

Note technique

Travaux financés par le ministère chargé de l'environnement

UTILISATION DES DONNEES DE MICRO-CAPTEURS EN MODELISATION ET CARTOGRAPHIE DE LA QUALITE DE L'AIR

Analyse bibliographique

Alicia Gressent, Laure Malherbe, Bertrand Bessagnet
(INERIS)

SYNTHESE

Malgré les incertitudes attachées à ces systèmes de mesure, le recours croissant aux micro-capteurs ouvre de nouvelles possibilités pour la représentation spatiale de la qualité de l'air. D'après la littérature scientifique récente, l'élaboration de cartographies à partir de données de micro-capteurs n'en est qu'à ses débuts. Deux catégories de méthodes susceptibles de se recouper ou d'être combinées émergent cependant : l'interpolation géostatistique par krigeage et les méthodes statistiques spécifiques (Land Use Regression, agrégation spatiale, et apprentissage statistique).

Les résultats associés à ces méthodes montrent le potentiel de l'utilisation de telles données pour améliorer les estimations de l'exposition jusqu'alors réalisées à partir de mesures de référence et de modèles de dispersion. Des travaux approfondis sont nécessaires pour rendre ces estimations plus précises et considérer des problématiques primordiales telles que (i) l'optimisation du déploiement des capteurs, (ii) la gestion de la représentativité spatiale des mesures, (iii) le traitement de données hétérogènes dans l'espace et dans le temps et (iv) la gestion de l'incertitude inhérente de ces mesures qui est plus élevée que celle des capteurs fixes homologués.

1. CONTEXTE

L'usage des micro-capteurs en qualité de l'air connaît un essor rapide, comme en témoignent la presse scientifique internationale ainsi que les discussions conduites au niveau européen (FAIRMODE et AQUILA notamment) et français (GT « Micro-Capteurs pour l'évaluation de la qualité de l'air » lancé en 2017 et animé par le LCSQA). Les expériences réalisées avec des micro-capteurs se multiplient (utilisation fixe ou mobile dans différents types d'environnement urbain, sur des moyens de transport, pour le suivi de l'exposition individuelle, comme outils de sensibilisation...), à l'initiative des acteurs traditionnels de la surveillance, des collectivités locales, ou d'opérateurs privés. L'abondance des informations ainsi collectées pose la question de l'exploitation des données et des possibilités offertes par ces nouvelles technologies pour la surveillance réglementaire de la qualité de l'air.

2. OBJECTIF

Dans ce contexte, en s'appuyant sur une analyse bibliographique, la présente note a pour objet d'étudier plus spécifiquement **l'utilisation qui est faite des micro-capteurs pour la modélisation et la cartographie** et de synthétiser les méthodologies mises en œuvre.

Les exploitations statistique et géostatistique de données de mesure (qu'il s'agisse de données manuelles ou automatiques de référence ou de campagnes de mesure par tubes à échantillonnage passif notamment) et de modélisation pour l'élaboration de cartographies sont des approches largement éprouvées en qualité de l'air (Xie, et al., 2017). La nouveauté avec les micro-capteurs réside dans la possibilité de réaliser des mesures à haute résolution spatiale et temporelle pour différents polluants gazeux et particulaires qui ouvre de nouvelles voies de développement, en particulier à l'échelle urbaine. Dans cette perspective, un certain nombre de caractéristiques et d'inconnues doivent être considérées :

- A l'intérieur d'un même domaine d'étude, les mesures des micro-capteurs peuvent être fixes et/ou mobiles et dans ce second cas, elles peuvent être effectuées sur différents supports : voiture, bus, tramway, vélo, individu...
- Selon la nature des mesures (fixes, embarquées), les plans d'échantillonnage peuvent différer sensiblement d'une étude à l'autre. En outre, les mesures n'obéissent pas nécessairement à des critères d'implantation ni à des exigences de représentativité spatiale comme c'est le cas des mesures réglementaires.
- Les incertitudes de mesure sont encore mal connues (travaux d'évaluation en cours, en laboratoire et sur le terrain).

3. INTRODUCTION

La synthèse ci-après repose sur une analyse de la littérature scientifique récente (cf. § 6.1). Notons que si les publications relatives aux micro-capteurs sont nombreuses, celles qui concernent la modélisation et la cartographie (au sens d'estimation d'un champ de concentration) constituent un sous-ensemble relativement réduit. Dans bon nombre de cas, le traitement cartographique se limite à la visualisation des données dans un système d'information géographique. A cela s'ajoutent des publications relevant des statistiques exploratoires et des analyses de performance des données de micro-capteurs par comparaison aux observations des stations fixes (Moltchanov, et al., 2014 ; Gao et al., 2015 ; Van den Bossche, 2015 ; Castell, et al., 2016 ; Borrego, et al., 2016 ; Zaldei, et al., 2016 ; Penza, et al., 2017). La problématique de l'optimisation du déploiement des capteurs est également étudiée (Boubrima et al., 2016). Des modèles ont été développés afin de sélectionner les positions de capteurs de manière à ce que les données recueillies par ces derniers permettent une estimation optimale des concentrations de polluants à chaque position de la zone de déploiement. Ces applications mettent en avant l'utilisation des capteurs à faible coût dans la mesure de la pollution de l'air mais ne proposent aucune cartographie des concentrations de polluants.

Deux catégories principales de méthodes pour l'estimation spatiale des concentrations ressortent :

- Interpolation géostatistique par krigeage ;
- Méthodes statistiques spécifiques (Land Use Regression, agrégation spatiale, et apprentissage statistique).

Ces approches de cartographie sont appliquées aux polluants réglementés tels que le dioxyde d'azote (NO₂) et l'ozone (O₃), principalement par l'interpolation géostatistique, et les particules PM₁₀ et PM_{2.5}, uniquement par les méthodes statistiques. Le tableau en annexe synthétise l'ensemble des publications consultées et donne ainsi le détail des polluants concernés.

Pour chaque catégorie, sont présentés les principes méthodologiques, les données d'entrée, les conditions de mises en œuvre et la nature des résultats obtenus.

4.1 Interpolation par krigeage

Le krigeage est une approche géostatistique amplement utilisée pour la cartographie de la qualité de l'air à l'échelle urbaine et régionale (Liu et al., 1996 ; Ferreira et al., 2000 ; Jerret et al., 2001 ; Künzli et al., 2005 ; Malherbe et al., 2005 ; Whitworth et al., 2011 ; Beauchamp et al., 2010, 2014, 2017 ; Xie et al., 2017). Le krigeage vise à estimer la valeur d'une fonction aléatoire (processus aléatoire décrivant des données d'observation) à certaines positions à partir des mesures réalisées aux points d'échantillonnage. L'idée principale du krigeage est que les points de mesure proches du point à estimer doivent avoir plus de poids dans l'estimation. Cela implique la connaissance d'une structure spatiale qui est représentée par un variogramme ou covariogramme (propriétés du second ordre) d'une fonction aléatoire (Goovaerts, 1997 ; Wackernagel, 2003 ; Chiles and Delfiner, 2012 ; Lichtenstern, 2013). Le krigeage consiste à déterminer la combinaison linéaire des données qui assure la variance d'estimation minimale (sous condition de non biais).

En un point s_0 , l'estimation de la concentration $\widehat{y}(s_0)$ s'écrit :

$$\widehat{y}(s_0) = \sum_{i=1}^N \lambda_i y(s_i)$$

où $y(s_i)$, $i = 1 \dots N$, sont les concentrations mesurées aux points d'échantillonnage sur l'ensemble du domaine (voisinage unique) ou dans les limites d'un certain voisinage de s_0 (voisinage glissant), et λ_i , $i = 1 \dots N$, sont les poids de krigeage associés.

Le variogramme permet également de calculer la variance de l'erreur d'estimation, encore appelée variance de krigeage, qui indique les endroits où cette erreur est plutôt faible et ceux où elle est plutôt élevée. L'une des forces du krigeage est ainsi d'apporter une information sur l'erreur et donc sur l'incertitude du résultat de cartographie obtenu (Goovaerts, 1997 ; Wackernagel, 2003). L'étude d'Alvear et al., 2016, présente des cartes de concentrations d'ozone réalisées à partir de données de micro-capteurs mobiles (plateforme Arduino, <https://www.arduino.cc/>) déployés à Valence, en Espagne. Ces cartes sont obtenues en appliquant un krigeage ordinaire (Figure 1).

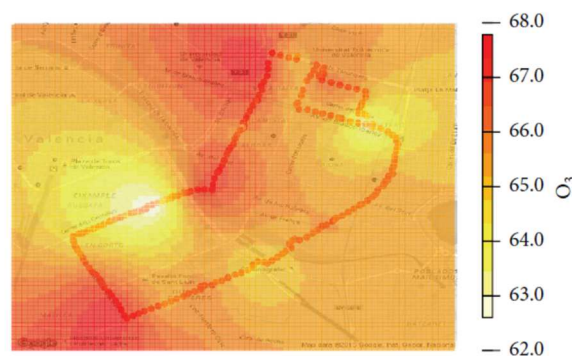


Figure 1 : Alvear, P., et al., 2016 : Niveaux d'ozone (ppb) estimés par krigeage ordinaire à partir de données de micro-capteurs mobiles.

La méthode d'interpolation spatiale utilisée dans ces travaux reste toutefois assez basique et pourrait être améliorée de façon à affiner la résolution spatiale de l'estimation et éviter les taches circulaires de fortes et faibles concentrations visibles sur la Figure 1. Pour cela, il existe d'autres variantes du krigeage, qui relèvent de la géostatistique multivariée (Wackernagel, 2003 ; Rivoirard, 2003) et permettent de considérer des informations additionnelles aux observations. Parmi ces variantes, le krigeage du résidu consiste en une régression linéaire multiple, suivie d'un krigeage du résidu de la régression. Cette méthode peut s'appliquer à une régression de type Land Use Regression, comme dans Mercer et al., 2011 (cf. 4).

Le krigeage avec dérive externe repose sur un modèle similaire (relation linéaire avec des variables auxiliaires et résidu spatialement corrélé) mais permet de combiner en une seule étape les données de mesure et les informations additionnelles considérées. Ce type de krigeage est notamment employé pour la fusion de données.

Les recherches traitant de la fusion de données issues des micro-capteurs avec les modèles n'en sont qu'à leurs débuts (voir Lahoz and Schneider, 2014, pour une revue bibliographique). Une seule étude utilisant une méthode géostatistique de fusion de données pour les micro-capteurs fixes a été publiée récemment (Schneider et al., 2017). Ce travail utilise le krigeage avec dérive externe (noté « krigeage universel » dans la publication) comme méthode d'interpolation spatiale afin de produire des cartes de concentration moyenne annuelle de NO₂ à Oslo, en Norvège. Les mesures par micro-capteurs ne permettent pas de combler totalement le manque d'observations et fournissent des mesures irrégulières tant dans l'espace que dans le temps. Une façon de combler ce manque d'observations, décrite par Schneider et al., 2017, est de combiner des données modélisées aux observations des micro-capteurs fixes.

A la différence du krigeage ordinaire, le krigeage avec dérive externe considère que la moyenne globale n'est pas constante à travers le domaine d'étude et que celle-ci dépend de variables explicatives (autrement dit, il existe une tendance ou une dérive de la moyenne). Cette technique de krigeage est employée depuis plusieurs années en qualité l'air par le dispositif national de surveillance pour une interpolation spatiale à l'échelle urbaine (Beauchamp et al., 2016) et à l'échelle nationale (PREV'AIR, Malherbe et Ung, 2009). Elle permet d'estimer les concentrations de polluants à des localisations où celles-ci sont inconnues en utilisant simultanément des données de mesure - la nouveauté est qu'il s'agit ici de micro-capteurs - et des données d'une ou plusieurs variables auxiliaires, qui décrivent la tendance spatiale sur le domaine d'étude. Soit $y(s_0)$, la concentration du polluant à estimer en un point s_0 ; l'hypothèse est celle d'une relation linéaire entre $y(s_0)$ et les variables auxiliaires prises en compte :

$$y(s_0) = m(s_0) + \varepsilon(s_0)$$

Avec :

$$m(s_0) = b_0 + b_1x_1(s_0) + b_2x_2(s_0) + \dots + b_px_p(s_0)$$

Où $m(s_0)$ représente la dérive de la moyenne avec b_0, b_1, \dots, b_p , les coefficients de la relation linéaire, x_1, x_2, \dots, x_p les variables auxiliaires. ε correspond au processus aléatoire stationnaire auquel est associé un semivariogramme. Si l'estimation est réalisée en voisinage glissant, les coefficients b_p dépendent du point s_0 .

Comme en krigeage ordinaire $y(s_0)$ s'estime par une combinaison linéaire des données de mesure :

$$\widehat{y(s_0)} = \sum_{i=1}^N \lambda_i y(s_i)$$

Cependant les poids de krigeage doivent aussi satisfaire à la condition de dérive :

$$\forall x_p : x_p(s_0) = \sum_{i=1}^N \lambda_i x_p(s_i)$$

Dans l'étude de Schneider et al., 2017, la dérive de la moyenne est estimée à partir d'une seule variable auxiliaire, qui est la carte des concentrations moyennes annuelles de NO₂ déterminées par un modèle urbain de dispersion atmosphérique (modèle EPISODE ; Slørðal, 2003). Cette carte, considérée comme représentative des concentrations du polluant sur le long terme, est combinée aux mesures de court terme (moyennes horaires) issues des micro-capteurs (plateforme AQMesh v3.5 fournies par Environmental Instruments Ltd, UK, www.aqmesh.com). La méthodologie a été établie et testée à partir de données simulées (Figure 2) puis appliquée à des données réelles.

Les cartes de l'exemple ci-après montrent l'apport respectif des pseudo-données de micro-capteurs (données extraites d'une modélisation horaire et perturbées par une erreur gaussienne) et des données issues d'une modélisation annuelle (produite pour l'année 2011) pour cartographier les concentrations horaires de NO₂.

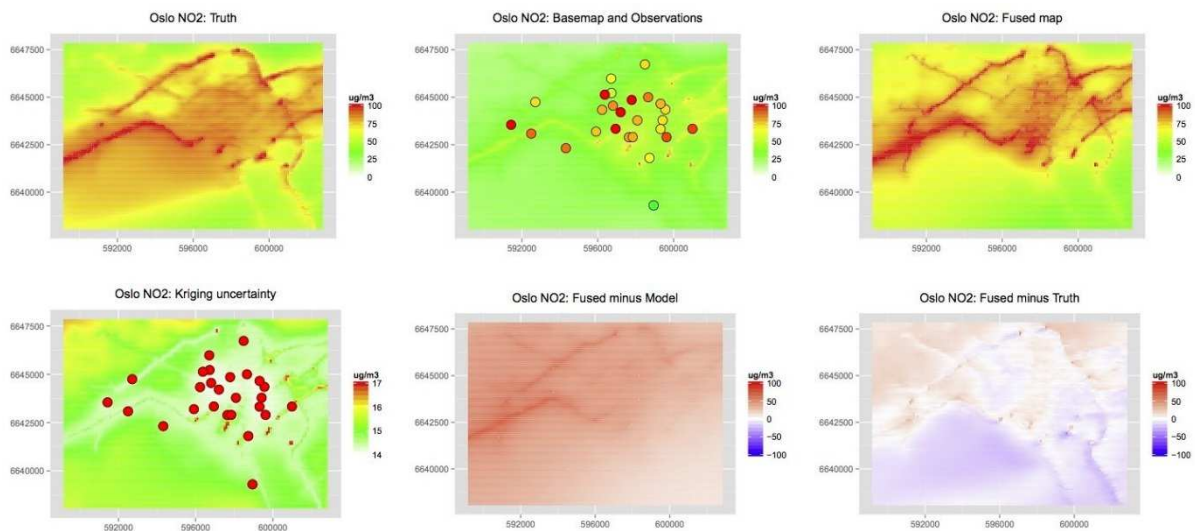


Figure 2 : Schneider et al., 2017 : Exemple de fusion de données de micro-capteurs avec des données modélisées, Oslo, Norvège. De haut en bas et de gauche à droite : carte simulée de la réalité (obtenue par modélisation à une heure de pointe du matin) ; mesures de micro-capteurs simulées pour cette même heure, avec en fond la carte des concentrations moyennes annuelles modélisées ; carte fusionnée par krigeage ; écart-type de l'erreur de krigeage ; différence entre la carte fusionnée et le champ de concentrations long-terme moyen modélisé ; différence entre la carte fusionnée et la réalité.

Une autre étude (Hamm et al., 2016) présente une analyse spatio-temporelle des observations de PM₁₀ et PM_{2.5} à Eindhoven aux Pays-Bas (Airbox, capteur de particules ultrafines, AeroSense Nanomonitor). Cette analyse est basée sur un modèle dynamique (décrit par des coefficients de régression qui varient dans le temps) pour la composante temporelle et un modèle géostatistique gaussien pour la composante spatiale. Les résultats présentés dans cette publication sont une ébauche des travaux en cours en donnant uniquement une estimation des concentrations dans le temps.

Bilan sur les conditions d'application et perspectives de développement :

Afin de rendre le calcul du variogramme possible et d'assurer la bonne qualité du résultat d'interpolation spatiale, un minimum de 25 à 50 micro-capteurs pour une ville ou une agglomération de taille et de densité de population similaire à celles d'Oslo, répartis sur le domaine d'étude semble être nécessaire. La qualité des résultats de cartographie par fusion de données en utilisant le krigeage est fortement dépendante à la fois des données modélisées et de l'incertitude associée aux mesures. Il est donc important de s'assurer du réalisme des champs de concentration issus du modèle de dispersion et du bon étalonnage des capteurs avant la mise en place d'une telle expérimentation. Il est possible d'appliquer cette approche pour obtenir des cartes de concentrations moyennes sur des périodes de temps plus réduites (moyennes horaires par exemple). Jusqu'à présent, le krigeage avec dérive externe n'a été appliqué qu'à des mesures de micro-capteurs fixes. Il est envisagé dans un futur proche d'adapter et de tester cette méthode pour des mesures issues de capteurs mobiles. Cependant, ces derniers sont associés à des incertitudes plus importantes que celles des appareils en site fixe du fait de leurs caractéristiques métrologiques moins performantes que celles des analyseurs de référence. L'incertitude associée à ces capteurs pourrait être compensée en augmentant leur nombre, on peut en effet concevoir que la moyenne des mesures de plusieurs capteurs à une localisation serait d'autant plus proche de la réalité qu'une mesure isolée. C'est un point qu'il serait intéressant d'investiguer. De façon générale, ces incertitudes sont donc à considérer avec attention dans de tels développements.

4.2 Méthodes statistiques

4.2.1 Land Use Regression (LUR)

Parmi les méthodes de modélisation et de cartographie recensées, les modèles de type LUR (Land Use Regression) sont les plus fréquemment utilisés pour traiter les données de capteurs mobiles embarqués sur véhicules (Xie et al., 2017). Ces modèles prennent en considération l'influence de l'utilisation des terrains et des caractéristiques environnementales de la zone avoisinante dans l'estimation de la pollution de l'air. L'idée est de construire des équations de régression linéaire multiple décrivant la relation entre les mesures de polluants aux stations de surveillance (ou micro-capteurs) et des variables prédictives obtenues par des systèmes d'information géographique (SIG).

Soit $y(s_0)$, la concentration du polluant à estimer en un point s_0 , cette relation est supposée s'écrire :

$$y(s_0) = b_0 + b_1x_1(s_0) + b_2x_2(s_0) + \dots + b_px_p(s_0) + \varepsilon(s_0)$$

où b_0, b_1, \dots, b_p , sont les coefficients de la relation linéaire, x_1, x_2, \dots, x_p sont les variables prédictives et ε est un résidu aléatoire. A la différence du krigeage avec dérive externe, la corrélation spatiale du résidu n'est généralement pas prise en compte.

Les variables utilisées sont communément la distance d'éloignement par rapport aux axes routiers majeurs ou mineurs, la distribution de la population, le trafic routier, les émissions liées au chauffage domestique, la vitesse du vent ou encore l'occupation des sols. Elles sont calculées à l'intérieur de zones tampons (dites « buffers »), qui correspondent à des « cercles d'influence » centrés sur les points de mesure. Les modèles LUR sont bien adaptés pour les estimations de la pollution à l'échelle urbaine. Cependant, le choix des variables prédictives utilisées dans ces modèles n'est souvent pas optimal pour le traitement de la pollution à l'échelle régionale puisqu'ils ignorent des facteurs importants de la pollution de l'air tels que les industries, l'activité agricole, les émissions maritimes ou encore la pollution longue distance. La fonction de régression décrivant le modèle est présentée ci-dessous :

$$\hat{y} = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n$$

où \hat{y} est la valeur estimée du niveau de pollution, x_1, \dots, x_n représentent les variables prédictives, et β_0, \dots, β_n , sont les coefficients estimés par régression, qui indiquent l'influence positive ou négative de ces variables sur le niveau de pollution. La régression linéaire vise à trouver des valeurs de β_0, \dots, β_n qui fournissent l'ajustement optimal dans les données - par exemple, en minimisant la somme des résidus au carré (différences quadratiques entre les valeurs observées et prédites du niveau de pollution) pour la régression des moindres carrés ordinaires. Les modèles résultants sont ensuite utilisés pour estimer les concentrations de polluants pour un lieu donné, en utilisant les variables prédictives associées à ce lieu.

La principale limitation des modèles LUR est la nécessité de disposer d'un très grand nombre de données. Il est requis un minimum de 40 à 80 positions d'échantillonnage différentes dans un domaine urbain, selon l'étude de Hoek et al., 2008. Toutefois, il ne ressort aucune méthode bien établie pour définir le nombre de sites de mesure, celle-ci sera dépendante des objectifs fixés pour l'expérimentation (estimation de la pollution urbaine, très locale ou associée à une source particulière), de la densité de population et de la taille du domaine considéré (taille de l'agglomération).

Les mesures réalisées par les micro-capteurs mobiles peuvent répondre à ces contraintes d'échantillonnage en permettant une couverture spatiale d'observation étendue. Récemment, les travaux d'Hankey et al., 2015, proposent des modèles LUR en utilisant des données de micro-capteurs mobiles embarqués sur des vélos et mesurant les particules dans la ville de Minneapolis, MN (USA). En particulier, le carbone suie - ou black carbon (AE51, AethLabs, San Francisco, CA), les $PM_{2.5}$ (DustTrak 8320, TSI, Inc.), le nombre de particules (CPC 3007, TSI, Inc., Shoreview, MN), et la taille des particules (NanoScan, TSI Inc.) sont mesurés. Trois parcours d'échantillonnage représentatifs des différentes intensités de trafic et occupations du sol ont été définis à travers de nombreux quartiers de la ville et ont fait l'objet de mesures répétées (août à octobre 2012). Les observations ainsi collectées (3×10^5 mesures par instrument soit ≈ 85 heures de mesures au total) ont été agrégées par portion de route et par tranche horaire (heures de pointe du matin, respectivement de l'après-midi). Pour chacune de ces tranches horaires, différents modèles LUR ont été construits et comparés, puis appliqués à l'ensemble de la ville (maillage d'estimation non détaillé dans la publication). La Figure 3 présente les estimations des concentrations en particules $PM_{2.5}$ à Minneapolis ainsi obtenues (Hankey et al., 2015).

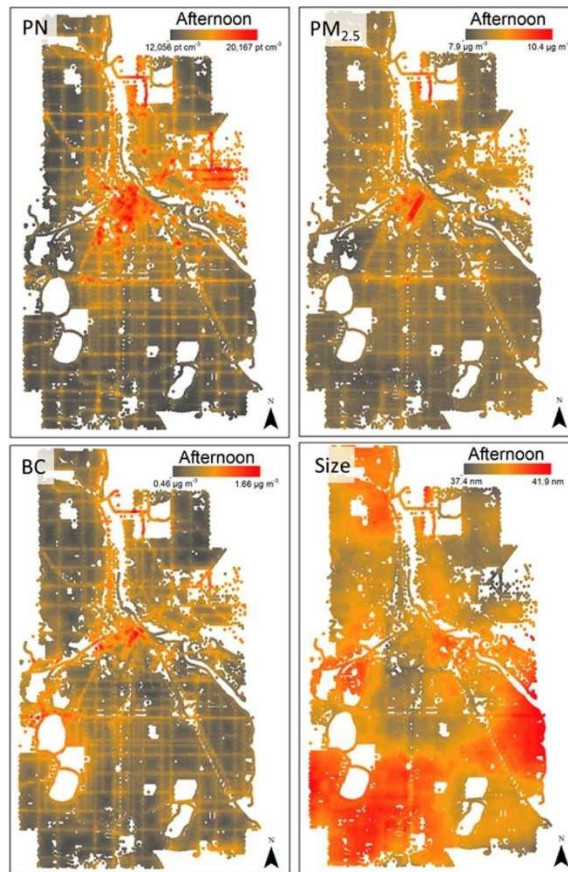


Figure 3 : Hankey et al., 2015 : Estimation des concentrations de $PM_{2.5}$ à Minneapolis, MN, aux heures de pointes de l'après-midi, par les modèles LUR calés sur des capteurs mobiles. Moyenne sur la période de campagne. Les zones blanches correspondent aux cours et plans d'eau et aux espaces verts.

La performance des modèles LUR a été évaluée en fonction de la résolution spatiale d'agrégation des mesures (de 50 m à 200 m) et de la période de temps pour laquelle la moyenne des concentrations est considérée (de 1 s à 60 s). L'étude d'Hankey et al., 2015, montre que la qualité des modèles LUR est davantage influencée par la période choisie pour la moyenne des concentrations que par la résolution spatiale d'agrégation. En outre, il s'avère que les variables prédictives les plus importantes (parmi 20 variables retenues) pour l'estimation sont celles liées à (i) l'éloignement des sources d'émissions : espaces ouverts, distance par rapport aux axes routiers principaux, longueur des routes caractérisées par un faible trafic, et (ii) celles liées à la proximité et à la densité des sources d'émissions : zones industrielles, densité d'habitations, longueur des routes caractérisées par un fort trafic. Cette étude souligne également l'importance du caractère répété des mesures. La performance des modèles a été évaluée en créant un jeu de données de validation (sélection de façon aléatoire de 1/3 des données) puis en comparant les résultats des modèles LUR aux observations. Le processus a été répété 100 fois pour chaque modèle. Cette analyse montre que la performance des modèles LUR augmente avec la fréquence d'échantillonnage des portions de routes jusqu'à atteindre un seuil. La Figure 4 issue de l'étude de Hankey, et al, 2015, illustre ce résultat en présentant le coefficient de détermination R^2 (ajusté) et le biais des modèles, qui tendent à augmenter, respectivement diminuer, avec le nombre d'heures de mesures.

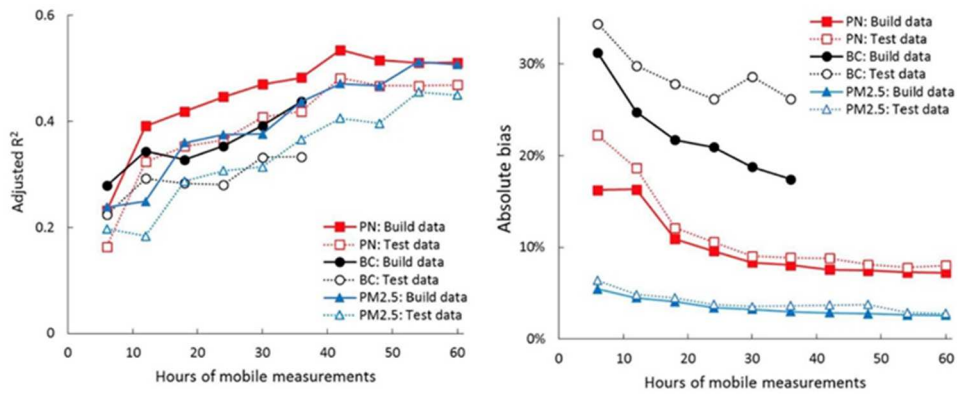


Figure 4 : Hankey et al., 2015 : Performance des modèles LUR dans le cas des estimations de concentrations de $PM_{2.5}$.

Un travail similaire et récent (Minet et al., 2017) réalisé à Montréal utilise également un modèle LUR pour cartographier les concentrations de polluants. Une campagne de surveillance mobile (vélos et piétons) a été menée en 2015 à Montréal ; les niveaux de dioxyde d'azote (NO_2 , Aeroqual Series 500 portable sensors) de 1395 segments de route ont été mesurés lors de « visites » répétées. Des modèles LUR appliqués sur des sous-ensembles de segments, catégorisés en fonction du nombre de visites, ont été développés. A chaque point de grille (résolution $100m \times 100m$) les prédicteurs ont été calculés et les concentrations ont été estimées par les modèles LUR. En appui aux travaux de Hankey et al., 2015, cette étude indique que les modèles LUR sont très sensibles au nombre de segments de route et au nombre de visites par segment routier. Les cartes de concentration associées à ces modèles varient sensiblement selon ces paramètres (Figure 5), rendant délicate leur exploitation à des fins de calcul d'exposition.

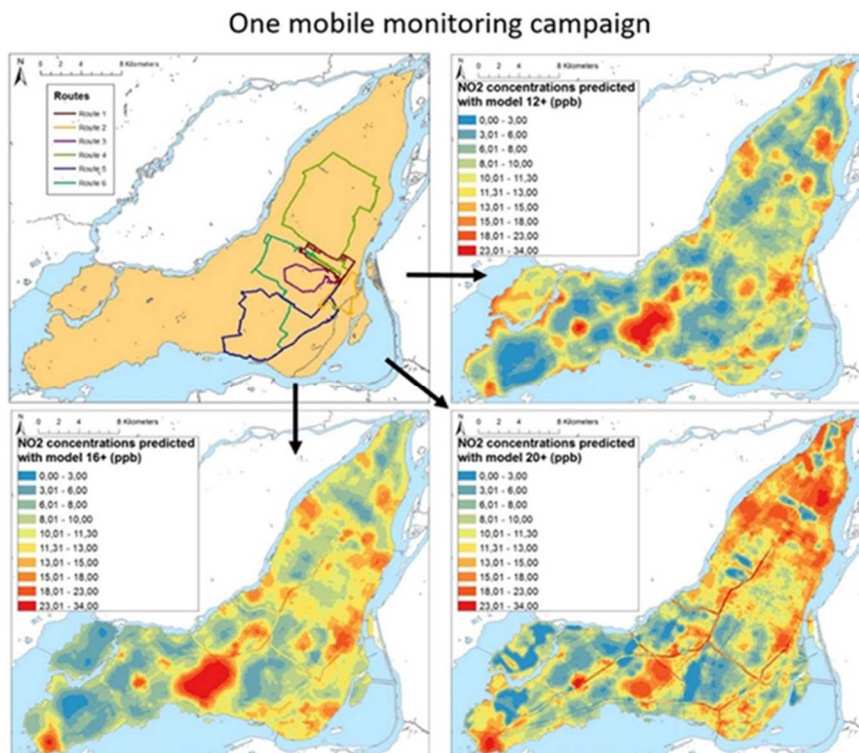


Figure 5 : Minet et al., 2017 : Cartographies des concentrations de NO_2 en fonction du nombre N de visites par segment ($N+$ avec $N=12, 16$ ou 20). Moyenne sur la période de campagne.

Bilan sur les conditions d'application et perspectives de développement :

La densité spatiale des mesures par micro-capteurs représente un réel avantage pour la construction des modèles LUR. Cela permet notamment de définir des « buffers » de taille plus petite, qui rendent possible l'investigation des variations de petite échelle dans les cadastres d'utilisation des sols et de leur influence sur les concentrations de polluants. La bonne qualité d'un modèle LUR, en plus d'être dépendante de la qualité des mesures micro-capteurs, sera directement influencée par les données SIG nécessaires à la construction des variables explicatives et à la définition du modèle. Particulièrement, la résolution spatiale des données SIG jouera un rôle déterminant.

La dépendance spatiale des résidus de la régression n'est généralement pas prise en compte dans les modèles LUR, ce qui distingue ces derniers du krigeage multivarié (cf. §3.1). Dans une étude sur l'exposition aux NO_x (Ogawa Sampler; Ogawa and Company, U.S.A., Inc.: Pompano Beach, FL, 1998) à Los Angeles (CA), Mercer et al., 2011, ont toutefois associé une estimation de la moyenne locale par régression multilinéaire (modèle LUR) avec un krigeage des résidus de la régression. Les résultats sont similaires à ceux d'un modèle LUR standard, voire un peu meilleurs. Les données exploitées dans cette étude étaient issues de tubes à diffusion passive (150 points d'échantillonnage) mais la méthodologie pourrait être testée avec des données de micro-capteurs.

D'un point de vue temporel, la possibilité de pouvoir disposer, pour chaque segment routier sélectionné, de données de mesure suffisamment nombreuses sur la période d'étude, constitue également un paramètre important. Dans la plupart des travaux considérés, la période d'estimation des concentrations est limitée à quelques mois de campagne. Une étude de faisabilité pour étendre cette période d'estimation à du plus long terme (moyenne annuelle) et l'appréciation des incertitudes associées sont à envisager pour des recherches futures.

4.2.2 Méthodes d'apprentissage statistique

Les méthodes d'apprentissage statistique sont de plus en plus développées dans le domaine de la qualité de l'air. Ces approches utilisent des données historiques pour « apprendre » de ces données et ainsi permettre des estimations plus réalistes et précises. Elles sont d'ores et déjà appliquées pour l'étalonnage des micro-capteurs notamment par le biais de réseaux de neurones artificiels (Spinelle et al., 2016 ; Zimmerman, et al., 2018). Récemment, des modèles statistiques d'apprentissage ont été appliqués dans le but de produire des cartes de concentration de polluant. L'étude de Mueller et al., 2016, utilise des données de capteurs embarqués sur des tramways à Zurich avec un modèle de type GAM (Generalized Additive Models ; Hastie and Tibshirani, 1996 ; Wood, 2006) afin de réaliser des cartes de PNC (Particle Number Concentration). La formulation du modèle est la suivante :

$$\log PNC = c + \sum_{i=1}^n s_i(\text{geo}_i(x, y)) + \varepsilon$$

où c est l'origine, s_i sont les splines de régression, c'est-à-dire des ajustements linéaires ou des fonctions polynomiales calculés pour des sous-divisions du jeu de données afin de représenter la non linéarité du système, geo_i sont les variables prédictives, x et y sont les coordonnées des mesures et ϵ l'erreur. Dans le cas particulier de Zurich, Mueller et al., 2016, ont montré que les variables prédictives dominantes sont l'altitude et l'intensité du trafic routier. Ils notent également que les variables météorologiques (vitesse et direction du vent) peuvent avoir un impact significatif sur les estimations des concentrations et qu'elles devront être considérées dans les travaux futurs.

L'étude de Hu, et al., 2017 propose un modèle nommé « HazeEst-a » qui combine les données de stations fixes et les données de capteurs mobiles (respectivement disponibles sur les périodes 2009-2016 et 2013-2016) afin d'estimer la pollution de l'air (en particulier les concentrations de monoxyde de carbone, CO) en surface pour chaque heure de chaque jour à Sydney (sur la période mai 2009 à mai 2016).

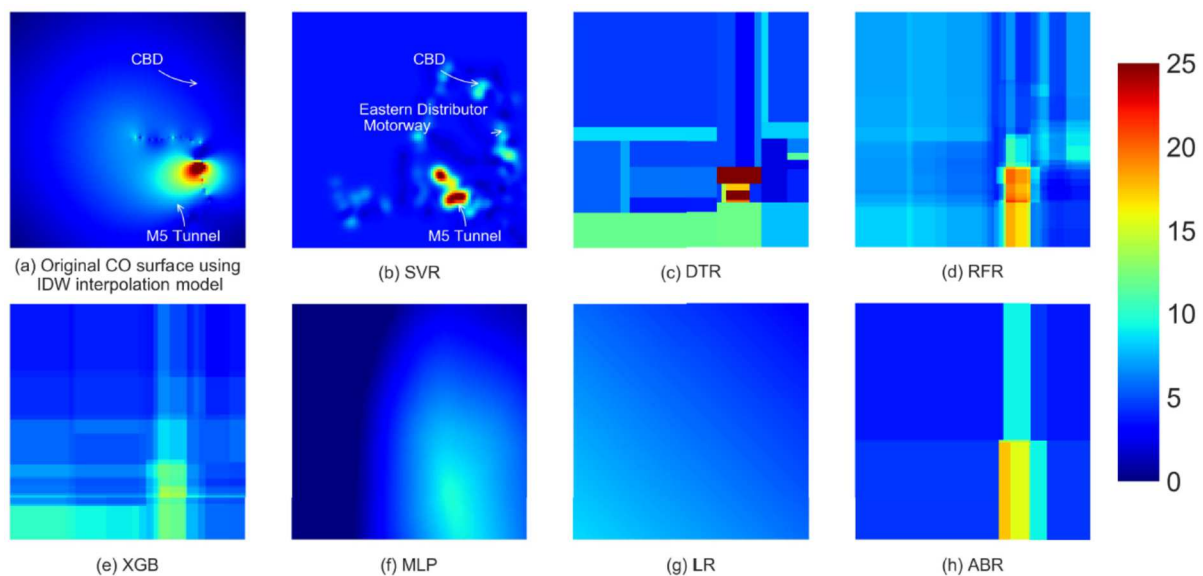


Figure 6 : Hu, K., et al., 2017 : Comparaison de cartes de la pollution de l'air (CO) à 16h le 03/05/2016 obtenues (a) par un modèle d'interpolation IDW (Inverse Distance Weighting) et (b)-(h) par des estimations issues des modèles de régression (Machine Learning) ; SVR : Support Vector Regression, DTR : Decision Tree Regression, RFR : Random Forest Regression, XGB : Extreme Gradient Boosting, MLP : Multi-Layer Perceptrons, LR : Linear Regression, ABR : Adaptive Boosting Regression.

L'évaluation des modèles par validation croisée et les cartes de pollution de l'air obtenues (Figure 6) montrent que les modèles SVR, DTR et RFR ont une performance d'estimation meilleure que celle des autres modèles de régression. Le modèle SVR en particulier semble bien reproduire les zones de plus forte pollution. Ce dernier définit une fonction non-linéaire à partir d'observations dites d'apprentissage, qui permet d'interpoler les concentrations sur un domaine à haute résolution spatiale. Ces estimations restent peu convaincantes en comparaison aux résultats de krigeage ou LUR. A nouveau, les variables météorologiques devront être considérées dans des travaux futurs en vue d'améliorer ces estimations.

Il est également intéressant de mentionner l'étude de Hu, et al., 2016, qui propose un modèle statistique dit PCEM (Probabilistic Concentration Estimation Method) afin de cartographier les concentrations de $PM_{2.5}$ (mesures réalisées par le micro-capteur DN7C3CA006, SHARP) à Hangzhou en Chine. Cet algorithme est composé de deux parties. La première correspond à la modélisation mathématique du transport de particules (mouvement brownien et modèle de dispersion par résolution de probabilités de transition d'un point de grille à un autre) suivie de l'estimation des concentrations des $PM_{2.5}$ sur les points de grille pour lesquelles les mesures sont inexistantes (Figure 7).

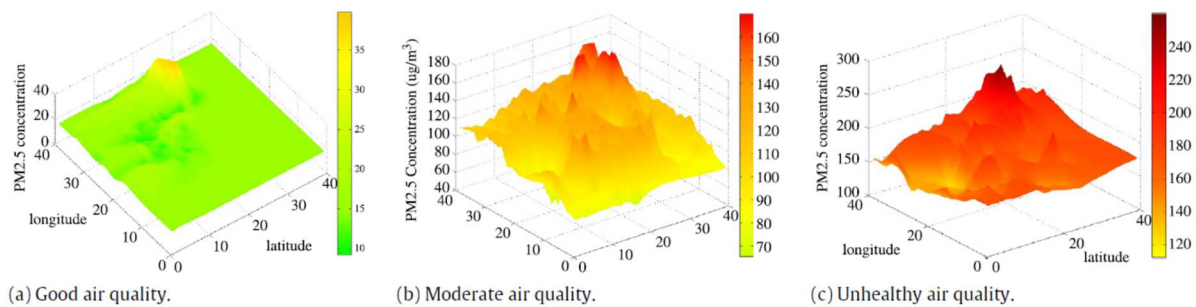


Figure 7 : Hu et al., 2016 : Exemple de distribution des concentrations de $PM_{2.5}$ entre 15h et 16h pour la zone d'étude à Hangzhou en Chine sous différentes conditions environnementales.

Une autre publication (Marjovi et al., 2015) présente trois modèles afin d'obtenir des cartes à haute résolution spatio-temporelle d'un paramètre d'exposition appelé LDSA (the Lung Deposited Surface Area) mesuré par les micro-capteurs. Les différents modèles sont caractérisés par (i) une régression log-linéaire, (ii) une régression log-linéaire basée sur un réseau et (iii) un PGM (Probabilistic Graphical Model). Les données utilisées sont issues de capteurs mobiles installés sur les toits de bus publics parcourant la ville de Lausanne en Suisse (de décembre 2013 à février 2015). L'estimation du LDSA est alors obtenue par extrapolation issue des différents modèles pour les portions de routes non échantillonnées (Figure 8). Pour cela, la ville a été divisée en sous-domaines pour lesquels les émissions et donc les estimations de trafic sont similaires. Chaque mesure est associée au segment de route le plus proche basé sur la distance euclidienne. Dans ce cas, il est important de noter que l'estimation des concentrations est réalisée uniquement pour les routes. La ville est donc considérée comme étant un réseau et non comme une surface continue. L'intégration de données relatives à l'occupation du sol dans ces modèles est un axe de travail futur primordial afin d'améliorer les estimations de pollution de l'air.

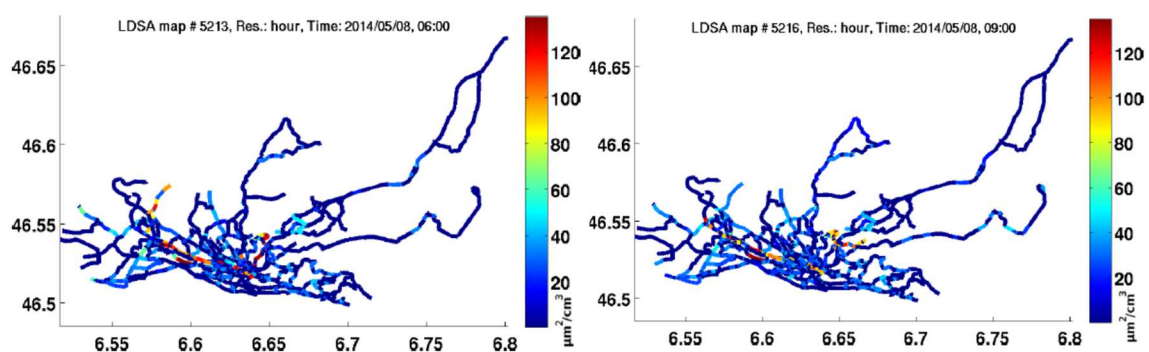


Figure 8 : Marjovi et al., 2015 : Deux exemples de cartes de pollution des rues de Lausanne (Suisse).

En plus de ces applications en cartographie, des méthodes d'apprentissage statistique sont fréquemment utilisées pour produire des prévisions de la qualité de l'air aux points de mesure (Gomez et al., 2014 ; Shaban et al., 2016 ; Liu and Wang, 2016). Cet usage dépasse le cadre de cette note et ne sera pas développé.

4.2.3 Agrégation spatiale pour des impacts ciblés

L'étude d'Apte et al, 2017 présente une méthode qui a pour but de caractériser l'hétérogénéité de l'exposition humaine à la pollution de l'air ainsi que l'identification de « hotspots » (ou localisation de fortes concentrations) de pollution pour le BC (Black Carbon), le NO et le NO₂ à partir de capteurs mobiles (capteurs Aclima Ei, <https://aclima.io/what/>) embarqués sur des véhicules Google à Oakland, CA, aux Etats-Unis. Ce travail présente une méthode dite de « snapping » permettant d'agrèger les observations sur des portions de routes de 30 m de longueur. La médiane des concentrations des polluants associées à chaque portion de route est retenue comme valeur centrale estimée. La précision des estimations de ces tendances centrales est évaluée grâce à un échantillonnage répété. Ce traitement de données a permis (i) de caractériser une distance d'influence des grands axes (fonction de décroissance exponentielle) à partir des portions de routes définies et d'estimer l'exposition, puis (ii) de définir un paramètre de pic de concentration afin d'identifier les « hotspots » de polluants dans le quartier étudié.

5. CONCLUSION

L'exploitation des données de micro-capteurs pour l'étude de la pollution atmosphérique est sans aucun doute un sujet d'actualité dans la communauté scientifique, comme l'illustrent la discussion sur les « Big Data » en matière de santé publique (Vayena et al., 2015) et l'accent mis sur l'utilisation de capteurs à bas coût dans les pays en développement (Austen et al., 2015).

Dans cette étude, nous nous sommes attachés à l'usage des micro-capteurs pour la cartographie et la modélisation de la qualité de l'air. La plupart des estimations de la pollution de l'air recensées sont réalisées à l'échelle urbaine (du quartier à la ville dans son ensemble).

Les méthodologies mises en œuvre sont généralement des approches statistiques. Les modèles LUR (Land Use Regression), qui définissent une relation linéaire multiple entre les concentrations mesurées en de nombreux endroits et des variables caractéristiques de l'occupation du sol et de l'activité urbaine, sont le plus employés. Récemment, le développement des méthodes d'apprentissage statistique a aussi permis de cartographier les concentrations de polluants. Ces modélisations considèrent des données historiques de la qualité de l'air et permettent ainsi une amélioration de la précision des cartes de polluants produites. L'intégration des conditions météorologiques dans de tels modèles est essentielle et devra être effectuée dans des travaux futurs.

L'interpolation géostatistique par krigeage, aisément utilisée pour la cartographie des polluants à partir des données des stations de référence (ou de campagnes par échantillonnage passif), est étendue aux données des micro-capteurs. Le krigeage avec dérive externe est notamment employé pour la fusion de données issues des micro-capteurs avec un modèle de dispersion et cette méthode est uniquement appliquée aux mesures fixes. Ce travail, qui repose sur une évaluation préalable approfondie des micro-capteurs, montre que c'est le réseau de micro-capteurs dans son ensemble plutôt que la donnée individuelle qui apporte une information pertinente pour les estimations. S'agissant de l'assimilation séquentielle ou variationnelle de données dans les modèles déterministes, aucune étude faisant usage de micro-capteurs n'a été mise en évidence.

Plusieurs défis subsistent pour la recherche sur la surveillance de la pollution atmosphérique et l'évaluation de l'exposition à partir de données de micro-capteurs (en particulier de capteurs mobiles) :

- Les micro-capteurs sont connus pour leurs faiblesses dans leurs performances météorologiques (JRC, 2017). Des méthodes ont été développées afin de réduire ces défauts de qualité des capteurs et transformer les données brutes en mesures exploitables (WMO/GAW, 2018). Ces méthodes de traitement des données et d'étalonnage, qui peuvent faire appel à des techniques d'apprentissage statistique, constituent une réelle problématique pour la qualité des observations et in fine des estimations de la pollution de l'air. Par ailleurs, de possibles redondances entre modélisation pour l'étalonnage et modélisation pour l'interpolation des données et la cartographie doivent être évaluées et évitées.
- Si les campagnes exploitées à ce jour obéissent généralement à des plans d'échantillonnage réfléchis, des approfondissements restent indispensables à l'élaboration de cartographies cohérentes et précises. Tant dans la mesure fixe que

mobile, l'optimisation du déploiement des micro-capteurs, en relation avec les réseaux de surveillance existants, est nécessaire afin de limiter les biais dans les estimations. Des travaux ont été engagés sur cette problématique (Boubrima et al., 2016) mais demandent à être poursuivis.

- La représentativité spatiale des mesures selon le type de support (piéton, vélo, toit des bus...) et la nature des voies empruntées (axes à fort ou faible trafic) ne fait encore l'objet d'aucun questionnement dans les publications analysées. Cet aspect nécessiterait d'être examiné.
- Le traitement de données hétérogènes, aussi bien dans l'espace que dans le temps (par exemple, peu de données dans des conditions de faible trafic ou la nuit), appelle également des orientations méthodologiques.

Dans ce contexte, le LCSQA prévoit un travail en collaboration étroite avec les AASQA pour le développement d'une méthodologie de cartographie qui utilise les mesures issues des micro-capteurs. Ce travail a pour objets de i) valoriser les nouvelles mesures de micro-capteurs qui seront obtenues lors des prochaines expérimentations dans les régions et de ii) produire des cartographies plus précises à l'échelle urbaine (et régionale) pour l'exposition des populations aux polluants réglementés.

6. REFERENCES

6.1 Micro-capteurs

- Alvear, O., Zamora, W., Calafate, C., Cano, Juan-Carlos and Manzoni, P., 2016. An Architecture Offering Mobile Pollution Sensing with High Spatial Resolution, *Journal of Sensors*, Article ID 1458147, <http://dx.doi.org/10.1155/2016/1458147>.
- Apte, J. S., Messier, K. P., Gani, S., Brauer, M., Kirchstetter, T. W., et al., 2017. High-Resolution Air Pollution Mapping with Google Street View Cars: Exploiting Big Data, *Environ. Sci. Technol.*, 51, 6999-7008, DOI: 10.1021/acs.est.7b00891.
- Bigazzi, A. Y., and Figliozzi, M. A., 2015. Roadway determinants of bicyclist exposure to volatile organic compounds and carbon monoxide, *Transportation Research Part D*, 41, 13-23, <http://dx.doi.org/10.1016/j.trd.2015.09.008>.
- Borrego, C., et al, 2016. Assessment of air quality microsensors versus reference methods: The EuNetAir joint exercise, *Atm. Env.*, 147, 246-263.
- Boubrima, A., Bechkit, W., Rivano, H., 2016. Error-Bounded Air Quality Mapping Using Wireless Sensor Networks. *LCN 2016 - The 41st IEEE Conference on Local Computer Networks*, Dubai, United Arab Emirates.
- Castell, N., Dauge, F. R., Schneider, P., Vogt, M., Lerner, U., et al, 2016. Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates, *Env. Int.*, 99, 293-302.
- Cavellin, L. D., Eichenthq, S., Tack, R., Ragettli, M. S., Smargiassi, A., Hatzopoulou, M., 2015. Investigating the Use Of Portable Air Pollution Sensors to Capture the Spatial Variability Of Traffic-Related Air Pollution, *Eviron. Sc. and Technol.*, 50, 313-320.
- Dong, W., Guan, G., Chen, Y., Guo, K., Gao, Y. Mosaic, 2015. Towards city scale sensing with mobile sensor networks. In *Proceedings of the 2015 IEEE 21st International Conference on Parallel and Distributed Systems (ICPADS)*, Melbourne, VIC, Australia, 14 December 2015; IEEE: Piscataway, NJ, USA, pp. 29–36.
- ETC/ACM, 2018. Air quality sensors. Possibilities and challenges. Bartonova A., Castell N., Colette A., Malherbe L., Schneider P., Viana M., Voogt M., Weijers E., Wesseling J., ETC/ACM Working Paper 2018. A paraître prochainement.
- Gao, Y., Dong, W., Guo, K., Liu, X., Chen, Y., Liu, X., Bu, J., Chen, C. Mosaic, 2016. A low-cost mobile sensing system for urban air quality monitoring. In *Proceedings of the IEEE INFOCOM 2016—The 35th Annual IEEE International Conference on Computer Communications*, San Francisco, CA, USA, 10–14 April 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1–9.
- Gao, M., Cao, J., and Seto, E., 2015. A distributed network of low-cost continuous reading sensors to measure spatiotemporal variations of PM_{2.5} in Xi'an, China, *Env. Pol.*, 199, 56-65.
- Gomez, M. and Hidalgo, D., 2014. New Strategy for a good Management and Control of Pollution Caused by urban Traffic, *International Journal of Environmental Research* 8(4): 1215-1222.
- Hamm, N. A. S., 2016. Spatial temporal modelling of particulate matter for health effects studies, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XLI-B8 2016, XXIII ISPRS Congress, 12–19 July 2016, Prague, Czech Republic.

Hankey, S., Marshall, J.D., 2015. Land use regression models of on-road particulate air pollution (particle number, black carbon, PM2.5, particle size) using mobile monitoring. *Environ. Sci. Technol.*, 49, 9194–9202.

Hatzopoulou, M.; Valois, M.F.; Levy, I.; Mihele, C.; Lu, G.; Bagg, S.; Minet, L.; Brook, J., 2017. Robustness of Land-Use Regression Models Developed from Mobile Air Pollutant Measurements. *Environ. Sci. Technol.*, 51, 3938–3947.

Hasenfratz, D.; Saukh, O.; Walser, C.; Hueglin, C.; Fierz, M.; Arn, T.; Beutel, J.; Thiele, L., 2015. Deriving high-resolution urban air pollution maps using mobile sensor nodes. *Pervasive Mob. Comput.*, 16, 268–285.

Hu, Y., Fan, J., Zhang, H., Chen, X., and Dai, G., 2016. An estimated method of urban PM2.5 concentration distribution for a mobile sensing system, *Pervasive and Mobile Computing*, 25, 88-103, <http://dx.doi.org/10.1016/j.pmcj.2015.06.004>.

Hu, K., and Rahman, A., 2017. HazeEst: Machine Learning Based Metropolitan Air Pollution Estimation From Fixed and Mobile Sensors, *IEEE Sensors Journal*, 17, 3517-3525.

JRC, 2017. Measuring air pollution with low-cost sensors. Thoughts on the quality of data measured by sensors. Gerboles M., Spinelle L., Borowiak A., Publication de la Commission Européenne.

Kingham, S., Longley, I., Salmond, J., Pattinson, W., et al., 2013. Variations in exposure to traffic pollution while travelling by different modes in a low density, less congested city, *Environmental Pollution*, 181, 211-218.

Liu J.F., Wang Q.M., 2016. Application of an improved svm algorithm for wireless sensor networks in the prediction of air pollution, *Chemical Engineering Transactions*, 51, 337-342 DOI:10.3303/CET1651057.

MacNaughton, P., Melly, S., Vallarino, J., Adamkiewicz, and Spengler, J. D., 2014. Impact of bicycle route type on exposure to traffic-related air pollution, *Science of the Total Environment*.

Marjovi, A., Arfire, A., and Martinoli, 2015. High Resolution Air Pollution Maps in Urban Environments Using Mobile Sensor Networks, 2015 International Conference on Distributed Computing in Sensor Systems (DCOSS): 11-20.

Minet, L., Liu, R., Valois, M.-F., Gehr, R.; Xu, J., Wachenthal, S., 2018. Development and comparison of air pollution exposure surfaces derived from on-road mobile monitoring and short-term stationary sidewalk measurements. *Environ. Sci. Technol.* 2018, 52, 3512–3519. Capturing the sensitivity of land-use regression models to short-term mobile monitoring campaigns using air pollution micro-sensors. *Environ. Pollut.* 2017, 230, 280–290.

Moltchanov, S., et al, 2014. On the feasibility of measuring urban air pollution by wireless distributed sensor networks, *Sc. of the total Env.*, 502, 537-547.

Mueller, M. D., Hasenfratz, D., Saukh, O., Fierz, M., and Hueglin, C., 2017. Statistical modelling of particle number concentration in Zurich at high spatio-temporal resolution utilizing data from a mobile sensor network, *Atmospheric Environment*, 126, 171-181.

Penza, M., et al, 2017. Urban Air Quality Monitoring with Networked Low-Cost Sensor-Systems, *Proceedings*, 1, 573; doi:10.3390/proceedings1040573.

Peters, J., Van den Bossche, J., Reggente, M., Van Poppel, M., De Baets, B., Theunis, J., 2014. Cyclist exposure to UFP and BC on urban routes in Antwerp, Belgium. *Atmos. Environ.*, 92, 31–43.

- Schneider, P., Castell, N., Vogt, M., Dauge, F. R. and Lahoz, W. A., 2017. Mapping urban air quality in near real-time using observations from low cost sensors and model information, *Environment International*. 106, 234-247.
- Shaban, K. B., A. Kadri, et al., 2016. Urban Air Pollution Monitoring System With Forecasting Models, *IEEE Sensors Journal* 16(8): 2598-2606.
- Shi, Y., Lau, K.K.L., and Ng, E., 2016. Developing street-level PM_{2.5} and PM₁₀ land use regression models in high-density Hong Kong with urban morphological factors. *Environ. Sci. Technol.*, 50, 8178–8187.
- Shirai, Y., Kishino, Y., Naya, F.; Yanagisawa, Y. Toward On-Demand Urban Air Quality Monitoring using Public Vehicles. In *Proceedings of the 2nd International Workshop on Smart, Trento, Italy, 12–16 December 2016*; ACM: New York, NY, USA, 2016; p. 1.
- Spinelle, L., Gerboles, M., Villani, M. G., Aleixandre, M., Bonavitacola, F., Field calibration of a cluster of low-cost available sensors for air quality monitoring. Part A: Ozone and nitrogen dioxide, *Sens. Actuators B Chem.* 215(2015) 249–257, <http://dx.doi.org/10.1016/j.snb.2015.03.031>.
- Van den Bossche, J., Peters, J., Verwaeren, J., Botteldooren, D., Theunis, J., De Baets, B., 2015. Mobile monitoring for mapping spatial variation in urban air quality: Development and validation of a methodology based on an extensive dataset. *Atmos. Environ.*, 105, 148–161.
- Wallace, J., Corr, D., Deluca, P., Kanaroglou, P., and McCarry B., 2009. Mobile monitoring of air pollution in cities: the case of Hamilton, Ontario, Canada, *Environ. Monit.*, 11, 998, 1003.
- Wang, M., Zhu, T., Zheng, J., Zhang, R. Y., Xie, X. X., et al., 2009. Use of a mobile laboratory to evaluate changes in on-road air pollutants during the Beijing 2008 Summer Olympics, *Atmos. Chem. Phys.* 9, 8247-8263.
- WMO/GAW, 2018. Low-cost sensors for the measurement of atmospheric composition: overview of topic and future applications. A paraître prochainement.
- Zaldei, A., et al, 2016. An integrated low-cost road traffic and air pollution monitoring platform for next citizen observatories, *Transp. Research Procedia*, 24, 531-538.
- Zimmerman, N., Presto, A.A., Kumar, S. P. N., Gu, J., Haurlyuk A., et al, 2018. A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring, *Atmos. Meas. Tech.*, 11,291-313.
- Zwack, Leonard M., Christopher Joseph Paciorek, John D. Spengler, and Jonathan Ian Levy. 2011. Modeling spatial patterns of traffic-related air pollutants in complex urban terrain, *Environmental Health Perspectives* 119(6): 852-859.

6.2 Géostatistique, cartographie et modélisation statistique

6.2.1 Références méthodologiques

- Chilès, J.-P., Delfiner, P.P., 2012. *Geostatistics: Modeling Spatial Uncertainty*. John Wiley & Sonspp. 726.
- Goovaerts, P., 1997. *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York, pp. 483.
- Hastie, T.J., Tibshirani, R.J., 1996. *Generalized Additive Models*. Chapman and Hall.

- Lahoz, W.A., Schneider, P.P., 2014. Data assimilation: making sense of earth observation. *Front. Environ. Sci.* 2 (16), 1–28. <http://dx.doi.org/10.3389/fenvs.2014.00016>.
- Lichtenstern, A., 2013. Kriging methods in spatial statistics, Bachelor's Thesis, Technische Universität München, Department of Mathematics.
- Rivoirard, 2003. Cours de géostatistique multivariable. Cours C-17 2. Centre de Géostatistique. Ecole des Mines de Paris.
- Wackernagel, H., 2003. *Multivariate Geostatistics*. Springer Berlin Heidelbergpp. 387.
- Wood, S.N., 2006. *Generalized Additive Models: An Introduction with R*. Chapman & Hall/CRC.

6.2.1 Publications diverses

- Ferreira, F., Tente, H., Torres, P.; Cardoso, S.; Palma-Oliveira, J.M., 2000. Air quality monitoring and management in Lisbon. *Environ. Monit. Assess.*, 65, 443–450.
- Jerrett, M., Burnett, R.T., Kanaroglou, P., Eyles, J., Finkelstein, N., Giovis, C., Brook, J.R., 2001. A GIS–environmental justice analysis of particulate air pollution in Hamilton, Canada. *Environ. Plan. A*, 33, 955–973.
- Künzli, N., Jerrett, M., Mack, W.J., Beckerman, B., LaBree, L., Gilliland, F., Thomas, D., Peters, J., Hodis, H.N., 2005. Ambient air pollution and atherosclerosis in Los Angeles. *Environ. Health Perspect.*, 113, 201.
- Liu, L.J.S., Rossini, A., 1996. Use of kriging models to predict 12-hour mean ozone concentrations in metropolitan Toronto—a pilot study. *Environ. Int.*, 22, 677–692.
- Mercer, L., Szpiro, A.A., Sheppard, L., Lindström, J., Adar, S.D., Allen, R.W., Avol, E.L., Oron, A.P., Larson, T., Liu, L.-J.S., Kaufman, J.D., 2011. Comparing universal kriging and land-use regression for predicting concentrations of gaseous oxides of nitrogen (NO_x) for the Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air). *Atmospheric Environment*, 45(26), 4412–4420.
- Vayena, E., Salathé, M., Madoff, L.C., Brownstein, J.S., 2015. Ethical challenges of BigData in public health. *PLoS Comput Biol.* 11 (2), e1003904. <http://dx.doi.org/10.1371/journal.pcbi.1003904>.
- Xie, X., Semanjski, I., Gautama, S., Tsiligianni, E., Deligiannis, N., Rajan, R.T., Pasveer, F., Philips, W. A., 2017. Review of Urban Air Pollution Monitoring and Exposure Assessment Methods. *ISPRS Int. J. Geo-Inf.*, 6, 389.

6.2.2 Rapports LCSQA

- Beauchamp M., Malherbe L., Létinois L., Drevet J., 2017. Note de synthèse méthodologique sur la caractérisation des situations de dépassement de seuil : délimitation des zones de dépassement et estimation des populations et écosystèmes exposés. Note technique LCSQA.
- Beauchamp M., Malherbe L., Meleux F., Létinois L., 2016. Note de synthèse sur les développements récents en matière de cartes analysées des résultats de modélisation. Cartographie des concentrations de PM₁₀ et de PM_{2.5}. Note technique LCSQA.

Beauchamp M., Malherbe L., Létinois L., 2014. Estimation de l'exposition des populations aux dépassements de seuils réglementaires. Echelle urbaine. Rapport LCSQA et annexe technique associée (2016).

Beauchamp M., Malherbe L. et Létinois, 2010. L. Application de méthodes géostatistiques pour la détermination de zones de représentativité en concentration et la cartographie des dépassements de seuils. Rapport LCSQA.

Malherbe L., Ung A., 2009. Travaux relatifs à la plate-forme nationale de modélisation PREV'AIR : réalisation de cartes analysées d'ozone. Rapport LCSQA.

Wroblewski A., Riffault V., Perdrix E., Malherbe L., 2007. Adaptation des plans d'échantillonnage aux objectifs des campagnes. Echantillonnage spatial – Guide de recommandations. Rapport LCSQA.

Malherbe L., Cárdenas G., 2005. Application des méthodes géostatistiques pour l'exploitation conjointe des mesures de fond et de proximité. Rapport LCSQA.

Annexe 1

Tableau 1 : Etudes sur les micro-capteurs. Synthèse des publications consultées

Référence	Polluant(s)	Période d'échantillonnage	Domaine d'étude	Support(s) des micro-capteurs	Méthode	Cartographie : Oui/Non/ ? Résolution temporelle des cartographies	Objectif
Alvear et al., 2016(a), 2016 (b)	O ₃	2008-2014	Valence (Espagne)	Capteurs mobiles : smartphones	Krigeage ordinaire	Oui	Estimer l'exposition Etudier l'influence de l'échantillonnage spatial et temporel sur l'interpolation
Apte et al., 2017	BC, NO, NO ₂	Mai 2015 – mai 2016	Oklahoma, CA (Etats-Unis)	Capteurs mobiles : voitures Google	« Snapping » (agrégation de données sur des portions de routes)	Oui (sur les routes uniquement) Journalière (médiane)	Identifier des hotspots et estimer l'exposition
Bigazzi and Figliozzi, 2015	CO, COV (composés organiques volatils)	9 jours, 2013	Portland, OR (Etats-Unis)	Capteurs mobiles : vélos	LUR	Non	Estimation (i) de la relation entre l'exposition des cyclistes et la densité de trafic, et (ii) de l'exposition au COVs en fonction de l'installation échantillonnée (axes majeurs, axes mineurs, ponts...)
Boubrima et al., 2017 Voir aussi Boubrima et al., 2016(a), 2016(b)	NO ₂	2012-2013	Lyon, France	Capteurs fixes virtuels (Utilisation de données simulées par un modèle de dispersion)	Interpolation par inverse des distances	Oui Résol. temp. non précisée ou moyenne annuelle dans les exemples traités	Mettre au point d'une méthodologie permettant d'optimiser le placement des micro-capteurs afin d'estimer un champ 2D ou 3D de concentration
Cavellin et al., 2016	O ₃ , NO ₂	Printemps, été et automne 2014	Montréal (Canada)	Capteurs fixes	LUR	Oui Moyenne saisonnière	Estimer l'exposition
Dong et al, 2015 ; Gao et al, 2016	PM _{2.5}	24 février – 3 avril 2015 (Hangzhou) 13 décembre 2014–2016 (Ningbo)	Hangzhou, Ningbo (Chine)	Capteurs mobiles : Bus (Hangzhou) Voitures (Ningbo)	LUR	Pas d'information	Pas d'information
Gomez et al., 2014	O ₃ , NO ₂ , CO, PM	Une année	Salamanque (Espagne)	Capteurs fixes	ARIMA	Non	Prévoir les concentrations à H+1 et H+3 pour la gestion du trafic
Hamm et al., 2016	PM ₁₀ , PM _{2.5}	1-14 octobre 2014	Eindhoven (Pays-Bas)	Capteurs fixes	Modèle dynamique (composante temporelle) et processus géostatistique gaussien (composante spatiale)	En projet	Estimer l'exposition
Hankey and Marshall, 2015	BC, PM _{2.5}	Heures de pointe, 14 août – 16 octobre 2012	Minneapolis, MN (Etats-Unis)	Capteurs mobiles : vélos	LUR	Oui Moyenne sur le matin et sur l'après-midi	Estimer l'exposition

Hasenfratz et al, 2015	UFP	2012-2015	Zürich (Suisse)	Capteurs mobiles : 10 tramways	LUR	Oui Moyenne sur plusieurs jours ou journalière	Estimer l'exposition
Hatzopoulou et al, 2017	NO ₂ , UFP	2009	Montréal, QC (Canada)	Capteurs mobiles : voitures	LUR	Non	Evaluer l'impact des protocoles d'échantillonnage sur les estimations de concentration de polluant
Hu K. et al., 2017	CO	Mai 2009 - mai 2016	Sydney (Australie)	Capteurs mobiles : citoyens et voitures	Machine learning (SVR, DTR, RFR, XGB, MLP, LR, ABR)	Oui Horaire	Information auprès des citoyens (Système HazeEst et application web)
Hu Y.D. et al., 2016	PM _{2.5}	Octobre 2014 – mars 2015	Quartier de HangZhou (Chine)	Capteurs mobiles	Machine Learning (PCEM)	Oui Résol. temp. non précisée	Estimer les concentrations dans les zones non instrumentées
Kingham et al, 2013	CO, PM, UFP	Pas d'information	Christchurch (Nouvelle Zélande)	Capteurs mobiles : bus, voitures, vélos	LUR	Non	Estimer l'exposition en fonction du mode de déplacement
Liu & Wang, 2016	Gaz et PM _{2.5}	2015	Pékin (Chine)	Capteurs fixes	Machine Learning (SVR)	Non	Prévision
MacNaughton et al, 2014	BC, CO, CO ₂ , NO ₂ , O ₃	2007-2011	Boston, MA (Etats-Unis)	Capteur mobile : vélo	LUR	Non	Estimer l'exposition
Marjovi et al., 2015	LDSA, NO ₂ et CO	14 mois	Lausanne (Suisse)	Capteurs mobiles : bus	Machine Learning (LLR, NBLLR), PGM	Uniquement le long des routes Moyenne horaire, journalière, hebdomadaire et mensuelle	Estimation des concentrations avec une haute résolution spatio-temporelle
Minet et al., 2018	BC, UFP	Mai-août 2016	Toronto, ON (Canada)	Capteurs mobiles (vélos) et fixes (piétons stationnant pendant 20 minutes)	LUR	Oui Moyenne estivale	Estimer l'exposition
Minet et al, 2017	NO ₂	Mai-août 2015	Montréal, QC (Canada)	Capteurs mobiles : vélos et piétons	LUR	Oui Moyenne estivale	Estimer l'exposition
Mueller et al., 2016	Particules (PNC)	Juillet-septembre 2013 et décembre 2013-février 2014	Zürich (Suisse)	Capteurs mobiles : tramways	GAM avec algorithme de sélection de variables	Oui 30 mn	Estimation de l'exposition, aménagement urbain
Peters et al, 2014	BC, UFP	Février-mars 2012	Antwerpen (Belgique)	Capteur mobile : vélo	LUR	Oui (uniquement sur les routes)	Estimer l'exposition
Schneider et al., 2017	NO ₂	Avril 2015	Oslo (Norvège)	Plateforme AQMesh	Krigeage Universel à dérive externe	Oui Annuelle	Comblent le manque d'information (variations spatiales et temporelles) ; méthode de fusion de données
Shaban et al., 2016	O ₃ , NO ₂ , SO ₂		Echelle urbaine	Capteurs fixes	Machine Learning (SVR, M5P, ANN). Modélisation univariée et multivariée.	Pas d'information	Prévision

Shi et al, 2016	PM _{2.5} , PM ₁₀	2014-2015	Hongkong, Chine	Capteur mobile : Voiture Toyota HiAce	LUR	Oui	Identifier des hotspots de concentration de polluants
Shirai et al, 2016	CO, NO ₂ , O ₃ , PM _{2.5} , ultraviolet, poussières, pollen, 2 types de contaminants	Janvier 2015- 2016	Fujisawa (Japon)	Capteurs mobiles : camions poubelles	LUR	Pas d'information	Estimer l'exposition
Van den Bossche et al, 2015	BC		Antwerpen (Belgique)	Capteur mobile : vélo	Statistiques exploratoires	Non	Validation des données de micro-capteurs pour la cartographie
Wallace et al., 2009	SO ₂ , NO ₂ , PM _{2.5}	2005-2013	Hamilton, ON, Canada	Capteur mobile : camionnette	LUR	Oui (sur les routes uniquement) Journalière	Estimer l'exposition et Identifier les hotspots sur les axes routiers
Wang et al., 2009	NO _x , CO, BC, surface PM ₁₀ , benzène, toluène, la somme des éthylbenzène, et m-, p-, et o-xylène (BTEX).	Août 2008	Pékin, Chine	Capteur mobile : camionnette	LUR	Oui Moyenne sur 4 sections (nord, est, sud et ouest de la route échantillonnée), moyenne sur plusieurs jours	Evaluer les variations de concentrations de polluants sur la portion de route échantillonnée
Zwack et al, 2011	PM _{2.5} , UFP	Juin 2007 (3 semaines)	Williamsburg, NY (Etats-Unis)	Capteurs mobiles : 6 piétons	LUR	Oui Moyenne sur la période d'échantillonnage	Evaluer l'impact à la fois de la météorologie et du trafic sur les concentrations de polluants

Signification des sigles :

Méthodes :

ABR: Adaptive Boosting Regression

ANN: Artificial Neural Network

DTR: Decision Tree Regression

GAM: General Additive Model

LR: Linear Regression

LLR: Log-Linear Regression

M5P: M5P model tree

MLP: Multi- Layer Perceptrons

NBLLR: Network-based Log-Linear Regression

PCEM: Probabilistic Concentration Estimation Model

PGM: Probabilistic Graphical Model

RFR: Random Forest Regression

SVR: Support Vector Regression (SVM: Support Vector Machine)

XGB: Extreme Gradient Boosting

Paramètres mesurés:

LDSA: Lung Deposited Surface Area

PNC: Particle Number Concentration (concentration en nombre de particules)

UFP : Ultrafine Particles