réalisée à partir des taux de mortalité lissés a fait apparaître sept corrélations supérieures à 0,35 chez les hommes, 16 chez les femmes et 23 pour l'ensemble de la population. 19 de ces corrélations portaient sur l'indicateur de défaveur et la valeur maximale observée pour celui-ci était de 0,7 (cancer du poumon chez les hommes et dans l'ensemble de la population et cancer du foie dans l'ensemble de la population). 11 de ces corrélations portaient sur la DJE au chrome et 9 sur la DJE au cadmium.

Les corrélations observées entre défaveur sociale et mortalité par cancer sont cohérentes avec la littérature sur les inégalités sociales de santé (30) et cohérentes également avec les résultats obtenus dans les deux autres régions où la démarche a été effectuée : la Bretagne et la Lorraine (32, 33). L'interprétation des autres corrélations appelle à la prudence. En effet, les biais associés aux études écologiques mentionnés plus haut ne peuvent être exclus. Les régressions aspatiales destinées à déterminer la robustesse des associations observées vont dans ce sens puisque seul le modèle incluant la défaveur et la mortalité par cancer du poumon chez les hommes était associé à une variance de plus de 50%. Les deux modèles proches de ce critère de 50% concernaient également la défaveur, associée à la mortalité par cancer du poumon et à la mortalité par cancer du foie, pour les deux sexes confondus. En revanche, aucune association entre DJE au chrome ou au cadmium ne satisfaisait les critères retenus pour être qualifiée de robuste.

Conformément aux critères retenus dans la démarche, la régression géographiquement pondérée a porté sur le modèle cherchant à « expliquer » la mortalité par cancer du poumon chez les hommes par la défaveur sociale. Cette régression a mis en évidence deux zones où le modèle est le plus performant : une zone nord-ouest sud-est centrée sur Paris et deux cantons au sud de la région : La Chapelle-la-Reine et Nemours. On peut donc considérer qu'il s'agit de territoires avec inégalités territoriales sociales et de santé caractérisées.

À l'inverse, le modèle n'était pas performant pour deux autres zones : le croissant nord et est de la région et le sud de l'Essonne autour du canton de La Ferté-Alais, zones essentiellement rurales. À ce stade, il est difficile d'interpréter ces résultats et de conclure

directement à l'absence d'inégalités territoriales sociales et de santé au sein de ces territoires. Les cartes de défaveur et de mortalité montrent qu'il s'agit de zones plutôt défavorisées par rapport à l'ensemble de la région.

Enfin les zones de points chauds et froids combinant performance du modèle et valeurs élevés (resp. faibles) de défaveur et de mortalité par cancer du poumon chez les hommes ont mis en évidence : pour les points « chauds » 8 cantons situés au nord de Paris, à l'est et dans le Val-de-Marne, ainsi que Nemours, au sud de la région (carte 14) ; pour les points « froids » 26 cantons figurant dans Paris ouest, dans l'ouest et le sud des Hauts-de-Seine auxquels il faut ajouter trois cantons du Val-de-Marne.

5.2 Limites

Les limites de l'étude relèvent à la fois du choix des indicateurs retenus et de la méthode statistique utilisée.

Pour les indicateurs environnementaux, les éléments traces métalliques (chrome, nickel, plomb, cadmium) ne représentent qu'une petite partie d'une multitude de substances parfois mal connues et auxquelles la population est susceptible d'être exposée. Il serait intéressant de pouvoir développer la démarche pour d'autres polluants connus pour leurs effets cancérigènes.

L'indicateur de défaveur (FDep08) prend en compte le contexte socio-économique mais ne prend pas en compte le contexte du milieu de vie. Il ne distingue pas les éléments qui reviennent à l'individu d'une part et à ceux de son environnement collectif d'autre part.

La mortalité par cancer est un indicateur de santé solide résultant à la fois de l'incidence du cancer considéré (occurrence de nouveaux cas de cancers) et de la létalité (taux de décès parmi les personnes atteintes). Cette dernière dépend elle-même des modalités de dépistage et de prise en charge de la maladie. On peut ainsi penser que la défaveur sociale joue sur les deux composantes : augmentation de l'incidence du fait des conditions de vie et de travail et du fait d'habitude de vie spécifiques et augmentation de la létalité associée à un recours moindre au dépistage et, peut-être à une moins bonne prise en charge. En revanche, si l'on cherche à mettre en évidence des associations entre facteurs de l'environnement physico-chimique et cancers il est plus pertinent de retenir un indicateur d'incidence ou un proxy de celui-ci (incidence médico-administrative par le biais des prises en charge en affections de longue durée ou des séjours hospitaliers).

Les associations entre expositions environnementales et survenue de cancers ont le plus souvent un temps de latence élevé, il est donc important de tenir compte de ce décalage dans les périodes retenues pour caractériser chaque variable. Si cet aspect est moins important pour les DJE aux métaux, qui sont biopersistants, il est non négligeable pour les concentrations atmosphériques qui ont beaucoup évolué au cours des dernières décennies et ceci de façon différentielle selon les territoires d'Île-de-France (évolution des axes de transport, ouvertures et fermetures d'usines). Par ailleurs, ces variables peuvent être corrélées au degré d'urbanisation et celui-ci peut être également un déterminant de certaines composantes de santé tant sur les facteurs de risque (sédentarité par exemple) que sur l'aval de la survenue de la pathologie : dépistage, accès aux soins.

Sur le plan de la méthode statistique, les croisements ont été réalisés de façon systématique entre les indicateurs environnementaux et les pathologies retenues sans formulation d'hypothèse préalable basée sur la littérature ou en fonction du contexte régional. Or, les croisements systématiques favorisent, par le nombre de tests réalisés, l'apparition d'associations simplement liées au hasard. Le choix des variables sur la base d'hypothèse portant sur les relations entre pathologies et polluants pourrait limiter ce risque et rendre la démarche plus opérationnelle.

Par ailleurs, comme nous l'avons mentionné dans la partie méthode, les études écologiques ont de nombreuses limites : biais écologique, problème de la mobilité des populations et prise en compte de la latence entre expositions et pathologies. Cette étude sur les expositions environnementales, sociales et la mortalité par cancers n'y échappe pas. L'exposition réelle des individus peut ainsi avoir été très différente de celle mesurée dans le canton de résidence au moment du décès.

Enfin, le choix de l'unité géographique peut considérablement influer sur les résultats d'analyse (1). Ce phénomène est appelé Modifiable areal unit problem (MAUP) (34) (35). Il entraîne deux effets statistiques sur les valeurs des données spatiales. Le premier est un effet d'échelle défini comme « les variations que subissent les données lorsque l'on change de niveau d'observation » (34). Lors d'une analyse sur de petites unités géographiques, les indicateurs sanitaires sont estimés sur de petits effectifs avec une variance importante. Ceci entraîne une surestimation de l'hétérogénéité spatiale. À l'inverse, l'agrégation des unités géographiques sur des partitions moins fines a pour conséquence de lisser les données et donc de diminuer l'hétérogénéité spatiale (36). Le second effet statistique, appelé effet de zonage, intervient quant à lui lorsque l'on change la forme de ces unités spatiales, et souligne ainsi « le rôle des découpages territoriaux sur les résultats ». De nombreuses études sur le MAUP, ont montré que le résultat des analyses qui s'appuient sur un découpage administratif sont plus fortement dépendants du niveau de découpage utilisé (37, 38, 39) .

Nous avons également mentionné les avantages de ce type d'étude (étude de population de taille importante pouvant faciliter la détection de risques faibles et utilisation de données existantes notamment). Ceci permet de bien insister sur le statut du protocole CIRCÉ : pertinent dans l'exploration d'associations voire dans la génération d'hypothèses et non dans une optique étiologique.

5.3 Perspectives

Les croisements de données sanitaires, sociales et environnementales réalisés dans le cadre de ce projet sont de type écologique et donc à visée essentiellement descriptive. Ils peuvent appeler à deux types de prolongement : par des approches quantitatives étiologiques pour confirmer ou infirmer les associations observées et explorer plus finement les déterminants de la survenue des cancers dans les territoires ou par des approches qualitatives d'investigation au plan local dans les zones identifiées comme particulièrement affectées par certains cancers et/ou par certaines expositions. Ces approches qualitatives « sur zone » pourraient rendre compte à la fois de la perception et des attentes des habitants et des acteurs de santé de ces territoires ainsi que des facteurs historiques et/ou socioéconomiques susceptibles d'avoir un impact sur les observations réalisées.

Les résultats pourront également être confrontés à ceux d'autres études visant à déterminer des zones de points chauds ou points noirs sur le plan environnemental et sanitaire. Ces éléments sont autant de critères permettant de guider l'action sur certains territoires que les décideurs auront considérés comme prioritaires.

Par ailleurs, l'approche exploratoire elle-même pourra être améliorée à mesure de l'utilisation de la démarche et du retour d'expérience des utilisateurs : détermination d'autres seuils pour passer à l'étape suivante de la démarche, choix des variables au sein des modèles par exemple. Des études de sensibilité pourraient utilement évaluer l'impact du choix de ces seuils sur les résultats.

Un élargissement de la méthodologie utilisée dans le cadre de ce projet à d'autres pathologies ou expositions pourrait être envisagé : bruit et pathologies cardio-vasculaires, profession domiciliée et cancers (sous réserve de la disponibilité des données dans le recensement), pesticides et maladies neurodégénératives pour ne citer que quelques exemples.

LISTE DES SIGLES ET ACRONYMES

ALD : Affection de longue durée

ARS : Agence régionale de santé

Cantons-villes : « À la différence du canton, le « canton-ou-ville » (ou pseudo-canton) est un regroupement d'une ou plusieurs communes entières. Dans les agglomérations urbaines, chaque canton comprend en général une partie de la commune principale et éventuellement une ou plusieurs communes périphériques. Dans ce cas, l'Insee considère la commune principale, entière, comme un pseudo-canton unique et distinct. Pour la ou les communes périphériques, le pseudo-canton considéré est alors identique au vrai canton amputé de la fraction de la commune principale que comprend le vrai canton. » (Insee : https://www.insee.fr/fr/metadonnees/definition/c1725, accédé le 02/11/2017

CépiDC : Centre d'épidémiologie sur les causes médicales décès

CIRC : Centre international de recherche sur le cancer

DJE : Dose journalière d'exposition

FDep08 : Indice de défaveur sociale selon la méthode de Rey (14) actualisé par Saïb (1) avec les données de 2008

GWR : Geographically weighted regression, régression géographiquement pondérée

Ineris : Institut national de l'environnement industriel et des risques

Insee : Institut national de la statistique et des études économiques

MAUP : Modifiable area unit problem, problème associé aux conséquences du choix de l'unité géographique en analyse spatiale

LBP : lèvres, cavité buccale et pharynx

ORS : Observatoire régional de santé

BIBLIOGRAPHIE

1. Saib MS. Construction d'une méthode de caractérisation des inégalites territoriales, environnementales et sociales de santé : développement d'un outil d'integration des phénomènes spatiaux pour la construction d'indicateurs composites à l'echelle régionale. Compiègne: Université de Picardie Jules Verne; 2015.
2. Caudeville J, Bonnard R, Boudet C, Denys S, Govaert G, Cicolella A. Development of a spatial stochastic multimedia exposure model to assess population exposure at a regional scale. The Science of the total environment. 2012;432:297-308.
3. Pépin P, Lefranc A, Dalbarade M, Grémy I. Atlas de la mortalité par cancer en Ile-de-France. Observatoire régional de santé d'Ile-de-France (ORSIF), 2007.
4. Chatignoux E, Pépin P. Atlas de la mortalité par cancer en Île-de-France 2000–2007. ORS Ile-de-France, 2012 978-2-7371-1804-3.
5. Ineris. Inégalités environnementales – Plaine Résultats pour l'Ile de France. 2012.
6. IARC. List of classification. Vol 1-120 2017 [updated 201702/11/2017]. Available from: http://monographs.iarc.fr/ENG/Classification/latest_classif.php.
7. Gouzy A, Ducos G. La connaissance des éléments traces métalliques: un défi pour la gestion de l'environnement. Air pur. 2008(75):6-10.
8. Ineris. Portail substances chimiques 2017 [updated 201602/11/2017]. Available from: https://substances.ineris.fr/fr/page/21.
9. Caudeville J. Exemples de dispositifs d'évaluation des inégalités environnementales de santé. Annales des Mines-Responsabilité et environnement. 2015(3):20-5.
10. Caudeville J, Boudet C, Govaert G, Bonnard R, Denys S, Cicolella A. Construction d'une plate-forme intégrée pour la cartographie de l'exposition des populations aux substances chimiques de l'environnement. Environnement, Risques & Santé. 2011;10(3):239-42.
11. Ineris. Exposition environnementale et vulnérabilité des populations : Les inégalités environnementales sur les territoires. Verneuil en Halatte: 2014.
12. Townsend P, Phillimore P, Beattie A. Health and deprivation: inequality and the North. 1988 1988.
13. Carstairs V, Morris R. Deprivation: explaining differences in mortality between Scotland and England and Wales. BMJ (Clinical research ed). 1989;299(6704):886-9.
14. Rey G, Rican S, Jougla E. Mesure des inégalités de mortalité par cause de décès. Approche écologique à l'aide d'un indice de désavantage social. Bull Epidémiol Hebdomadaire. 2011:8-9.
15. Lang T, Cases C, Chauvin P, Jougla E, Jusot F, Laporte A, et al. Indicateurs de suivi de l'évolution des inégalités sociales de santé dans les systèmes d'information en santé. Haut Conseil de la santé publique, Paris, 2013.
16. Caudeville J, Rican S. Étude de l'inégalité socio-environnementale en France: associations spatiales entre désavantage social des populations et proximité à un site potentiellement dangereux. Environnement, Risques & Santé. 2016;15(1):39-47.
17. Sénécal N, Telle-Lamberton M, Camard J-P. Repères sur la santé en Île-de-France. 2015 978-2-7371-1897-5.

18. Goovaerts P. Geostatistical analysis of disease data: estimation of cancer mortality risk from empirical frequencies using Poisson kriging. Int J Health Geogr. 2005;4:31.

19. Wakefield J. A critique of statistical aspects of ecological studies in spatial epidemiology. Environmental and Ecological Statistics. 2004;11(1):31-54.

20. Elliott P, Savitz DA. Design issues in small-area studies of environment and health. Environmental health perspectives. 2008;116(8):1098.

21. Anselin L, Lozano N, Koschinsky J. Rate transformations and smoothing. Urbana. 2006;51:61801.

22. Goovaerts P. Geostatistical analysis of health data: state-of-the-art and perspectives. Geoenv VI–Geostatistics for Environmental Applications. 2008:3-22.

23. Lawson A, Biggeri A, Boehning D, Lesaffre E, Viel J, Clark A, et al. Disease mapping models: an empirical evaluation. Disease Mapping Collaborative Group. Statistics in medicine. 2000;19(17-18):2217-41.

24. Fotheringham AS, Brunsdon C, Charlton M. Geographically weighted regression: the analysis of spatially varying relationships: John Wiley & Sons; 2003.

25. Brunsdon C, Fotheringham AS, Charlton M. Spatial nonstationarity and autoregressive models. Environment and Planning A. 1998;30(6):957-73.

26. Brunsdon C, Fotheringham S, Charlton M. Geographically weighted regression. Journal of the Royal Statistical Society: Series D (The Statistician). 1998;47(3):431-43.

27. Goovaerts P, Gebreab S. How does Poisson kriging compare to the popular BYM model for mapping disease risks? International journal of health geographics. 2008;7(1):6.

28. Goovaerts P. Medical geography: a promising field of application for geostatistics. Mathematical Geosciences. 2009;41(3):243.

29. Palm R. La régression géographique pondérée. Principes et applications. Gembloux. Belgique.: Faculté des sciences agronomiques, 1994.

30. Menvielle G, Luce D, Geoffroy-Perez B, Chastang JF, Leclerc A. Social inequalities and cancer mortality in France, 1975-1990. Cancer causes & control : CCC. 2005;16(5):501-13.

31. Brunsdon C, Fotheringham S, Charlton M. Geographically weighted regression as a statistical model. 2000.

32. Gardeur E. CIRCÉ : cancers, inégalités régionales, cantonales et environnement. Croisement de données sanitaires, environnementales et socio-économiques en Lorraine : approche géostatistique. 2017.

33. Pénnognon L. Projet CIRCÉ : Cancer, facteurs environnementaux et facteurs socio-économiques. Croisement des données par l'approche géostatistique. Résultats pour la région Bretagne. 2017.

34. Fotheringham AS, Wong DW. The modifiable areal unit problem in multivariate statistical analysis. Environment and planning A. 1991;23(7):1025-44.

35. Sabel CE, Kihal W, Bard D, Weber C. Creation of synthetic homogeneous neighbourhoods using zone design algorithms to explore relationships between asthma and deprivation in Strasbourg, France. Social Science & Medicine. 2013;91:110-21.

36. Oliveau S. Autocorrélation spatiale: leçons du changement d'échelle. L'Espace géographique. 2010;39(1):51-64.

37. Flowerdew R, Manley DJ, Sabel CE. Neighbourhood effects on health: does it matter where you draw the boundaries? Social science & medicine. 2008;66(6):1241-55.

38. Amrhein CG. Searching for the elusive aggregation effect: evidence from statistical simulations. Environment and planning A. 1995;27(1):105-19.

39. Holt D, Steel D, Tranmer M. Area homogeneity and the modifiable areal unit problem. Geographical Systems. 1996;3(2/3):181-200.

40. Darby S, Hill D, Auvinen A, Barros-Dios JM, Baysson H, Bochicchio F, et al. Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 European case-control studies. BMJ (Clinical research ed). 2005;330(7485):223.

41. Kafadar K. Choosing among two-dimensional smoothers in practice. Computational Statistics & Data Analysis. 1994;18(4):419-39.

42. Talbot TO, Kulldorff M, Forand SP, Haley VB. Evaluation of spatial filters to create smoothed maps of health data. Stat Med. 2000;19(17-18):2399-408.

43. Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. Annals of the institute of statistical mathematics. 1991;43(1):1-20.

44. Best N, Richardson S, Thomson A. A comparison of Bayesian spatial models for disease mapping. Statistical methods in medical research. 2005;14(1):35-59.

45. Christensen OF, Waagepetersen R. Bayesian prediction of spatial count data using generalized linear mixed models. Biometrics. 2002;58(2):280-6.

46. Pickle LW. Exploring spatio-temporal patterns of mortality using mixed effects models. Statistics in Medicine. 2000;19(1718):2251-63.

47. Schabenberger O, Gotway CA. Statistical methods for spatial data analysis: CRC press; 2017.

48. Leyland AH, Davies CA. Empirical Bayes methods for disease mapping. Statistical methods in medical research. 2005;14(1):17-34.

49. Johnson GD. Small area mapping of prostate cancer incidence in New York State (USA) using fully Bayesian hierarchical modelling. International Journal of Health Geographics. 2004;3(1):29.

50. Clayton D, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. Biometrics. 1987:671-81.

51. Marshall RJ. Mapping disease and mortality rates using empirical Bayes estimators. Applied Statistics. 1991:283-94.

52. Chiles J-P, Delfiner P. Geostatistics: modeling spatial uncertainty: John Wiley & Sons; 2009.

53. Goovaerts P. Geostatistics for natural resources evaluation: Oxford University Press on Demand; 1997.

54. Waller LA, Gotway CA. Applied spatial statistics for public health data: John Wiley & Sons; 2004.

55. Pickle L. Spatial analysis of disease. Biostatistical Applications in Cancer Research. 7. Boston, Kluwer Academic Publishers: Beam C. ; 2002. p. 113-50.

56. Matheron G. Traite de Geotatistique Appliquee, Tome I: Memoires du Bureau de Recherches Geologiques et Minieres1962.

57. Matheron G. Le krigeage universel. Paris: Ecole supérieure des mines de Paris; 1969.

58. Goovaerts P. Geostatistical tools for characterizing the spatial variability of microbiological and physico-chemical soil properties. Biology and Fertility of soils. 1998;27(4):315-34.

59. Shao C. Approaches to the spatial modelling of cancer incidence and mortality in metropolitan Perth, Western Australia, 1990-2005: Edith Cowan University Western Australia; 2011.

60. Oliver M, Webster R, Lajaunie C, Muir K, Parkes S, Cameron A, et al. Binomial cokriging for estimating and mapping the risk of childhood cancer. Mathematical Medicine and Biology: A Journal of the IMA. 1998;15(3):279-97.

61. Nielsen AA. Geostatistics and analysis of spatial data. Informatics and Mathematical Modelling, Technical University of Denmark, DTU. 2009:7-12.

62. Monestiez P, Dubroca L, Bonnin E, Durbec J-P, Guinet C. Comparison of model based geostatistical methods in ecology: application to fin whale spatial distribution in northwestern Mediterranean Sea. Geostatistics Banff 2004: Springer; 2005. p. 777-86.

63. Monestiez P, Dubroca L, Bonnin E, Durbec J-P, Guinet C. Geostatistical modelling of spatial distribution of Balaenoptera physalus in the Northwestern Mediterranean Sea from sparse count data and heterogeneous observation efforts. Ecological Modelling. 2006;193(3):615-28.

64. Goovaerts P, Jacquez GM, Greiling D. Exploring Scale-Dependent Correlations Between Cancer Mortality Rates Using Factorial Kriging and Population-Weighted Semivariograms. Geographical Analysis. 2005;37(2):152-82.

65. Cloutier M-S, Apparicio P. L'environnement autour des écoles a-t-il un impact sur le risque routier impliquant des enfants piétons à Montréal?. Apport de la régression de Poisson géographiquement pondérée. Territoire en mouvement Revue de géographie et aménagement Territory in movement Journal of geography and planning. 2008(1):25-38.

66. Fotheringham AS, Charlton ME, Brunsdon C. Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. Environment and planning A. 1998;30(11):1905-27.

67. LeSage JP. A family of geographically weighted regression models. Advances in spatial econometrics: Springer; 2004. p. 241-64.

68. Nakaya T, Fotheringham AS, Brunsdon C, Charlton M. Geographically weighted Poisson regression for disease association mapping. Statistics in medicine. 2005;24(17):2695-717.

69. Apparicio P, Séguin A, Leloup X. Modélisation spatiale de la pauvretéà Montréal: apport méthodologique de la régression géographiquement pondérée. The Canadian Geographer/Le Géographe canadien. 2007;51(4):412-27.

ANNEXES

Annexe 1 : Biais écologique, données fictives illustratives, radon domestique et cancer du poumon

La littérature montre qu'il existe une relation dose-effet entre l'exposition au radon domestique et la mortalité par cancer du poumon (40). Nous avons choisi cet exemple pour illustrer le biais écologique qui peut survenir lors de la mise en relation de données groupées. Le nuage de point figurant dans la Figure 3 montre clairement une association linéaire décroissante entre les niveaux d'exposition domestique au radon et le taux de mortalité par cancer du poumon.

Figure 3. Biais écologique : données fictives mais plausibles. Radon domestique et cancer du poumon dans quatre régions fictives. Données groupées.

Supposons que chaque région ait un niveau de tabagisme différent et inversement proportionnel au niveau de radon domestique. Alors, si l'on examine au sein de chaque région, par une étude basée sur des données individuelles la relation entre l'exposition au radon et la mortalité par cancer du poumon, on obtiendra des nuages de points analogues à ceux de la Figure 4. Au sein de chaque région, une relation dose-effet est observée, mais à partir d'un niveau de base différent de mortalité par cancer du poumon dans chaque région. Ce biais est d'autant plus important que la relation entre le facteur étudié et la pathologie est faible par rapport à celle entre le facteur de confusion et la pathologie, comme dans cet exemple.

Figure 4. Biais écologique : données fictives mais plausibles. Radon domestique et cancer du poumon dans quatre régions fictives. Données individuelles.

Chaque courbe de couleur différente correspond à l'analyse de la relation dose-effet à partir de données individuelles dans la région.

Annexe 2 : les 258 cantons d'Île-de-France

Carte 15. Les cantons d'Île-de-France de Paris et sa proche couronne

Carte 16. Les cantons d'Île-de-France en grande couronne

Figure 5. Distribution de la population des cantons d'Île-de-France en 2006

Annexe 3 : Compléments méthodologiques

Lissage spatial

Cette partie est en grande partie reprise de l'article de Goovaerts (18)

Les avantages principaux des études dites écologiques par rapport aux études sur des données individuelles sont qu'elles utilisent des données collectées de manière routinière et sont généralement effectuées sur des zones géographiques plus vastes avec des contrastes d'exposition plus élevés (19). La taille optimale de ces zones doit combiner les deux contraintes suivantes : elles doivent être suffisamment petites pour permettre une vue assez précise des caractéristiques géographiques de l'incidence de la pathologie, mais de taille suffisante pour limiter l'incertitude causée par des valeurs extrêmes non fiables sur des zones peu peuplées. Cet effet est connu sous le nom de "problème du petit nombre".

Différents types de lissage statistique ont été développés pour filtrer les variations locales à petite échelle au sein des cartes de mortalité (18, 23, 41) . Ces méthodes diffèrent grandement dans leurs exigences informatiques, ainsi que dans les hypothèses sous-jacentes sur les modèles spatiaux et la distribution des valeurs de risque. Les lissages les plus simples sont déterministes et sont réalisés à partir d'une moyenne pondérée simple des taux voisins. Les poids peuvent être basés, par exemple, sur l'inverse du carré de la distance par rapport à la surface à lisser et/ou sur la taille de la population de ces zones (42). Une limitation de ces lissages simples est qu'en l'absence de toute modélisation probabiliste, l'incertitude attachée aux taux lissés ne peut être quantifiée.

Les approches fondées sur un modèle traitent le nombre de décès ou de cas observés comme la réalisation d'une variable aléatoire avec une distribution de probabilité spécifique, de Poisson ou binomiale. Au fil des années, les statisticiens ont développé des modèles de complexité croissante, combinant des effets fixes avec des effets aléatoires non corrélés et spatialement structurés, conduisant à des modèles mixtes ou à des modèles hiérarchiques (43-47). La plupart de ces méthodes ont été développées dans un cadre bayésien où des distributions a priori sont affectées aux paramètres du modèle, les paramètres de ces distributions a priori étant qualifiés d'hyperparamètres. Ceci permet de prendre en compte toutes les sources d'incertitude du modèle. En revanche, l'estimation des paramètres de ce type de modèle est complexe. Elle est réalisée à l'aide de procédures itératives, telles que les méthodes « Markov chain Monte Carlo » (MCMC), qui requièrent beaucoup d'informatisation et nécessitent des ajustements, ce qui rend leur application et leur interprétation difficiles pour les non-statisticiens (18,48,49).

Les méthodes bayésiennes empiriques simplifient grandement la procédure d'estimation en attribuant des estimations ponctuelles aux hyperparamètres, obtenues à partir des données par la méthode du maximum de vraisemblance (50) ou la méthode des moments (51). Bien que les méthodes empiriques négligent la variabilité associée à l'estimation des paramètres et ne permettent que le calcul d'erreurs types approximatives pour le risque, elles sont plus faciles à mettre en œuvre et sont préférées par les praticiens (18).

Une modélisation probabiliste des données spatiales sanitaires a également été réalisée dans la littérature géostatistique, en dehors du courant dominant en épidémiologie. La géostatistique fournit un ensemble d'outils statistiques pour l'analyse des données distribuées dans l'espace et le temps. Elle permet la description de structures spatiales dans les données, l'incorporation de multiples sources d'information dans la cartographie, et surtout la modélisation de l'incertitude spatiale et son utilisation dans la prise de décision (52, 53). Depuis son développement dans l'industrie minière, la géostatistique est devenue le principal outil d'analyse de données spatiales dans divers domaines allant des sciences de la terre et de l'atmosphère à l'agriculture, à la pédologie, aux études environnementales puis à l'évaluation épidémiologique (54). La mise en œuvre traditionnelle des méthodes géostatistiques ne tient toutefois pas compte de l'hétéroscédasticité des taux et dénombrements de la maladie, c'est-à-dire du fait que leur variance dans chaque lieu varie

en fonction de la taille de la population (55). Ses outils de base, le semivariogramme et les algorithmes de krigeage développés par Matheron (56, 57) doivent donc être développés pour tenir compte de la nature spécifique des données de santé (18).

Variogramme et krigeage de Poisson

Avant de préciser en quoi consiste le krigeage de Poisson, nous décrivons les notions de scattergrammes et de semivariogramme introduits par la géostatistique et nécessaires à la compréhension du krigeage.

Scattergrammes et variogrammes

Soit z(u), une fonction associant à un point u de l'espace ou du plan, une grandeur particulière. En épidémiologie, z(u) peut être, par exemple, le taux de mortalité dans un canton dont le barycentre, ou centroïde, est u, défini par ses coordonnées cartésiennes.

On associe à z(u), la valeur de z en u+h, z(u+h), ou h est un vecteur de déplacement, caractérisable par ses coordonnées dans le plan mais également par sa norme |h| et sa direction. On peut tracer le graphe du nuage de points (z(u),(u+h)), appelé h-scattergramme dans la littérature géostatistique (de scatter : dispersion) (58). Un point donné correspond donc à une valeur de z en u et à la valeur de z à une distance |h| de u. On a N(h) points, N(h) correspondant à la somme pour tous les u, des voisins à une distance |h|. Goovaerts donne un exemple avec les mesures de pH sur un sol présentée à la Figure 6.

On peut calculer le coefficient de corrélation linéaire associé aux observations pour chaque distance, on voit que celle-ci diminue à mesure que la distance de séparation augmente, ce qui est cohérent avec l'existence d'une auto-corrélation spatiale.

Figure 6. Scattergrammes entre les valeurs de pH du sol séparées par une distance de 1 m, 2 m, 3 m et 4 m (58)

En géostatistique, les caractéristiques spatiales sont plutôt décrites en termes de dissimilarité entre observations en fonction de la distance de séparation qu'en similarité (corrélations). La dissimilarité moyenne entre les données séparées par un vecteur h est mesurée par le semi-variogramme expérimental g (h), égal à la moitié de l'écart quadratique moyen entre les composantes de chaque paire (z(u), z(u+h)) (58). Il s'écrit :

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} (z(u_\alpha) - z(u_\alpha + h))^2 \quad (2.1)$$

Où α indice le nombre de paires à la distance |h|et N(h) le nombre de paires total.

Le problème est formalisé mathématiquement par la notion de fonction aléatoire, notée Z. il s'agit d'une application de la région étudiée à une famille de variables aléatoires spatialement dépendantes (59).

Soit u un point de la zone étudiée, caractérisé par ses coordonnées cartésiennes, Z(u) désigne alors la variable aléatoire au point u.

Comme pour toute variable aléatoire, l'espérance mathématique et la variance de Z(u) sont notées E[Z(u)] et Var[Z(u)]. Elles ne dépendent que de u et peuvent donc s'écrire :

$$m(u) = E[Z(u)] \qquad (2.2)$$

$$s(u) = Var[Z(u)] = E[(Z(u)-m(u))^2] \qquad (2.3)$$

Les relations spatiales sont analysées en examinant les variables aléatoires Z(u), Z(u+h) et leur différence Z(u) - Z(u+h).

Z(u) - Z(u+h) représente l'écart de la variable aléatoire entre deux points distants de |h|

Si l'on suppose que l'espérance de Z est constante localement, l'espérance de cette différence est égale à zéro et la variance caractérise le carré de l'écart moyen entre les deux points u et u+h.

Pour des raisons de symétrie, la variance au point u est égale à la moitié de cette variance et est dénommée semi-variogramme (60). Le semi-variogramme, théorique cette fois, souvent abrégé variogramme, est donc défini par :

$$\gamma(u, u+h) = \frac{Var[Z(u) - Z(u+h)]}{2} \qquad (2.4)$$

L'application des techniques utilisées en géostatistique suppose que γ ne dépend que de h, dite hypothèse intrinsèque (61).

On a donc

$$\gamma(h) = \frac{Var[Z(u) - Z(u+h)]}{2} \qquad (2.5)$$

Le semi-variogramme expérimental défini en (2.1) est un estimateur du semi-variogramme théorique défini en (2.5).

En épidémiologie, pour tenir compte de la variabilité des indicateurs associée aux variations de la taille de la population selon les unités géographiques, on définit un estimateur pondéré par celle-ci de la façon suivante :

$$\hat{\gamma}(h) = \frac{1}{2\sum_{\alpha=1}^{N(h)} n(u_\alpha)n(u_\alpha+h)} \sum_{\alpha=1}^{N(h)} n(u_\alpha)n(u_\alpha + h)(z(u_\alpha) - z(u_\alpha + h))^2 \quad (2.5)$$

Où $n(u_\alpha)$ et $n(u_\alpha + h)$ sont les tailles des populations des unités géographiques de barycentre ou centroïde u_α et $u_\alpha + h$.

L'algorithme de régression par les moindres carrés pondérés permet alors d'ajuster un modèle aux données expérimentales, comme l'illustre la Figure 7 portant sur la mortalité par cancer du cerveau et de déterminer également des estimations lissées par $\widehat{z_{MPP}}$ estimateur par la moyenne pondérée sur la population

$$\widehat{z_{MPP}}(u_\alpha) = \sum_{i=1}^{K} \lambda_i(u_\alpha)z(u_i) \quad \lambda_i(u_\alpha) = \frac{n(u_i)}{\sum_{i=1}^{K} n(u_i)} \quad (2.6)$$

Cet estimateur n'est toutefois pas un estimateur sans biais de variance minimale. Ceci a conduit Monestiez et ses collaborateurs à élaborer l'estimateur par krigeage de Poisson (62, 63).

Figure 7. Semi-variogramme des taux de mortalité par cancer du cerveau entre 50 et 69 ans calculés à l'aide de l'estimateur pondéré en fonction de la population, femmes blanches, États-Unis, Nouvelle-Angleterre, 295 comtés, période 1950-1969 (64).

Les points désignent le variogramme expérimental et la courbe l'ajustement par moindres carrés pondérés

Estimation par krigeage de Poisson

L'estimateur par krigeage de Poisson (KP) est tel que

$$\widehat{z_{KP}}(u_\alpha) = \sum_{i=1}^{K} \lambda_i(u_\alpha) z(u_i) \quad (2.7)$$

Et la condition d'estimateur sans biais de variance minimale se traduit par un système d'équation linéaires portant sur les poids $\lambda_i(u_\alpha)$ et dénommé « système d'équations de krigeage de Poisson » Goovaerts (18)

Le variogramme correspondant est :

$$\hat{\gamma}(h) = \frac{1}{2\sum_{\alpha=1}^{N(h)} \frac{n(u_\alpha)n(u_\alpha+h)}{n(u_\alpha)+n(u_\alpha+h)}} \sum_{\alpha=1}^{N(h)} \frac{n(u_\alpha)n(u_\alpha+h)}{n(u_\alpha)+n(u_\alpha+h)} [(z(u_\alpha) - z(u_\alpha + h))^2 - m*] \quad (2.8)$$

Où m* est la moyenne des taux pondérée par la taille de la population

Goovaerts a comparé les techniques de lissage précitées. La Figure 8 ci-dessous compare les différents variogrammes obtenus sur l'exemple de la mortalité par cancer du sein en Nouvelle-Angleterre entre 1970 et 1994 (18)

Figure 8. Semivariogrammes des taux de mortalité par cancer du sein en Nouvelle-Angleterre et modèle ajusté associé

Le semi-variogramme traditionnel correspond à l'équation (2.1) $n(u_\alpha) = 1 \ \forall \ \alpha$, celui dit pondéré par la population semivariogramme correspond à l'équation (2.5). Le semi-variogramme de krigeage correspond à l'équation (2.8). La courbe pleine indique le modèle ajusté par régression des moindres carrés pondérée.

Régression géographiquement pondérée

L'article de Cloutier et Apparicio (65) définit de façon synthétique la notion de régression géographiquement pondérée et renvoie aux principales références utiles pour une compréhension détaillée[1] :

« Afin de contrer ce problème de l'instabilité spatiale du modèle, mais aussi afin d'analyser localement la relation entre la variable dépendante et ses prédicteurs, les géographes ont recours de plus en plus fréquemment à ce que Fotheringham et ses collègues ont appelé la régression géographiquement pondérée (Geographically weighted regression – GWR) (24, 66, 67). Succinctement, les modèles GWR sont en fait des extensions des modèles de régression classique. Par exemple, un modèle GWR Poisson s'écrit de la façon suivante, où (u_i, v_i) représente les coordonnées géographiques du point i (ou du centroïde du polygone i) et où les paramètres $ß_0$ et $ß_k$ peuvent varier dans l'espace (24, 66, 68).

$$\log\left(\left[E(y_i)\right]\right) = \beta_0(u_i, v_i) + \sum_{j=1}^{p} \beta_j(u_i, v_i) x_{ij} + \varepsilon_i$$

L'objectif ici n'est pas de décrire en détail la GWR. Pour une description plus détaillée des paramètres du modèle, on pourra consulter en anglais (24) ou en français (69). Cependant, retenons que la résolution de cette équation de régression sous une distribution de Poisson est basée sur une méthode itérative des moindres carrés (iteratively reweighted least squares) et sur une matrice de pondération qui varie en fonction de la distance séparant les unités i et j. Ainsi, les modèles GWR produisent une équation de régression pour chaque entité spatiale, et des valeurs locales pour les paramètres ($ß_0$, $ß_j$), les résidus, les T de Student, etc. Ces valeurs peuvent donc être cartographiées et analysées dans leur contexte spécifique. »

---

[1] Les numéros associés aux références sont ceux de la bibliographie du présent rapport.

Annexe 4 : Corrélations entre les expositions et les taux de mortalité des cantons d'Île-de-France

Tableau 12. Coefficients de corrélations linéaires entre les taux de mortalité standardisés des cantons d'Île-de-France (hommes) et les facteurs sociaux et environnementaux

	Code CIM10	Défaveur	Particules	Concentrations atmosphériques				Doses journalières d'exposition				Indices spatialisés relatifs			
Hommes		FDep08	PM10	Cd	Pb	Ni	Cr	Cd	Pb	Ni	Cr	Cd	Pb	Ni	Total
Tous cancers	C00-C97	0,56	-0,20	-0,21	-0,23	-0,30	-0,05	0,23	0,06	-0,02	0,21	-0,08	-0,03	-0,30	-0,15
Lèvre cavité buccale pharynx	C00-C14	0,31	0,01	-0,03	-0,03	-0,01	-0,06	0,22	-0,07	-0,08	0,06	0,12	-0,09	-0,04	-0,02
Œsophage	C15	0,28	-0,19	-0,19	-0,19	-0,17	-0,10	0,20	0,03	0,04	0,19	-0,07	-0,04	-0,16	-0,12
Estomac	C16	0,28	0,03	0,01	0,01	-0,06	0,02	0,13	0,00	-0,10	0,03	0,10	0,01	-0,09	0,03
Colon-rectum	C18-C21	0,15	-0,08	-0,06	-0,07	-0,07	0,04	0,00	0,03	-0,01	0,09	-0,06	0,00	-0,07	-0,05
Foie	C22	0,47	-0,02	-0,02	-0,04	-0,11	-0,04	0,10	-0,07	0,15	0,07	0,03	-0,09	-0,06	-0,07
Pancréas	C25	0,01	0,00	0,01	0,01	-0,01	0,04	-0,05	0,11	0,01	-0,03	-0,02	0,13	0,00	0,09
Larynx	C32	0,19	0,00	-0,04	-0,04	-0,07	-0,02	0,08	-0,06	-0,07	-0,01	0,00	-0,08	-0,09	-0,09
Poumon	C33-C34	0,65	-0,13	-0,10	-0,13	-0,21	0,00	0,15	0,09	-0,11	0,15	-0,02	0,05	-0,23	-0,05
Plèvre	C384, C450, C459	0,17	-0,07	-0,06	-0,07	-0,15	0,02	0,04	0,06	0,02	0,00	-0,05	0,03	-0,14	-0,05
Sein	C50	0,10	0,07	0,06	0,06	0,03	-0,08	0,05	-0,06	-0,05	-0,08	0,10	-0,04	0,02	0,03
Ovaire	C56, C570-C574, C578														
Prostate	C61	0,16	-0,20	-0,24	-0,25	-0,27	-0,09	0,21	0,04	0,05	0,19	-0,11	-0,06	-0,25	-0,18
Vessie	C67	0,26	-0,07	-0,08	-0,11	-0,13	0,02	0,12	0,09	-0,05	0,09	-0,01	0,05	-0,14	-0,01
Rein	C64, C66-68	0,22	-0,07	-0,06	-0,07	-0,13	0,01	0,04	0,06	-0,02	0,03	-0,04	0,04	-0,13	-0,03
Système nerveux central	C70-C72	-0,10	0,01	0,00	0,01	-0,03	-0,12	-0,04	-0,08	-0,02	-0,02	-0,04	-0,08	-0,04	-0,09
Lymphome non Hodgkinien	C82-C85, C96	-0,06	-0,07	-0,10	-0,10	-0,06	-0,07	0,01	0,02	0,10	0,08	-0,10	-0,01	-0,03	-0,07
Toutes leucémies	C91 -C95	0,03	-0,11	-0,11	-0,11	-0,12	-0,05	0,01	0,09	0,14	0,10	-0,11	0,06	-0,07	-0,03

Taux standardisé pour 100 000 habitants, population de référence : population française de 2006 tous sexes confondus. 21 classes d'âge : [0-1[, [1-5[, [5-10[, [10-15[, …[90-95[, ≥95 ans
FDep : indice de défaveur sociale (Rey, 2011), PM10 : particules de diamètre inférieur au millième de millimètre, Cd : cadmium, Pb : plomb, Ni : nickel, Cr : chrome
En rouge : coefficients de corrélation linéaire > 0,35
Sources : expositions : Ineris. Décès : CépiDC. Population : Insee. Exploitation ORS Île-de-France

Tableau 13. Coefficients de corrélations linéaires entre les taux de mortalité standardisés des cantons d'Île-de-France (femmes) et les facteurs sociaux et environnementaux

Localisations cancéreuses	Code CIM10	Défaveur	Particules	Concentrations atmosphériques				Doses journalières d'exposition				Indices spatialisés relatifs			
Femmes		FDep08	PM10	Cd	Pb	Ni	Cr	Cd	Pb	Ni	Cr	Cd	Pb	Ni	Total
Tous cancers	C00-C97	0,33	-0,20	-0,26	-0,27	-0,22	0,06	0,29	0,14	-0,09	0,32	-0,08	0,04	-0,24	-0,08
Lèvre cavité buccale pharynx	C00-C14	0,26	-0,06	-0,09	-0,08	-0,08	0,26	0,05	0,07	0,00	0,04	-0,07	0,04	-0,08	-0,03
Œsophage	C15	0,13	0,00	-0,05	-0,04	-0,03	-0,10	0,10	0,02	-0,04	0,08	0,02	0,00	-0,05	0,00
Estomac	C16	0,31	-0,05	-0,04	-0,06	-0,07	0,03	0,04	-0,01	-0,16	0,08	-0,02	-0,04	-0,11	-0,08
Colon-rectum	C18-C21	0,18	-0,19	-0,20	-0,21	-0,21	-0,01	0,17	0,08	-0,07	0,25	-0,09	0,00	-0,22	-0,11
Foie	C22	0,32	0,08	0,04	0,02	0,00	0,09	0,09	-0,02	0,00	0,02	0,09	-0,01	0,00	0,04
Pancréas	C25	-0,02	0,05	0,01	0,00	0,00	-0,07	0,09	-0,12	-0,02	0,04	0,07	-0,14	0,00	-0,07
Larynx	C32	0,09	-0,06	-0,09	-0,08	-0,03	0,05	0,04	0,04	0,00	0,06	-0,08	0,01	-0,03	-0,04
Poumon	C33-C34	0,12	0,13	0,14	0,13	0,15	0,08	-0,06	-0,05	-0,03	-0,12	0,11	0,00	0,13	0,09
Plèvre	C384, C450, C459	0,09	-0,05	-0,03	-0,05	-0,07	-0,02	0,02	-0,01	0,03	-0,04	-0,03	-0,04	-0,06	-0,06
Sein	C50	-0,02	-0,22	-0,24	-0,25	-0,23	-0,06	0,13	0,18	-0,01	0,31	-0,17	0,09	-0,23	-0,09
Ovaire	C56, C570-C574, C578	-0,23	-0,03	-0,05	-0,05	0,01	0,07	0,07	0,11	-0,03	0,05	0,00	0,10	0,01	0,07
Prostate	C61														
Vessie	C67	0,09	-0,05	-0,05	-0,04	-0,02	0,09	0,09	0,18	-0,11	0,02	0,01	0,18	-0,05	0,13
Rein	C64, C66-68	0,18	-0,14	-0,11	-0,10	-0,11	0,10	0,08	0,27	-0,06	0,05	-0,07	0,26	-0,12	0,12
Système nerveux central	C70-C72	-0,09	0,05	0,03	0,04	0,06	-0,02	-0,02	-0,02	0,07	-0,02	0,03	-0,01	0,08	0,04
Lymphome non Hodgkinien	C82-C85, C96	0,09	-0,02	-0,03	-0,05	-0,04	0,04	0,02	0,00	-0,08	0,08	-0,02	-0,03	-0,07	-0,05
Toutes leucémies	C91 -C95	-0,05	-0,17	-0,19	-0,20	-0,11	0,17	0,14	0,21	-0,06	0,19	-0,11	0,15	-0,13	0,02

Taux standardisé pour 100 000 habitants, population de référence : population française de 2006 tous sexes confondus. 21 classes d'âge : [0-1[, [1-5[, [5-10[, [10-15[, …[90-95[, ≥95 ans
FDep : indice de défaveur sociale (Rey, 2011), PM10 : particules de diamètre inférieur au millième de millimètre, Cd : cadmium, Pb : plomb, Ni : nickel, Cr : chrome
En rouge : coefficients de corrélation linéaire > 0,35
Sources : expositions : Ineris. Décès : CépiDC. Population : Insee. Exploitation ORS Île-de-France

Tableau 14.Coefficients de corrélations linéaires entre les taux de mortalité standardisés des cantons d'Île-de-France (hommes et femmes) et les facteurs sociaux et environnementaux

Localisations cancéreuses	Code CIM10	Défaveur	Particules	Concentrations atmosphériques				Doses journalières d'exposition				Indices spatialisés relatifs			
Tous		FDep08	PM10	Cd	Pb	Ni	Cr	Cd	Pb	Ni	Cr	Cd	Pb	Ni	Total
Tous cancers	C00-C97	0,60	-0,35	-0,39	-0,41	-0,40	0,00	0,38	0,20	-0,09	0,41	-0,16	0,05	-0,42	-0,18
Lèvre cavité buccale pharynx	C00-C14	0,49	-0,04	-0,08	-0,08	-0,06	0,03	0,17	-0,03	-0,06	0,07	0,03	-0,07	-0,07	-0,06
Œsophage	C15	0,33	-0,21	-0,25	-0,26	-0,21	-0,13	0,30	0,08	-0,01	0,26	-0,05	-0,01	-0,21	-0,10
Estomac	C16	0,47	-0,04	-0,05	-0,07	-0,11	0,06	0,09	0,04	-0,17	0,09	0,01	0,02	-0,16	-0,03
Colon-rectum	C18-C21	0,31	-0,26	-0,27	-0,29	-0,29	0,02	0,21	0,15	-0,05	0,30	-0,14	0,04	-0,30	-0,13
Foie	C22	0,54	-0,03	-0,06	-0,08	-0,12	0,00	0,15	-0,04	0,12	0,12	0,03	-0,07	-0,07	-0,07
Pancréas	C25	-0,01	-0,02	-0,03	-0,04	-0,06	0,00	0,04	0,04	-0,01	0,02	0,00	0,03	-0,06	0,00
Larynx	C32	0,31	-0,12	-0,17	-0,16	-0,12	0,02	0,22	-0,03	-0,11	0,16	-0,03	-0,09	-0,14	-0,13
Poumon	C33-C34	0,61	-0,15	-0,12	-0,15	-0,19	0,02	0,17	0,08	-0,10	0,17	-0,02	0,03	-0,21	-0,06
Plèvre	C384, C450, C459	0,18	-0,13	-0,12	-0,13	-0,19	-0,01	0,05	0,05	0,04	0,02	-0,10	0,01	-0,17	-0,10
Sein	C50	-0,08	-0,12	-0,15	-0,16	-0,16	-0,06	0,06	0,14	-0,04	0,21	-0,11	0,09	-0,16	-0,04
Ovaire	C56, C570-C574, C578														
Prostate	C61														
Vessie	C67	0,30	-0,24	-0,25	-0,25	-0,24	0,09	0,24	0,27	-0,16	0,21	-0,10	0,20	-0,28	0,01
Rein	C64, C66-68	0,30	-0,36	-0,35	-0,35	-0,36	0,08	0,23	0,32	-0,11	0,30	-0,22	0,21	-0,38	-0,07
Système nerveux central	C70-C72	-0,22	0,06	0,05	0,05	0,05	-0,14	-0,05	-0,08	0,03	-0,06	0,02	-0,07	0,06	-0,03
Lymphome non Hodgkinien	C82-C85, C96	0,02	-0,06	-0,09	-0,10	-0,07	-0,01	0,06	0,01	0,00	0,11	-0,05	-0,03	-0,07	-0,07
Toutes leucémies	C91 -C95	-0,07	-0,25	-0,27	-0,28	-0,21	0,08	0,14	0,27	0,00	0,26	-0,20	0,18	-0,20	-0,03

Taux standardisé pour 100 000 habitants, population de référence : population française de 2006 tous sexes confondus. 21 classes d'âge : [0-1[, [1-5[, [5-10[, [10-15[, …[90-95[, ≥95 ans
FDep : indice de défaveur sociale (Rey, 2011), PM10 : particules de diamètre inférieur au millième de millimètre, Cd :cadmium, Pb : plomb, Ni : nickel, Cr : chrome
En rouge : coefficients de corrélation linéaire > 0,35
Sources : expositions : Ineris. Décès : CépiDC. Population : Insee. Exploitation ORS Île-de-France

Tableau 15. Coefficients de corrélations linéaires entre les taux de mortalité standardisés des cantons d'Île-de-France lissés (hommes) et les facteurs sociaux et environnementaux

Localisations cancéreuses	Code CIM10	Défaveur	Particules	Concentrations atmosphériques				Doses journalières d'exposition				Indices spatialisés relatifs			
Hommes		FDep08	PM10	Cd	Pb	Ni	Cr	Cd	Pb	Ni	Cr	Cd	Pb	Ni	Total
Tous cancers	C00-C97	0,60	-0,24	-0,25	-0,28	-0,35	-0,05	0,25	0,06	-0,03	0,25	-0,10	-0,04	-0,35	-0,19
Lèvre cavité buccale pharynx	C00-C14	0,39	0,01	-0,06	-0,05	-0,03	-0,07	0,31	-0,07	-0,09	0,11	0,16	-0,10	-0,05	-0,01
Œsophage	C15	0,33	-0,39	-0,41	-0,40	-0,36	-0,14	0,39	0,08	0,00	0,42	-0,17	-0,08	-0,35	-0,26
Estomac	C16	0,38	0,04	0,00	-0,01	-0,12	0,01	0,19	-0,01	-0,14	0,05	0,14	-0,01	-0,16	0,01
Colon-rectum	C18-C21	0,18	-0,11	-0,09	-0,11	-0,11	0,03	0,04	0,04	-0,03	0,12	-0,08	0,00	-0,11	-0,08
Foie	C22	0,60	-0,05	-0,06	-0,08	-0,17	-0,07	0,14	-0,08	0,15	0,12	0,04	-0,12	-0,12	-0,11
Pancréas	C25	0,01	0,05	0,06	0,07	0,03	0,10	-0,07	0,10	-0,01	-0,06	0,02	0,14	0,02	0,12
Larynx	C32	0,21	-0,01	-0,05	-0,05	-0,08	-0,02	0,08	-0,06	-0,08	0,00	0,01	-0,08	-0,10	-0,09
Poumon	C33-C34	0,71	-0,18	-0,15	-0,19	-0,27	0,02	0,18	0,10	-0,11	0,21	-0,04	0,04	-0,30	-0,09
Plèvre	C384, C450, C459	0,33	-0,16	-0,14	-0,16	-0,31	0,07	0,04	0,13	-0,01	0,05	-0,13	0,08	-0,30	-0,10
Sein	C50	0,16	0,10	0,06	0,06	0,02	-0,24	0,17	-0,20	-0,06	-0,06	0,19	-0,20	0,00	-0,06
Ovaire	C56, C570-C574, C578														
Prostate	C61	0,21	-0,31	-0,35	-0,37	-0,39	-0,11	0,27	0,04	0,03	0,28	-0,19	-0,10	-0,37	-0,29
Vessie	C67	0,28	-0,07	-0,09	-0,12	-0,15	0,02	0,11	0,07	-0,06	0,10	-0,01	0,03	-0,16	-0,04
Rein	C64, C66-68	0,24	-0,08	-0,06	-0,08	-0,14	0,01	0,04	0,06	-0,02	0,04	-0,04	0,03	-0,15	-0,04
Système nerveux central	C70-C72	-0,22	0,01	0,02	0,04	-0,01	-0,24	-0,08	-0,18	-0,05	-0,08	-0,04	-0,18	-0,03	-0,16
Lymphome non Hodgkinien	C82-C85, C96	-0,11	-0,20	-0,25	-0,25	-0,17	-0,16	0,10	0,03	0,08	0,23	-0,20	-0,07	-0,14	-0,20
Toutes leucémies	C91 -C95	0,03	-0,21	-0,21	-0,20	-0,21	-0,06	0,02	0,08	0,22	0,18	-0,21	0,00	-0,13	-0,15

Taux standardisé pour 100 000 habitants, population de référence : population française de 2006 tous sexes confondus. 21 classes d'âge : [0-1[, [1-5[, [5-10[, [10-15[, ...[90-95[, ≥95 ans
FDep : indice de défaveur sociale (Rey, 2011), PM10 : particules de diamètre inférieur au millième de millimètre, Cd :cadmium, Pb : plomb, Ni : nickel, Cr : chrome
En rouge : coefficients de corrélation linéaire > 0,35
Sources : expositions : Ineris. Décès : CépiDC. Population : Insee. Exploitation ORS Île-de-France

Tableau 16. Coefficients de corrélations linéaires entre les taux de mortalité standardisés des cantons d'Île-de-France lissés (femmes) et les facteurs sociaux et environnementaux

Localisations cancéreuses	Code CIM10	Défaveur	Particules	Concentrations atmosphériques				Doses journalières d'exposition				Indices spatialisés relatifs			
Femmes		FDep08	PM10	Cd	Pb	Ni	Cr	Cd	Pb	Ni	Cr	Cd	Pb	Ni	Total
Tous cancers	C00-C97	0,45	-0,27	-0,36	-0,38	-0,29	0,03	0,43	0,16	-0,14	0,45	-0,09	0,03	-0,33	-0,13
Lèvre cavité buccale pharynx	C00-C14	0,39	-0,09	-0,15	-0,14	-0,17	0,41	0,16	0,26	-0,15	0,11	-0,04	0,23	-0,21	0,09
Œsophage	C15	0,08	0,11	-0,01	0,02	0,11	-0,12	0,21	-0,06	-0,19	0,10	0,15	-0,06	0,04	0,05
Estomac	C16	0,51	-0,10	-0,11	-0,16	-0,18	-0,08	0,15	0,04	-0,16	0,20	-0,01	-0,02	-0,22	-0,09
Colon-rectum	C18-C21	0,35	-0,45	-0,47	-0,51	-0,47	-0,06	0,38	0,16	-0,09	0,49	-0,25	-0,03	-0,48	-0,30
Foie	C22	0,44	0,21	0,15	0,11	0,04	0,12	0,04	0,01	-0,06	-0,02	0,19	0,06	0,02	0,15
Pancréas	C25	-0,12	0,11	0,03	0,01	0,07	-0,09	0,20	-0,14	-0,02	0,08	0,17	-0,15	0,06	-0,01
Larynx	C32	0,28	0,07	0,02	0,04	0,10	0,11	-0,07	-0,01	-0,03	-0,06	-0,03	0,01	0,09	0,02
Poumon	C33-C34	0,11	0,38	0,42	0,41	0,41	0,18	-0,29	-0,14	-0,07	-0,37	0,25	0,02	0,37	0,25
Plèvre	C384, C450, C459	0,23	-0,11	-0,07	-0,10	-0,14	0,05	0,00	0,06	0,08	-0,08	-0,07	0,02	-0,11	-0,05
Sein	C50	0,08	-0,41	-0,45	-0,46	-0,44	-0,08	0,26	0,27	-0,05	0,50	-0,31	0,11	-0,44	-0,21
Ovaire	C56, C570-C574, C578	-0,34	-0,05	-0,08	-0,08	0,11	0,17	0,11	0,16	-0,06	0,10	-0,01	0,14	0,09	0,13
Prostate	C61														
Vessie	C67	0,14	-0,04	-0,06	-0,04	0,00	0,08	0,17	0,06	-0,16	0,05	0,06	0,05	-0,05	0,05
Rein	C64, C66-68	0,25	-0,32	-0,26	-0,26	-0,27	0,15	0,16	0,30	-0,10	0,17	-0,17	0,22	-0,30	-0,01
Système nerveux central	C70-C72	-0,17	0,07	0,10	0,10	0,11	-0,22	-0,17	-0,10	0,15	-0,08	-0,01	-0,07	0,16	-0,01
Lymphome non Hodgkinien	C82-C85, C96	0,16	-0,01	-0,02	-0,06	-0,04	0,12	-0,01	-0,05	-0,15	0,11	-0,03	-0,08	-0,09	-0,11
Toutes leucémies	C91 -C95	-0,09	-0,38	-0,40	-0,42	-0,23	0,17	0,30	0,26	-0,08	0,47	-0,22	0,12	-0,25	-0,10

Taux standardisé pour 100 000 habitants, population de référence : population française de 2006 tous sexes confondus. 21 classes d'âge : [0-1[, [1-5[, [5-10[, [10-15[, …[90-95[, ≥95 ans
FDep : indice de défaveur sociale (Rey, 2011), PM10 : particules de diamètre inférieur au millième de millimètre, Cd : cadmium, Pb : plomb, Ni : nickel, Cr : chrome
En rouge : coefficients de corrélation linéaire > 0,35
Sources : expositions : Ineris. Décès : CépiDC. Population : Insee. Exploitation ORS Île-de-France

Tableau 17. Coefficients de corrélations linéaires entre les taux de mortalité standardisés des cantons d'Île-de-France lissés (hommes et femmes) et les facteurs sociaux et environnementaux

Localisations cancéreuses	Code CIM10	Défaveur	Particules	Concentrations atmosphériques				Doses journalières d'exposition				Indices spatialisés relatifs			
Tous		FDep08	PM10	Cd	Pb	Ni	Cr	Cd	Pb	Ni	Cr	Cd	Pb	Ni	Total
Tous cancers	C00-C97	0,65	-0,41	-0,46	-0,48	-0,47	-0,01	0,44	0,20	-0,10	0,48	-0,19	0,02	-0,49	-0,23
Lèvre cavité buccale pharynx	C00-C14	0,59	-0,05	-0,11	-0,11	-0,08	0,04	0,25	0,01	-0,08	0,13	0,06	-0,03	-0,10	-0,03
Œsophage	C15	0,42	-0,41	-0,49	-0,49	-0,37	-0,16	0,59	0,12	-0,11	0,54	-0,11	-0,06	-0,40	-0,23
Estomac	C16	0,63	-0,07	-0,10	-0,13	-0,21	0,01	0,17	0,06	-0,18	0,18	0,02	0,01	-0,26	-0,06
Colon-rectum	C18-C21	0,42	-0,47	-0,50	-0,53	-0,52	-0,03	0,38	0,20	-0,10	0,50	-0,27	0,01	-0,54	-0,30
Foie	C22	0,70	-0,03	-0,07	-0,10	-0,18	-0,04	0,18	-0,03	0,10	0,16	0,05	-0,07	-0,14	-0,08
Pancréas	C25	-0,06	0,02	-0,01	-0,02	-0,03	0,09	0,11	0,08	-0,03	0,04	0,07	0,08	-0,03	0,08
Larynx	C32	0,49	-0,17	-0,27	-0,26	-0,17	0,01	0,44	-0,02	-0,16	0,26	0,03	-0,12	-0,21	-0,15
Poumon	C33-C34	0,70	-0,19	-0,16	-0,19	-0,23	0,04	0,20	0,10	-0,11	0,21	-0,03	0,03	-0,26	-0,08
Plèvre	C384, C450, C459	0,34	-0,23	-0,21	-0,24	-0,36	0,07	0,04	0,16	0,04	0,06	-0,21	0,08	-0,34	-0,15
Sein	C50	-0,03	-0,23	-0,28	-0,29	-0,30	-0,08	0,14	0,18	-0,08	0,34	-0,20	0,08	-0,32	-0,14
Ovaire	C56, C570-C574, C578														
Prostate	C61														
Vessie	C67	0,44	-0,41	-0,43	-0,46	-0,46	0,09	0,39	0,29	-0,18	0,36	-0,19	0,14	-0,50	-0,15
Rein	C64, C66-68	0,39	-0,62	-0,61	-0,65	-0,63	0,09	0,38	0,41	-0,12	0,54	-0,39	0,19	-0,65	-0,25
Système nerveux central	C70-C72	-0,35	0,09	0,10	0,12	0,12	-0,38	-0,13	-0,26	0,07	-0,18	0,02	-0,23	0,14	-0,12
Lymphome non Hodgkinien	C82-C85, C96	0,03	-0,13	-0,19	-0,21	-0,15	-0,01	0,12	-0,02	-0,10	0,25	-0,12	-0,11	-0,17	-0,19
Toutes leucémies	C91 -C95	-0,03	-0,47	-0,48	-0,50	-0,36	0,10	0,22	0,28	0,00	0,50	-0,37	0,11	-0,35	-0,22

Taux standardisé pour 100 000 habitants, population de référence : population française de 2006 tous sexes confondus. 21 classes d'âge : [0-1[, [1-5[, [5-10[, [10-15[, ...[90-95[, ≥95 ans
FDep : indice de défaveur sociale (Rey, 2011), PM10 : particules de diamètre inférieur au millième de millimètre, Cd :cadmium, Pb : plomb, Ni : nickel, Cr : chrome
En rouge : coefficients de corrélation linéaire > 0,35
Sources : expositions : Ineris. Décès : CépiDC. Population : Insee. Exploitation ORS Île-de-France

Annexe 5 : Régression géographiquement pondérée entre la défaveur sociale et le taux de mortalité par cancer du poumon chez les hommes d'Île-de-France. Résultats détaillés.

```
*******************************************************************************
*          Semiparametric Geographically Weighted Regression          *
*                    Release 1.0.90 (GWR 4.0.90)                      *
*                         12 May 2015                                 *
*             (Originally coded by T. Nakaya: 1 Nov 2009)             *
*                                                                     *
*          Tomoki Nakaya(1), Martin Charlton(2), Chris Brunsdon (2)   *
*          Paul Lewis (2), Jing Yao (3), A Stewart Fotheringham (4)   *
*                    (c) GWR4 development team                        *
* (1) Ritsumeikan University, (2) National University of Ireland, Maynooth, *
*       (3) University of Glasgow, (4) Arizona State University       *
*******************************************************************************

Program began at 23/01/2018 16:58:53

*******************************************************************************
Session:
Session control file: G:\ORS\Thematiques\Etudes\CANCERS\CIRCÉ\GWR\GWR4\gwr4_kpoumh_fdep_23012018.ctl
*******************************************************************************
Data filename: G:\ORS\Thematiques\Etudes\CANCERS\CIRCÉ\GWR\centroids_datasets_vf_modifie.txt
Number of areas/points: 258

Model settings--------------------------------
Model type: Gaussian
Geographic kernel: fixed bi-square
Method for optimal bandwidth search: interval search
```

Criterion for optimal bandwidth: AICc

Number of varying coefficients: 2

Number of fixed coefficients: 0

Modelling options----------

Standardisation of independent variables: OFF

Testing geographical variability of local coefficients: OFF

Local to Global Variable selection: OFF

Global to Local Variable selection: OFF

Prediction at non-regression points: OFF

Variable settings----------

Area key: field5: codeCanton

Easting (x-coord): field13 : X

Northing (y-coord): field14: Y

Cartesian coordinates: Euclidean distance

Dependent variable: field11: lissage

Offset variable is not specified

Intercept: varying (Local) intercept

Independent variable with varying (Local) coefficient: field10: Fdep08

****

****

Global regression result

****

< Diagnostic information >

Residual sum of squares: 17993,357219

Number of parameters: 2

(Note: this num does not include an error variance term for a Gaussian model)

ML based global sigma estimate: 8,351149

Unbiased global sigma estimate: 8,383708

-2 log-likelihood: 1827,330255

Classic AIC: 1833,330255

AICc: 1833,424743

BIC/MDL: 1843,989133

CV: 71,262336

R square: 0,511085

Adjusted R square: 0,507250

Variable	Estimate	Standard Error	t(Est/SE)
Intercept	80,316685	0,522206	153,802754
Fdep08	4,281196	0,261707	16,358736

****

GWR (Geographically weighted regression) bandwidth selection

****

Bandwidth search <interval search> min, max, step

1000, 100000, 1000

Bandwdith: 100000,000 Dev: 1826,017 trace(Hat): 2,774 Criterion: 1833,708 Valid_fit

Bandwdith: 99000,000 Dev: 1825,963 trace(Hat): 2,798 Criterion: 1833,704 Valid_fit

Bandwdith: 98000,000 Dev: 1825,906 trace(Hat): 2,823 Criterion: 1833,698 Valid_fit

Bandwdith: 97000,000 Dev: 1825,844 trace(Hat): 2,849 Criterion: 1833,690 Valid_fit

Bandwdith: 96000,000 Dev: 1825,778 trace(Hat): 2,877 Criterion: 1833,681 Valid_fit

Bandwdith: 95000,000 Dev: 1825,707 trace(Hat): 2,906 Criterion: 1833,671 Valid_fit

Bandwdith: 94000,000 Dev: 1825,631 trace(Hat): 2,937 Criterion: 1833,659 Valid_fit

Bandwdith: 93000,000 Dev: 1825,551 trace(Hat): 2,969 Criterion: 1833,645 Valid_fit

Bandwdith: 92000,000 Dev: 1825,466 trace(Hat): 3,003 Criterion: 1833,631 Valid_fit

Bandwdith: 91000,000 Dev: 1825,375 trace(Hat): 3,039 Criterion: 1833,615 Valid_fit

Bandwdith: 90000,000 Dev: 1825,280 trace(Hat): 3,077 Criterion: 1833,598 Valid_fit

Bandwdith: 89000,000 Dev: 1825,181 trace(Hat): 3,117 Criterion: 1833,581 Valid_fit

Bandwdith: 88000,000 Dev: 1825,077 trace(Hat): 3,159 Criterion: 1833,563 Valid_fit

Bandwdith:	87000,000	Dev:	1824,968	trace(Hat):	3,203	Criterion:	1833,546	Valid_fit
Bandwdith:	86000,000	Dev:	1824,854	trace(Hat):	3,249	Criterion:	1833,528	Valid_fit
Bandwdith:	85000,000	Dev:	1824,736	trace(Hat):	3,297	Criterion:	1833,512	Valid_fit
Bandwdith:	84000,000	Dev:	1824,614	trace(Hat):	3,349	Criterion:	1833,495	Valid_fit
Bandwdith:	83000,000	Dev:	1824,485	trace(Hat):	3,403	Criterion:	1833,478	Valid_fit
Bandwdith:	82000,000	Dev:	1824,348	trace(Hat):	3,459	Criterion:	1833,459	Valid_fit
Bandwdith:	81000,000	Dev:	1824,201	trace(Hat):	3,519	Criterion:	1833,436	Valid_fit
Bandwdith:	80000,000	Dev:	1824,044	trace(Hat):	3,581	Criterion:	1833,410	Valid_fit
Bandwdith:	79000,000	Dev:	1823,876	trace(Hat):	3,647	Criterion:	1833,378	Valid_fit
Bandwdith:	78000,000	Dev:	1823,698	trace(Hat):	3,716	Criterion:	1833,344	Valid_fit
Bandwdith:	77000,000	Dev:	1823,512	trace(Hat):	3,789	Criterion:	1833,310	Valid_fit
Bandwdith:	76000,000	Dev:	1823,319	trace(Hat):	3,865	Criterion:	1833,276	Valid_fit
Bandwdith:	75000,000	Dev:	1823,119	trace(Hat):	3,946	Criterion:	1833,244	Valid_fit
Bandwdith:	74000,000	Dev:	1822,912	trace(Hat):	4,030	Criterion:	1833,213	Valid_fit
Bandwdith:	73000,000	Dev:	1822,698	trace(Hat):	4,118	Criterion:	1833,183	Valid_fit
Bandwdith:	72000,000	Dev:	1822,472	trace(Hat):	4,211	Criterion:	1833,150	Valid_fit
Bandwdith:	71000,000	Dev:	1822,230	trace(Hat):	4,308	Criterion:	1833,111	Valid_fit
Bandwdith:	70000,000	Dev:	1821,970	trace(Hat):	4,410	Criterion:	1833,065	Valid_fit
Bandwdith:	69000,000	Dev:	1821,694	trace(Hat):	4,516	Criterion:	1833,012	Valid_fit
Bandwdith:	68000,000	Dev:	1821,401	trace(Hat):	4,627	Criterion:	1832,952	Valid_fit
Bandwdith:	67000,000	Dev:	1821,089	trace(Hat):	4,744	Criterion:	1832,885	Valid_fit
Bandwdith:	66000,000	Dev:	1820,758	trace(Hat):	4,865	Criterion:	1832,809	Valid_fit
Bandwdith:	65000,000	Dev:	1820,406	trace(Hat):	4,992	Criterion:	1832,724	Valid_fit
Bandwdith:	64000,000	Dev:	1820,030	trace(Hat):	5,125	Criterion:	1832,629	Valid_fit
Bandwdith:	63000,000	Dev:	1819,630	trace(Hat):	5,264	Criterion:	1832,520	Valid_fit
Bandwdith:	62000,000	Dev:	1819,204	trace(Hat):	5,409	Criterion:	1832,400	Valid_fit
Bandwdith:	61000,000	Dev:	1818,759	trace(Hat):	5,560	Criterion:	1832,275	Valid_fit
Bandwdith:	60000,000	Dev:	1818,291	trace(Hat):	5,720	Criterion:	1832,145	Valid_fit
Bandwdith:	59000,000	Dev:	1817,797	trace(Hat):	5,887	Criterion:	1832,006	Valid_fit
Bandwdith:	58000,000	Dev:	1817,277	trace(Hat):	6,063	Criterion:	1831,858	Valid_fit
Bandwdith:	57000,000	Dev:	1816,730	trace(Hat):	6,247	Criterion:	1831,702	Valid_fit
Bandwdith:	56000,000	Dev:	1816,152	trace(Hat):	6,440	Criterion:	1831,536	Valid_fit

Bandwdith: 55000,000 Dev: 1815,531 trace(Hat): 6,644 Criterion: 1831,349 Valid_fit

Bandwdith: 54000,000 Dev: 1814,859 trace(Hat): 6,859 Criterion: 1831,136 Valid_fit

Bandwdith: 53000,000 Dev: 1814,131 trace(Hat): 7,085 Criterion: 1830,892 Valid_fit

Bandwdith: 52000,000 Dev: 1813,347 trace(Hat): 7,325 Criterion: 1830,621 Valid_fit

Bandwdith: 51000,000 Dev: 1812,512 trace(Hat): 7,579 Criterion: 1830,331 Valid_fit

Bandwdith: 50000,000 Dev: 1811,624 trace(Hat): 7,849 Criterion: 1830,025 Valid_fit

Bandwdith: 49000,000 Dev: 1810,693 trace(Hat): 8,137 Criterion: 1829,714 Valid_fit

Bandwdith: 48000,000 Dev: 1809,720 trace(Hat): 8,444 Criterion: 1829,404 Valid_fit

Bandwdith: 47000,000 Dev: 1808,699 trace(Hat): 8,771 Criterion: 1829,093 Valid_fit

Bandwdith: 46000,000 Dev: 1807,626 trace(Hat): 9,122 Criterion: 1828,782 Valid_fit

Bandwdith: 45000,000 Dev: 1806,493 trace(Hat): 9,496 Criterion: 1828,464 Valid_fit

Bandwdith: 44000,000 Dev: 1805,300 trace(Hat): 9,895 Criterion: 1828,144 Valid_fit

Bandwdith: 43000,000 Dev: 1804,032 trace(Hat): 10,325 Criterion: 1827,819 Valid_fit

Bandwdith: 42000,000 Dev: 1802,678 trace(Hat): 10,790 Criterion: 1827,488 Valid_fit

Bandwdith: 41000,000 Dev: 1801,239 trace(Hat): 11,290 Criterion: 1827,155 Valid_fit

Bandwdith: 40000,000 Dev: 1799,701 trace(Hat): 11,829 Criterion: 1826,812 Valid_fit

Bandwdith: 39000,000 Dev: 1798,053 trace(Hat): 12,408 Criterion: 1826,455 Valid_fit

Bandwdith: 38000,000 Dev: 1796,257 trace(Hat): 13,034 Criterion: 1826,061 Valid_fit

Bandwdith: 37000,000 Dev: 1794,284 trace(Hat): 13,714 Criterion: 1825,622 Valid_fit

Bandwdith: 36000,000 Dev: 1792,131 trace(Hat): 14,455 Criterion: 1825,147 Valid_fit

Bandwdith: 35000,000 Dev: 1789,777 trace(Hat): 15,261 Criterion: 1824,632 Valid_fit

Bandwdith: 34000,000 Dev: 1787,225 trace(Hat): 16,138 Criterion: 1824,094 Valid_fit

Bandwdith: 33000,000 Dev: 1784,490 trace(Hat): 17,098 Criterion: 1823,580 Valid_fit

Bandwdith: 32000,000 Dev: 1781,605 trace(Hat): 18,151 Criterion: 1823,152 Valid_fit

Bandwdith: 31000,000 Dev: 1778,598 trace(Hat): 19,300 Criterion: 1822,853 Valid_fit

Bandwdith: 30000,000 Dev: 1775,503 trace(Hat): 20,543 Criterion: 1822,715 Valid_fit

Bandwdith: 29000,000 Dev: 1772,301 trace(Hat): 21,879 Criterion: 1822,727 Valid_fit

Bandwdith: 28000,000 Dev: 1768,977 trace(Hat): 23,318 Criterion: 1822,906 Valid_fit

Bandwdith: 27000,000 Dev: 1765,503 trace(Hat): 24,887 Criterion: 1823,299 Valid_fit

Bandwdith: 26000,000 Dev: 1761,869 trace(Hat): 26,597 Criterion: 1823,943 Valid_fit

Bandwdith: 25000,000 Dev: 1758,073 trace(Hat): 28,477 Criterion: 1824,925 Valid_fit

Bandwdith: 24000,000 Dev: 1754,052 trace(Hat): 30,565 Criterion: 1826,302 Valid_fit

```
Bandwdith: 23000,000  Dev:  1749,771  trace(Hat):  32,888  Criterion:  1828,146 Valid_fit
Bandwdith: 22000,000  Dev:  1745,184  trace(Hat):  35,446  Criterion:  1830,452 Valid_fit
Bandwdith: 21000,000  Dev:  1740,302  trace(Hat):  38,309  Criterion:  1833,476 Valid_fit
Bandwdith: 20000,000  Dev:  1734,980  trace(Hat):  41,626  Criterion:  1837,582 Valid_fit
Bandwdith: 19000,000  Dev:  1729,128  trace(Hat):  45,376  Criterion:  1842,741 Valid_fit
Bandwdith: 18000,000  Dev:  1722,632  trace(Hat):  49,403  Criterion:  1848,519 Valid_fit
Bandwdith: 17000,000  Dev:  1715,531  trace(Hat):  53,819  Criterion:  1855,437 Valid_fit
Bandwdith: 16000,000  Dev:  1707,752  trace(Hat):  58,231  Criterion:  1862,290 Valid_fit

Best bandwidth size 30000,000
Minimum AICc     1822,715

*******************************************************************************
  GWR (Geographically weighted regression) result
*******************************************************************************
  Bandwidth and geographic ranges
Bandwidth size:              30000,000000
Coordinate              Min             Max          Range
--------------- --------------- --------------- ---------------
X-coord          595069,600000   728223,100000   133153,500000
Y-coord         6786134,000000  6895151,000000   109017,000000

  Diagnostic information
Residual sum of squares:        14718,761381
Effective number of parameters (model: trace(S)):                20,543368
Effective number of parameters (variance: trace(S'S)):            13,988089
Degree of freedom (model: n - trace(S)):                       237,456632
Degree of freedom (residual: n - 2trace(S) + trace(S'S)):        230,901353
ML based sigma estimate:           7,553109
Unbiased sigma estimate:           7,984034
-2 log-likelihood:              1775,503297
Classic AIC:                    1818,590032
```

AICc: 1822,715294

BIC/MDL: 1895,132747

CV: 63,433581

R square: 0,600062

Adjusted R square: 0,552921

****

<< Geographically varying (Local) coefficients >>

****

Estimates of varying coefficients have been saved in the following file.

Listwise output file: G:\ORS\Thematiques\Etudes\CANCERS\CIRCÉ\GWR\GWR4\gwr4_kpoumh_fdep_23012018_listwise.csv

Summary statistics for varying (Local) coefficients

Variable	Mean	STD
Intercept	80,521475	2,888667
Fdep08	4,054831	1,396085

Variable	Min	Max	Range
Intercept	78,302706	99,811664	21,508958
Fdep08	-3,772586	5,569137	9,341724

Variable	Lwr Quartile	Median	Upr Quartile
Intercept	78,841751	79,810315	80,960573
Fdep08	3,798797	4,519313	4,898037

Variable	Interquartile R	Robust STD
Intercept	2,118822	1,570661

```
Fdep08                    1,099240        0,814855

 (Note: Robust STD is given by (interquartile range / 1.349) )

*******************************************************************************

 GWR ANOVA Table

*******************************************************************************

Source                      SS         DF           MS          F

----------------- -------------------- ---------- --------------- ----------

Global Residuals            17993,357    256,000

GWR Improvement              3274,596     25,099         130,469

GWR Residuals               14718,761    230,901          63,745   2,046740

*******************************************************************************

Program terminated at 23/01/2018 16:58:56
```

Observatoire régional de santé Île-de-France
15, rue Falguière - 75015 PARIS - Tél. (33) 01 77 49 78 60 - www.ors-idf.org

L'ORS Île-de-France, département autonome de l'IAU Île-de-France, est un observatoire scientifique indépendant financé par l'Agence régionale de santé et le Conseil régional d'Île-de-France.

ISBN : 978-2-7371-1914-9