



HAL
open science

Caractérisation de la pollution urbaine en Île-de-France par une synergie de mesures de surface et de modélisation fine échelle

Delphy Rodriguez

► **To cite this version:**

Delphy Rodriguez. Caractérisation de la pollution urbaine en Île-de-France par une synergie de mesures de surface et de modélisation fine échelle. Océan, Atmosphère. Sorbonne Université, 2019. Français. NNT : 2019SORUS341 . tel-03016033

HAL Id: tel-03016033

<https://theses.hal.science/tel-03016033>

Submitted on 20 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Sorbonne Université

Ecole Doctorale des Sciences de l'Environnement d'Île-de-France (ED 129)

LATMOS/ Équipe TROPO et LMD/ Équipe InTro

Caractérisation de la pollution urbaine en Île-de-France par une synergie de mesures de surface et de modélisation fine échelle

Par Delphy Rodriguez

Thèse de doctorat en Sciences

Spécialité : Physique de l'Atmosphère

Dirigée par Sébastien Payan et Myrto Valari

Présentée et soutenue publiquement à Paris, le 24 juin 2019

Devant un jury composé de :

<i>Président :</i>	François RAVETTA (LATMOS)
<i>Rapporteurs :</i>	Bérengère LEBENTAL (IFSTTAR) Cathy LIOUSSE (Laboratoire d'Aérodologie)
<i>Examineurs :</i>	Augustin COLETTE (INERIS) Valérie GROS (LSCE) Olivier SANCHEZ (AIRPARIF)
<i>Directeur :</i>	Sébastien PAYAN (LATMOS)
<i>Co-directeur :</i>	Myrto VALARI (LMD)
<i>Invités :</i>	Laurence EYMARD (LATMOS) Eric PARENT (AgroParisTech)

Titre : Caractérisation de la pollution urbaine en Île-de-France par une synergie de mesures de surface et de modélisation fine échelle

Mots clés : pollution atmosphérique urbaine, capteurs mobiles à bas coût, réseau de surveillance, modèle urbain haute résolution, représentativité spatiale, estimation bayésienne

Résumé :

L'impact sanitaire lié à la pollution de l'air nécessite une estimation précise de celle-ci.

Les réseaux de stations de mesures des agences de surveillance de la qualité de l'air (AIRPARIF en Île-de-France) ne sont pas suffisamment denses pour renseigner sur l'hétérogénéité de la pollution en ville. Et, les modèles haute résolution simulant les champs de concentration de polluants en 3D ont une large couverture spatiale mais sont limités par leurs incertitudes. Ces deux sources d'information exploitées indépendamment ne permettent pas d'évaluer finement l'exposition d'un individu. Nous proposons deux approches pour résoudre ce problème :

(1) par la mesure directe des polluants avec des capteurs mobiles à bas coût et des instruments de référence. Des niveaux de pollution très variables ont été constatés entre les microenvironnements et dans une même pièce. Ces capteurs devraient être déployés en grand nombre pour palier à leurs contraintes techniques. Les instruments de référence, très coûteux et volumineux, ne peuvent être utilisés que ponctuellement.

(2) en combinant les concentrations simulées par le modèle Parallel Micro-SWIFT-SPRAY (PMSS) à Paris avec une résolution horizontale de 3 mètres et les mesures des stations de surface AIRPARIF.

Nous avons déterminé des « zones de représentativité » - zones géographiques où les concentrations sont très proches de celle de la station - uniquement à partir des sorties du modèle PMSS. Ensuite, nous avons développé un modèle bayésien pour propager la mesure des stations dans ces zones.

Title: Characterization of urban pollution in Parisian region by a synergy of surface measurements and high-resolution modelling

Keywords: Urban air pollution; Mobile low cost sensors; Surface monitoring networks; High-resolution air quality modeling; Spatial representativeness; Bayesian estimation

Abstract:

The harmful effects of air pollution need a high-resolution concentration estimate. Ambient pollutant concentrations are routinely measured by surface monitoring sites of local agencies (AIRPARIF in Paris area, France). Such networks are not dense enough to represent the strong horizontal gradients of pollutant concentrations over urban areas. And, high-resolution models that simulate 3D pollutant concentration fields have a large spatial coverage but suffer from uncertainties. Those both information sources exploited independently are not able to accurately assess an individual's exposure. We suggest two approaches to solve this problem:

(1) direct pollution measurement by using low cost mobile sensors and reference instruments. A high variability across pollution levels is shown between microenvironments and also in the same room. Mobile sensors should be deployed on a large scale due to their technical constraints. Reference instruments are very expensive, cumbersome, and can only be used occasionally.

(2) by combining concentration fields of the Parallel Micro-SWIFT-SPRAY (PMSS) model over Paris at a horizontal resolution of 3 meters with AIRPARIF local ground stations measurements. We determined "representativeness areas" - perimeter where concentrations are very close to the one of the station location – only from PMSS simulations. Next, we developed a Bayesian model to extend the stations measurements within these areas.

Remerciements

Je profite de ces quelques lignes pour remercier les personnes qui ont contribué à la réussite de cette thèse.

Je remercie tout d'abord Philippe Keckhut, le directeur du LATMOS, de m'avoir accueillie dans son laboratoire.

Je remercie les membres du jury. François Ravetta d'avoir accepté de présider ma soutenance. Bérengère Lebental et Cathy Liousse, dont j'ai intégré la plupart des commentaires et remarques très pertinentes de leurs rapports dans la version finale de mon manuscrit. Olivier Sanchez, Augustin Colette et Valérie Gros, les examinateurs. Augustin et Valérie ont suivi l'avancée de ma thèse lors de deux comités de thèse et Olivier lors de réunions dans les locaux d'AIRPARIF. Je les remercie pour leurs précieux conseils lors de nos rencontres.

Je remercie mes directeurs de thèse : Sébastien Payan, qui m'a accordé toute sa confiance dès le début pour mener à bien cette thèse. Myrto Valari qui m'a aidée à définir de nouveaux axes de recherche lorsqu'on s'est aperçu que les capteurs à bas coût - en cours de développement au LATMOS - ne seraient pas prêts à temps pour ma thèse. Et je la remercie également de m'avoir poussé à faire des « talks » dans des conférences internationales.

Je remercie chaleureusement les deux invités de mon jury : Eric Parent et Laurence Eymard. Ces invités – que je suis allée chercher lors d'une période de remise en question – ont joué un rôle clé dans la réussite de ma thèse.

Un grand merci à Eric Parent, expert en statistiques bayésiennes, pour sa disponibilité et sa transmission de savoir. Nos nombreuses discussions, toujours plus intéressantes les unes que les autres, ont largement contribué à ma compréhension sur le sujet. J'ai également eu la chance de suivre le cours qu'il enseigne à AgroParisTech sur son domaine de prédilection, que je recommande fortement !

Un immense merci à Laurence Eymard, qui m'a accompagnée au quotidien pendant ma thèse. Je la remercie pour sa pédagogie, son partage de connaissances et son optimisme. C'est une grande chance d'avoir travaillé à ses côtés. Grâce à Laurence, j'ai réussi à donner le meilleur de moi. Et, elle m'a poussé à valoriser davantage mon travail avec la publication de deux articles. Laurence est un puits de savoir et une ressource extraordinaire. J'ai une profonde gratitude à son égard. Ces mots sont peu de chose pour la remercier.

Je remercie tous mes collègues du LATMOS, qui vont me manquer. Je fais un petit clin d'œil à Maya. J'ai eu grand plaisir à partager son bureau : c'est un rayon de soleil et un condensé de bonne humeur.

Je pense également aux collègues qui sont passés au LATMOS et qui n'y sont plus.

Je remercie maintenant mes parents. J'ai réalisé l'exploit d'avoir deux enfants pendant ma thèse : Alice (2 ans et demi) et Hugo (4 mois) sans pour autant renoncer à aller jusqu'au

bout de celle-ci. J'ai pu réaliser cela grâce au soutien sans faille de mes parents, depuis toujours. Ils ont été d'une grande disponibilité et aux petits soins pour leurs petits-enfants. Merci 'Mouchka' d'avoir toujours cru en moi et de m'avoir redonné espoir dans les moments de doute.

John, le papa de mes enfants a été un élément moteur dans la réalisation de cette thèse. Je le remercie pour son positivisme dans les moments difficiles et pour son éternel soutien. Il m'a aidé à maintenir le cap en me rappelant mes objectifs moyen et long terme lorsque cela était nécessaire.

Ces quelques de mots de la fin vont à mes deux étoiles : Alice et Hugo. Merci à vous deux, mes petits anges, pour tout ce bonheur que vous nous apportez.

*J'espère que ce manuscrit apportera une modeste contribution à ce que l'on respire tous
un air plus sain.*

Table des matières

Remerciements	1
Table des matières	4
Table des figures	6
Liste des tableaux	9
Glossaire	10
Introduction	11
Chapitre 1. État de l'art : généralités sur la pollution atmosphérique	15
1.1 Les principaux polluants	15
1.1.1 Impacts sanitaires	15
1.1.2 Seuils réglementés par la législation européenne et française	17
1.2 Cas de la région parisienne	19
1.2.1 Dépassements des seuils	19
1.3 Processus physico-chimiques	20
1.3.1 Sources d'émissions	22
1.3.2 Transformations chimiques	24
1.3.3 Transport horizontal	27
1.3.4 Mélange dans la couche limite	27
1.3.5 Dépôts humides (précipitation) et dépôts secs	29
1.4 Qualité de l'air intérieur	30
Chapitre 2. Outils utilisés	33
2.1 Instruments de mesures	33
2.2 Modélisation	36
2.2.1 Généralités	36
2.2.2 Cadastres d'émissions en entrée des modèles de qualité de l'air	37
2.2.3 PMSS : un modèle urbain de dispersion atmosphérique haute résolution	38
2.3 AIRPARIF : réseau de mesures de la pollution de l'air en région parisienne	41
Chapitre 3. Approche expérimentale de l'exposition aux PM ₁₀ et PM _{2,5} en Île-de-France	45
3.1 Introduction	45
3.2 Campagne de mesures à Paris	46
3.2.1 Caractéristiques de la période d'étude : conditions météorologiques et niveaux de pollution de fond	46
3.2.2 Protocole expérimental	47
3.2.3 Résultats et discussion	49
3.3 Campagne de mesures dans le démonstrateur de Sense-City, à Champs-sur-Marne (77)	56
3.3.1 Sense-City : un équipement d'excellence pour concevoir la ville de demain	56
3.3.2 Protocole expérimental	58
3.3.3 Résultats et discussion	63
3.4 Retour d'expérience en vue de l'élaboration des capteurs mobiles	66
3.4.1 Principaux résultats	66
3.4.2 Contraintes à intégrer dans les protocoles de mesures	67
Chapitre 4. Représentativité spatiale des stations de mesures des NO _x et PM ₁₀ à Paris /« On the spatial representativeness of NO _x and PM ₁₀ monitoring-sites in Paris, France »	69
4.1 Notion de représentativité spatiale d'une mesure	69
4.2 Résumé de l'article	69
4.3 Article publié	72

Chapitre 5. Estimation bayésienne des concentrations de NO _x et PM ₁₀ en utilisant des simulations numériques haute résolution et des mesures de surface à Paris /« NO _x and PM ₁₀ Bayesian concentration estimates using high-resolution numerical simulations and ground measurements over Paris, France ».....	89
5.1 Concept bayésien.....	89
5.2 Résumé de l'article.....	90
5.2.1 Description du modèle bayésien	90
5.2.2 À la recherche d'un modèle probabiliste optimal au niveau des stations de mesures	92
5.2.3 Propager la mesure locale dans la zone de représentativité	93
5.2.4 Principaux résultats	94
5.2.5 Discussion	95
5.3 Article publié.....	96
Conclusion et perspectives	107
Bibliographie	111
Annexe : Support de la soutenance du 24 juin 2019, à Paris (Campus Pierre et Marie Curie, Sorbonne Université).....	118

Table des figures

Figure 1 : Schéma de variation des concentrations de particules fines (PM _{2,5}) dans l'air ambiant. Source : AIRPARIF (rapport, 2011)	12
Figure 2 : Répartition des dépassements de seuils d'information et d'alerte pour (a) les particules PM ₁₀ et (b) l'ozone (O ₃), de janvier 2015 à janvier 2018. Le chiffre indique le nombre de jours de dépassement de seuils pour les mois de l'année concernés. Figure réalisée à partir des données AIRPARIF.....	20
Figure 3 : Représentation schématique des principaux processus de la chimie atmosphérique illustrant le cycle : émission en surface des polluants par différentes sources naturelles ou anthropiques, diffusion verticale, transport et transformations photochimiques dans l'atmosphère, dépôt sec par interaction directe avec la surface ou dépôt humide par les précipitations. Source : Atmosphère, océan et climat (R. Delmas, <i>et al.</i> , 2007).....	21
Figure 4 : Comparaison des temps de vie et des échelles spatiales pour quelques espèces. Source : Physique et chimie de l'atmosphère (R. Delmas, 2005)	22
Figure 5 : Représentation schématique des principaux mécanismes impliqués dans la formation d'ozone troposphérique. Source : Camredon <i>et al.</i> (2007)	25
Figure 6 : Évolution de la stabilité de la couche limite atmosphérique au cours d'une journée	28
Figure 7 : Exemple d'un épisode de pollution anticyclonique à Paris, Notre Dame, le 13 mars 2014 (à gauche) versus un jour sans pollution particulaire le 21 octobre 2013 (à droite)	29
Figure 8 : Sources des polluants de l'air intérieur. Source : ADEME, « un air sain chez soi »	31
Figure 9 : Photographie de l'OSIRIS (a), du DustMate (b) et du capteur e-PM (c)	33
Figure 10 : Schéma de fonctionnement d'un compteur optique. Source : rapport INERIS (Blanchard <i>et al.</i> , 2004)	34
Figure 11 : Radar d'évaluation d'un système « capteurs » pour chaque polluant. Source : rapport LCSQA. MAPE=Pourcentage Moyen des Ecarts en Valeur Absolue et IPI=Indice de performance intégré.....	36
Figure 12 : Différentes étapes de la chaîne de calcul HEAVEN pour estimer les émissions horaires liées au trafic en Île-de-France. Source : AIRPARIF.....	38
Figure 13 : Schéma d'une maille lagrangienne. Source : cours sur la modélisation de la chimie atmosphérique (Aumont).....	39
Figure 14 : Déploiement instrumental du réseau AIRPARIF en région Île-de-France. Source : AIRPARIF.....	42
Figure 15 : Plan de Paris avec l'emplacement des stations AIRPARIF étudiées, la Seine et les routes (en blanc et orange). Le rouge représente la limite de Paris avec le boulevard périphérique.....	43
Figure 16 : Nuage de points qui représente les concentrations mesurées des stations AIRPARIF versus les concentrations simulées par le modèle PMSS (en µg/m ³). La ligne continue noire correspond à y=x. La ligne en pointillée représente la régression linéaire y=ax+b. Le RMSE est exprimé en µg/m ³	44
Figure 17 : Évolution temporelle des concentrations de PM ₁₀ et PM _{2,5} (en µg/m ³) du 07 au 16 mars 2016, au niveau de la station AIRPARIF PA04C, située Place Igor Stravinsky (Paris 4 ^{ème} arrondissement). Les lignes horizontales indiquent les seuils à ne pas dépasser pour les PM ₁₀ (ligne rouge) et les PM _{2,5} (ligne bleue) selon les recommandations de l'OMS. En effet, l'OMS recommande de ne pas dépasser ces moyennes journalières plus de 35 jours par an pour les PM ₁₀ et cette moyenne annuelle pour les PM _{2,5}	46

Figure 18 : Localisation des sites de mesures : (1) station de mesures de fond AIRPARIF PA04C (4 ^{ème} arrondissement de Paris), (2) local technique du LATMOS – Sorbonne Université (5 ^{ème} arrondissement de Paris), (3) domicile : 39 avenue des Gobelins (13 ^{ème} arrondissement de Paris)	48
Figure 19 : Concentrations enregistrées par l’OSIRIS positionné au LATMOS (référence) versus concentrations enregistrées par l’OSIRIS positionné à domicile (« Home ») ou du DustMate pour les (a) PM _{2,5} et (b) PM ₁₀ (en µg/m ³)	49
Figure 20 : Évolution temporelle des concentrations de PM ₁₀ et PM _{2,5} (en µg/m ³) enregistrées par l’OSIRIS positionné au LATMOS (avec des maxima à 17.3 µg/m ³ (PM _{2,5}) et 203.8 µg/m ³ (PM ₁₀)). La ligne horizontale rouge indique le seuil à ne pas dépasser pour les PM ₁₀ selon les recommandations de l’OMS. Ici, les valeurs de PM _{2,5} ne dépassent pas le seuil recommandé par l’OMS.....	50
Figure 21 : Évolution temporelle des concentrations de PM ₁₀ et PM _{2,5} (en µg/m ³) enregistrées par l’OSIRIS positionné au domicile. Les données ont été étalonnées par rapport à l’OSIRIS positionné au LATMOS (avec des maxima à 206.7 µg/m ³ (PM _{2,5}) et 4458.9 µg/m ³ (PM ₁₀)). Les lignes horizontales indiquent les seuils à ne pas dépasser pour les PM ₁₀ (ligne rouge) et les PM _{2,5} (ligne bleue) selon les recommandations de l’OMS.....	51
Figure 22 : Plan du quartier du campus Pierre et Marie Curie de Sorbonne Université. Axes routiers : quai St Bernard (en bleu) : trafic fort tout au long de la journée, rue Cuvier (en rouge) : trafic généralement limité (~2/3 voitures par minute), place Jussieu (en jaune) : zone ouverte avec trafic de proximité faible. Source : Stephane Pawlak, Colloque international "Capteurs Individuels de Pollution Atmosphérique : Innovation ou Révolution ?", 28 et 29 novembre 2018.	53
Figure 23 : Boîte à moustache avec la moustache qui indique le 5 ^{ème} et 95 ^{ème} percentile. Les autres valeurs extrêmes ne sont pas représentées. Concentrations en (a) PM ₁₀ et (b) PM _{2,5} des environnements intérieurs. Concentrations en (c) PM ₁₀ et (d) PM _{2,5} lors de déplacements à l’extérieur (en µg/m ³)	56
Figure 24 : Localisation de Sense-City dans la région Île-de-France, plan Google map (a). Photographie de la première mini-ville de Sense-City et la chambre climatique (b) et du démonstrateur de Sense-City (c)	58
Figure 25 : Positionnement des instruments de référence (OSIRIS) dans le démonstrateur de Sense-City : à l’extérieur du chalet et à l’intérieur du chalet (a). Photographie de l’OSIRIS au niveau de son emplacement à l’extérieur du chalet (b)	59
Figure 26 : Photographie à l’intérieur d’un des deux chalets du démonstrateur de Sense-City avec les capteurs e-PM accrochés autour du cou et le positionnement de l’instrument de référence (OSIRIS).....	59
Figure 27 : Photographies (a),(b),(c) des scénarios de cuisine, bricolage et aspirateur et (d) de la phase d’aération.....	60
Figure 28 : Déploiement instrumental au sol pendant l’inter-étalonnage	61
Figure 29 : Histogramme sur les données brutes des capteurs e-PM superposées à celles de l’OSIRIS (en µg/m ³).....	61
Figure 30 : Données brutes des concentrations des PM ₁₀ et PM _{2,5} (en µg/m ³), le 22 juin 2016	62
Figure 31 : Concentrations en PM ₁₀ (en µg/m ³) supérieures à 5 µg/m ³ : OSIRIS versus capteurs e-PM, moyennes par pas de 35 minutes. La ligne noire représente une droite y=x. La ligne rouge représente la droite de régression linéaire y=ax+b pour chaque e-PM.	63
Figure 32 : Concentrations en PM ₁₀ (en µg/m ³) de l’OSIRIS et des capteurs e-PM post étalonnage moyennées par pas de 35 minutes pour les concentrations en PM ₁₀ de l’OSIRIS supérieures à 5 µg/m ³	63
Figure 33 : Concentrations en PM ₁₀ (en µg/m ³) étalonnées à partir des coefficients de régression calculées sur des moyennes de 35 minutes pour les concentrations supérieures à	

5 $\mu\text{g}/\text{m}^3$. Les lignes horizontales noires représentent le minimum et le maximum de la gamme de mesures de l'e-PM. La ligne horizontale violet clair représente le seuil qu'il est recommandé de ne pas dépasser plus de 35 jours par an (en moyenne journalière) selon l'OMS..... 64

Figure 34 : Rapport des concentrations intérieure/extérieure des deux OSIRIS (intérieur et extérieur) à partir des données étalonnées, le 22 juin 2016 65

Figure 35 : Exemple de mise à jour bayésienne du logarithme de la concentration de NO_x au niveau de la station AIRPARIF située Place de l'Opéra à 15 h le 2 mars 2016 avec la distribution *a priori* (orange), la vraisemblance (bleu) et la distribution *a posteriori* (rouge), la mesure AIRPARIF (ligne verticale verte en pointillé) et la concentration simulée par le modèle PMSS (ligne verticale noire en pointillé). 91

Liste des tableaux

Tableau 1 : Seuils à ne pas dépasser pour les (a) dioxydes et oxydes d'azote (NO ₂ et NO _x), les particules PM ₁₀ et le dioxyde de soufre (SO ₂), (b) l'ozone (O ₃) et (c) les particules fines PM _{2,5} selon la réglementation française. Source : AIRPARIF.....	19
Tableau 2 : Répartition des émissions par secteur d'activité pour les (a) NO _x (en équivalent NO ₂), (b) en Île-de-France. Source : AIRPARIF (inventaire sur l'année 2015).....	24
Tableau 3 : Origine et nature des polluants de l'air intérieur. Source : thèse Le Behec (2016).....	32
Tableau 4 : Caractéristiques techniques des instruments de référence : OSIRIS, DustMate et capteur e-PM	35
Tableau 5 : Caractéristiques des jours étudiés, pression de surface, direction et vitesse de vent (en moyenne journalière) sont extraits des ré-analyses d'ERA 5 du modèle ECMWF (0.25°).....	47
Tableau 6 : Nombre de mesures (« Nb mesures ») avec une fréquence d'échantillonnage d'une minute, moyenne, écart type, médiane, minimum et maximum des concentrations en PM ₁₀ et PM _{2,5} (en µg/m ³) à l'intérieur (a et b), lors de trajets à pied et en métro (c et d).....	55
Tableau 7 : Temps d'acquisition de l'OSIRIS et des capteurs e-PM.....	59
Tableau 8 : Définition des scénarios dans un de chalet du démonstrateur de Sense-City	60
Tableau 9 : Temps d'acquisition de l'OSIRIS et des capteurs e-PM pendant la phase d'étalonnage	61
Tableau 10 : Coefficient de corrélation des concentrations en PM ₁₀ de l'OSIRIS versus celles des capteurs e-PM pour plusieurs pas de temps et nombre de points (« nb points »).....	62
Tableau 11 : Les deux premières colonnes indiquent le type de station (colonne 1) et l'identifiant de la station (colonne 2). Les colonnes 3 et 4 indiquent le rapport entre l'aire la plus probable de NRMSE et l'aire d'homogénéité la plus probable pour les NO _x (colonne 3) et pour les PM ₁₀ (colonne 4).....	71
Tableau 12 : La première colonne indique le type de station et l'identifiant de la station (colonne 2) et les colonnes 3 et 4 indiquent le rapport de l'aire partagée les jours sans vent versus les jours avec pour l'ensemble des stations et les deux polluants NO _x (colonne 3) et PM ₁₀ (colonne 4).....	72
Tableau 13 : Différence de RMSE (en pourcentage %) sur les concentrations horaires des PM ₁₀ pour les dix jours de l'étude, avec pour référence la mesure des stations AIRPARIF. Zone grisée et résultats en gras, pour la vraisemblance avec les données regroupées par tranches de 2 heures pour les priors « trafic » et « météo »	93

Glossaire

AASQA Association Agrée pour la Surveillance de la Qualité de l’Air
ADEME Agence de l’Environnement et de la Maîtrise de l’Énergie
AEE Agence Européenne de l’Environnement
ANR Agence Nationale de Recherche
CEREA Centre d’Enseignement et de Recherche en Environnement Atmosphérique
CIRC Centre International de Recherche sur le Cancer
CLA Couche Limite Atmosphérique
EMEP European Monitoring and Evaluation Programme
EPA Environmental Protection Agency
GUAPO Global Urban Air Pollution Observatory
HEAVEN Healthier Environment through the Abatement of Vehicle Emissions and Noise
IFSTTAR Institut Français des Sciences et Technologies des Transports, de l’aménagement et des Réseaux
FIMEA Fédération Interprofessionnelle de Métiers de l’Environnement Atmosphérique
LATMOS Laboratoire Atmosphères Milieux Observations Spatiales
LAURE Loi sur l’Air et l’Utilisation Rationnelle de l’Énergie
LCSQA Laboratoire Central de Surveillance de la Qualité de l’Air
LMD Laboratoire de Météorologie Dynamique
NO_x Oxydes d’azote (NO_x= NO+NO₂)
NO Oxyde d’azote
NO₂ Dioxyde d’azote
OQAI Observatoire de la Qualité de l’Air Intérieur
OMS Organisation Mondiale de la Santé
PCIT Pôle de Coordination des Inventaires Territoriaux
PM_{2,5} Particulate Matter 2.5
PM₁₀ Particulate Matter 10
PMSS Parallel Micro-SWIFT-SPRAY
PRIMEQUAL Programme de Recherche Interorganisme pour une Meilleure Qualité de l’air
WCRP World Climate Research Programme
WRF Weather Research and Forecasting

Introduction

Selon l'Organisation Mondiale de la Santé (OMS), en 2016, 8 millions de personnes meurent prématurément dans le monde sous l'effet de la pollution de l'air intérieur et extérieur. La pollution atmosphérique peut raccourcir l'espérance de vie des personnes exposées et contribuer à l'apparition de maladies graves, telles que les accidents vasculaires cérébraux, cancers, maladies cardiaques, ou encore des maladies chroniques respiratoires telles que l'asthme.

Selon l'Agence Européenne de l'Environnement (rapport AEE¹, 2015), la pollution de l'air cause 430 000 décès prématurés chaque année en Europe. La plupart des citoyens européens respire un air nocif : 87 % de la population urbaine européenne a été exposée à des niveaux de particules fines (PM_{2,5}) supérieures au seuil recommandé par l'OMS en 2013.

En 2017, AIRPARIF, Association Agréée pour la Surveillance de la Qualité de l'Air (AASQA) en Île-de-France, évalue à 1.3 million les franciliens exposés à des niveaux de pollution 2 fois supérieurs à la valeur limite annuelle en dioxyde d'azote (NO₂) à proximité des axes routiers, dont un parisien sur deux².

À ce lourd bilan sanitaire, s'ajoute un coût économique et financier de 101.3 milliards d'euros (rapport du Sénat, 2015) imputable à une mauvaise qualité de l'air chaque année en France. Ce montant intègre le coût sanitaire : mortalité, hospitalisation, baisse de productivité et le coût environnemental : baisse des rendements agricoles et forestiers, dégradation du bâti et des écosystèmes, nuisances.

La région parisienne, deuxième mégapole européenne, connaît une forte pression démographique avec ses 12 millions d'habitants et un PIB par habitant le plus élevé de France. Elle centralise une forte activité économique qui favorise un trafic routier intense et des encombrements significatifs sur ses axes routiers, notamment sur son autoroute urbaine : le « boulevard périphérique ».

Ces spécificités justifient les nombreux travaux de recherche sur la pollution de l'air dans cette région depuis une dizaine d'années tels que :

- le projet européen MEGAPOLI³ (2008-2011) : *Megacities : Emissions, urban, regional and Global Atmospheric POLLution and climate effects, and Integrated tools for assesment and mitigation*. Deux campagnes de mesures ont été menées : en été 2009 et en hiver 2010. Les objectifs étaient de déterminer les niveaux et la variabilité des concentrations de composés gazeux et particulaires à Paris, ainsi que leurs sources d'émission et origines géographiques. Et aussi de déterminer l'impact de la pollution régionale sur la pollution locale. Puis développer des outils d'intégrations pour prédire les émissions polluantes des mégapoles (thèse Baudic⁴ 2016, Baklanov *et al.*, 2010, Beekmann *et al.*, 2015).

- le programme PARTICULES, une campagne de mesures d'une année complète a été menée de septembre 2009 à septembre 2010, sur cinq sites de mesures représentatifs des fonds urbains, suburbains et ruraux en Île-de-France. Les résultats de ces campagnes ont permis de constituer une base de données sur la composition chimique, les sources et origines

¹ <https://www.eea.europa.eu/fr/pressroom/newsreleases/de-nombreux-europeens-restent-exposes>

² https://www.airparif.asso.fr/_pdf/publications/bilan-2017.pdf

³ <http://megapoli.dmi.dk/>

⁴ <https://tel.archives-ouvertes.fr/tel-01546311/document>

géographiques des aérosols PM_{2,5} en région parisienne (thèse Bressi⁵ 2012, Bressi *et al.*, 2013, Bressi *et al.*, 2014).

- le projet FRANCIPOL⁶ (2009-2013) : *Impact de la pollution longue distance sur les particules et leurs précurseurs gazeux à Paris et en Île-de-France*. Ce projet avait pour objectif de caractériser les précurseurs gazeux des particules fines en déterminant leurs niveaux, leurs variabilités, leurs sources et origines géographiques. Des mesures ont été enregistrées pendant un an dans une station urbaine de fond à Paris. Ces mesures et celles du programme PARTICULES ont permis d'évaluer le modèle de prévision de qualité de l'air CHIMERE.

- des thèses : Favez⁷, 2008 (sources et processus de formation des particules), Petetin, 2014 (simulation des sources d'aérosols).

Ces travaux de recherche s'intéressent à la pollution régionale ou à la pollution de fond urbain (**Figure 1**). Pour les études épidémiologiques, ces échelles peuvent être considérées suffisantes pour associer les niveaux de pollution à des données agrégées sur l'ensemble de la population. Seulement, ces échelles ne permettent pas de mener des études plus fines, sur des groupes de populations vulnérables ou sur l'« effet cocktail » de polluants (Valari *et al.*, 2011). Or en ville, la présence de nombreuses sources d'émissions locales liées à l'activité humaine, la grande réactivité des espèces chimiques, les saisons et conditions météorologiques et la densité du bâti génèrent une dispersion des polluants près des sources et engendrent une importante variabilité spatio-temporelle de ces derniers (Harrison, 2018).

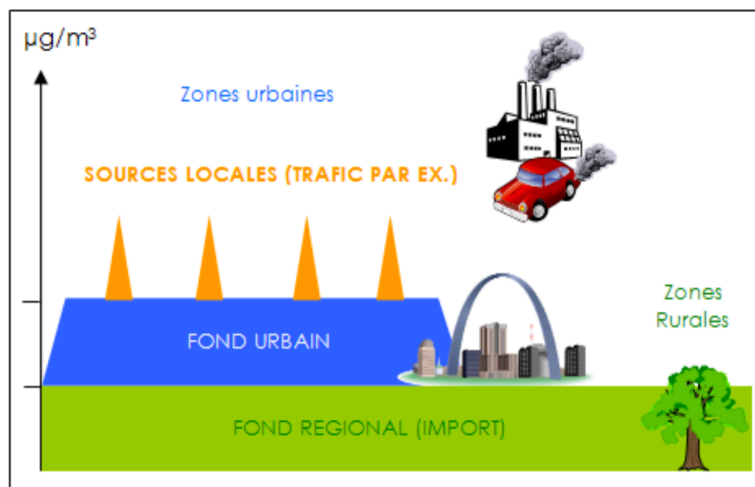


Figure 1 : Schéma de variation des concentrations de particules fines (PM_{2,5}) dans l'air ambiant. Source : AIRPARIF⁸ (rapport, 2011)

⁵<http://www.uvsq.fr/les-aerosols-fins-en-ile-de-france-chimie-sources-et-origines-geographiques-par-michael-bressi-242217.kjsp>

⁶ http://www.rst.developpement-durable.gouv.fr/IMG/pdf/Primequal_FRANCIPOL_cle0219e7.pdf

⁷ <https://www.theses.fr/2008PA077007>

⁸ https://www.airparif.asso.fr/_pdf/publications/rapport-particules-110914.pdf

Caractériser cette variabilité locale de la pollution urbaine est un enjeu majeur avec plusieurs objectifs :

- (1) réduire les incertitudes des études épidémiologiques en intégrant les spécificités du milieu urbain : variabilité spatio-temporelle des polluants, mobilité des individus (intégrer le « budget espace-temps »), « effet cocktail » des polluants entre eux,
- (2) définir des outils d'aide à la décision avec la mise en place des politiques locales de réduction des émissions à la source pour améliorer le niveau de fond et en cas de pic de pollution, préconiser un itinéraire plutôt qu'un autre aux personnes vulnérables.

En 2019, le réseau AIRPARIF dispose de quarante stations en Île-de-France qui mesurent les oxydes d'azote (NO_x), de vingt-quatre stations pour les particules dont le diamètre est inférieur à 10 µm (PM₁₀) et de treize stations pour les particules fines (PM_{2,5}). Ce réseau n'est pas suffisant pour étudier la pollution à l'échelle de la rue.

Le projet POLLUSCOPE⁹ (2016-2020) a pour objectif d'évaluer finement les risques sanitaires induits au niveau individuel grâce à un système de « crowdsensing » mobile (thèse Languille, en cours). L'objectif est d'équiper un grand nombre de citoyens de capteurs mobiles à bas coût et ainsi caractériser leur exposition.

Dans le cadre de cette thèse, nous utilisons des mesures de polluants et également des simulations de modèle afin de caractériser le plus finement possible la pollution à l'échelle du quartier dans la région Île-de-France.

Ce manuscrit de thèse s'organise en cinq chapitres suivis d'une conclusion générale qui présente une synthèse des résultats obtenus lors de ces trois années de recherche et propose des perspectives à ce travail.

Dans le chapitre 1, un état de l'art sur le sujet est présenté en ciblant les principaux polluants et les particularités de la région parisienne. Les processus physico-chimiques associés à la pollution de l'air sont rappelés. Des éléments succincts sur la qualité de l'air intérieur sont également mentionnés.

Dans le chapitre 2, une description sur les outils utilisés dans ce travail de thèse est donnée, à savoir les instruments de mesure, les simulations du modèle urbain PMSS (Parallel Micro-SWIFT-SPRAY), les mesures des stations AIRPARIF à Paris.

Dans le chapitre 3, la variabilité de l'exposition est abordée de façon expérimentale. Pour mesurer l'exposition d'un individu, on peut déployer des instruments de référence à l'intérieur des microenvironnements dans lesquels séjournent les individus : au domicile, au bureau, à l'école et sur les trajets empruntés par les individus. Un tel dispositif requiert un budget important et ne peut être mis en place que de manière ponctuelle sur une portion limitée du territoire. De plus, il exige un emploi du temps détaillé de chaque individu.

Une alternative à cela est d'équiper des volontaires de capteurs mobiles (idéalement à bas coût) fonctionnant 24 heures sur 24 (Steinle *et al.*, 2014). L'utilisation de ces capteurs est en plein essor, cependant la communauté scientifique se questionne fortement sur la fiabilité des données obtenues.

Dans ce travail de thèse, une campagne de mesures pour évaluer l'exposition aux particules d'une parisienne (en l'occurrence, l'auteurice de ces lignes) a été réalisée dans des conditions de

⁹ <https://anr.fr/Projet-ANR-15-CE22-0018>

vie « réalistes ». Trois instruments de référence ont été déployés : un instrument disponible dans sa version mobile était utilisé lors de déplacements. Les deux autres ont été positionnés à mon domicile et dans le local technique du LATMOS (Campus Pierre et Marie Curie de Sorbonne Université).

La seconde campagne de mesures s'est déroulée lors d'une journée dans le démonstrateur de Sense-City (IFSTTAR) avec des capteurs mobiles à bas coût et deux instruments de référence. Les capteurs ont été testés en effectuant des scénarios de la vie quotidienne (cuisine, bricolage, ménage, aération, bougie). Les limites de capteurs à bas coût (200 €) ont ainsi été identifiées et les caractéristiques techniques nécessaires pour établir des recommandations dans les protocoles de mesures ont été établies. Ces recommandations intègrent entre autres les contraintes liées à la mesure en mobilité et en environnement intérieur.

Initialement, des micro-capteurs avec des propriétés proches de celles des capteurs e-PM – en phase de développement au LATMOS – devaient être portés par une centaine de volontaires. En raison du retard dans la conception de ces micro-capteurs, nous avons proposé une autre stratégie que celle initialement prévue pour remonter à la caractérisation la plus fine de l'air respiré par un francilien.

Dans le chapitre 4, nous recherchons des zones géographiques dites « zones de représentativité » autour de chaque station AIRPARIF avec des propriétés semblables à l'air sondé au niveau de la station. Pour définir ces zones, nous avons utilisé les champs de concentrations horaires des NO_x et PM_{10} simulés par le modèle PMSS. C'est un modèle de dispersion urbain haute résolution (3 mètres de résolution horizontale dans notre cas) avec une couverture spatiale sur l'ensemble de Paris. Bien que le modèle PMSS soit performant, les simulations restent biaisées par rapport aux mesures de référence AIRPARIF.

Dans le chapitre 5, nous avons développé un modèle bayésien pour propager la mesure de la station dans les zones de représentativité de chaque station. L'approche bayésienne met à jour notre connaissance *a priori* (avant d'observer les données) à partir des données (« vraisemblance ») pour obtenir une connaissance *a posteriori* (après avoir observé les données) plus précise. Ainsi, les concentrations simulées par le modèle PMSS sont combinées aux mesures AIRPARIF pour obtenir les concentrations les plus probables dans les zones préalablement définies.

Chapitre 1. État de l'art : généralités sur la pollution atmosphérique

1.1 Les principaux polluants

1.1.1 Impacts sanitaires

La pollution atmosphérique est un problème de santé publique majeur. Selon l'OMS, la pollution de l'air est le principal risque environnemental avec des impacts négatifs sur la santé tels que les accidents vasculaires cérébraux, les maladies du cœur, les cancers du poumon, et les maladies respiratoires chroniques ou aiguës, incluant l'asthme¹⁰. En 2016, 4.2 millions de personnes sont mortes prématurément dans le monde sous l'effet de la pollution atmosphérique extérieure dans les zones urbaines, périurbaines et rurales. 91 % de ces morts prématurés sont recensés dans les pays à revenus faibles et modestes, et la plupart en Asie du Sud Est et dans le Pacifique occidental.

Une exposition chronique à l'ozone serait responsable de près de 500 décès par an pour causes respiratoires en France, d'après une étude de Santé Publique France¹¹ (ex-InVS).

La pollution aux particules fines PM_{2,5} est la troisième cause de mortalité en France après le tabac et l'alcool, soit 48 000 décès prématurés chaque année. Cet impact représente 9 % de la mortalité en France, avec une perte d'espérance de vie pouvant dépasser 2 ans dans les villes les plus exposées.

La pollution de l'air intérieur nuit également gravement à la santé avec près de 3 milliards de personnes dans le monde qui utilisent des combustibles à base de kérosène, de biomasse, de fuel et de charbon pour cuisiner, se chauffer et éclairer leur logement.

Jusqu'à présent, les études épidémiologiques menées sur Paris considéraient la pollution homogène à l'échelle de l'agglomération, en utilisant des données des stations de mesures AIRPARIF. Cette hypothèse d'exposition homogène engendre des incertitudes sur les fonctions dose/réponse des modèles épidémiologiques et ne permet pas de distinguer l'influence des plusieurs polluants lorsqu'ils coexistent, ce qu'il est possible de faire si nous raffinons l'échelle de l'étude, par exemple avec un modèle de qualité de l'air, tout en prenant en compte la mobilité des populations (Valari, *et al.* 2011). Pour ce faire, il est nécessaire d'intégrer la variabilité locale de la pollution urbaine.

L'OMS préconise des « valeurs guides » définies par des lignes directrices applicables dans le monde entier (lignes directrices OMS relatives à la qualité de l'air de 2005¹²). Ces valeurs guides se basent sur l'évaluation d'experts pour que le polluant n'ait pas ou peu d'impact sur

¹⁰ [http://www.who.int/news-room/fact-sheets/detail/ambient-\(outdoor\)-air-quality-and-health](http://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health)

¹¹ <http://invs.santepubliquefrance.fr/Dossiers-thematiques/Environnement-et-sante/Air-et-sante/Ozone-chaaleur-et-sante>

¹² http://apps.who.int/iris/bitstream/handle/10665/69476/WHO_SDE_PHE_OEH_06.02_fre.pdf;jsessionid=6B786591616BC95B18B079DDDC5CE638?sequence=1

la santé humaine. En 2016, 91 % de la population mondiale vivait dans des lieux où la valeur guide n'était pas respectée.

Dans la suite de ce travail, nous décrirons plus spécifiquement les polluants concernés par les lignes directrices de l'OMS à savoir : les particules en suspension dans l'air (PM), l'ozone (O_3), le dioxyde d'azote (NO_2) et le dioxyde de soufre (SO_2). Nous élargirons l'étude du NO_2 à la famille des oxydes d'azote (NO_x) qui inclut également le NO ($NO_x=NO_2+NO$).

Les particules dont le diamètre est inférieur à 10 et 2.5 μm (PM_{10} et $PM_{2.5}$ respectivement), ont un impact sanitaire même à de très faibles concentrations. On distingue les particules primaires (directement émises) et secondaires (produites par réactions chimiques ou agglomération de particules plus fines). La nocivité des particules est à la fois liée : (1) à leur taille, les particules les plus fines pénétreront plus profondément dans l'appareil respiratoire et (2) aux polluants qu'elles transportent (métaux, hydrocarbures, etc.).

Les PM_{10} et $PM_{2.5}$ sont à l'origine de maladies respiratoires et cardio-vasculaires pouvant aller jusqu'au décès. L'OMS préconise donc de ne pas dépasser une concentration de 10 $\mu g/m^3$ en moyenne annuelle pour les $PM_{2.5}$ et 25 $\mu g/m^3$ en moyenne sur 24 heures. Pour les PM_{10} , l'OMS recommande de ne pas dépasser une concentration de 20 $\mu g/m^3$ en moyenne annuelle et 50 $\mu g/m^3$ en moyenne sur 24 heures.

À de fortes concentrations, l'ozone affecte les muqueuses oculaires et engendre des problèmes respiratoires, le déclenchement de crises d'asthme, une diminution de la fonction pulmonaire pouvant entraîner la mort (Jerrett, *et al.* 2009).

L'OMS recommande de ne pas dépasser une exposition de 100 $\mu g/m^3$ en moyenne sur 8 heures. L'ozone a également des effets phytotoxiques qui impactent les principaux processus physiologiques des plantes, ce qui peut réduire la productivité des cultures agricoles.

Une exposition de longue durée au dioxyde d'azote (NO_2) entraîne une augmentation de la fréquence et de la gravité des crises d'asthme chez les adultes et favorise l'infection pulmonaire chez les enfants. Les oxydes d'azotes (NO_2 et NO) mélangés à de l'eau forment de l'acide nitrique, composant majoritaire des précipitations acides (pluies, brouillards et neige) à l'origine de l'altération des sols, de la déforestation, de l'acidification des eaux de surface (lacs, rivières) et des dommages aux écosystèmes aquatiques ce qui entraîne la disparition progressive des espèces aquatiques. L'OMS a ainsi préconisé de ne pas dépasser une exposition de 40 $\mu g/m^3$ en moyenne annuelle et 200 $\mu g/m^3$ en moyenne horaire.

Le dioxyde de soufre (SO_2) entraîne une altération de la fonction respiratoire (toux, production de mucus, asthme, bronchite chroniques, infection respiratoire), le fonctionnement des poumons et des irritations oculaires et cutanées. Mélangé à de l'eau, il réagit pour former de l'acide sulfurique, composant des précipitations acides. L'OMS recommande de ne pas dépasser une concentration moyenne de 20 $\mu g/m^3$ sur 24 heures et 500 $\mu g/m^3$ en moyenne sur 10 minutes.

Dans le rapport REVIHAAP "Review of evidence on health aspects of air pollution" publié en 2013, l'OMS interpelle sur l'« effet cocktail » ou « effet synergique » lorsque les polluants sont présents simultanément (versus polluants séparés), au niveau des tissus biologiques, entre les particules ultrafines et : (1) des métaux de transition ou (2) des composés organiques volatils (COV).

1.1.2 Seuils réglementés par la législation européenne et française

L'ozone (O₃), les particules : PM₁₀ et PM_{2,5}, les oxydes d'azote (NO_x) (en particulier, le dioxyde d'azote NO₂), le dioxyde de soufre (SO₂) sont réglementés en raison de leur impact sur la santé et sur l'écosystème soulignés par l'OMS. Ils sont surveillés 24 heures sur 24 et 7 jours sur 7 par les Associations Agréées pour la Surveillance de la Qualité de l'Air en France, en Europe grâce à la mise en place d'un programme de surveillance européen EMEP (European Monitoring and Evaluation Programme). Aux Etats-Unis, le Clean Air Act (1970) est une loi qui charge l'Agence de Protection de l'Environnement Américaine (Environmental Protection Agency, EPA¹³) de protéger et améliorer la qualité de l'air des Etats-Unis. Un nouvel observatoire mondial de la pollution de l'air : Global Urban Air Pollution Observatory – GUAPO¹⁴ diffuse les bonnes pratiques à travers le monde.

La stratégie communautaire de surveillance de la qualité de l'air se base sur la directive européenne (2008/50/CE) du 21 mai 2008 et sur la directive n° 2004/107/CE du 15 décembre 2004. Les directives européennes sont transposées dans la réglementation française et peuvent être déclinées à l'échelle nationale et locale avec la définition des objectifs de qualité tels qu'on les retrouve sur le site d' AIRPARIF¹⁵.

- **Valeur limite** : niveau à atteindre dans un délai donné et à ne pas dépasser, et fixé sur la base des connaissances scientifiques afin d'éviter, de prévenir ou de réduire les effets nocifs sur la santé humaine ou sur l'environnement dans son ensemble.
- **Valeur cible** : niveau à atteindre, dans la mesure du possible, dans un délai donné, et fixé afin d'éviter, de prévenir ou de réduire les effets nocifs sur la santé humaine ou l'environnement dans son ensemble.
- **Objectif de qualité** : niveau à atteindre à long terme et à maintenir, sauf lorsque cela n'est pas réalisable par des mesures proportionnées, afin d'assurer une protection efficace de la santé humaine et de l'environnement dans son ensemble.
- **Niveau critique** : niveau fixé sur la base des connaissances scientifiques, au-delà duquel des effets nocifs directs peuvent se produire sur certains récepteurs, tels que les arbres, les autres plantes ou écosystèmes naturels, à l'exclusion des êtres humains.
- **Seuil d'information et de recommandation** : niveau au-delà duquel une exposition de courte durée (1 heure à 1 jour) présente un risque pour la santé humaine de groupes particulièrement sensibles au sein de la population et qui rend nécessaire l'émission d'informations immédiates et adéquates à destination de ces groupes et des recommandations pour réduire certaines émissions.
- **Seuil d'alerte** : niveau au-delà duquel une exposition de courte durée présente un risque pour la santé de l'ensemble de la population ou de dégradation de l'environnement, justifiant l'intervention de mesures d'urgence.

Le **tableau 1** indique l'ensemble de ces valeurs pour les oxydes d'azote (NO_x), le dioxyde d'azote (NO₂), le dioxyde de soufre (SO₂), les particules PM₁₀ et PM_{2,5} et l'ozone (O₃). Une nouvelle loi est en cours de préparation pour réglementer les PM1 encore plus nocives.

¹³ <https://www.epa.gov/>

¹⁴ <http://www.who.int/phe/news/e-News96Dec2017.pdf>

¹⁵ <https://www.airparif.asso.fr/reglementation/normes-francaises>

a)

	Valeurs limites	Objectif de qualité	Seuils de recommandation et d'information	Seuils d'alerte	Niveau critique
NO ₂	Moyenne annuelle : depuis le 01/01/10, 40 µg/m ³ Moyenne horaire : depuis le 01/01/10, 200 µg/m ³ à ne pas dépasser plus de 18 heures par an.	Moyenne annuelle : 40 µg/m ³	Moyenne horaire : 200 µg/m ³	Moyenne horaire : 400 µg/m ³ dépassé sur 3 heures consécutives. 200 µg/m ³ si dépassement de ce seuil la veille, et risque de dépassement de ce seuil le lendemain.	
NO _x					Moyenne annuelle (pour la protection de la végétation) : 30 µg/m ³ (équivalent NO ₂)
PM ₁₀	Moyenne annuelle : depuis le 01/01/05, 40 µg/m ³ Moyenne journalière : depuis le 01/01/05, 50 µg/m ³ à ne pas dépasser plus de 35 jours par an.	Moyenne annuelle : 30 µg/m ³	Moyenne journalière : 50 µg/m ³	Moyenne journalière : 80 µg/m ³	
SO ₂	Moyenne journalière : 125 µg/m ³ à ne pas dépasser plus de 3 jours par an. Moyenne horaire : depuis le 01/01/05, 350 µg/m ³ à ne pas dépasser plus de 24 heures par an.	Moyenne annuelle : 50 µg/m ³	Moyenne annuelle : 300 µg/m ³	Moyenne horaire sur 3 heures consécutives : 500 µg/m ³	Moyenne annuelle et hivernale (pour la protection de la végétation) : 50 µg/m ³

b)

	Objectif de qualité	Seuils de recommandation et d'information	Seuils d'alerte	Valeurs cibles
O ₃	Seuil de protection de la santé , pour le maximum journalier de la moyenne sur 8 heures : 120 µg/m ³ pendant une année civile. Seuil de protection de la végétation , AOT 40* de mai à juillet de 8 h à 20 h : 6 000 µg/m ³ .h	Moyenne horaire : 180 µg/m ³	Seuil d'alerte pour une protection sanitaire pour toute la population , en moyenne horaire : 240 µg/m ³ sur 1 heure Seuils d'alerte pour la mise en œuvre progressive de mesures d'urgence , en moyenne horaire : 1 ^{er} seuil : 240 µg/m ³ dépassé pendant trois heures consécutives. 2 ^{ème} seuil : 300 µg/m ³ dépassé pendant trois heures consécutives. 3 ^{ème} seuil : 360 µg/m ³	Seuil de protection de la santé : 120 µg/m ³ pour le max journalier de la moyenne sur 8 h à ne pas dépasser plus de 25 jours par année civile en moyenne calculée sur 3 ans. Cette valeur cible est appliquée depuis 2010. Seuil de protection de la végétation : AOT 40 ¹⁶ de mai à juillet de 8 h à 20 h : 18 000 µg/m ³ .h en moyenne calculée sur 5 ans. Cette valeur cible est appliquée depuis 2010.

¹⁶ AOT 40 (exprimé en µg/m³.heure) signifie la somme des différences entre les concentrations horaires supérieures à 80 µg/m³ et le seuil de 80 µg/m³ durant une période donnée en utilisant uniquement les valeurs sur 1 heure mesurées quotidiennement entre 8 heures et 20 heures. (40 ppb ou partie par milliard=80 µg/m³).

c)

Valeurs limites	Objectif de qualité	Valeur cible	Objectif de réduction de l'exposition par rapport à l'IEM 2011* , qui devrait être atteint en 2020		Obligation en matière de concentration relative à l'exposition qui doit être respectée en 2015	
			Concentration initiale	Objectif de réduction		
PM_{2,5}	Moyenne annuelle : depuis le 01/01/15, 25 µg/m ³	Moyenne annuelle : 10 µg/m ³	Moyenne annuelle : 20 µg/m ³	<= à 8,5 µg/m ³	0%	20 µg/m ³ pour l'IEM 2015 ¹⁷
				>8,5 et <13 µg/m ³	10%	
				>=13 et <18 µg/m ³	15%	
				>=18 et <22 µg/m ³	20%	
				>= à 22 µg/m ³	Toute mesure appropriée pour atteindre 18 µg/m ³	

Tableau 1 : Seuils à ne pas dépasser pour les (a) dioxydes et oxydes d'azote (NO₂ et NO_x), les particules PM₁₀ et le dioxyde de soufre (SO₂), (b) l'ozone (O₃) et (c) les particules fines PM_{2,5} selon la réglementation française. Source : AIRPARIF¹⁸

1.2 Cas de la région parisienne

1.2.1 Dépassements des seuils

La pollution de l'air est un problème sanitaire majeur y compris en région parisienne. L'exposition chronique à la pollution de l'air en Île-de-France est responsable annuellement de l'ordre de 1 400 décès anticipés chez les adultes, soit 4 % de la mortalité totale correspondant à une perte de 6 mois d'espérance de vie à l'âge de 30 ans. Cette pollution francilienne a également un coût économique non négligeable de 2.3 milliards d'euros (Host *et al.*¹⁹, 2012). Un rapport du Sénat chiffre le coût économique de la pollution de l'air au niveau national à 101.3 milliards d'euros par an (rapport²⁰, 2015).

Un historique des dépassements de seuils en Île-de-France est disponible en ligne²¹ pour le dioxyde d'azote (NO₂), l'ozone (O₃), les particules PM₁₀ et le dioxyde de soufre (SO₂). De janvier 2015 à janvier 2018, 33 jours de dépassements de seuils en PM₁₀ ont été recensés dans la région Île-de-France avec 22 jours de dépassements du seuil et 11 jours de dépassement du seuil d'alerte. La **Figure 2** représente la répartition des jours de dépassements de seuils d'information et d'alerte pour les PM₁₀ de janvier 2015 à janvier 2018.

¹⁷ IEM 2015 : Indicateur d'exposition moyenne de référence, correspondant à la concentration moyenne annuelle en µg/m³ sur les années 2013, 2014 et 2015.

¹⁸ <https://www.airparif.asso.fr/reglementation/normes-francaises>

¹⁹ http://www.ors-idf.org/dmdocuments/2012/Synthese_erpurs_EIS_trafic.pdf

²⁰ http://www.senat.fr/fileadmin/Fichiers/Images/commission/enquete/pollution_air/Synthese_CE_Pollution_de_l_air.pdf

²¹ <https://www.airparif.asso.fr/alertes/historique>

Le nombre de dépassement de seuils était en baisse en 2017 par rapport à 2015. 94 % des dépassements des seuils d'information et d'alerte ont lieu en période hivernale (entre décembre et mars) de janvier 2015 à janvier 2018 avec un maximum en janvier.

De janvier 2015 à janvier 2018, 13 jours de dépassements du seuil d'information ont été détectés et aucun dépassement du seuil d'alerte. En 2017, le nombre de jours de dépassement était à la hausse avec 6 jours de dépassements, alors qu'en 2016 on dénombrait 3 jours de dépassement et 4 jours en 2015. 92 % de ces jours de dépassement de seuil pour l'ozone ont lieu en été.

Un seul jour de dépassement du seuil d'information a été identifié (en décembre 2016) pour le dioxyde d'azote (NO₂), et aucun jour de dépassement de seuil n'a été enregistré pour le dioxyde de soufre (SO₂).

a) Particules PM₁₀

b) Ozone (O₃)

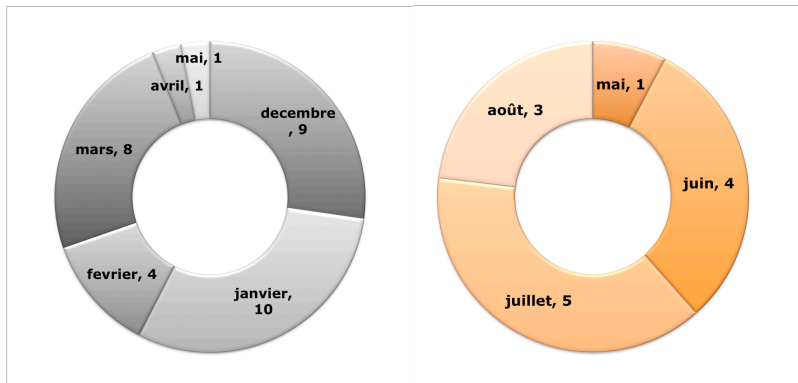


Figure 2 : Répartition des dépassements de seuils d'information et d'alerte pour (a) les particules PM₁₀ et (b) l'ozone (O₃), de janvier 2015 à janvier 2018. Le chiffre indique le nombre de jours de dépassement de seuils pour les mois de l'année concernés. Figure réalisée à partir des données AIRPARIF.

Ces niveaux élevés de pollution lors des jours de dépassement de seuils s'étendent sur plus de 100 km² sur la région Île-de-France et touchent ainsi plus de 10 % de la population d'un département d'Île-de-France.

1.3 Processus physico-chimiques

La pollution de l'air en ville s'explique par les processus physico-chimiques : émissions, transport, mélange turbulent, transformation chimique et dépôt, agissant sur différentes échelles spatiales et temporelles (**Figure 3**). Le cycle émission-transport-transformation-dépôt explique l'origine des polluants (sources naturelles et anthropiques), leur évolution dans l'atmosphère (transport et transformation) et leur élimination (transformation et dépôt).

On distingue les polluants primaires des polluants secondaires. Les polluants primaires comme le dioxyde de soufre (SO₂), sont émis directement dans l'atmosphère par des sources et transportés horizontalement et verticalement à différentes altitudes sous l'effet des mouvements d'air et sur des distances plus ou moins grandes selon leur durée de vie. Au cours de ce transport, une partie de ces polluants se transforme en polluants secondaires comme l'ozone (O₃). Les polluants secondaires ne sont pas directement émis par des sources

mais résultent de réactions chimiques entre plusieurs espèces et de réactions d'oxydations supérieures. En effet, l'atmosphère est un milieu oxydant avec une forte présence d'oxygène. La plupart de ces réactions sont initiées par le rayonnement solaire. On parle dans ce cas de réactions photochimiques.

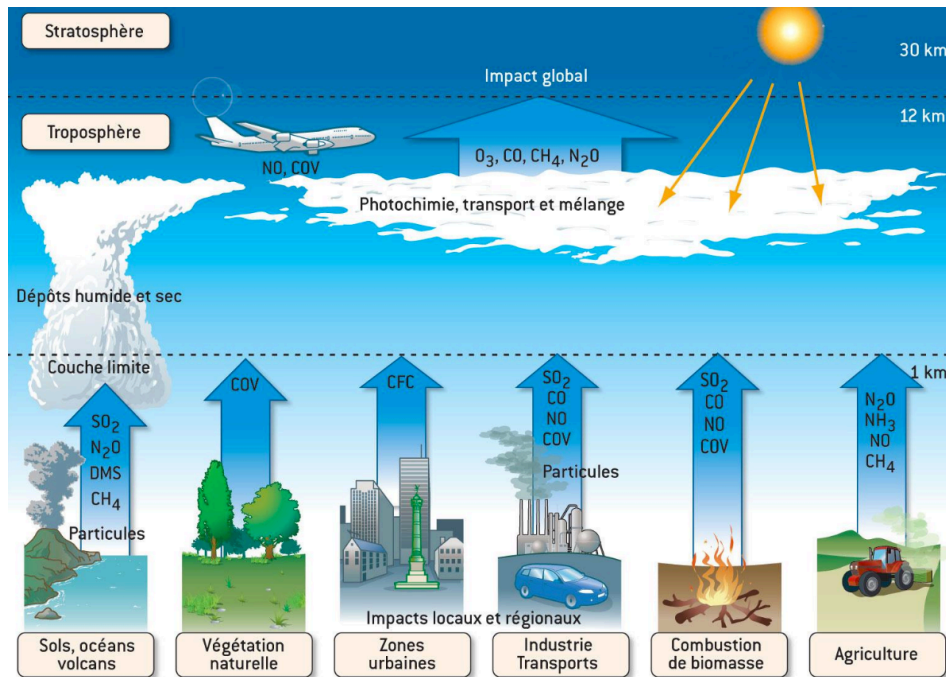


Figure 3 : Représentation schématique des principaux processus de la chimie atmosphérique illustrant le cycle : émission en surface des polluants par différentes sources naturelles ou anthropiques, diffusion verticale, transport et transformations photochimiques dans l'atmosphère, dépôt sec par interaction directe avec la surface ou dépôt humide par les précipitations. Source : Atmosphère, océan et climat (R. Delmas, *et al.*, 2007)

Selon leur durée de vie dans l'atmosphère, les polluants auront un impact à des échelles spatiales plus ou moins grandes (**Figure 4**). On distingue l'échelle locale (< 1 km), l'échelle régionale (1 à 100 km) et l'échelle globale (> 100 km). La durée de vie des polluants dans l'atmosphère est très variable et dépend de leurs propriétés physico-chimiques. Par exemple, les polluants solubles peuvent être rapidement éliminés par lessivage dans les précipitations. Les espèces à courte durée de vie (quelques minutes à quelques heures) sont très réactives et auront un impact local près des sources de rejets (trafic automobile et industries). La qualité de l'air résultante, à cette échelle, sera fortement liée aux conditions météorologiques et aux phénomènes de petite échelle résultant de la densité des bâtiments en ville. La dispersion de ces émissions est difficile avec la formation de tourbillons dans les rues « canyons » où le rapport entre la hauteur du bâti et la largeur de la rue est supérieur à 1. Les espèces dont la durée de vie est supérieure à la journée pourront être transportées sur de grandes distances et être dispersées à l'échelle régionale ou planétaire. Les espèces à longue durée de vie dans la troposphère (> 5 à 7 ans) peuvent également se déplacer dans la stratosphère, la couche au-dessus de la troposphère. Ces espèces peuvent avoir un impact global sur le climat.

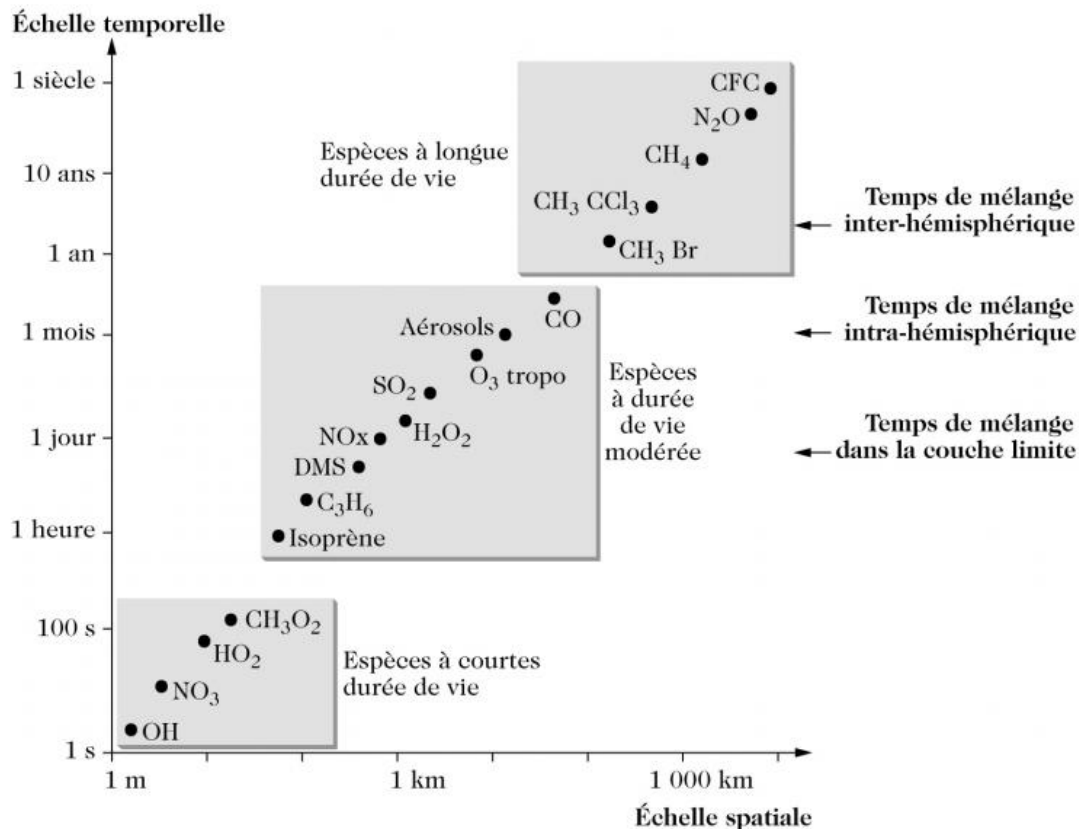


Figure 4 : Comparaison des temps de vie et des échelles spatiales pour quelques espèces.
Source : Physique et chimie de l'atmosphère (R. Delmas, 2005)

Les oxydes d'azote (NO_x) et le dioxyde de soufre (SO₂) ont un temps de résidence moyen dans l'atmosphère d'un jour. Ces polluants réagiront à proximité des sources d'émissions. L'ozone (O₃) a un temps de résidence de une à quelques semaines l'été et de quelques mois l'hiver. L'ozone peut se former et agir sur de plus longues distances et être transporté dans des régions plus lointaines.

1.3.1 Sources d'émissions

Les émissions dans l'atmosphère proviennent à la fois de sources naturelles liées à l'activité naturelle et des sources anthropiques rejetées par l'activité humaine. La contribution des émissions naturelles est plus ou moins importante d'un polluant à l'autre. Par exemple, à l'échelle mondiale, la végétation est responsable d'environ 90 % des émissions de composés organiques volatils (COV) contre 10 % pour les activités humaines. Alors que les émissions d'oxydes d'azote (NO_x) sont principalement dues à l'activité humaine avec 80 % des émissions anthropiques telles que le trafic routier ou les feux de biomasse.

Les sources naturelles principales sont :

- les éruptions volcaniques qui peuvent émettre une grande quantité de gaz (dioxyde de soufre : SO₂, oxydes d'azote : NO_x, ...) et particules (métaux lourds : éléments constitutifs de la croûte terrestre). Par exemple, l'Etna émet en moyenne 3000 tonnes de SO₂ par jour et jusqu'à 10 000 tonnes en cas de forte activité.

- l'érosion éolienne, c'est à dire l'action du vent sur la surface terrestre et l'océan, qui conduit à l'émission de particules. Les particules en suspension émises par les poussières désertiques du Sahara atteignent souvent l'Europe. Des particules de sels marins (« sea-salt ») sont également générées au niveau des océans.
- les émissions biogéniques, c'est à dire les gaz produits par réactions biologiques dans l'océan, les sols et la végétation terrestre suite à la dégradation de la matière organique. Il s'agit entre autres de l'émission des COV due à la photosynthèse.
- la foudre qui produit des oxydes d'azote (NO_x),
- la végétation qui émet des composés organiques volatils (COV) dus à la photosynthèse ... et les grains de pollen – allergisants, dispersés par l'action des insectes du vent.

Les sources anthropiques peuvent être décomposées en plusieurs secteurs d'activité et le nombre de secteurs varie selon l'agence qui effectue l'inventaire. L'inventaire d'AIRPARIF s'appuie sur les méthodologies du Pôle de Coordination des Inventaires Territoriaux (PCIT) et propose 9 secteurs anthropiques :

- la production d'énergie (centrales thermiques, installations d'extraction du pétrole, ...),
- l'industrie manufacturière avec l'émission de polluant lors des procédés de production et le chauffage des locaux des entreprises,
- les chantiers et carrières, lors de la construction de bâtiments, travaux publics et carrière,
- le traitement des déchets : incinération des déchets et centre de stockage,
- le résidentiel et tertiaire (à l'intérieur des bâtiments) qui proviennent principalement de la climatisation des bâtiments, des appareils de combustions fixes et des sources de pollution intérieure (peintures, solvants, ...),
- le trafic routier où les émissions sont liées à la combustion des combustibles fossiles, à l'évaporation de l'essence et aux abrasions (freins, pneus, route, caténaires),
- le trafic ferroviaire et fluvial,
- les plateformes aéroportuaires avec l'émission des avions et l'activité au sol,
- les feux de forêt qui émettent des oxydes d'azotes (NO_x) et des métaux lourds, particules,
- l'agriculture avec la croissance de l'agriculture intensive et l'utilisation massive d'engrais, de pesticides (...) et la décomposition de matières organiques.

Si l'on considère les émissions anthropiques en région parisienne, la part du trafic routier est plus de deux fois plus importante pour les émissions des oxydes d'azote (NO_x) que pour les particules PM₁₀ (56 vs 23 %, **Tableau 2**). Pour les PM₁₀, le secteur résidentiel et tertiaire occupe une part équivalente à celle du trafic routier (23 %). Les émissions des particules PM_{2,5} sont à 27 % associées au trafic routier et 49 % au secteur tertiaire.

Secteurs	NO _x	PM ₁₀	PM _{2.5}
Production d'énergie	5%	1%	1%
industrie manufacturière	6%	4%	2%
chantiers et carrières	3%	16%	10%
traitement des déchets	3%	1%	1%
résidentiel et tertiaire	15%	33%	49%
trafic routier	56%	23%	27%
trafic ferroviaire et fluvial	1%	4%	2%
plateformes aéroportuaires	8%	1%	1%
agriculture	2%	17%	6%

Tableau 2 : Répartition des émissions par secteur d'activité pour les (a) NO_x (en équivalent NO₂), (b) en Île-de-France. Source : AIRPARIF (inventaire sur l'année 2015²²)

Notons que les épisodes de pollution particulaire sont majoritairement dus à de l'import de particules secondaires provenant d'Europe continentale. Les sources locales, majoritairement combustion de biomasse et transport routier ne contribuent qu'au quart des niveaux mesurés (thèse Bressi, 2012).

1.3.2 Transformations chimiques

Dans un contexte de pollution urbaine aux échelles locales et régionales, les principaux mécanismes chimiques mis en jeu sont la photochimie avec la production d'ozone troposphérique et la formation d'aérosols secondaires.

1.3.2.1 Pollution photo-oxydante, formation de l'ozone

L'ozone est un polluant secondaire formé à la suite de plusieurs réactions chimiques. La production d'ozone est à l'origine d'une pollution photo-oxydante issue de processus d'oxydation des composés organiques volatils (COV) et du monoxyde de carbone (CO) en présence des oxydes d'azote (NO_x), et du rayonnement solaire (**Figure 5**).

²² https://www.airparif.asso.fr/_pdf/publications/inventaire_emissions_idf_2015_20190418.pdf

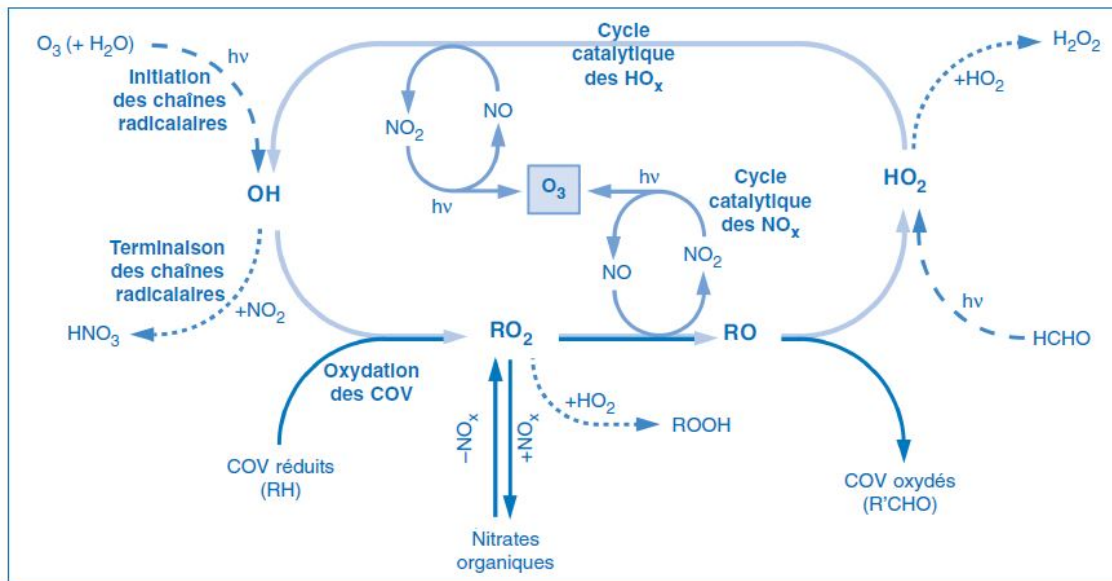
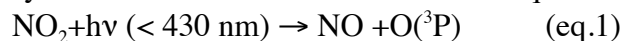


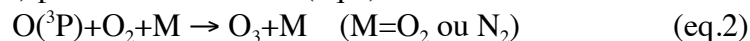
Figure 5 : Représentation schématique des principaux mécanismes impliqués dans la formation d'ozone troposphérique. Source : Camredon *et al.* (2007)

L'ozone se forme suite à la dissociation du NO_2 par réaction photochimique : une liaison est cassée en présence du rayonnement solaire suivant la réaction eq.1.



avec $\text{O}(^3\text{P})$: le niveau fondamental de O

L'oxygène libre nouvellement formé va réagir avec le dioxygène de l'air en présence d'un troisième corps (M) pour former l'ozone (eq.2).



L'ozone formé est immédiatement consommé par le monoxyde d'azote (NO) émis par les émissions dues au trafic automobile. Lors de cette réaction, du dioxyde d'azote (NO_2) est reformé (eq.3).



Un équilibre photochimique s'établit entre le dioxyde d'azote (NO_2), le monoxyde d'azote (NO) et l'ozone (O_3). Ce cycle de réaction, « cycle de Chapman », n'est pas suffisant pour expliquer la production nette d'ozone car en présence de NO, l'ozone est consommé et le bilan est nul.

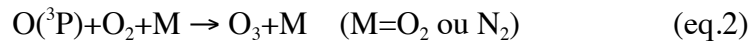
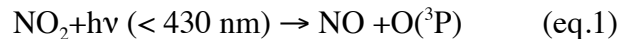
Deux régimes chimiques caractérisent l'équilibre entre l'ozone, les oxydes d'azote (NO_x) et les composés organiques volatils (COV). Lorsque la teneur en NO_x est faible on parle d'un « régime limité en NO_x » ou « régime saturé en COV », propice à la production d'ozone. À l'inverse, lorsque les niveaux en NO_x sont élevés, on parle de « régime saturé en NO_x » ou « régime limité en COV » propice à la destruction d'ozone. La transition entre les deux régimes est de l'ordre de quelques ppb de NO_x (1 ppb est équivalent à $1.88 \mu\text{g}/\text{m}^3$ pour le NO_2 et $1.25 \mu\text{g}/\text{m}^3$ pour le NO).

« Régime limité en NO_x » ou « régime saturé en COV » : production d'ozone (Seinfeld et Pandis, 1998)

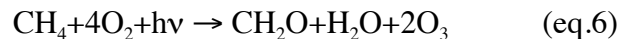
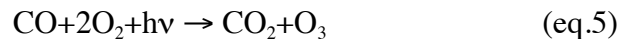
Ce régime est caractéristique des milieux ruraux. Le rapport COV/ NO_x est supérieur à 15.

La production nette d'ozone a lieu lorsque le cycle de Chapman est perturbé par l'ajout de composés organiques volatils (COV) ou de monoxyde de carbone (CO). En effet, l'oxydation

de COV et CO par le radical OH conduit à la formation de radicaux peroxydes (RO₂ et HO₂) qui permettent de convertir NO en NO₂ sans consommer d'ozone (eq.4).



Les équations bilan des chaînes d'oxydation du monoxyde de carbone et du méthane, conduisant à la production d'ozone sont les suivantes :



La formation de peroxydes domine les réactions de terminaison suivant la réaction (eq.7) :



Les niveaux d'ozone augmentent avec ceux des NO_x de façon quasi linéaire et sont peu perturbés par les variations des concentrations en COV.

« Régime saturé en NO_x » ou « régime limité en COV » : destruction d'ozone

Ce régime est caractéristique des zones urbaines. Les concentrations en NO_x sont supérieures à quelques ppb. Le rapport COV/NO_x est inférieur à 4.

La formation d'acide nitrique domine les réactions de terminaison en raison de la forte teneur en NO_x (eq.8) et entraîne une perte du radical OH. Les niveaux d'ozone diminuent lorsque les concentrations en NO_x augmentent.



En région parisienne on se trouve en général dans le « régime saturé en NO_x ». La vitesse de production d'ozone est donc lente et les maxima n'excèdent que rarement la centaine de ppb. Des épisodes estivaux de pollution à l'ozone peuvent avoir lieu en cas de recirculation des masses d'air au-dessus de la zone urbaine. Lorsqu'on s'éloigne du centre de Paris et qu'on se dirige vers la périphérie, la parcelle d'air passe progressivement d'un régime saturé en NO_x à un régime limité en NO_x. La diminution en NO_x (diminution des émissions de trafic) et l'augmentation du rapport COV/NO_x (émissions complémentaires de COV par la végétation) induisent une rapide augmentation de la vitesse de production d'ozone. On retrouve ainsi fréquemment des niveaux élevés d'ozone à plusieurs dizaines de kilomètres des zones urbaines. Le vent (direction et intensité) et la hauteur de la couche limite (section 1.1.3.4) déterminent l'intensité, la forme et la localisation du panache.

1.3.2.2 Aérosols

Un aérosol atmosphérique est « une particule solide ou liquide suspendue dans l'air à l'exception des gouttelettes nuageuses » (World Climate Research Programme : WCRP). Il s'agit de particules de quelques nanomètres à plusieurs micromètres de diamètre.

Les aérosols primaires (dont le diamètre est supérieur à 0.001 µm) sont directement émis dans l'atmosphère par les activités humaines (usure de pneu, pots catalytiques, feux de biomasse) ou des sources naturelles (particules minérales, sels de mer, cendre volcanique, aérosols biologiques primaires) et sont remis en suspension.

Les aérosols secondaires (dont le diamètre est inférieur à 1 µm) se forment suite à la conversion de gaz en particules qui s'effectue si les conditions thermodynamiques sont

favorables. Les gaz chauds émis par les différentes sources refroidissent au fur et à mesure qu'ils s'élèvent dans l'atmosphère. Ils peuvent donc subir des processus de condensation ou nucléation hétérogène, pour former des nouveaux aérosols ou transformer des aérosols existants²³.

La composition chimique des aérosols est très complexe et dépend de nombreux facteurs tels que la proximité des sources principales. L'aérosol peut être constitué d'une fraction organique et du carbone suie (« Black Carbon ») et d'une fraction inorganique essentiellement composée de minéraux tels que les nitrates, les sulfates, l'ammonium, le calcium.

1.3.3 Transport horizontal

Les polluants sont dispersés par : (1) le transport horizontal par l'advection selon le vent et (2) le brassage vertical dû aux phénomènes de turbulence dans la couche limite atmosphérique (voir section 1.1.3.4).

Un vent faible entraîne l'accumulation des polluants et explique certains épisodes de pollution locale. Alors qu'un vent fort diffuse la pollution locale et transporte de la pollution sur de plus ou moins longues distances. Un vent fort amène aussi de la pollution d'autres zones polluées. La dispersion et le transport des polluants dépendent de la vitesse et de la direction du vent. Le vent peut être perturbé par les obstacles environnants et aussi par la rugosité des terrains. La rugosité des terrains étant la perturbation des écoulements d'air en fonction des micro-reliefs (bâtiments et arbres). Ainsi, le vent sera davantage freiné lorsqu'il passe au-dessus d'un bâtiment qu'au-dessus d'un parc car la rugosité est plus faible pour un parc.

La topographie d'un site influence l'écoulement de l'air. En effet, les bords de mer et les vallées peuvent influencer la dispersion des polluants. À Marseille, l'effet de brise de mer concentre les polluants sur la côte en milieu de journée. Et à Grenoble, les effets de la brise de montagne et de canalisation par la vallée sont capables d'advecter des polluants sur les massifs montagneux la journée et sur la vallée la nuit. La région Île-de-France est une plaine éloignée de la mer et principalement soumise à un régime météorologique moyen : des dépressions venant du sud ou de l'ouest avec des fronts qui favorisent une rapide dispersion des polluants dans la région et des anticyclones qui accumulent les polluants près de la surface, sur une large zone.

1.3.4 Mélange dans la couche limite

La couche limite atmosphérique (CLA) est la première couche atmosphérique et comprise entre 0 et 2 km au-dessus du sol. La CLA est directement soumise aux effets de la surface terrestre. La CLA est caractérisée par des transferts de quantité de mouvement, vapeur d'eau, de polluants et chaleur entre la surface et l'atmosphère libre. L'épaisseur de la CLA dépend de la rugosité de la surface. Elle varie au cours de la journée (**Figure 6**) car elle dépend également de la stabilité de l'atmosphère. On distingue la CLA stable de la CLA instable ou « convective ».

²³ <http://omer7a.obs-mip.fr/mallette/fiches/Chimie-des-aerosols>

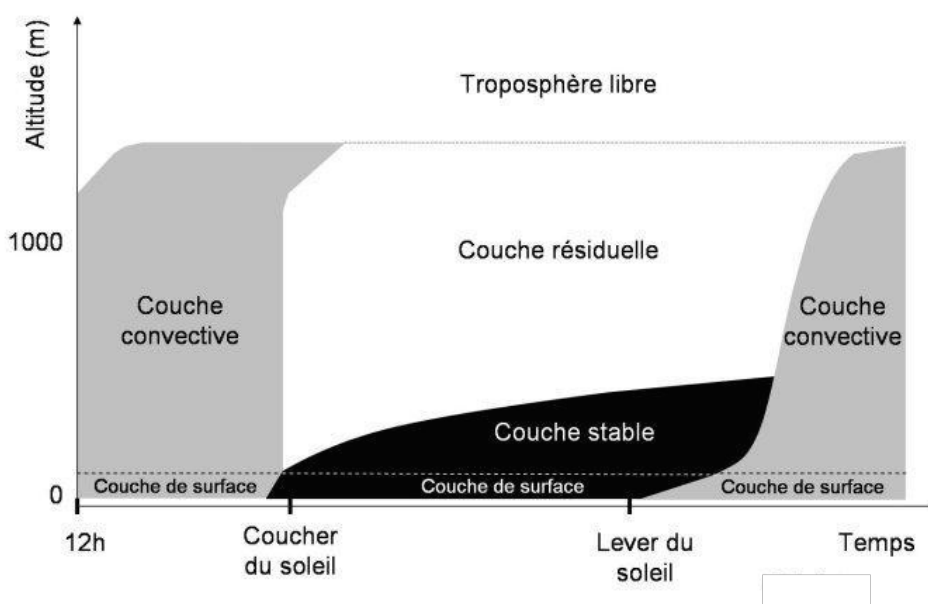


Figure 6 : Évolution de la stabilité de la couche limite atmosphérique au cours d'une journée²⁴

Dans la CLA instable la température décroît plus rapidement avec l'altitude que le gradient adiabatique. Le sol est plus chaud que l'air et surchauffe l'air qui se trouve au-dessus. L'air surchauffé, de moins en moins dense, s'élève et d'importants phénomènes de turbulence thermique sont créés. L'épaisseur de la CLA instable varie de 1 à 2 km. Plus la CLA est épaisse et plus les structures turbulentes prennent de l'ampleur et diffusent les émissions. Pour un même niveau d'émission, on retrouvera des concentrations deux fois plus faibles avec une CLA deux fois plus épaisse. Les mouvements verticaux rapides (de l'ordre de 15 minutes) homogénéisent les forts gradients verticaux. Une fois les masses d'air chaudes arrivées au sommet de la CLA, elles sont arrêtées par l'atmosphère libre (thermiquement stable) mais y pénètrent partiellement avant de redescendre par gravité dans la CLA. Au cours de cette redescente, les masses d'air entraînent avec elles des masses d'air de l'atmosphère libre. Ce « phénomène d'entraînement » explique le développement de la CLA et génère un flux qui tend à la réchauffer.

La CLA est particulièrement instable le matin avec le réchauffement du sol par le soleil. Les épisodes estivaux de pollution à l'ozone sont favorisés par le développement de la CLA instable où les oxydes d'azote (NO_x), précurseurs de l'ozone (O_3), sont dilués.

Dans la CLA stable, la température décroît moins rapidement avec l'altitude que le gradient adiabatique : l'atmosphère stable est stratifiée. C'est le cas du soir ou en hiver en cas de régime anticyclonique (forte pression, faible vent, ciel dégagé). L'épaisseur de la CLA stable varie entre quelques dizaines de mètres et quelques centaines de mètres.

Le sol est plus froid que l'air au-dessus. Les forces de flottabilité empêchent l'ascension des masses d'air, les masses d'air plus chaudes étant au-dessus des plus froides. Les mouvements verticaux des particules d'air sont inhibés, ce qui empêche le développement de la turbulence et donc le développement de la CLA. Lorsqu'un anticyclone stationne pendant plusieurs jours

²⁴http://sup.upstlse.fr/uved/Ozone/BasesScientifiques/projet/site/html/CoucheLimite_StratificationStabilite_2.html

au-dessus d'une zone géographique, il joue le rôle de couvercle et bloque les polluants. L'air froid descend et la pollution atmosphérique reste bloquée dans les basses couches près du sol, surplombé en altitude par l'air plus chaud, on parle alors d'inversion thermique. En hiver, cette situation est caractéristique des épisodes de pollutions aux polluants primaires (oxydes d'azote (NO_x), monoxyde de carbone (CO), particules primaires) pouvant être accompagnés de brouillard (**Figure 7**).

La dispersion des polluants dans la CLA stable dépend de :

- (1) la vitesse du vent : plus les vents sont forts et moins le phénomène d'accumulation se produit,
- (2) la stratification verticale : plus le gradient vertical de température potentielle est fort, plus les mouvements verticaux et la turbulence sont inhibés.



Figure 7 : Exemple d'un épisode de pollution anticyclonique à Paris, Notre Dame, le 13 mars 2014 (à gauche) versus un jour sans pollution particulaire le 21 octobre 2013 (à droite)²⁵

1.3.5 Dépôts humides (précipitation) et dépôts secs

Les dépôts humides par précipitation lessivent une partie des polluants présents dans l'atmosphère. Ce lessivage a lieu pour les polluants gazeux qui sont solubles dans l'eau et pour les particules qui entrent en collision avec les gouttes de pluie. Un polluant sera d'autant plus lessivé qu'il est soluble dans l'eau. Si les espèces chimiques en présence sont des acides forts (acide sulfurique (H_2SO_4), acide nitrique (HNO_3), acide chlorhydrique (HCl)), il peut en résulter des précipitations acides (voir section 1.1.1.1).

Les principaux processus qui mènent à des dépôts secs pour les particules sont la sédimentation qui est l'effet de gravité terrestre sur les particules, les impacts par inertie ou l'interception. Pour les polluants gazeux ou particuliers il s'agit de la diffusion.

Les polluants atmosphériques peuvent se déposer à la surface des bâtiments (dépôt de particules sur les revêtements du bâtiment, thèse Abadie²⁶ 2005), sur la végétation, sur les sols et les eaux de surface par des processus « secs », qui ne dépendent pas des précipitations. Ces dépôts secs et humides ont des effets néfastes sur l'environnement (voir section 1.1.1.1). Des polluants émis par le trafic routier, à proximité des voies de transport en Île-de-France, se déposent sur les espaces agricoles, maraîchers et horticoles (Petit, 2013).

²⁵ <https://www.bfmtv.com/planete/avant-apres-paris-brumes-pollution-732424.html>

²⁶ <https://tel.archives-ouvertes.fr/tel-00010527/document>

1.4 Qualité de l'air intérieur

Cette section, bien que brève, permet d'appréhender les grandes lignes de ce sujet complexe qu'est la qualité de l'air intérieur.

Il y a une vingtaine d'années, des études au Canada et aux Etats-Unis ont montré que les concentrations extérieures en particules ne donnent pas d'indication sur le niveau de concentration à l'intérieur (Pellizzari *et al.*, 1999). De nombreuses études ont montré que l'air à l'intérieur des maisons, des bâtiments pouvait être fortement pollué (Owen, 1992 ; Stetzenbach, 2004 ; Koivisto, 2019 ; Chalvatzaki,²⁷ 2019, thèse Abadie, 2005). Or nous passons 80 % à 90 % de notre temps à l'intérieur des locaux. De plus ce sont les personnes les plus sensibles (personnes âgées, enfants, personnes souffrantes) qui sont les plus exposées à cette pollution.

Si l'on s'intéresse à l'exposition, à savoir au cumul des concentrations auquel un individu a été exposé au cours de sa journée, il est indispensable de caractériser l'air que l'on respire à l'intérieur des locaux (chez soi, au travail, dans les transports etc.).

En 2001, l'Observatoire de la Qualité de l'Air Intérieur (OQAI), a été créé à l'initiative du ministère du logement pour pallier au manque d'information sur l'exposition des populations à la pollution intérieure et a mené des campagnes de mesures dans différents locaux. En 2013, le Centre International de Recherche sur le Cancer (CIRC) a classé la pollution de l'air intérieur comme cancérigène certain pour l'homme.

De nombreux programmes de recherche traitent de la qualité de l'air intérieur :

- le Programme de Recherche Interorganisme pour une Meilleure QUALité de l'air, PRIMEQUAL²⁸. Lancé en 1995, ce programme a été mis en œuvre par le Ministère chargé de l'Environnement et par l'Agence de l'Environnement et de la Maîtrise de l'Énergie (ADEME). Différentes études sont financées par ce programme : l'étude ADOQ²⁹ *Activités Domestiques et Qualité de l'air intérieur : émissions, réactivité et produits secondaires*, l'étude ECOS-HABITAT³⁰ : Expositions cumulées aux Composés Organiques Semi-volatils dans l'habitat : risques pour le développement de l'enfant / Contribution des poussières sédimentées à l'exposition : contamination des logements français.
- le programme « Qualité de l'air, impacts sur la santé et l'environnement » de la stratégie Recherche 2014-2020 de l'ADEME apporte de nouvelles solutions et afin d'améliorer la qualité de l'air intérieur et extérieur. Ce programme intègre le programme « COonnaissances, Réduction à la source et Traitement des Émissions dans l'Air » (CORTEA³¹) et un programme d'aide pour les actions des collectivités territoriales et locales en faveur de l'air : Aact-air³².
- le programme National de Recherche Environnement-Santé-Travail, PNREST³³, pour appuyer les politiques publiques et travaux d'évaluation des risques sanitaires.

²⁷ <https://www.mdpi.com/2073-4433/10/2/96/pdf>

²⁸ www.primequal.fr

²⁹ <http://www.primequal.fr/sites/default/files/68c21c037f19ed0d.pdf>

³⁰ <http://www.primequal.fr/sites/default/files/2c6d3519dbf25ada.pdf>

³¹ <https://www.ademe.fr/recherche-innovation/programmes-projets-recherche>

³² <https://www.ademe.fr/expertises/air-bruit/passer-a-l'action/dossier/programmes-faveur-qualite-lair-aact-air-cortea-primequal/aact-air-actions-faveur-qualite-lair-territoires>

³³ <https://www.anses.fr/fr/content/le-programme-national-de-recherche-environnement-santé-travail-pnrest>

Ces programmes ont pour objectif principal de fournir les bases scientifiques et les outils nécessaires aux décideurs et aux gestionnaires de territoires afin d'améliorer la qualité de l'air et réduire les risques pour la santé. Ces travaux étudient entre autre la caractérisation des composés organiques volatils et semi volatils en air intérieur mais ne permettent pas de remonter à l'exposition individuelle.

Les sources de pollution à l'intérieur des bâtiments sont multiples et essentiellement dues (1) à la pollution en provenance de l'extérieur, (2) aux pollutions ponctuelles liées à l'activité et à la présence humaine, (3) aux pollutions émises en continu par l'équipement intérieur : meubles, peintures, parquets, chauffage (**Figure 8**).

Les sources de pollution liées aux activités humaines englobent le tabagisme, les opérations de cuisson et de chauffage au gaz, au bois et au charbon, les feux de cheminée au bois ou poêle et tout particulièrement les chauffages d'appoint, les appareils de production d'eau chaude au gaz, l'utilisation des produits d'entretien, de nettoyage, de pesticides, de cosmétique, les activités de bricolage, les épurateurs d'air, les photocopieurs, les imprimantes laser, les ioniseurs peuvent produire de l'ozone.

L'importance de ces différentes sources dépend du type d'émission, de la taille et de la nature du polluant. L'absence de ventilation naturelle ou une ventilation mécanique contrôlée mal entretenue, une température ou un niveau d'humidité élevés peuvent augmenter ce niveau de pollution.

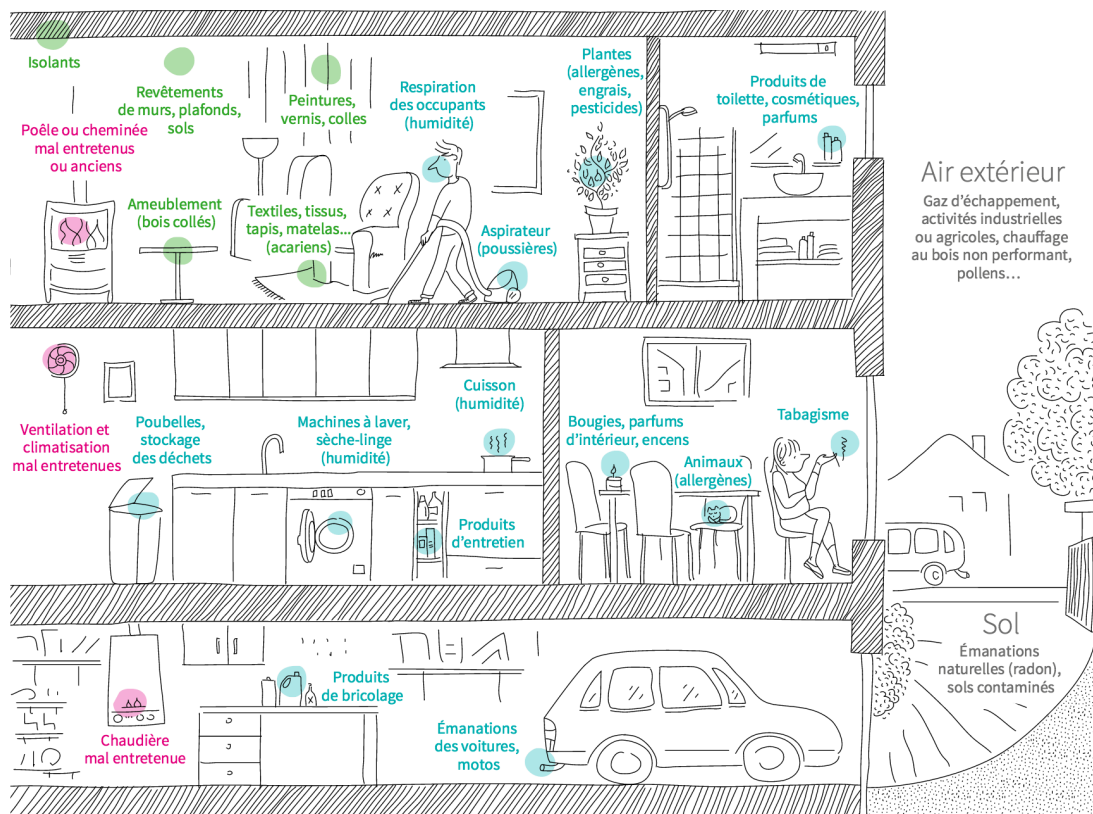


Figure 8 : Sources des polluants de l'air intérieur. Source : ADEME, « un air sain chez soi »³⁴

³⁴ <https://www.ademe.fr/sites/default/files/assets/documents/guide-pratique-un-air-sain-chez-soi.pdf>

La pollution de l'air intérieur fait intervenir des molécules variées et différentes de celles rencontrées en air extérieur (**Tableau 3**). Comme mentionné dans la thèse de Le Behec³⁵, 113 composés organiques volatils (COV) ont été identifiés dans l'air intérieur d'une habitation. Ces COV sont issus de solvants, de colles ou de peintures utilisés dans les matériaux d'ameublement et de construction.

Les particules organiques et biologiques sont émises par les appareils de combustion, le nettoyage et le bricolage, les animaux domestiques et l'activité humaine.

Origine	Polluants
Air extérieur	<ul style="list-style-type: none"> • Particules organiques et minérales • Gaz (CO₂, CO, Radon) • Hydrocarbure Aromatiques Polycycliques (HAP)
Matériaux de construction	<ul style="list-style-type: none"> • HAP (solvants) • COV (revêtements plastifiés, peintures, colles)
Ameublement	<ul style="list-style-type: none"> • HAP (solvants) • COV (vernis, colles, peintures, matières plastiques)
Appareils de combustion	<ul style="list-style-type: none"> • Particules organiques • Gaz (CO, CO₂) • COV
Nettoyage / bricolage	<ul style="list-style-type: none"> • Particules organiques • Composés chimiques (solvants, HAP) • COV
Animaux domestiques	<ul style="list-style-type: none"> • Particules organiques (poils, virus, bactéries, acariens) • Gaz (CO₂)
Activité humaine	<ul style="list-style-type: none"> • Particules biologiques (virus, bactéries, moisissures) • Gaz (CO₂) • COV (cosmétique)

Tableau 3 : Origine et nature des polluants de l'air intérieur. Source : thèse Le Behec (2016)

Dans ce premier chapitre, nous avons évoqué l'impact sanitaire et les seuils des polluants surveillés et réglementés. Nous avons également spécifié les particularités de la région parisienne avant d'aborder les processus physico-chimiques sur différentes échelles spatiales et temporelles. Pour finalement introduire la qualité de l'air intérieur. Dans le chapitre suivant (chapitre 2), nous décrirons les outils de mesures et de modélisation utilisés dans le cadre de cette thèse afin de répondre à notre objectif principal, à savoir caractériser finement la pollution de l'air en région parisienne.

³⁵ <http://www.theses.fr/2016PAUU3012>

Chapitre 2. Outils utilisés

2.1 Instruments de mesures

Dans le cadre de ce travail de thèse, nous avons effectué des mesures avec les analyseurs de particules (OSIRIS et DustMate) et un capteur de particules e-PM (**Figure 9**).

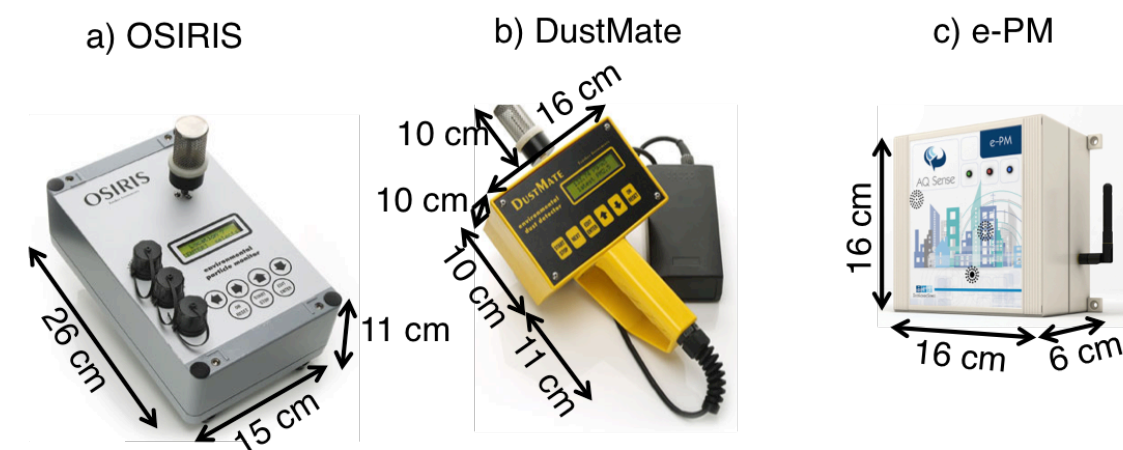


Figure 9 : Photographie de l'OSIRIS (a), du DustMate (b) et du capteur e-PM (c)

OSIRIS, DustMate et capteur e-PM (**Figure 9**) sont des compteurs optiques qui mesurent la concentration en particules pour différentes fractions granulométriques (PM_{10} , $PM_{2,5}$ et parfois PM_1). Le compteur optique de particules est basé sur la mesure de l'intensité diffusée par une particule (**Figure 10**). Le principe de fonctionnement est le suivant : les particules présentes dans l'air sont aspirées à travers le capteur par un flux forcé. Ces particules passent devant un faisceau laser, généré par une LED infrarouge. La quantité de lumière diffusée à un angle θ est mesurée par un détecteur. Cette quantité est proportionnelle à l'intensité incidente au diamètre optique, et à la concentration en nombre des particules. Le rayonnement incident sera plus ou moins diffracté selon la taille de la particule rencontrée. La concentration massique en particules est alors déterminée en faisant l'hypothèse que les particules mesurées sont sphériques et en attribuant une valeur de densité constante aux différentes fractions granulométrique de l'aérosol.

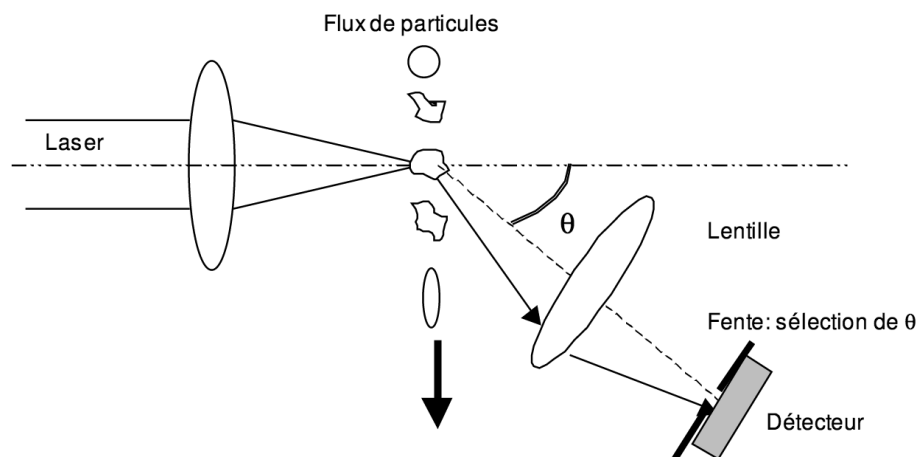


Figure 10 : Schéma de fonctionnement d'un compteur optique. Source : rapport INERIS (Blanchard *et al.*, 2004)

L'OSIRIS ou le DustMate sont considérés comme des instruments de référence au LATMOS. Dans certaines campagnes de mesures de particules, la référence peut être la méthode gravimétrique qui consiste à peser des filtres avant et après prélèvement. Ne disposant pas de filtres au LATMOS, nous avons uniquement utilisé l'OSIRIS et le DustMate.

Un instrument de référence est 34 fois (OSIRIS) à 22 fois (DustMate) plus cher que le capteur e-PM (200 €). L'OSIRIS est un instrument fait pour une mesure en point fixe alors que le DustMate et le capteur e-PM sont suffisamment légers pour être mobiles pesant respectivement 1.2 kg et 0.8 kg. Les instruments de référence : OSIRIS et DustMate peuvent mesurer des niveaux très faibles de pollution avec une limite de détection de $0.01 \mu\text{g}/\text{m}^3$ alors que le capteur e-PM a une limite de détection de $5 \mu\text{g}/\text{m}^3$. Une autre différence majeure entre les instruments de référence et le capteur e-PM est la gamme de mesure (**Tableau 4**).

	OSIRIS	DustMate	e-PM
Constructeur	Turnkey		EcoLogicSense
Technologie	compteur optique		diffraction de la lumière
Polluants	TPS, PM ₁₀ , PM _{2.5} et PM ₁		PM ₁₀ et PM _{2.5}
Gamme d'utilisation	-5 à 50°C		
Gamme de mesure	0-6000 $\mu\text{g}/\text{m}^3$		0-100 $\mu\text{g}/\text{m}^3$ (PM _{2.5}) 0-200 $\mu\text{g}/\text{m}^3$ (PM ₁₀)
Débit de la pompe	600 cc/min	500 cc/min	pas de pompe
Limite de détection	0.01 $\mu\text{g}/\text{m}^3$		5 $\mu\text{g}/\text{m}^3$
Incertitudes	5 $\mu\text{g}/\text{m}^3$		min 5 $\mu\text{g}/\text{m}^3$
Taux d'échantillonnage minimum	1 s		2 min
Autonomie de la batterie	10 h	4 h	1 semaine avec périodicité 10 min

Mémoire interne	128 ko	32 ko	5 mois de mesure (avec périodicité de 10 min)
Dimensions boîtier	26×20×15 cm ³ (avec filtre) 26×11×15 cm ³ (sans filtre)	16×10×10 cm ³ (sans le manche et le filtre) le manche mesure 5 cm et le filtre 10 cm	16×16×6 cm ³
Poids	3.5 kg	1.2 kg	800 g
Prix	6850 €	4500 €	200 €
Commentaires	bryant	bryant, portatif et léger	silencieux, portatif et très léger

Tableau 4 : Caractéristiques techniques des instruments de référence : OSIRIS, DustMate et capteur e-PM

Les capteurs e-PM étaient les capteurs mobiles les plus performants à l'époque où ces mesures ont été réalisées en 2016 et nous avaient été prêtés par l'IFSTTAR. Depuis le développement de ces capteurs est en plein essor et de nombreux capteurs à bas coût sont sur le marché. De nombreux colloques (FIMEA, novembre 2016, colloque de l'APPA³⁶, novembre 2018, Sorbonne Université³⁷, Avril 2019, LCSQA, mai 2019) ou encore l'ANR Polluscope s'intéressent à ce sujet.

Le LCSQA a mené une étude d'inter-comparaison de ces capteurs sur la base de volontariat des participants pour proposer un cadre d'évaluation commun à tous ces instruments. En effet, il n'existe à l'heure actuelle aucun cadre normatif national ou européen permettant de comparer les performances de ces différents appareils commercialisés aux appareils de mesures de référence. Pour chaque appareil, cette étude fournit une fiche d'évaluation par polluant mesuré (NO₂, PM_{2.5}, PM₁₀ et O₃) pour huit critères (Energie, Fiabilité, Versatilité, Mise en œuvre, Pente, R2, MAPE, variabilité) avec des notes allant de 0 à 5 (radar d'évaluation, **Figure 11**).

AIRPARIF a déjà mené un premier « challenge AIRLAB - Microcapteurs 2018 »³⁸ pour tester ces capteurs – toujours sur la base du volontariat des constructeurs - et un deuxième challenge est en cours.

³⁶ <http://lille2018.appa.asso.fr/communications/>

³⁷ <https://caspa.sciencesconf.org/>

³⁸ <https://www.airparif.asso.fr/actualite/detail/id/226>

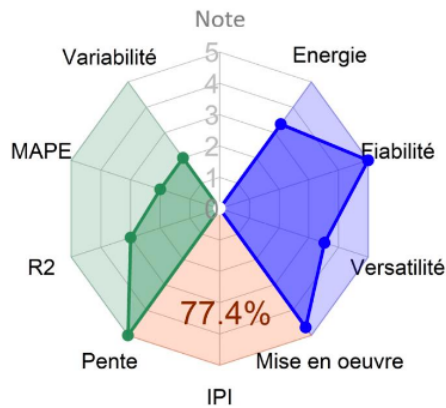


Figure 11 : Radar d'évaluation d'un système « capteurs » pour chaque polluant. Source : rapport LCSQA³⁹. MAPE=Pourcentage Moyen des Ecart en Valeur Absolue et IPI=Indice de performance intégré

2.2 Modélisation

2.2.1 Généralités

Les modèles numériques de qualité de l'air régionaux et urbains permettent de : (1) cartographier les zones de dépassements et estimer l'exposition des individus à la pollution (Markakis et Valari, 2015⁴⁰), (2) prévoir les niveaux de concentration des polluants à court terme et à long terme (+100 ans) afin de mettre en œuvre des mesures pour éviter la prolongation de certains pics de pollution et informer le public, (3) analyser la contribution des sources : connaître l'origine de dépassement et orienter les stratégies de planification, (4) évaluer l'impact des scénarios de réduction des émissions et (5) optimiser le réseau de mesures, comme outil d'aide à la décision dans les choix des lieux d'implantation des stations fixes (Rouïl, 2012).

Par exemple, le modèle CHIMERE⁴¹ est un modèle régional 3D de chimie transport. Il présente donc l'avantage de simuler les processus physico-chimiques (section 1.1.3) qui ont lieu dans les différentes couches de l'atmosphère. Une première limite pour la modélisation par un tel modèle est la difficulté à modéliser les particules en raison des incertitudes sur les inventaires d'émissions et sur la paramétrisation des phénomènes physico-chimiques mis en jeu. De plus, des modèles régionaux eulériens de chimie transport tels que CHIMERE ne peuvent pas descendre en deçà de 1 km de résolution horizontale car la turbulence est paramétrisée et non calculée (Stein *et al.*, 2007 et Valari⁴², 2009). Cette résolution n'est pas suffisamment fine pour simuler l'hétérogénéité de la pollution en ville.

³⁹ https://www.lcsqa.org/system/files/rapport/LCSQA2017-CILmicrocapteurs-synthese_resultats.pdf

⁴⁰ <http://www.acceptedera.eu/wp-content/uploads/ACCEPTED-Final-Report.pdf>

⁴¹ <http://www.lmd.polytechnique.fr/chimere>

⁴² http://www.lmd.polytechnique.fr/~intro/Files/2009-These_Myrto_Valari.pdf

Différents outils de modélisation peuvent être utilisés pour cartographier la pollution à l'échelle de la rue. Voici une liste non exhaustive des modèles qui existent afin de donner un panorama global : (1) des modèles paramétrisés : les modèles gaussiens comme ADMS-Urban⁴³ (développé par Numtech) ou des modèles de rues comme MUNICH, *The Model of Urban Network of Intersecting Canyons and Highways*⁴⁴ (développé au CEREAA), (2) des modèles CFD (Computational Fluid Dynamics) comme le code SATURNE⁴⁵ (développé au CEREAA) ou (3) des modèles lagrangiens comme le modèle PMSS *Parallel Micro-SWIFT-SPRAY*⁴⁶ (développé par ARIA-Technologies).

Dans le cadre du projet MOUVIE : Mobilité et qualité de vie en milieu urbain, ARIA Technologies nous a fourni des simulations provenant du modèle urbain PMSS que nous avons utilisées dans cette thèse. Les modèles urbains à haute résolution de dispersion atmosphérique, tels que le modèle PMSS, résolvent les équations de Navier Stokes avec la diffusion-advection des parcelles d'air et une paramétrisation de la turbulence d'ordre supérieure. De cette manière, ces modèles intègrent les processus à fine échelle (< 1 km) imposées par la structure 3D de la ville en prenant en compte la hauteur et la géométrie des bâtiments. Ces modèles permettent ainsi de caractériser la pollution de l'air à l'échelle de la rue, avec une résolution allant jusqu'au mètre. Le modèle PMSS est décrit dans la suite de ce manuscrit (section 2.2.3).

2.2.2 Cadastres d'émissions en entrée des modèles de qualité de l'air

Les cadastres d'émissions sont des inventaires d'émissions spatialisés sur un territoire. L'une des applications des cadastres d'émissions est l'alimentation de modèles de prévision. La qualité des modèles de la qualité de l'air dépend en grande partie de la qualité et de la précision de ces cadastres.

La constitution des cadastres peut s'élaborer selon deux approches possibles. La première approche, dite « top-down » collecte les données générales : échelle administrative, de la région, par secteurs d'activité pour l'industrie. Les émissions sont calculées à cette grande échelle puis distribuées avec l'utilisation des « proxys » (par exemple le réseau routier, l'occupation du sol) pour atteindre la résolution spatiale requise. La deuxième approche, dite « bottom-up » collecte les données les plus fines disponibles (logement, industrie, axe routier) puis les rassemble pour atteindre la résolution spatiale voulue⁴⁷. Au niveau local, la deuxième approche : « bottom-up » est privilégiée avec par exemple l'utilisation de données telles que le comptage routier, des données de production par site ou la composition du parc de logements à l'échelle de la commune⁴⁸. C'est d'ailleurs cette approche qui est utilisée pour réaliser le cadastre d'émissions d'AIRPARIF. Chaque type d'émission fait l'objet d'une base de données et d'une représentation particulière.

La base de données des émissions surfaciques recense les émissions diffuses non affectables à un point précis comme par exemple le chauffage résidentiel, l'agriculture ou les petites industries avec peu d'informations sur leurs émissions.

⁴³ https://www.numtech.fr/contenu/adms_urban_fr.pdf

⁴⁴ <http://cerea.enpc.fr/munich/index.html>

⁴⁵ http://cerea.enpc.fr/fich/mercure_plaquette.pdf

⁴⁶ http://www.aria.fr/projets/aircity/pdf/AIRCITY_General.4_juillet_2013.J.Moussafir.pdf

⁴⁷ http://www.cnrs.fr/cw/dossiers/dosclim1/biblio/pigb13/03_inventaires.htm

⁴⁸ http://www.airparif.fr/_pdf/publications/inventaire-emissions-idf-2010-rapport-130731.pdf

Les résolutions temporelles des inventaires varient de l'heure à l'année selon la résolution requise pour le phénomène étudié.

Le modèle de qualité de l'air utilisé dans cette thèse (le modèle PMSS) intègre les émissions liées au trafic routier, fournies par la chaîne de calcul HEAVEN - *Healthier Environment through the Abatement of Vehicle Emissions and Noise*⁴⁹. Les émissions sont estimées en quasi-temps réel pour chaque voie routière du réseau francilien (40 000 portions de voies routières, soit 10 000 km d'axes routiers). Les conditions de trafic sont déterminées sur chaque voie et intègrent : le nombre de véhicules, leurs caractéristiques et la longueur de la portion de voie. Les émissions horaires de trafic sont calculées en combinant les données issues du modèle trafic (flux de véhicules roulants, vitesse moyenne des véhicules roulants, pourcentage de véhicules circulant avec un moteur froid) et des facteurs d'émissions (type de véhicules, vitesse, si oui ou non les véhicules roulent « à froid », à température ambiante) (Figure 12).

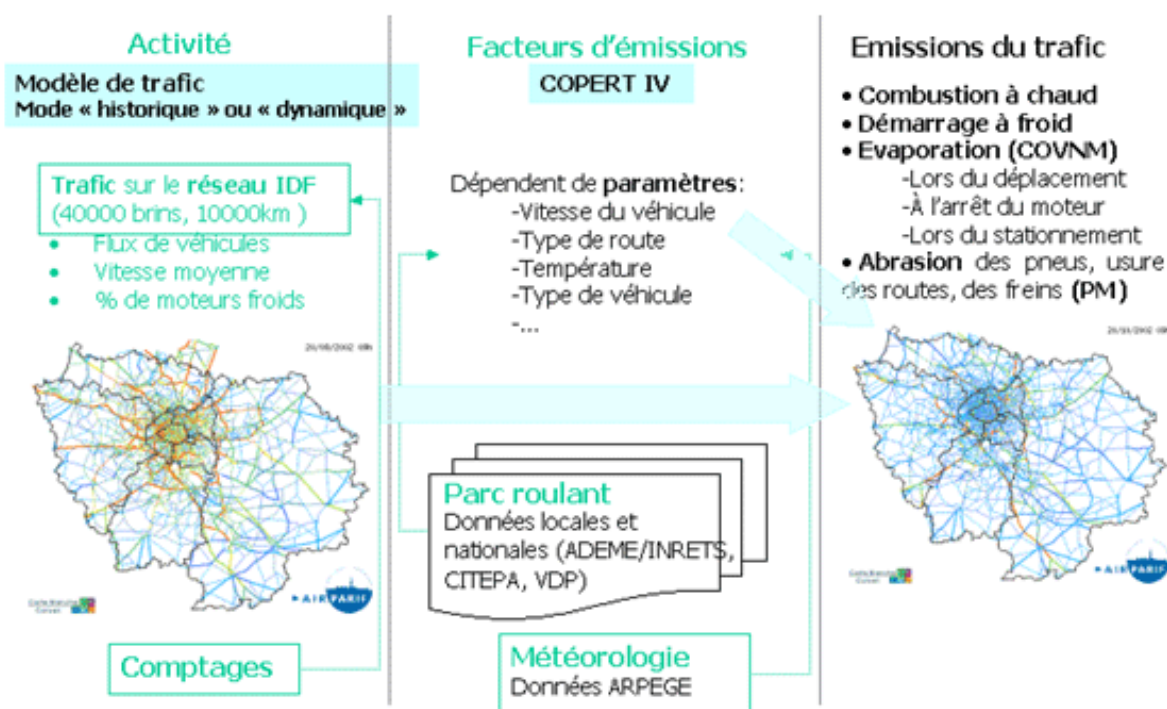


Figure 12 : Différentes étapes de la chaîne de calcul HEAVEN pour estimer les émissions horaires liées au trafic en Île-de-France. Source : AIRPARIF

2.2.3 PMSS : un modèle urbain de dispersion atmosphérique haute résolution

Le modèle PMSS, co-développé par ARIA Technologies, ARIANET, le CEA, MOKILI et ISAC-CNR (Italie), permet de modéliser la dispersion des polluants en milieu urbain ou industriel à petite échelle. Les dix jours de simulations fournies par ARIA Technologies ont été réalisés dans les mêmes configurations que le projet AIRCITY qui a pour mission le

⁴⁹ <http://www.airparif.asso.fr/etat-air/air-et-climat-emissions-heaven>

développement d'un système de simulation 3D de haute résolution de la qualité de l'air sur l'ensemble de la ville de Paris⁵⁰.

Le modèle PMSS simule les champs de concentrations en trois dimensions avec une résolution horizontale de 3 mètres et couvre l'ensemble du domaine de Paris : 12 km × 10 km. Le modèle comprend 35 couches verticales allant du sol à 800 mètres d'altitude avec une épaisseur de 2 mètres près du sol (Moussafir *et al.*, 2014).

Ici, nous avons utilisé les champs de concentrations horaires des NO_x et PM₁₀ simulés dans la première couche du modèle PMSS, de 0 à 2 mètres du sol, au-dessus de Paris, pour dix jours non consécutifs de mars 2016. En annexe de ce manuscrit, figure une carte des concentrations de NO_x à une résolution dégradée, moyennée sur la période étudiée.

Le modèle PMSS est la version parallèle du code MSS, combinant :

- un modèle météo/CFD (Computational Fluid Dynamic) simplifié (Micro-SWIFT). Micro-SWIFT est un modèle diagnostique à conservation de la masse qui calcule le vent, la température et la turbulence ;
- un modèle de dispersion des gaz et particules (Micro-SPRAY). Le modèle Micro-SPRAY calcule le transport de diffusion de dépôt des gaz-particules à partir des données météorologiques simulées par Micro-SWIFT. Il calcule également les champs de pression pour l'intégration des bâtiments dans le modèle.

Le modèle Micro-SPRAY est un modèle lagrangien. Les modèles lagrangiens résolvent un système d'équation dans un repère lié au déplacement d'une colonne d'air dans la trajectoire d'un panache (**Figure 13**).

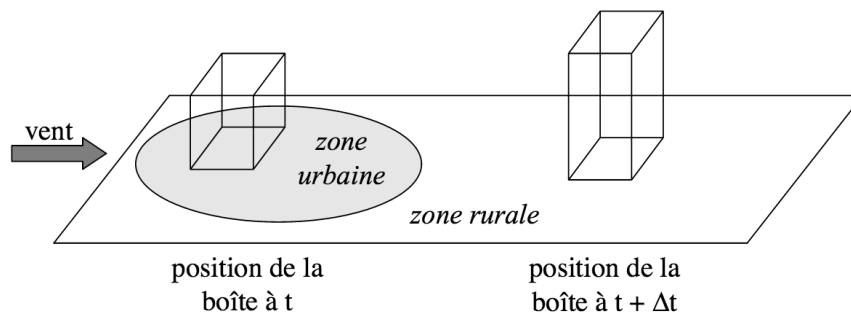


Figure 13 : Schéma d'une maille lagrangienne. Source : cours sur la modélisation de la chimie atmosphérique (Aumont⁵¹)

La parcelle d'air se déplace selon les conditions météorologiques (vent, etc.) : il n'y a donc pas d'échange de matière entre la parcelle considérée et son environnement. Les seuls échanges de matière se font par la surface (émissions, chimie et dépôt). Les émissions injectées dans la colonne sont soumises à la dispersion verticale et aux réactions chimiques.

Le système d'équation d'une maille lagrangienne est construit sur la conservation de la masse à l'intérieur et résout l'équation suivante (eq.9) :

$$\frac{dC_i}{dt} = \text{émission} + \text{dépôt} + \text{chimie} \quad (\text{eq.9})$$

⁵⁰ http://www.aria.fr/projets/aircity/pdf/AIRCITY_General.4%20juillet%202013.J.Moussafir.pdf

⁵¹ http://www.lisa.u-pec.fr/~aumont/contenu/enseignement/M2_SGE/modelisation_short.pdf

À la différence d'une maille eulérienne, ici il n'y a pas de terme d'advection.

La conservation de la masse à l'intérieur de la maille à une seule couche s'écrit selon l'équation (eq.10)

$$\frac{dC_i}{dt} = \frac{q_i}{h} - \frac{v_d}{h} C_i + R_i \quad (\text{eq.10})$$

avec :

– $\frac{dC_i}{dt}$: la variation temporelle de concentration de l'espèce chimique i à l'intérieur de la maille (C_i)

– $\frac{q_i}{h}$: le terme d'émission

– q_i : quantité émise par unité de temps et de surface dans la maille et h l'épaisseur de la maille.

– le terme de dépôt : $-\frac{v_d}{h} C_i$

– v_d : la vitesse de dépôt

– R_i : le terme de chimie

R_i : représente le terme net lié aux réactions de production et de consommation de l'espèce i .

Micro-SPRAY ne peut simuler qu'une seule réaction chimique : la photodissociation du NO_2 (cycle de Chapman). Les autres espèces sont considérées comme des traceurs passifs.

En entrée, le modèle PMSS intègre les données urbaines, d'émissions et de météorologie dans les configurations des simulations étudiées:

- la description fine échelle du relief des bâtiments (fichier de description 3D). Les bâtiments sont vus comme des obstacles pour la dispersion de la pollution. Un préprocesseur intègre en entrée un fichier de description 3D des bâtiments et génère en sortie des prismes à base triangulaire. Ces prismes sont ensuite intégrés en entrée du modèle PMSS pour prendre en compte les bâtiments.

- le réseau routier ajusté à la cartographie des bâtiments (fourni par l'IGN).

- la détection des rues « canyons » en calculant automatiquement le rapport entre la largeur de la rue et la hauteur des bâtiments. Une correction des simulations initiales est alors appliquée en cas de rue « canyons ». En effet, dans les rues « canyons » la vitesse du vent et le mélange turbulent peuvent être faibles et entraîner l'accumulation des polluants.

Ce modèle haute résolution est couplé : (1) aux modèles de météorologie régionaux WRF (Weather Research and Forecasting), (2) à la pollution de fond assimilée (anthropique : activité industrielle, émissions résidentielles, aéroport, trafic automobiles, fluvial et maritime, et d'origine naturelle) fournit par le modèle CHIMERE (la sortie CHIMERE ayant été réajustée par des données observées sur les six dernières heures). Le modèle CHIMERE est lui-même alimenté par l'inventaire d'émissions réalisé par AIRPARIF (inventaire 2012 dans le cas de ce projet), (3) aux émissions linéaires liées au trafic (chaîne de calcul HEAVEN).

Atouts et limites du modèle PMSS:

Les processus fines échelles (< 1 km) spécifiques des villes, tels que les flux turbulents intégrant la géométrie des bâtiments et la météorologie, sont ainsi pris en compte et permettent ainsi de caractériser la pollution de l'air à l'échelle de la rue avec une résolution allant jusqu'au mètre.

Le modèle ne simule qu'une seule réaction chimique : la photodissociation du NO_2 (cycle de Chapman), or l'atmosphère est le siège de nombreuses réactions chimiques (section 1.1.3) notamment pour la pollution photo-oxydante à l'origine de la production d'ozone fait

intervenir plusieurs espèces : COV, NO_x et HO_x et ceci peut avoir un impact sur les simulations.

2.3 AIRPARIF : réseau de mesures de la pollution de l'air en région parisienne

Le 30/12/1996, la Loi sur l'Air et l'Utilisation Rationnelle de l'Énergie (LAURE) donne le droit à chacun de respirer un air sain qui ne nuise pas à sa santé et le droit à l'information sur la qualité de l'air et ses effets. L'Etat confie ainsi la mise en œuvre de la surveillance de la qualité de l'air et l'information du public à des Associations Agréées de Surveillance de la Qualité de l'Air (AASQA). Depuis janvier 2012, dix-huit AASQA sont réparties sur l'ensemble du territoire français et mesurent en continu les concentrations des polluants.

Les polluants surveillés par les AASQA sont principalement les polluants réglementés par le code de l'environnement et les directives européennes ainsi que certains polluants spécifiques issus d'activités industrielles et d'autres sources. Il s'agit du dioxyde de soufre (SO₂), du dioxyde d'azote (NO₂), des oxydes d'azote (NO_x), l'ozone (O₃), des particules (PM₁₀ et PM_{2,5}), du monoxyde de carbone (CO), du benzène (C₆H₆), du plomb (Pb), de l'arsenic (As), du cadmium (Cd), du nickel (Ni), du mercure (Hg), des hydrocarbures, des aromatiques polycycliques (benzo(a)pyrène principalement). Les AASQA sont chargées d'alerter les autorités en cas de dépassement d'un seuil d'information et recommandation ou d'alerte.

Le dispositif national comptait six cent douze stations fixes de surveillance en 2018. Chaque station de mesure est équipée d'un ou plusieurs instruments (analyseurs) selon les espèces chimiques surveillées (en fonction de leur effet sur la santé et de la réglementation). Ces mesures s'effectuent de manière automatique, sept jours sur sept, par tranche de quinze minutes. Le dispositif compte : (1) des stations de fond urbain qui suivent l'exposition moyenne de la population aux phénomènes de pollution atmosphérique dits de « fond » dans les centres urbains, (2) des stations trafic, représentatives du niveau maximum d'exposition auquel la population située à proximité d'une infrastructure routière est susceptible d'être exposée et (3) des stations de fond dans les zones rurales, éloignées de sources industrielles.

L'agence AIRPARIF assure le suivi et la surveillance du réseau dans la région Île-de-France. Environ soixante-dix stations de mesures sont déployées en Île-de-France (**Figure 14**). Parmi ces stations, on compte plus d'une cinquantaine de stations automatiques permanentes et plus d'une dizaine de stations semi-permanentes. Ce réseau de mesure est très dense et les données y sont archivées depuis 1999 pour certains polluants et disponibles en ligne⁵². Des instruments supplémentaires sont parfois déployés lors de mesures de campagnes ponctuelles.

⁵² <https://www.airparif.asso.fr>

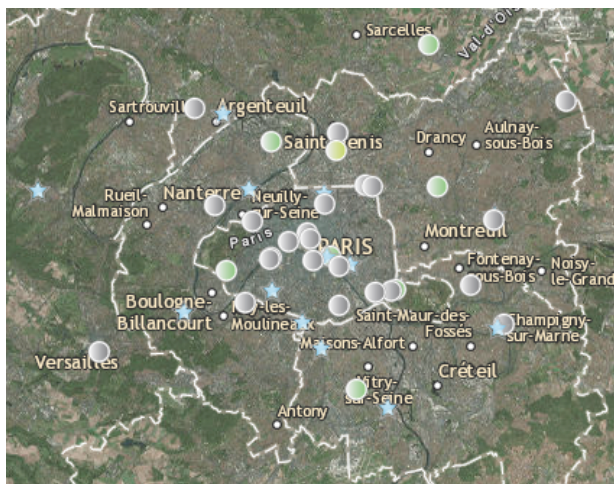


Figure 14 : Déploiement instrumental du réseau AIRPARIF en région Île-de-France.
Source : AIRPARIF⁵³

Dans le cadre de cette thèse, nous avons utilisé les concentrations horaires des NO_x et PM_{10} des stations de surface AIRPARIF de trafic et de fond urbain à Paris en mars 2016. Nous avons étudié les mesures de dix sur treize stations de NO_x et les mesures de cinq sur huit stations de PM_{10} (**Figure 15**). Les stations mesurant les NO_x et PM_{10} , localisées sur le boulevard périphérique est (BP_EST) et sur le boulevard périphérique près de la porte d'Auteuil (AUT), n'ont pas pu être exploitées. De même que les stations de NO_x localisées sur le boulevard Sault (SOULT) et rue Bonaparte (BONAP). En effet, des brins d'émissions étaient manquants dans les simulations du modèle PMSS au niveau de ces stations et à proximité de celles-ci. Deux autres stations, avec des brins d'émission manquants sur une plus courte distance, ont pu être « déplacées » de quelques mètres (stations virtuelles BASCH_v et CELES_v) car des brins d'émissions étaient présents à proximité de la station et le déplacement de ces stations permettait de conserver la même typologie de site que celle de la station originale (décrit dans l'article sur la représentativité, chapitre 4).

Nous n'avons pas retenu les mesures des stations AIRPARIF pour les particules fines $\text{PM}_{2,5}$ qui étaient suivies uniquement par trois stations à cette période, dont deux stations étaient non exploitables (AUT et BP_EST) en raison des brins d'émission manquants dans le modèle PMSS.

⁵³ <https://www.airparif.asso.fr/stations/index>

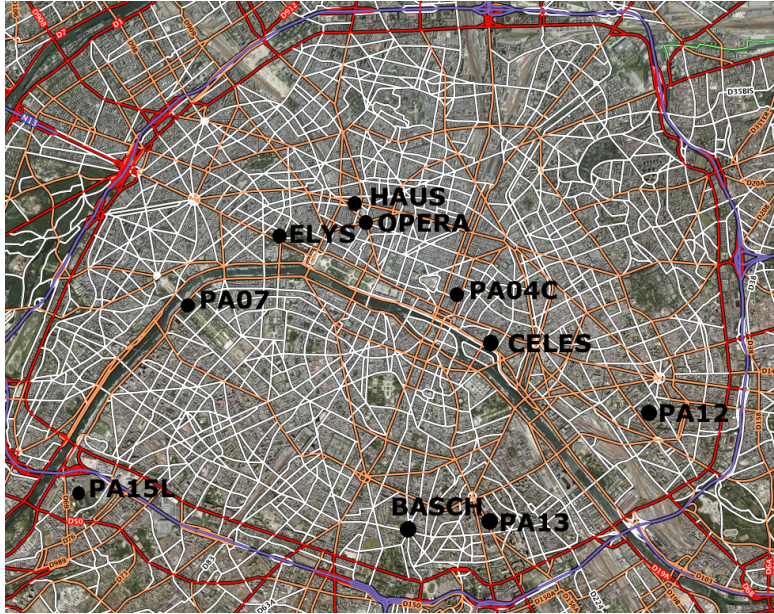


Figure 15 : Plan de Paris avec l'emplacement des stations AIRPARIF étudiées, la Seine et les routes (en blanc et orange). Le rouge représente la limite de Paris avec le boulevard périphérique.

La comparaison entre les sorties du modèle PMSS et les mesures AIRPARIF montre que le modèle PMSS est très performant pour simuler les champs de concentration de surface au niveau des stations, particulièrement pour les PM_{10} . Pour les NO_x , le modèle a parfois du mal à reproduire certains pics de concentration (**Figure 16**).

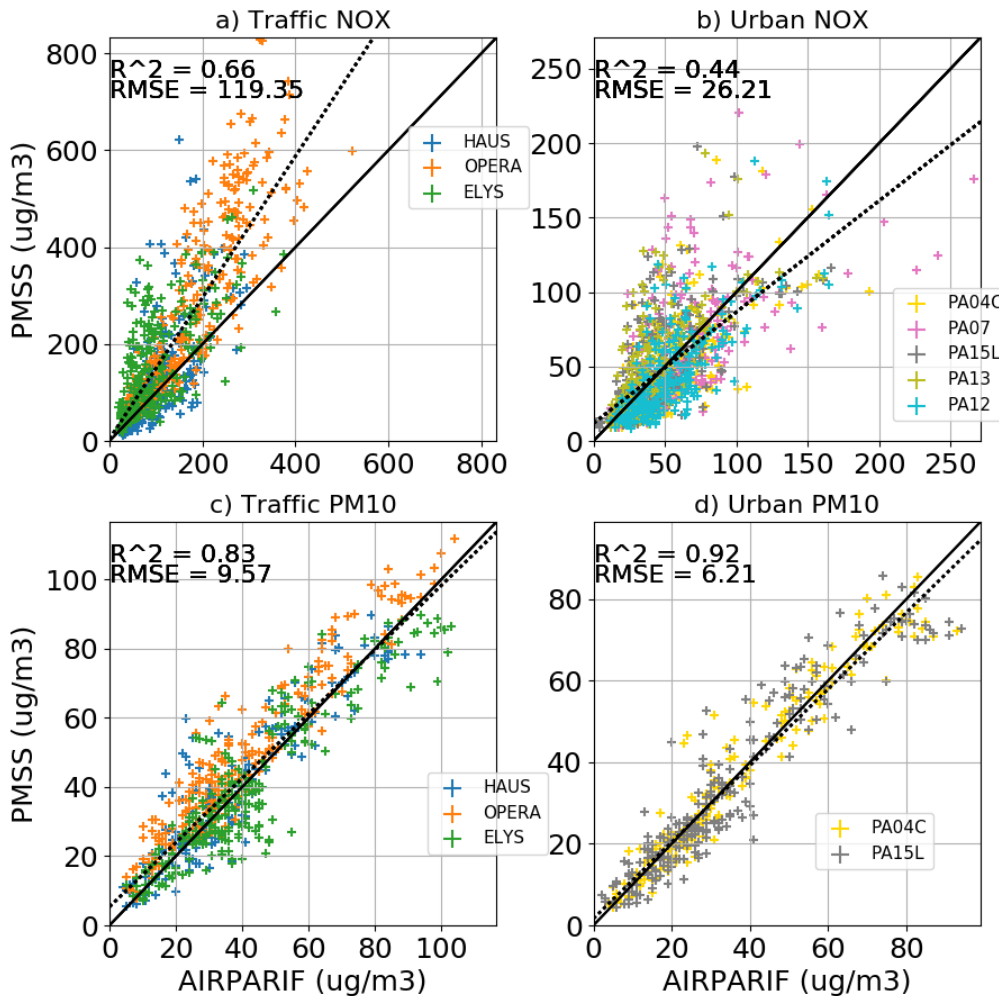


Figure 16 : Nuage de points qui représente les concentrations mesurées des stations AIRPARIF versus les concentrations simulées par le modèle PMSS (en $\mu\text{g}/\text{m}^3$). La ligne continue noire correspond à $y=x$. La ligne en pointillée représente la régression linéaire $y=ax+b$. Le RMSE est exprimé en $\mu\text{g}/\text{m}^3$.

Dans ce deuxième chapitre, nous avons décrit les outils utilisés : (1) les instruments de référence du LATMOS pour les mesures de particules (OSIRIS et DustMate) et des capteurs mobiles à bas coût (capteurs e-PM), (2) les simulations du modèle PMSS à 3 mètres de résolution horizontale pour cartographier la pollution à Paris, (3) les mesures des stations AIRPARIF à Paris.

Nous avons réalisé une étude exploratoire de l'exposition en menant des campagnes de mesures avec des instruments de mesure et des capteurs à bas coût (chapitre 3).

Nous avons également cherché des zones autour des stations AIRPARIF à Paris, avec des caractéristiques physiques proches de celles de la station. Ces zones ont été déterminées uniquement à partir des simulations du modèle PMSS (chapitre 4).

Puis nous avons propagé la mesure AIRPARIF dans ces zones préalablement établies (chapitre 5).

Chapitre 3. Approche expérimentale de l'exposition aux PM₁₀ et PM_{2,5} en Île-de-France

3.1 Introduction

Les réseaux de mesures des agences locales (AIRPARIF, à Paris) ne sont pas suffisamment denses pour renseigner de l'hétérogénéité des concentrations en ville (Gozzi *et al.*, 2016, Monn *et al.*, 1997). De plus, la contribution des sources de pollution varie très fortement entre l'intérieur et l'extérieur. Pour les particules, cette contribution dépend de leur taille et de leur nature (Ramalho *et al.*, 2012). Pour les particules, les niveaux sont bien plus importants à l'intérieur qu'à l'extérieur (étude de Sexton et Ryan (1988) citée dans une revue M.R Ashmore (2009)).

Des solutions doivent être mises en œuvre pour pallier ce manque de données afin de mieux estimer l'exposition des individus à la pollution. Des instruments peuvent être déployés dans différents microenvironnements lors de campagnes ponctuelles mais ce type de dispositif est très coûteux sachant le prix à l'unité d'un tel instrument (de 5000 € à 30 000 €, Castell *et al.*, 2017). Une autre solution est d'équiper les citoyens volontaires, membres d'observatoires participatifs, de nombreux capteurs à bas coût (< 500 €). L'objectif est de couvrir la région étudiée avec le plus grand nombre de capteurs afin de cartographier le plus finement possible l'exposition des franciliens à la pollution de l'air et ainsi pouvoir informer les citoyens qui ne sont pas munis de capteurs.

Alors que le développement de ces capteurs est en plein essor (Gozzi *et al.*, 2016, Merbitz *et al.*, 2012, Patton *et al.*, 2014, Van den Bosshe *et al.*, 2015), une question persiste : peut-on se fier aux mesures fournies par ce type de capteurs ? Sont-elles aussi fiables que les mesures des stations fixes équipées d'instrument de référence ? Certes, ces capteurs sont peu encombrants, légers, ergonomiques, avec une grande autonomie ; mais leurs contraintes techniques, « humaines » et financières permettent-elles d'intégrer la variabilité spatio-temporelle de la mesure de la pollution en ville ? Qu'en est-il de la reproductibilité de ces mesures ? Ces capteurs peuvent-ils finalement nous renseigner sur le niveau de pollution auquel un individu est exposé au cours de sa journée ? Ou *a minima* peuvent-ils mieux quantifier la pollution ?

Des capteurs à bas coût seront mis à disposition de l'observatoire participatif PartiCitaE⁵⁴ en cours de développement à l'institut de la transition environnementale de Sorbonne Université (SU-ITE). Ces micro-capteurs sont actuellement en phase de développement au LATMOS. Le premier volet de cette étude a pour but de contribuer à établir une stratégie d'utilisation de tels capteurs.

Dans cet objectif, nous avons réalisé des campagnes de mesures de particules PM₁₀ et PM_{2,5} à partir d'instruments de référence afin d'étudier la variabilité spatio-temporelle de cette pollution en ville. Et, nous avons également mené une campagne de mesure avec des capteurs mobiles à bas coût (capteurs e-PM). Le retour de ces expériences nous a permis d'identifier

⁵⁴ <http://www.particitae.upmc.fr/fr/participez.html>

les avantages et les limites de capteurs à bas coût (avec des contraintes similaires à celles des capteurs en phase de développement au LATMOS) si l'on souhaite caractériser finement la pollution urbaine. Ces mesures ont été effectuées à l'extérieur et à l'intérieur, en mobilité, dans le métro et également une journée dans un milieu semi-contrôlé : le démonstrateur de Sense-City (décrit en section 3.3.1). Nous exploiterons également les résultats du stage de master 1 de Stéphane Pawlak qui a réalisé des mesures à proximité du campus de la faculté des sciences et ingénierie de Sorbonne Université.

3.2 Campagne de mesures à Paris

3.2.1 Caractéristiques de la période d'étude : conditions météorologiques et niveaux de pollution de fond

Nous avons choisi de mener cette campagne exploratoire neuf jours en mars 2016 car de nombreux épisodes de pollution aux particules ont lieu à cette période. Ces épisodes de pollution sont favorisés par l'utilisation massive du chauffage au bois des ménages et l'accumulation de cette pollution au sol est favorisée par des conditions anticycloniques assez fréquentes à cette période. Notons qu'en mars 2016, les particules issues de l'agriculture, en provenance du Nord de l'Europe (voir plateforme PREV'AIR⁵⁵) ont été transportées jusque dans la région parisienne et de nombreuses alertes particules étaient signalées par AIRPARIF.

Une grande variabilité des particules PM₁₀ et PM_{2.5} a eu lieu pendant la période d'étude, du 7 mars au 16 mars 2016 (**Figure 17**). Les concentrations des PM₁₀ (de 5.7 µg/m³ à 93.8 µg/m³) et PM_{2.5} (2.2 µg/m³ à 73.4 µg/m³) sont très proches sur la période étudiée : les PM₁₀ sont en très grande majorité des particules PM_{2.5}.

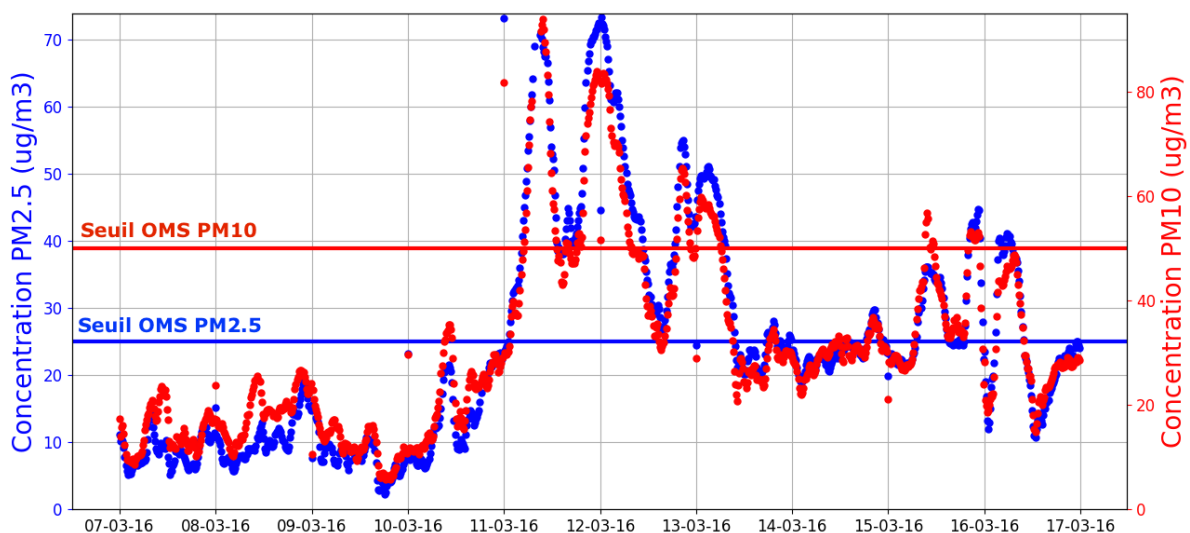


Figure 17 : Évolution temporelle des concentrations de PM₁₀ et PM_{2.5} (en µg/m³) du 07 au 16 mars 2016, au niveau de la station AIRPARIF PA04C, située Place Igor Stravinsky (Paris 4^{ème} arrondissement). Les lignes horizontales indiquent les seuils à ne pas dépasser pour les PM₁₀ (ligne rouge) et les PM_{2.5} (ligne bleue) selon les

⁵⁵ <http://www2.prevoir.org/>

recommandations de l’OMS. En effet, l’OMS recommande de ne pas dépasser ces moyennes journalières plus de 35 jours par an pour les PM₁₀ et cette moyenne annuelle pour les PM_{2,5}.

Le passage d’un système cyclonique à un système anticyclonique du 10 au 11 mars, est caractérisé par de fortes pressions (de 1001 hPa le 10/03/16 à 1012 hPa le 11/03/16) et une absence de vent les 11/03/16 et 12/03/16 (**Tableau 5**). Ces conditions sont favorables à l’accumulation de polluants. Du 10 au 11 mars, la moyenne journalière des particules au niveau de la station de fond AIRPARIF est multipliée par un facteur 2.4 pour les PM₁₀ (22.2 µg/m³ vs 52.8 µg/m³) et par un facteur 3.2 pour les PM_{2,5} (15.2 µg/m³ vs 48.7 µg/m³).

Jours	[PM ₁₀] µg/m ³	[PM _{2,5}] µg/m ³	Pression de surface (hPa)	Direction du vent (vitesse vent (m/s))
07-03-16 lundi	15.7	8.2	993	NO (3.45)
08-03-16 mardi	18.3	10.9	1003	
09-03-16 mercredi	12.2	7.1	989	S(6.04)
10-03-16 jeudi	22.2	15.2	1001	
11-03-16 vendredi	52.8	48.7	1012	
12-03-16 samedi	54	49	1015	
13-03-16 dimanche	36.8	29.7	1016	NE (4.45)
14-03-16 lundi	28.1	23.5	1014	NE (4.52)
15-03-16 mardi	37.2	30.9	1012	NE (3.27)
16-03-16 mercredi	25.9	21.6	1010	NE (5.37)

Tableau 5 : Caractéristiques des jours étudiés, pression de surface, direction et vitesse de vent (en moyenne journalière) sont extraits des ré-analyses d'ERA 5 du modèle ECMWF (0.25°).

3.2.2 Protocole expérimental

Deux des trois instruments déployés pour mesurer les PM₁₀ et PM_{2,5} dans Paris étaient fixes (OSIRIS) : un était positionné dans le local technique du LATMOS de Jussieu (Paris 5^{ème}, Place Jussieu, tour 45-46, 4^{ème} étage) et l’autre à mon domicile (Paris 13^{ème}, Avenue de Gobelins, 2^{ème} étage) (**Figure 18**). Le DustMate, dans sa version mobile était utilisé à domicile, au LATMOS lors de trajets LATMOS-domicile, lors de promenades dans Paris les jours de pic de pollution ou encore dans le métro. Les instruments étaient réglés sur une fréquence d’échantillonnage d’une minute.



Figure 18 : Localisation des sites de mesures : (1) station de mesures de fond AIRPARIF PA04C (4^{ème} arrondissement de Paris), (2) local technique du LATMOS – Sorbonne Université (5^{ème} arrondissement de Paris), (3) domicile : 39 avenue des Gobelins (13^{ème} arrondissement de Paris)

Une phase d'inter-étalonnage a été réalisée avant de commencer cette campagne de mesures pendant une heure (le 07/03/16 de 18h15 à 19h15) avec les trois instruments positionnés dans le local technique du LATMOS. L'OSIRIS du LATMOS servira de référence pour ajuster les données.

L'étalonnage des instruments a été effectué sur des faibles valeurs ; ce qui assure qu'il n'y a pas de problème de filtre encrassé. On présumera par la suite qu'il n'y a pas de problème majeur d'étalonnage du DustMate et de l'OSIRIS placé à domicile (« OSIRIS HOME ») (Figure 19).

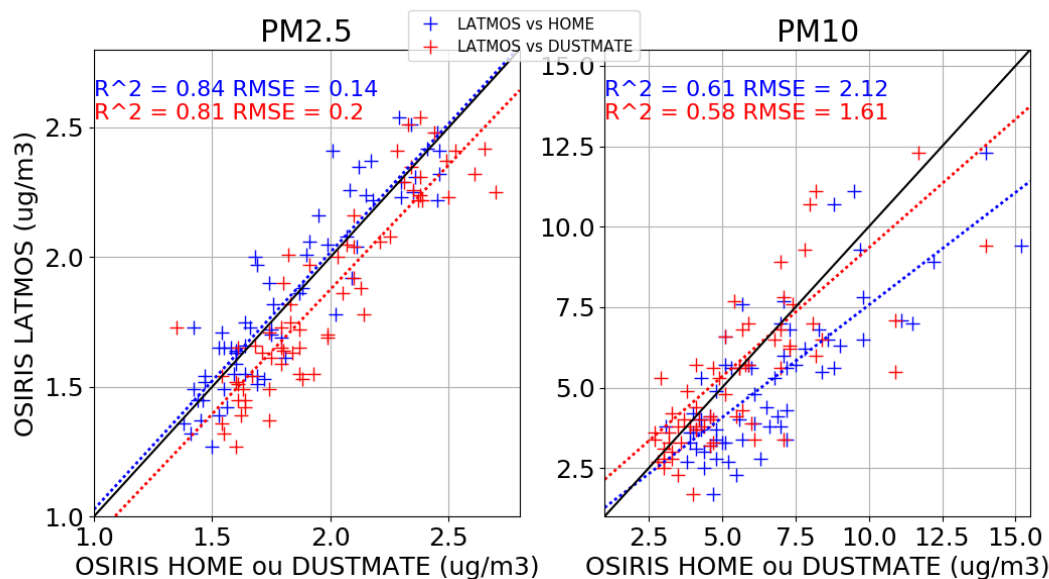


Figure 19 : Concentrations enregistrées par l’OSIRIS positionné au LATMOS (référence) versus concentrations enregistrées par l’OSIRIS positionné à domicile (« Home ») ou du DustMate pour les (a) $PM_{2.5}$ et (b) PM_{10} (en $\mu g/m^3$)

Les concentrations en particules sont reproductibles d’un instrument à l’autre avec des coefficients de corrélation au carré (R^2) très élevés et de la racine carré moyenne des erreurs (root mean square error, RMSE) très faible à la fois pour les PM_{10} et les $PM_{2.5}$:

- $R^2(PM_{2.5})=[0.84 ; 0.81]$
- $R^2(PM_{10})=[0.61 ; 0.58]$
- $RMSE(PM_{2.5})=[0.14 ; 0.2 \mu g/m^3]$
- $RMSE(PM_{2.5})=[2.12 ; 1.61 \mu g/m^3]$.

Lors de ses expériences, Stéphane Pawlak a déployé un OSIRIS et un DustMate côte à côte dans différents endroits à proximité du Campus de la faculté des sciences et ingénierie de Sorbonne Université, lieux représentatifs d’un trafic assez dense afin de réaliser un inter-étalonnage sur des fortes concentrations. La fréquence d’échantillonnage des instruments était paramétrée à une minute. Les concentrations maximales de PM_{10} et $PM_{2.5}$ étaient respectivement $73.8 \mu g/m^3$ et $31.5 \mu g/m^3$. Après filtrage des valeurs aberrantes (moyenne de l’échantillon ± 4 fois l’écart type), les résultats montrent des mesures reproductibles d’un instrument à l’autre :

- $R^2(PM_{10})=0.45$ et $R^2(PM_{2.5})=0.77$
- $RMSE(PM_{10}) = 21.2 \mu g/m^3$ et $RMSE(PM_{2.5})=4.2 \mu g/m^3$.

Ces instruments pourront ainsi être utilisés sur une large gamme de valeurs.

3.2.3 Résultats et discussion

3.2.3.1 Mesures dans le local technique du LATMOS

Dans le local technique du LATMOS les concentrations en PM_{10} ont oscillé entre $0.7 \mu g/m^3$ et $203.8 \mu g/m^3$ et pour les $PM_{2.5}$ entre $0.4 \mu g/m^3$ et $17.3 \mu g/m^3$ (**Figure 20**). Cette variabilité des

concentrations est liée à l'activité humaine dans le local : (1) atelier de soudure avec un maximum pour les PM_{10} à $50.3 \mu\text{g}/\text{m}^3$ (le 08-03-16 à 17 heures) et pour les $PM_{2.5}$ à $17.3 \mu\text{g}/\text{m}^3$ (le 10-03-16 un peu avant 16 heures) et (2) déplacement de personnes. En effet, des niveaux élevés de particules (PM_{10} : $203.8 \mu\text{g}/\text{m}^3$ et $PM_{2.5}$: $14.06 \mu\text{g}/\text{m}^3$ soit respectivement $\times 8$ et $\times 3$ par rapport au niveau de fond) pourraient être attribués à la remise en suspension des particules lors du déplacement de personnes dans la salle car aucune activité particulière n'a été notée sur le planning.

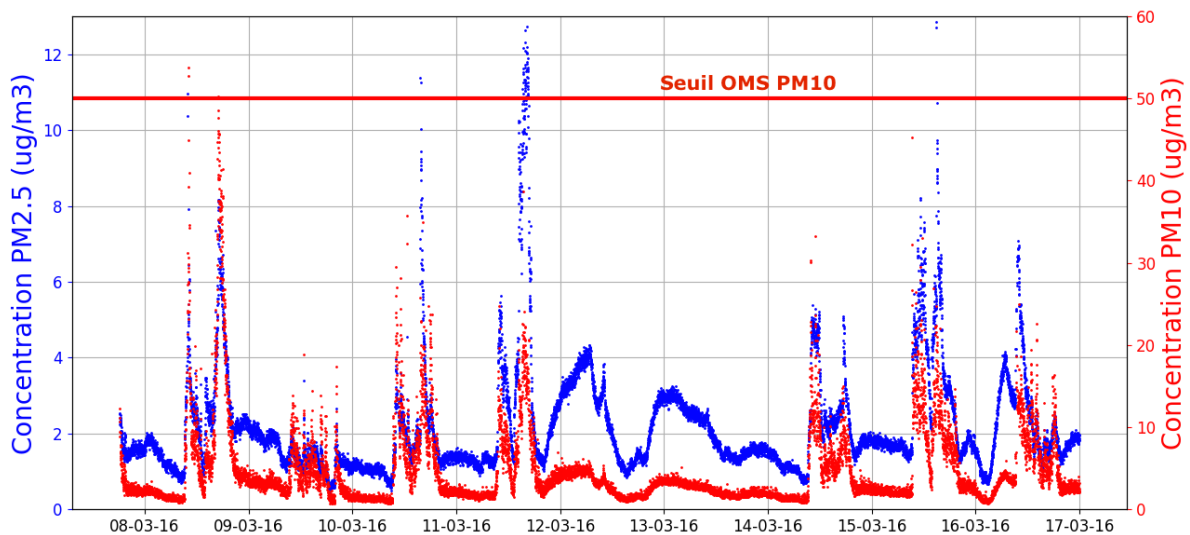


Figure 20 : Évolution temporelle des concentrations de PM_{10} et $PM_{2.5}$ (en $\mu\text{g}/\text{m}^3$) enregistrées par l'OSIRIS positionné au LATMOS (avec des maxima à $17.3 \mu\text{g}/\text{m}^3$ ($PM_{2.5}$) et $203.8 \mu\text{g}/\text{m}^3$ (PM_{10})). La ligne horizontale rouge indique le seuil à ne pas dépasser pour les PM_{10} selon les recommandations de l'OMS. Ici, les valeurs de $PM_{2.5}$ ne dépassent pas le seuil recommandé par l'OMS.

Les jours de week-end (samedi 12 et dimanche 13), la variabilité des $PM_{2.5}$ était caractérisée par un cycle diurne avec des concentrations les plus importantes la nuit et les plus faibles la journée. Lorsqu'aucune activité humaine n'a lieu, les niveaux de $PM_{2.5}$ proviennent essentiellement de l'extérieur où le niveau de particules est plus important la nuit.

En France, la plupart des mesures dans les bureaux ont été réalisées avant février 2007, période pendant laquelle il était autorisé de fumer dans les bureaux. Le tabac étant une source non négligeable d'émissions de particules, on ne mentionnera pas les résultats de ces campagnes. De juin 2013 à novembre 2017, l'OQAI a engagé une campagne nationale dans 129 immeubles de bureaux. Les résultats sont en cours d'exploitation⁵⁶.

En Écosse, en novembre 2012 et mai 2013, deux campagnes de mesures de $PM_{2.5}$ ont été réalisées pour mesurer l'exposition de 17 volontaires (Steinle, 2015). La moyenne des concentrations de $PM_{2.5}$ au travail était de $3 \mu\text{g}/\text{m}^3$, très proche des valeurs observées lors de notre campagne de mesures ($3.6 \mu\text{g}/\text{m}^3$).

⁵⁶ <http://www.inrs-airinterieur2017.fr/wp-content/uploads/2017/12/09-MANDIN.pdf>

3.2.3.2 Mesures à domicile

Les niveaux de pollution ont été enregistrés dans des conditions ordinaires d'exposition aux émissions intérieures. L'idée ici est de caractériser l'exposition intérieure d'un individu dans des conditions réalistes. Notons qu'il s'agit d'un habitat non-fumeur.

Une grande variabilité temporelle des niveaux de pollution de PM_{10} et $PM_{2.5}$ a été observée à l'intérieur du domicile (**Figure 21**) avec des valeurs comprises entre 1.1 et $4458.9 \mu\text{g}/\text{m}^3$ pour les PM_{10} et entre 0.6 et $206.7 \mu\text{g}/\text{m}^3$ pour les $PM_{2.5}$. Les plus fortes concentrations des particules enregistrées sont très supérieures aux maxima enregistrés dans le local technique du LATMOS avec un ratio supérieur à 11 entre les maxima pour les $PM_{2.5}$ ($206.7 \mu\text{g}/\text{m}^3$ vs $18 \mu\text{g}/\text{m}^3$) et un ratio de 22 pour les PM_{10} ($4458.9 \mu\text{g}/\text{m}^3$ vs $206.7 \mu\text{g}/\text{m}^3$). Ces valeurs sont bien au-delà de seuils recommandés de ne pas dépasser par l'OMS. En effet, l'OMS recommande de ne passer en moyenne journalière $50 \mu\text{g}/\text{m}^3$ (pour les PM_{10}) pour les plus de 35 jours par an et en moyenne annuelle $25 \mu\text{g}/\text{m}^3$ (pour les $PM_{2.5}$).

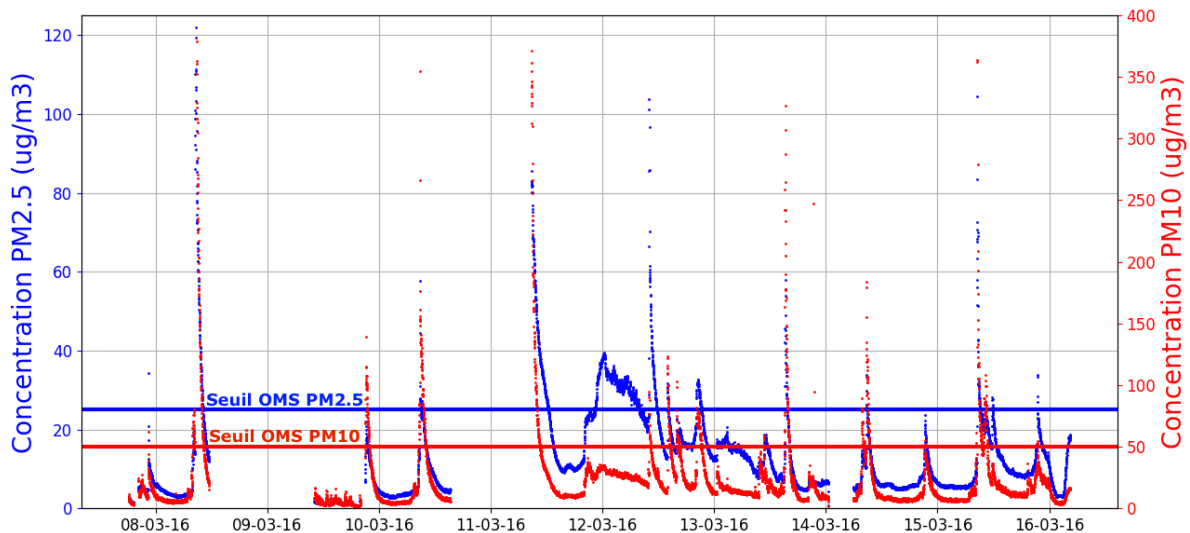


Figure 21 : Évolution temporelle des concentrations de PM_{10} et $PM_{2.5}$ (en $\mu\text{g}/\text{m}^3$) enregistrées par l'OSIRIS positionné au domicile. Les données ont été étalonnées par rapport à l'OSIRIS positionné au LATMOS (avec des maxima à $206.7 \mu\text{g}/\text{m}^3$ ($PM_{2.5}$) et $4458.9 \mu\text{g}/\text{m}^3$ (PM_{10})). Les lignes horizontales indiquent les seuils à ne pas dépasser pour les PM_{10} (ligne rouge) et les $PM_{2.5}$ (ligne bleue) selon les recommandations de l'OMS.

Les niveaux de pollution les plus importants ont été observés (1) lors de l'application d'un spray coiffant (le 08-03-16, $PM_{10} > 4400^* \mu\text{g}/\text{m}^3$ soit 90 fois le niveau de fond et $PM_{2.5}$: $206 \mu\text{g}/\text{m}^3$ soit 41 fois le niveau de fond avant application du spray (avant étalonnage, la mesure était de $6000 \mu\text{g}/\text{m}^3$ ce qui correspond à la valeur de saturation de l'appareil) et (2) lorsqu'on a soufflé sur une bougie (le 07-03-16) avec un maximum pour les PM_{10} à $60 \mu\text{g}/\text{m}^3$ ($\times 12$ par au fond avant de souffler la bougie) et pour les $PM_{2.5}$ à $35 \mu\text{g}/\text{m}^3$ ($\times 7$ par rapport au fond). La cuisson des aliments n'a pas été enregistrée comme une source majeure de pollution. Ceci pourrait s'expliquer par l'emplacement de l'instrument dans le salon et non dans la cuisine et aussi par le fait qu'aucune grillade n'a été réalisée pendant cette période. Des problèmes de pompe de l'OSIRIS expliquent l'absence de mesures les 9 et 10 mars.

Une campagne nationale de mesures dans les logements français a été réalisée de septembre 2003 à décembre 2005 par l'OQAI sur 567 logements⁵⁷ (Kirchner, 2007). Des données valides ont été obtenues pour à peine plus de 50 % des logements, soit 300 logements.

Selon ce rapport, 50 % des logements ont des teneurs en particules supérieures à 31.3 µg/m³ pour les PM₁₀ et 19.1 µg/m³ pour les PM_{2.5} et en moyenne les niveaux étaient de 53.6 µg/m³ pour les PM₁₀ et 37.1 µg/m³ pour les PM_{2.5}. Des variations saisonnières ont été constatées avec un maximum au mois de mars. Les niveaux moyens de particules à domicile obtenus lors de notre campagne sont assez proches des niveaux médians lors de la campagne de l'OQAI (PM₁₀ : 25.7 µg/m³ vs 31.3 µg/m³ et pour les PM_{2.5} : 13.1 µg/m³ vs 10.96 µg/m³).

3.2.3.3 Mesures en mobilité (à l'extérieur et métro) et stationnaire à l'intérieur des bâtiments

Le DustMate a permis d'effectuer des mesures en mobilité à l'extérieur, et stationnaire à l'intérieur. L'instrument, tenu à la main était orienté vers le haut avec le filtre à peu près à hauteur du menton. La plupart des mesures ont été réalisées au LATMOS (les 7, 8, 9, 10 mars, avec un total de 22 heures de mesures collectées), à domicile (le 7, 8, 10, 12 et 13 mars avec près de 17 heures de mesures), lors de déplacements dans Paris – le quartier des Halles et du jardin du Luxembourg – pendant les jours de pics de pollution (2 heures, le samedi et dimanche 12 mars), au restaurant universitaire (2 heures, du 8 au 10 mars), lors des trajets à pied LATMOS-domicile (moins de 2 heures du 7 au 9 mars) puis des trajets LATMOS-restaurant universitaire (les 8, 9, 10 mars, 45 min) et dans le métro (22 minutes les 10 et 12 mars).

Environnements intérieurs : restaurant universitaire, domicile, bureau :

Sur la période étudiée, le restaurant universitaire est en moyenne l'environnement intérieur le plus pollué en PM₁₀ et PM_{2.5} avec respectivement une moyenne à 32.5 µg/m³ et 11.8 µg/m³ (**Tableau 6a et Figure 23**) et des *minima* de concentrations supérieurs au domicile et au bureau. En revanche les *maxima* de concentrations sont observés au domicile (2212.4 µg/m³ pour les PM₁₀ et 237.5 µg/m³ pour les PM_{2.5}) suivi du restaurant universitaire (586.8 µg/m³ pour les PM₁₀ et 43.0 µg/m³ pour les PM_{2.5}). En considérant la moyenne des mesures, le domicile arrive en deuxième place (11.4 µg/m³ pour les PM₁₀ et 3.6 µg/m³ pour les PM_{2.5}). Le local technique du LATMOS avec des niveaux moyens très proches du domicile arrive en troisième place (niveau moyen PM₁₀ : 9.3 µg/m³ et PM_{2.5} : 2.4 µg/m³). La plus grande variabilité des concentrations est enregistrée à la maison avec un écart type de 74.8 µg/m³ (PM₁₀) et 10.5 µg/m³ (PM_{2.5}).

Mobilité : trajets LATMOS-restaurant universitaire, LATMOS-domicile et promenade à pied dans Paris un week-end pollué ainsi que trajet en métro :

Le niveau moyen de pollution le plus élevé lors des trajets à pied a été enregistré lors d'une promenade dans Paris le week-end (pendant un pic de pollution) avec un niveau moyen pour les PM₁₀ de 35 µg/m³ et pour les PM_{2.5} de 30.9 µg/m³ (**Tableau 6c et d**).

Pour les PM₁₀, des niveaux moyens de pollution similaires ont été enregistrés lors des trajets LATMOS-domicile (17.4 µg/m³) et LATMOS-restaurant universitaire (18.7 µg/m³). Alors que pour les PM_{2.5}, le trajet LATMOS-restaurant universitaire était en moyenne plus pollué que le trajet LATMOS-domicile (6.6 µg/m³ vs 3.9 µg/m³). Notons que pour les trajets

⁵⁷ http://www.oqai.fr/userdata/documents/Document_133.pdf

LATMOS-domicile, une partie du trajet du matin s'effectuait dans le Jardin des Plantes de Paris alors que le soir, le parcours emprunté contournait le Jardin des Plantes (fermé) en passant sur le trottoir le long d'un axe routier très fréquenté (rue Geoffroy-Saint-Hilaire).

—
Aparté sur les résultats d'une étude menée par Stéphane Pawlak à partir d'instruments de référence (OSIRIS et DustMate) sur une période de douze jours non successifs de mai à août 2018. Des mesures de particules en statique et en mobilité (marche, trottinette, vélo) ont été réalisées sur une petite zone urbaine à proximité du campus de Jussieu (Figure 22).

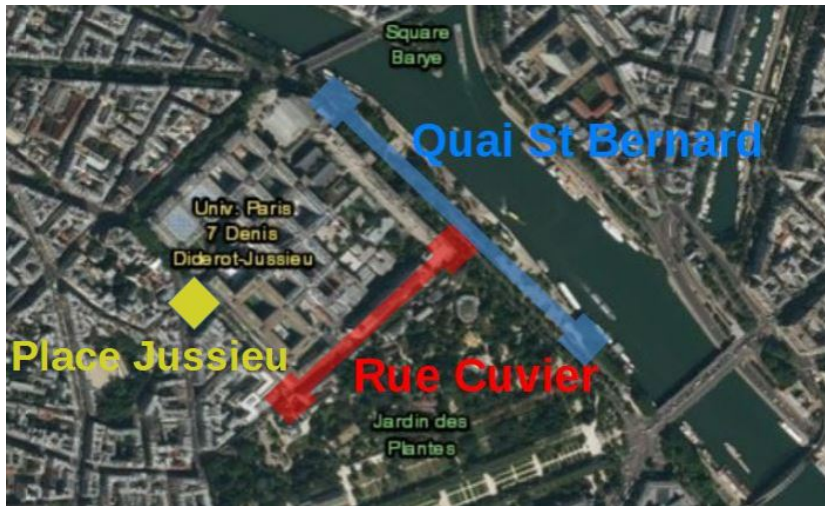


Figure 22 : Plan du quartier du campus Pierre et Marie Curie de Sorbonne Université. Axes routiers : quai St Bernard (en bleu) : trafic fort tout au long de la journée, rue Cuvier (en rouge) : trafic généralement limité (~2/3 voitures par minute), place Jussieu (en jaune) : zone ouverte avec trafic de proximité faible. Source : Stephane Pawlak, Colloque international "Capteurs Individuels de Pollution Atmosphérique : Innovation ou Révolution ?", 28 et 29 novembre 2018.

Des événements ponctuels survenus lors des mesures ne permettaient pas d'exploiter les données d'un seul trajet. Des moyennes sur plusieurs allers-retours sur la même portion de route ont dû être effectuées.

Pour des mesures statiques, la concentration en particules PM_{10} et $PM_{2.5}$ est relativement homogène sur un trottoir d'une dizaine de mètres de largeur, avec un trafic assez dense type Quai Saint-Bernard (5^{ème} arrondissement de Paris).

Les niveaux moyens de concentrations en PM_{10} enregistrés le long du Quai Saint-Bernard en mobilité et statique ($44 \mu\text{g}/\text{m}^3$ avec un écart type de $13 \mu\text{g}/\text{m}^3$) sont plus élevés que lors de notre promenade à Paris en plein pics de pollution aux particules en mars 2016 ($35 \mu\text{g}/\text{m}^3$) ou lors des trajets LATMOS-domicile et LATMOS-restaurant universitaire ($17.4 \mu\text{g}/\text{m}^3$ et $18.7 \mu\text{g}/\text{m}^3$). Pour les $PM_{2.5}$, les niveaux moyens de concentrations enregistrés le long du Quai Saint-Bernard ($13 \mu\text{g}/\text{m}^3$ avec un écart type de $4 \mu\text{g}/\text{m}^3$) sont plus faibles que lors de notre promenade à Paris en plein pics de pollution aux particules en mars 2016 ($30.9 \mu\text{g}/\text{m}^3$) mais sont 2 à 3 fois supérieurs aux niveaux moyens enregistrés lors des trajets LATMOS-restaurant universitaire et LATMOS-domicile ($6.6 \mu\text{g}/\text{m}^3$ et $3.9 \mu\text{g}/\text{m}^3$).

Des forts gradients de PM_{10} ont été détectés lorsqu'on s'éloigne du Quai Saint-Bernard par une rue perpendiculaire à faible trafic (rue Cuvier), avec des écarts de concentration allant de 15 % à 30 % dans la rue Cuvier à 10 m et 40 m du Quai Saint-Bernard. Le gradient pour les $PM_{2.5}$ était négligeable, les émissions trafic étaient probablement recouvertes par les émissions

liées au secteur résidentiel et tertiaire qui sont plus homogènes spatialement. Selon l'inventaire d'AIRPARIF de 2015, le secteur résidentiel et tertiaire représente 33 % des émissions totales pour les PM₁₀ vs 49 % pour les PM_{2,5} et le trafic routier 23 % des émissions totales de PM₁₀ vs 27 % des PM_{2,5}. De plus, au fur et à mesure qu'on s'éloigne de l'axe routier, les PM_{2,5} peuvent s'agglomérer avec d'autres particules pour former des particules plus grosses.

Les niveaux moyens de PM₁₀ et PM_{2,5} les plus élevés (tous milieux confondus : intérieur et mobilité) ont été enregistrés dans le métro avec respectivement des concentrations de 92.3 µg/m³ et 45.3 µg/m³.

Les mesures collectées dans le métro ont été effectuées lors de deux trajets sur uniquement 22 minutes de données au total. Ce temps intègre un temps d'attente sur la voie, un temps à l'intérieur de la rame (3 minutes pour le premier trajet Gobelins – Jussieu et 9 minutes pour le second trajet Gobelins-Châtelet) et une période à la sortie de la rame pour se diriger vers la sortie. Cet échantillonnage ne peut pas être considéré comme représentatif car les mesures ont été enregistrées sur une très courte période. Il s'agit de donner une estimation approximative du niveau de pollution dans le métro et non de quantifier avec précision.

En collaboration avec la RATP, AIRPARIF avait réalisé une campagne de mesures dans le métro du 1^{er} au 15 décembre 2008 dans la station Faidherbe-Chaligny⁵⁸. Un instrument de la RATP et un instrument d'AIRPARIF étaient positionnés de part et d'autre des voies. Une station permanente de mesures de la RATP du réseau SQUALES était aussi positionnée à la station Châtelet ligne 4.

Les moyennes enregistrées sur les quais de la station Faidherbe-Chaligny sont inférieures à celles enregistrées lors de notre expérience (pour les PM₁₀ : 61 µg/m³ et 58 µg/m³ vs 92.3 µg/m³, pour les PM_{2,5} : 25 µg/m³ et 26 µg/m³ vs 45.3 µg/m³). L'écart entre les mesures observées à la station Châtelet et nos mesures pour les PM₁₀ est plus resserré (107 µg/m³ vs 92.3 µg/m³). Ces niveaux importants de particules sont émis dans le métro lors de l'arrivée d'une rame provoqués par le système de freinage et la remise en suspension des particules au sol.

a)

Statistiques PM ₁₀ (µg/m ³)	Restaurant universitaire	Domicile	Bureau LATMOS-Jussieu
Moyenne	32.5	11.4	9.3
Ecart type	51.0	74.8	18.1
Médiane	23.8	2.8	5.8
Minimum	13.6	1.8	1.8
Maximum	586.8	2212.4	481.8
Nb mesures	129	1019	1313

b)

Statistiques PM _{2,5} (µg/m ³)	Restaurant universitaire	Domicile	Bureau LATMOS-Jussieu
Moyenne	11.8	3.6	2.4
Ecart type	6.8	10.5	1.9
Médiane	9.3	1.6	1.9
Minimum	4.7	0.7	0.5
Maximum	43.0	237.5	24.1
Nb mesures	129	1019	1313

⁵⁸ http://www.airparif.fr/pdf/publications/Rratp_20090701.pdf

c)				
Statistiques PM₁₀ (µg/m³)	LATMOS/ restaurant universitaire	LATMOS/ Domicile	Promenade Week-end	Métro
Moyenne	17.4	18.7	35.0	92.3
Ecart type	7.3	38.3	14.9	61.8
Médiane	16.1	5.3	33.1	89.8
Minimum	6.9	1.9	10.9	18.9
Maximum	40.6	286.8	96.8	292.5
Nb mesures	45	115	134	22

d)				
Statistiques PM_{2,5} (µg/m³)	LATMOS/ restaurant universitaire	LATMOS/ Domicile	Promenade Week-end	Métro
Moyenne	6.6	3.9	30.9	45.3
Ecart type	3.5	3.8	13.7	30.2
Médiane	6.2	1.7	29.2	40.3
Minimum	1.6	0.7	10.1	14.1
Maximum	20.3	24.0	55.9	147.8
Nb mesures	45	115	134	22

Tableau 6 : Nombre de mesures (« Nb mesures ») avec une fréquence d'échantillonnage d'une minute, moyenne, écart type, médiane, minimum et maximum des concentrations en PM₁₀ et PM_{2,5} (en µg/m³) à l'intérieur (a et b), lors de trajets à pied et en métro (c et d).

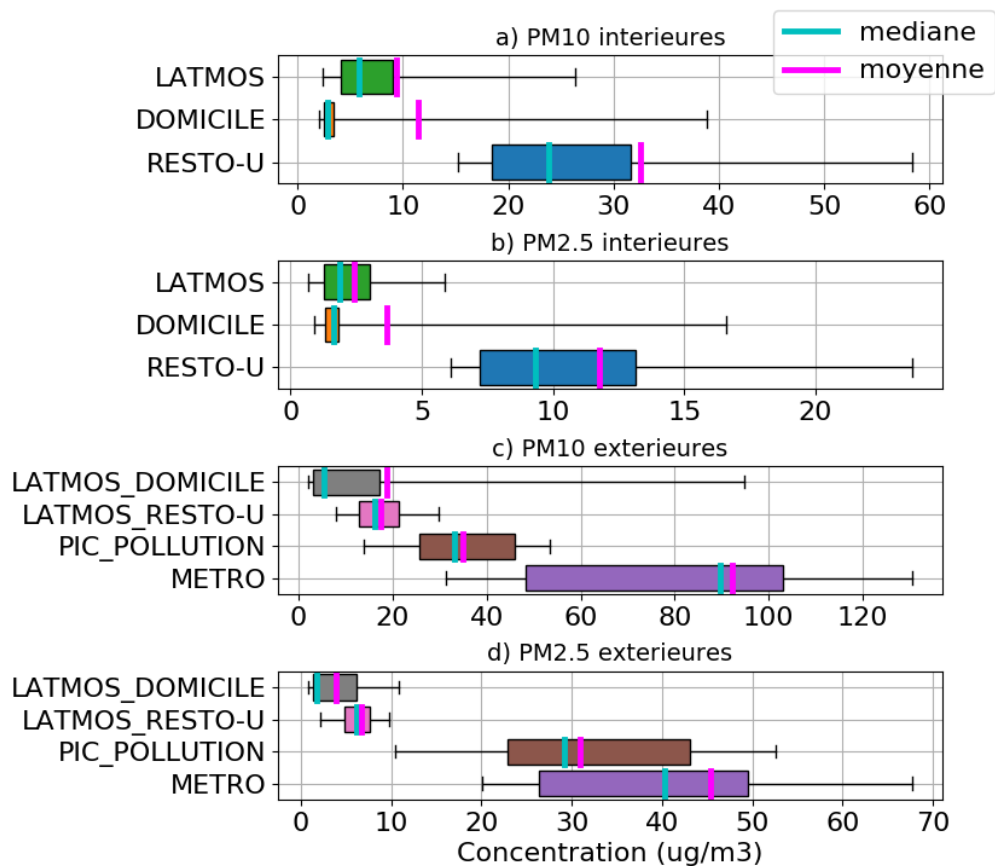


Figure 23 : Boîte à moustache avec la moustache qui indique le 5^{ème} et 95^{ème} percentile. Les autres valeurs extrêmes ne sont pas représentées. Concentrations en (a) PM_{10} et (b) $\text{PM}_{2.5}$ des environnements intérieurs. Concentrations en (c) PM_{10} et (d) $\text{PM}_{2.5}$ lors de déplacements à l'extérieur (en $\mu\text{g}/\text{m}^3$)

3.3 Campagne de mesures dans le démonstrateur de Sense-City, à Champs-sur-Marne (77)

3.3.1 Sense-City : un équipement d'excellence pour concevoir la ville de demain

Sense-City est un projet d'Équipement d'Excellence du Programme d'Investissement d'Avenir de l'Agence Nationale de la Recherche (ANR) unique en Europe. Ce projet qui a débuté en 2011 a reçu un financement de 9 millions d'euros pour 8 ans. Ce projet implique l'Institut Français des Sciences et Technologie des Transports et l'Aménagement et des Réseaux (IFSTTAR), l'ESIEE Paris, le laboratoire CNRS-Polytechnique sur les nanomatériaux (LPICM), le Centre Scientifique et Technique de Bâtiment (CSTB), l'Institut

National de Recherche en Information et en Automatique (INRIA) et l'Université Paris-Est-Marne-la-Vallée.

Sense-City est une chambre climatique mobile pouvant recouvrir deux mini-villes de 400 m². Ces deux mini-villes sont complémentaires et n'ont pas la même géométrie ce qui permet de mettre en œuvre des projets scientifiques différents. La première de ces mini-villes (**Figure 24**) inaugurée le 3 avril 2018, est composée de maisons en terre cuite, mobilier urbain innovant, de végétation et de route, équipés de nombreux capteurs⁵⁹. Dans l'autre mini-ville, un sous-sol sera équipé d'une nappe phréatique afin d'étudier la détection de réseaux souterrains (câbles, canalisations), la géothermie et la pollution des sols. Cet équipement a pour objectif de tester de nouveaux matériaux urbains, d'étudier la pollution en ville (air, eau, sols) avec entre autre la conception et l'amélioration des micros et nano-capteurs. Le dispositif est équipé d'une chambre climatique mobile, ce qui permet de réaliser à la fois des expériences en condition réelle à l'air libre ou bien en condition contrôlée. En condition contrôlée, des événements météorologiques ordinaires ou extrêmes (tels que les canicules) pourront être reproduites dans ces mini-villes et ainsi évaluer leur impact sur la pollution, les végétaux ou la calibration des capteurs par exemple.

Dans le cadre de ce travail de thèse, une campagne de mesures d'une journée en juin 2016 a été réalisée dans le démonstrateur de Sense-City (**Figure 24c**). La chambre climatique et les deux mini-villes étaient en phase de construction lors de ma campagne de mesures. Le démonstrateur de Sense-City, en condition climatique réelle, est une mini-ville de 250 m² équipée de 80 capteurs déployés sur les routes et dans deux cabanes.

⁵⁹ <http://sense-city.ifsttar.fr/>

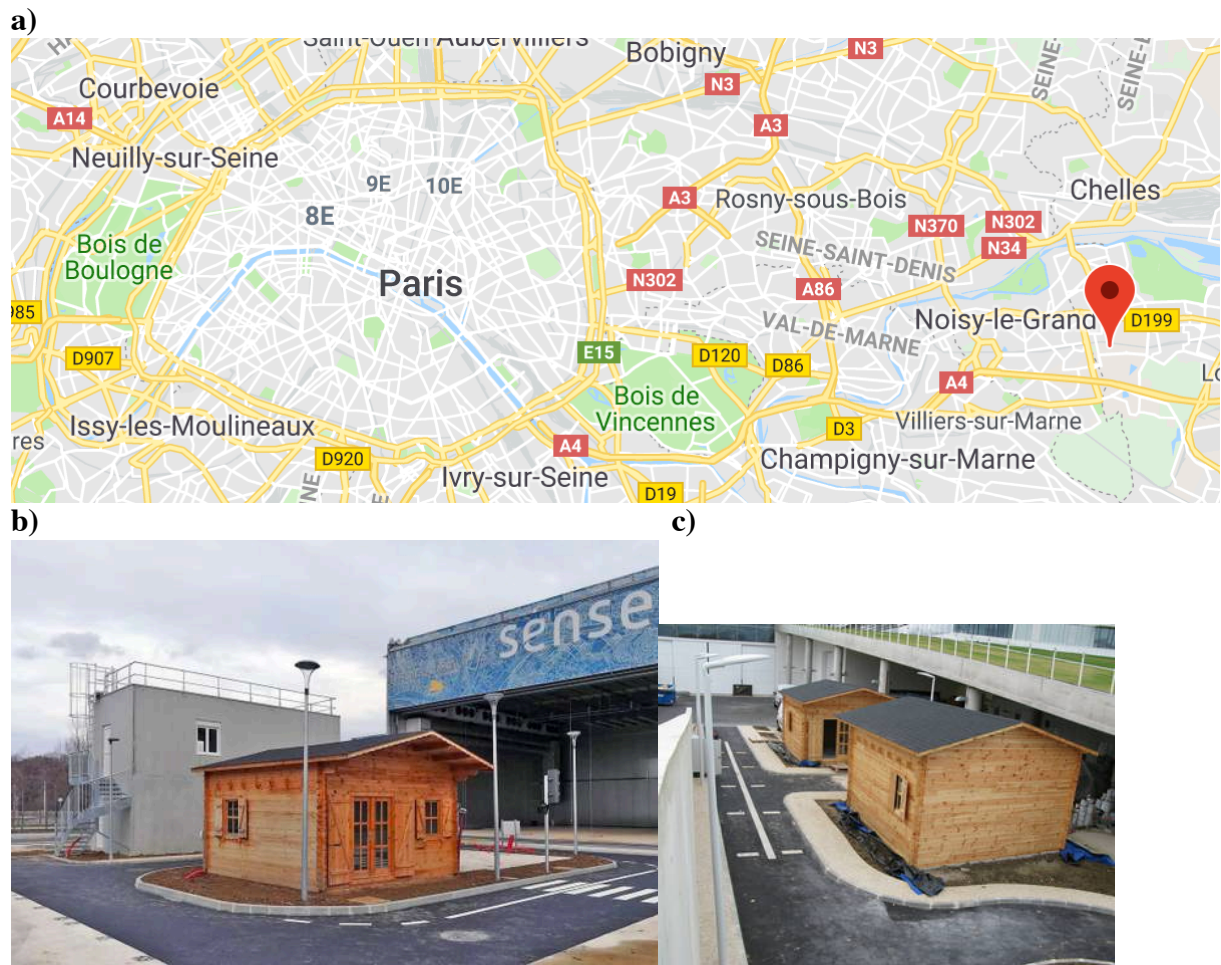


Figure 24 : Localisation de Sense-City dans la région Île-de-France, plan Google map (a). Photographie de la première mini-ville de Sense-City et la chambre climatique (b) et du démonstrateur de Sense-City (c)

3.3.2 Protocole expérimental

Deux instruments de référence (deux OSIRIS) ont été déployés lors de cette campagne de mesures, le 22 juin 2016. Un des instruments était placé à l'intérieur d'un des deux chalets du démonstrateur de Sense-City. Un deuxième instrument était placé à l'extérieur, sur la route du démonstrateur (**Figure 25**).

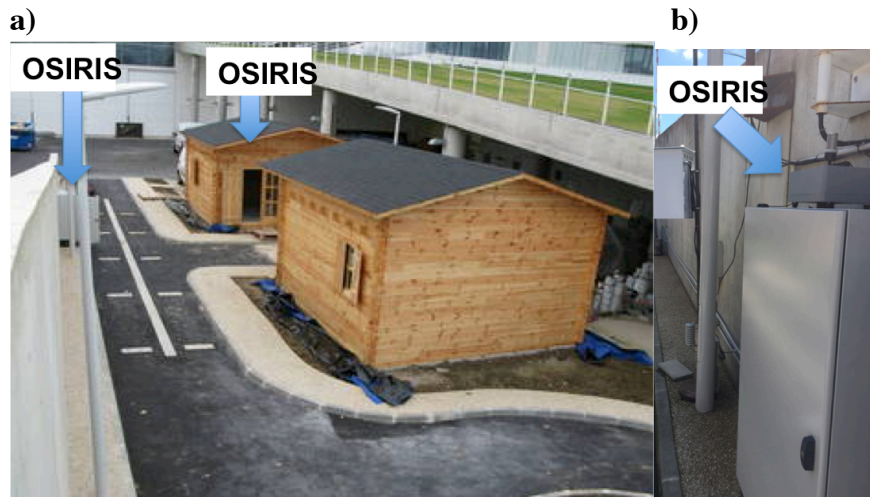


Figure 25 : Positionnement des instruments de référence (OSIRIS) dans le démonstrateur de Sense-City : à l'extérieur du chalet et à l'intérieur du chalet (a). Photographie de l'OSIRIS au niveau de son emplacement à l'extérieur du chalet (b)

De plus, des capteurs e-PM (voir les caractéristiques de ces capteurs dans le tableau descriptif des instruments, section 1.3.1) prêtés par l'IFSTTAR, dans le cadre de Sense-City, ont été accrochés autour du cou de trois volontaires du LATMOS - Simon Guillier, Jules Fillette et moi-même - à l'aide de lacets/cordons (**Figure 26**). Le temps d'acquisition des instruments de mesures est répertorié dans le **Tableau 7**.



Figure 26 : Photographie à l'intérieur d'un des deux chalets du démonstrateur de Sense-City avec les capteurs e-PM accrochés autour du cou et le positionnement de l'instrument de référence (OSIRIS)

Instruments	Temps d'acquisition
OSIRIS	1 mesure/min
e-PM	1 mesure/2min

Tableau 7 : Temps d'acquisition de l'OSIRIS et des capteurs e-PM

Nous avons défini plusieurs scénarios de mesures (**Tableau 8** et **Figure 27**). Dans le chalet utilisé lors de notre expérience, la VMC grande vitesse fonctionnait avec un renouvellement d'air de 160 m³/h. La durée de chaque scénario était d'environ 15 minutes.

Scénarios

1/ Combustion domestique : bougie et encens

2/ Cuisson

3/ Arrêt de la VMC

4/ Chauffage puissance maximale puis aération

5/ Combustion domestique : encens puis aération

6/ Bricolage puis aération

7/ Ménage : aspirateur, serpillère ; carreau vitre puis aération

8/ Ventilateur puis aération

9/ Chauffage puissance maximale

Tableau 8 : Définition des scénarios dans un de chalet du démonstrateur de Sense-City



Figure 27 : Photographies (a),(b),(c) des scénarios de cuisine, bricolage et aspirateur et (d) de la phase d'aération

Les capteurs e-PM n'étant pas étalonnés juste avant la campagne de mesures, un inter-étalonnage des instruments a dû être réalisé ultérieurement, du 25 au 29/07/2016. Les capteurs e-PM étaient placés au côté d'un OSIRIS, celui placé à l'intérieur du chalet lors de la campagne de mesure du 22/06/2016, et d'un LOAC (Light Optical Aerosol Counter⁶⁰) (Figure 28). Les trois capteurs e-PM étaient surélevés sur des cartons afin d'effectuer le prélèvement d'air au même endroit pour l'ensemble du dispositif instrumental. Le Tableau 9 rappelle la fréquence d'échantillonnage des instruments.

Instruments	Temps d'acquisition
OSIRIS	1 mesure/min
e-PM	1 mesure/5min

Tableau 9 : Temps d'acquisition de l'OSIRIS et des capteurs e-PM pendant la phase d'étalonnage

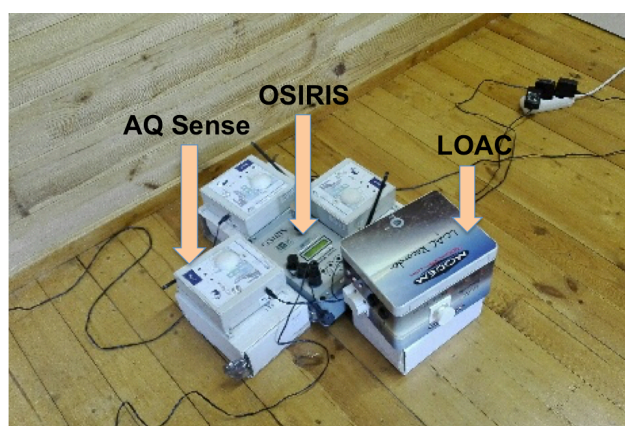


Figure 28 : Déploiement instrumental au sol pendant l'inter-étalonnage

Les capteurs e-PM sont plus bruités que l'OSIRIS avec des distributions plus larges que celles de l'OSIRIS et des plus fortes valeurs pour les e-PM (Figure 29).

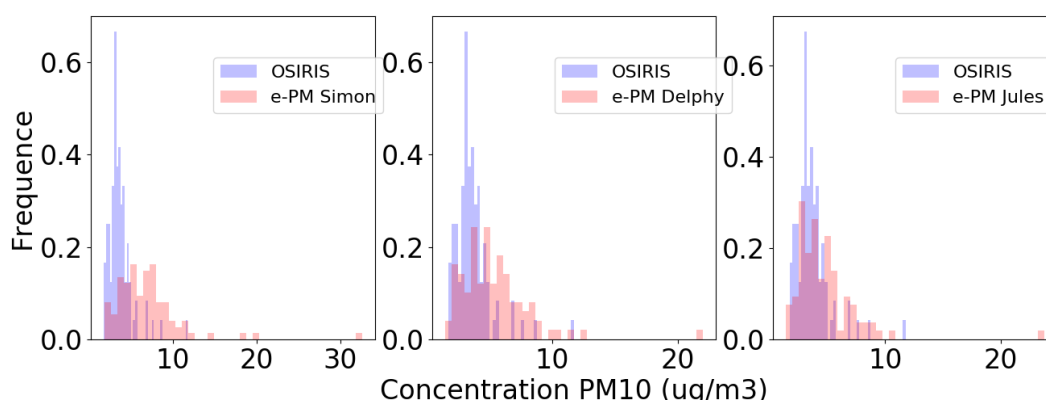


Figure 29 : Histogramme sur les données brutes des capteurs e-PM superposées à celles de l'OSIRIS (en $\mu\text{g}/\text{m}^3$)

⁶⁰ <https://loac.fr/principe.html>

Nous avons réalisé l'inter-étalonnage pour les concentrations supérieures à la limite de quantification des capteurs e-PM (concentrations supérieures à $5 \mu\text{g}/\text{m}^3$). Pour diminuer l'impact du bruit sur les données des capteurs e-PM, nous avons calculé les régressions linéaires OSIRIS (référence) versus capteurs e-PM à partir des données moyennées sur différents pas de temps de [5, 10, 15, 20, 25, 30,35] minutes (**Tableau 10**).

e-PM PM ₁₀	5 min	10 min	15 min	20 min	25 min	30 min	35 min
Delphy	0.36	0.50	0.56	0.69	0.61	0.79	0.79
Simon	0.41	0.49	0.71	0.71	0.74	0.71	0.73
Jules	0.56	0.64	0.71	0.76	0.72	0.81	0.81
Nb points	131	65	40	31	23	20	18

Tableau 10 : Coefficient de corrélation des concentrations en PM₁₀ de l'OSIRIS versus celles des capteurs e-PM pour plusieurs pas de temps et nombre de points (« nb points »)

L'OSIRIS n'a mesuré aucune concentration en PM_{2,5} supérieure à $5 \mu\text{g}/\text{m}^3$ (correspondant à la limite de quantification des e-PM) lors de cette phase post-étalonnage. De plus, les capteurs e-PM semblent avoir des difficultés à faire la distinction entre PM₁₀ et PM_{2,5} alors que l'OSIRIS faisait nettement la différence entre les deux (**Figure 30**). Pour ces deux raisons, nous exploiterons uniquement les mesures de PM₁₀ des capteurs e-PM.

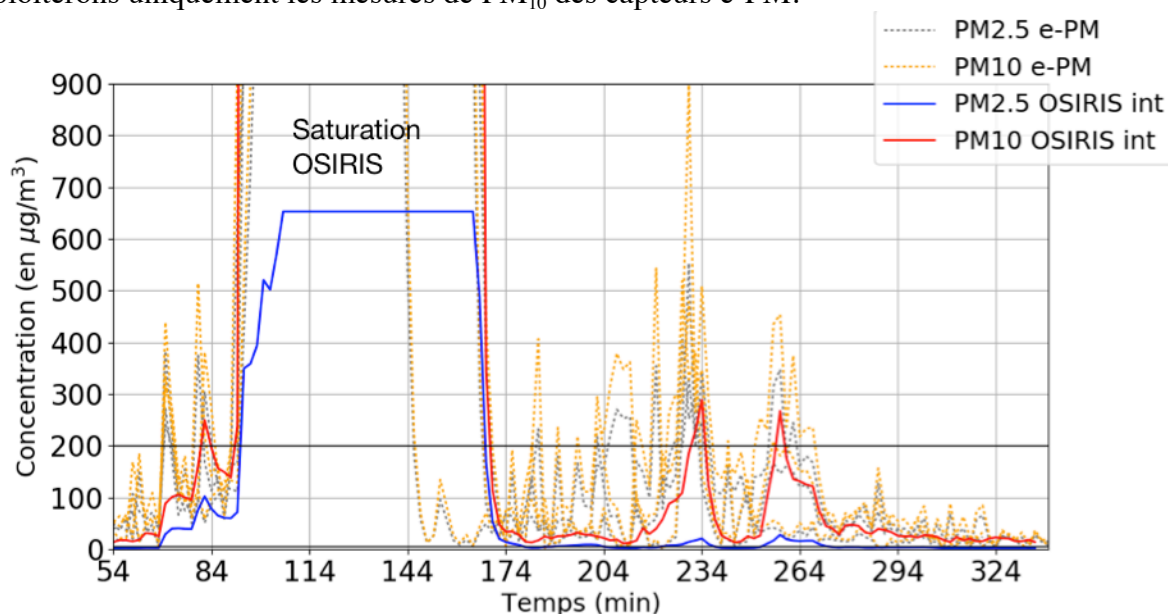


Figure 30 : Données brutes des concentrations des PM₁₀ et PM_{2,5} (en $\mu\text{g}/\text{m}^3$), le 22 juin 2016

L'étalonnage des capteurs e-PM à partir des moyennes toutes les 35 minutes des concentrations en PM₁₀ de l'OSIRIS montre la plus forte corrélation (**Figure 31** et **Tableau 10**) calculée à partir de dix-huit points dans la régression OSIRIS vs capteur e-PM. Nous retiendrons donc les coefficients de régressions obtenus à partir de valeurs moyennées toutes les 35 minutes pour étalonner les données obtenues lors de la journée de campagne de mesures.

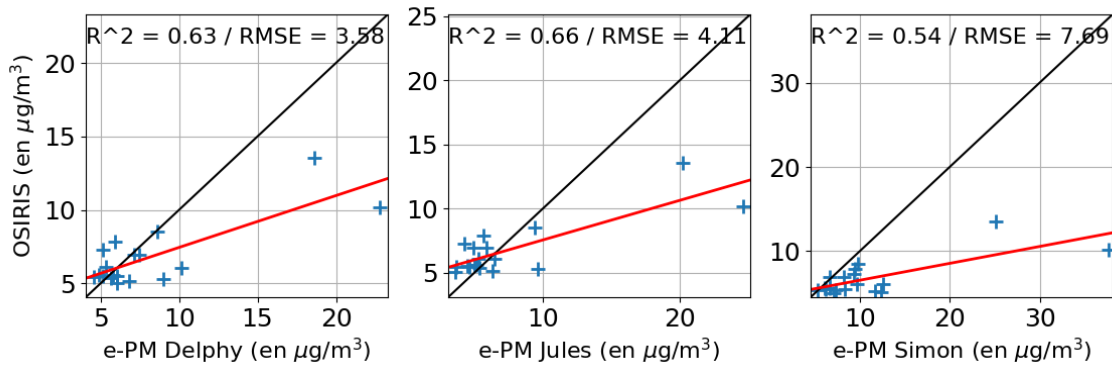


Figure 31 : Concentrations en PM_{10} (en $\mu g/m^3$) supérieures à $5 \mu g/m^3$: OSIRIS versus capteurs e-PM, moyennes par pas de 35 minutes. La ligne noire représente une droite $y=x$. La ligne rouge représente la droite de régression linéaire $y=ax+b$ pour chaque e-PM.

Les concentrations PM_{10} post-étalonnage des capteurs e-PM sont représentées **Figure 32** pour des concentrations supérieures à la limite de quantification des e-PM. Les concentrations PM_{10} des trois capteurs e-PM sont très proches les unes des autres. Les capteurs e-PM semblent reproduire la variabilité des concentrations PM_{10} mesurée par l'OSIRIS.

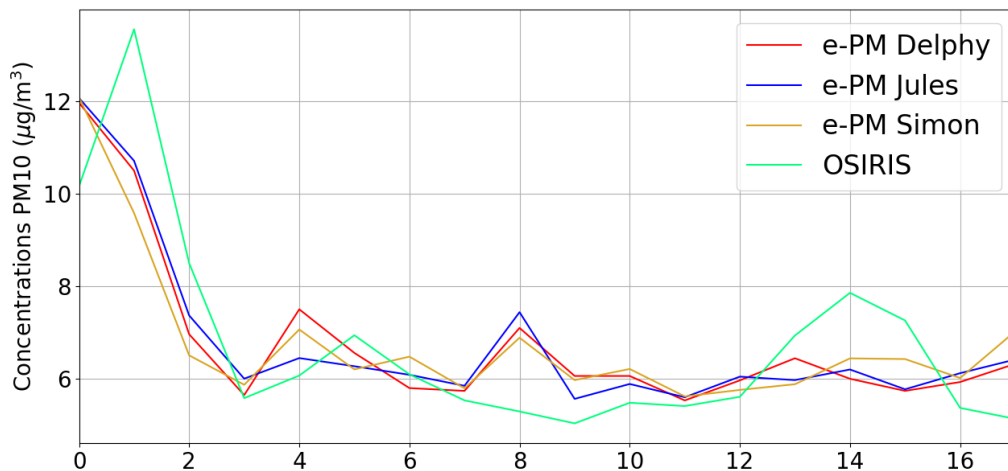


Figure 32 : Concentrations en PM_{10} (en $\mu g/m^3$) de l'OSIRIS et des capteurs e-PM post étalonnage moyennées par pas de 35 minutes pour les concentrations en PM_{10} de l'OSIRIS supérieures à $5 \mu g/m^3$

3.3.3 Résultats et discussion

Cuisine

La cuisine (en plus de la fumée de tabac) est une source majeure de pollution aux particules à l'intérieur des logements. La cuisson du steak est bien en dehors de la gamme de mesures des concentrations en PM_{10} des capteurs e-PM de $200 \mu g/m^3$ ($> 6000 \mu g/m^3$ donnée brute

enregistrée par l'OSIRIS). Nous analyserons donc les séquences de mesures avant et après la cuisson du steak (**Figure 33**).

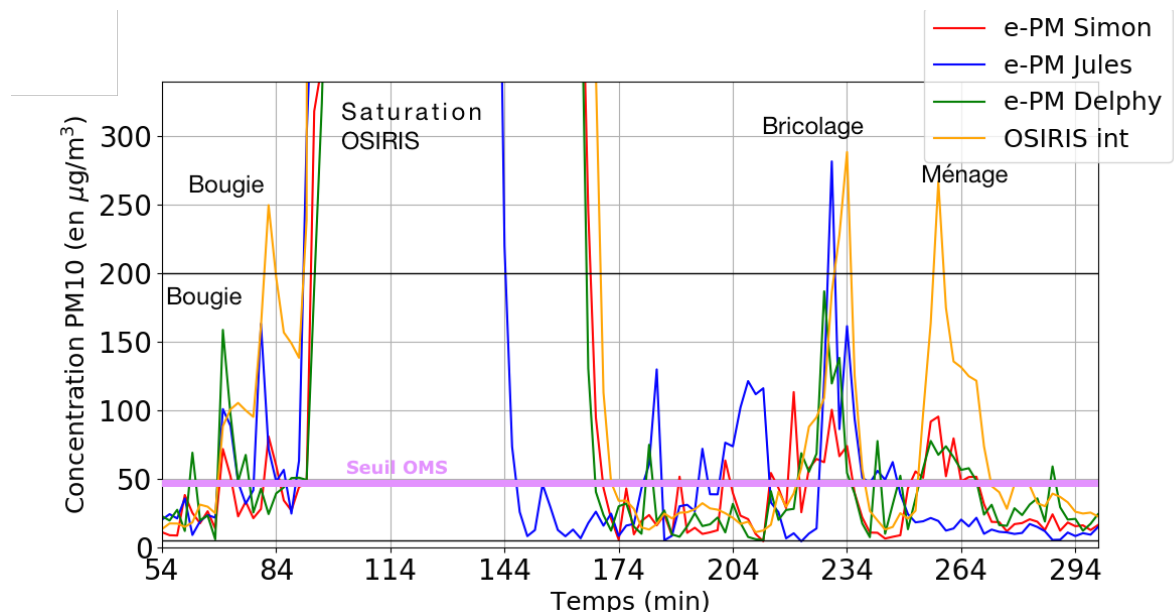


Figure 33 : Concentrations en PM_{10} (en $\mu g/m^3$) étalonnées à partir des coefficients de régression calculées sur des moyennes de 35 minutes pour les concentrations supérieures à $5 \mu g/m^3$. Les lignes horizontales noires représentent le minimum et le maximum de la gamme de mesures de l'e-PM. La ligne horizontale violet clair représente le seuil qu'il est recommandé de ne pas dépasser plus de 35 jours par an (en moyenne journalière) selon l'OMS.

Bougie

Les deux premiers pics observés (**Figure 33**) ont été enregistrés au moment où l'on a soufflé sur la bougie, et non lorsque la bougie se consumait. La première fois, à $t=70$ min c'est moi-même qui ai soufflé sur la bougie. J'étais la plus exposée aux PM_{10} , avec une concentration à $157 \mu g/m^3$ ($\times 26$ avant/après). Jules mesurait $100 \mu g/m^3$ de PM_{10} et Simon $72 \mu g/m^3$ alors que l'instrument de référence à l'intérieur du chalet enregistrait $88 \mu g/m^3$.

Dix minutes plus tard, Jules a soufflé sur la bougie et a été exposé à $164 \mu g/m^3$ de PM_{10} ($\times 4$ avant/après). Au même moment, j'étais exposée à $43 \mu g/m^3$, Simon $28 \mu g/m^3$, et l'instrument de référence/OSIRIS enregistrait $162 \mu g/m^3$ pour atteindre un maximum 2 minutes plus tard à $250 \mu g/m^3$. Le premier pic de l'OSIRIS (à $t=70$ min) est redescendu moins vite que celui des capteurs e-PM et n'est pas redescendu à la même valeur de fond. Ce qui explique que le second pic de l'OSIRIS, légèrement décalé par rapport aux capteurs e-PM surestime ceux des capteurs e-PM (n'étant pas parti de la même valeur initiale).

Aération

Jules enregistre deux pics en PM_{10} à $129 \mu g/m^3$ ($t=184$ min) et à $121 \mu g/m^3$ ($t=208$ min). Ces pics ne sont pas associés à des activités mais apparaissent à l'issue de période d'aération. Une hypothèse est que Jules a peut-être fermé les fenêtres à ce moment-là, et que lors de son déplacement il aurait soulevé des particules avec lui.

Bricolage

Simon enregistrerait $113 \mu g/m^3$ ($t=220$ min) lorsqu'il sciait une planche en bois. Des fortes concentrations en PM_{10} ont été enregistrées pendant 10 minutes lors des activités de bricolage

(Jules : $281 \mu\text{g}/\text{m}^3$ à $t=230 \text{ min}$, Simon et moi : $100 \mu\text{g}/\text{m}^3$ et $120 \mu\text{g}/\text{m}^3$ à $t=228 \text{ min}$, OSIRIS : $288 \mu\text{g}/\text{m}^3$ à $t=234 \text{ min}$).

Ménage

Un nouvel épisode de concentration en PM_{10} était enregistré lors de l'activité de ménage ($t=258 \text{ min}$) avec la plus forte concentration enregistrée par l'OSIRIS ($267 \mu\text{g}/\text{m}^3$) suivi par Simon ($94 \mu\text{g}/\text{m}^3$) alors que moi-même et Jules enregistrions respectivement des concentrations à $68 \mu\text{g}/\text{m}^3$ et $18 \mu\text{g}/\text{m}^3$.

À partir du scénario 4 – chauffage à puissance maximale - nous avons choisi d'aérer pour atteindre une valeur de fond plus basse entre chaque scénario et éviter d'avoir des résidus de l'activité précédente.

Les scénarios d'arrêt de la VMC et le premier scénario de chauffage à puissance maximale n'ont pas pu être évalués car les appareils étaient encore à saturation suite aux fortes concentrations accumulées lors de la cuisson du steak. L'utilisation du ventilateur et le second épisode de chauffage à puissance maximale n'ont pas généré de pic de pollution aux PM_{10} .

Nous constatons une grande hétérogénéité des concentrations en PM_{10} mesurées par l'OSIRIS et entre chaque capteur e-PM. La plupart du temps, le sujet le plus proche des sources aux PM_{10} enregistre les plus fortes concentrations.

Rapport des concentrations en PM_{10} intérieure/extérieure

Les données de l'OSIRIS positionné à l'extérieur ont été étalonnées à partir des coefficients de régression obtenus lors de la phase de calibration de mars 2016 (voir section 3.2.2). Le **Figure 34** est la série temporelle du rapport des concentrations en particules intérieure et extérieure.

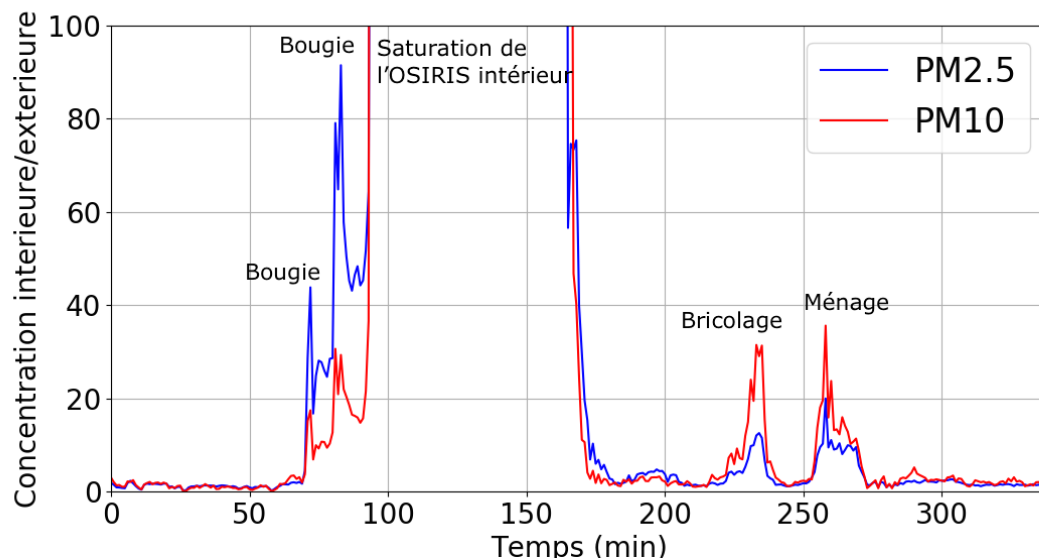


Figure 34 : Rapport des concentrations intérieure/extérieure des deux OSIRIS (intérieur et extérieur) à partir des données étalonnées, le 22 juin 2016

En moyenne, en retirant le scénario de cuisson, les concentrations en PM_{10} sont 3.4 fois supérieures dans le chalet par rapport à l'extérieur ($49.8/14.5 \mu\text{g}/\text{m}^3$) et le ratio intérieur/extérieur est de 4.7 pour les $\text{PM}_{2.5}$ ($9.5/2.0 \mu\text{g}/\text{m}^3$).

3.4 Retour d'expérience en vue de l'élaboration des capteurs mobiles

3.4.1 Principaux résultats

Lors de la campagne de mars 2016, l'OSIRIS qui était positionné à mon domicile a eu plusieurs problèmes de pompe et s'est donc arrêté de fonctionner à plusieurs reprises. Il fallait remettre en service l'instrument, ce qui ne permettait pas une utilisation en continu sans un suivi régulier. Nous avons également constaté que le DustMate dans sa version mobile est un instrument très bruyant qui ne peut être allumé en continu pendant plusieurs heures à cause de la gêne sonore.

Les concentrations les plus fortes en particules ont été enregistrées à l'intérieur des locaux pendant les neuf jours des campagnes en mars 2016 avec un maximum enregistré à domicile, lors de l'utilisation de spray ($> 4400 \mu\text{g}/\text{m}^3$ pour les PM_{10}) alors que l'OMS recommande de ne pas dépasser 35 jours par an les moyennes journalières de $50 \mu\text{g}/\text{m}^3$ pour les PM_{10} et de ne pas dépasser la moyenne annuelle de $25 \mu\text{g}/\text{m}^3$ pour les $\text{PM}_{2.5}$. L'utilisation de bougies, le taux de présence élevé des occupants dans le local du LATMOS ont également une influence sur le taux de particules.

Parmi les différents environnements intérieurs sondés (local du LATMOS, restaurant universitaire, domicile), en moyenne les concentrations les plus importantes sont enregistrées dans le restaurant universitaire (PM_{10} : $32.5 \mu\text{g}/\text{m}^3$ et $\text{PM}_{2.5}$: $11.8 \mu\text{g}/\text{m}^3$). Le métro utilisé dans cette étude (se limitant à quelques stations) s'avère être un moyen de transport très pollué.

Dans le démonstrateur de Sense-City, le 22 juin 2016, la cuisson du steak était identifiée comme étant l'activité la plus polluante alors que le chauffage à puissance maximale n'a pas eu d'impact sur le niveau en particules. Les capteurs e-PM sont limités dans leur utilisation en raison de leur gamme instrumentale de mesure (de $5 \mu\text{g}/\text{m}^3$ à $200 \mu\text{g}/\text{m}^3$). Bien que le maximum ($200 \mu\text{g}/\text{m}^3$) corresponde à 4 et 8 fois les seuils recommandés de ne pas dépasser respectivement pour le PM_{10} (en moyenne journalière, 35 jours par an) et $\text{PM}_{2.5}$ (en moyenne annuelle) par l'OMS. Cette gamme de mesure ne couvre que certaines activités d'intérieur et pas la cuisson d'un steak ou encore certaines activités de bricolage.

Nous avons observé une grande variabilité entre les capteurs e-PM et l'instrument de référence expliquée par la distance entre le capteur et la source de pollution et la dérive thermique de ces capteurs dans le temps.

L'OSIRIS, qui est pourtant un instrument de référence, arrive aussi à saturation lors de la cuisson d'un steak ou de l'utilisation de spray fixant pour cheveux.

Étude menée par Stéphane Pawlak à partir d'instruments de référence en mai-août 2018 :

Plusieurs allers-retours ont été nécessaires pour établir une cartographie correcte sur les différents tronçons de route étudiés : Quai Saint-Bernard (trafic dense) et rue Cuvier (faible trafic et perpendiculaire au Quai Saint-Bernard).

Les mesures obtenues en mobilité : marche à 5 km/h ou trottinette à 10 km/h ne sont pas bruitées pour un pas d'échantillonnage à 5 secondes et permettent d'intégrer la variabilité spatio-temporelle sur une longue distance et donc estimer l'exposition.

Pour un pas d'échantillonnage d'1 minute, correspondant à la résolution temporelle caractéristique d'un capteur à bas coût, la mesure collectée à pied (5 km/h) détecte la variabilité spatio-temporelle. En revanche, si l'on est en trottinette (10 km/h) ou en vélo

(20 km/h), cet échantillonnage n'est pas suffisamment fin pour détecter la variabilité spatio-temporelle rencontrée sur le trajet sauf si on est à l'arrêt. L'information collectée indiquera tout de même un niveau moyen d'exposition de l'individu qui porte le capteur sur le chemin traversé.

3.4.2 Contraintes à intégrer dans les protocoles de mesures

Le développement de capteurs à bas coût est en plein essor et s'explique par les besoins : (1) d'urbanisme et de développement de la ville en cartographiant finement la pollution dans la ville afin de valider les modèles numériques et les cadastres d'émissions et (2) sanitaires en caractérisant l'exposition des individus. Les personnes vulnérables ou des patients atteints de maladies respiratoires pourraient adapter leur comportement en cas de pics de pollution ou encore lors de niveaux quotidiens élevés de pollution en leur suggérant d'emprunter un itinéraire moins pollué.

L'objectif de ce retour d'expérience est d'anticiper les limites des capteurs à bas coût en vue de définir des protocoles de mesures pour ce type de capteurs. Nous pouvons également nous appuyer sur un rapport complémentaire à notre étude publié par l'organisation mondiale de météorologie (WMO, 2018⁶¹). Ce rapport précise l'utilisation qui peut être faite de tels capteurs et préconise de s'assurer des bonnes conditions d'utilisation de ces capteurs. En effet, l'utilisation de ces capteurs se heurte à des contraintes techniques, « humaines » et financières.

- **Des contraintes métrologiques**

Nous avons remarqué qu'avec des événements de la vie courante tels qu'une grillade ou encore l'utilisation d'un spray pour cheveux, les capteurs et instrument de référence arrivent facilement à saturation. Il est indispensable de contrôler la gamme de mesure et la limite de quantification (en cas de faible pollution) des instruments utilisés. Deux capteurs avec des gammes de mesures différentes pourraient être envisagés pour caractériser la pollution à l'intérieur et celle à l'extérieur.

Le temps d'échantillonnage minimum des capteurs e-PM, de 2 minutes, est une grosse contrainte (celui des instruments de référence est d'une seconde). Cet échantillonnage ne permet pas de caractériser la représentativité spatio-temporelle d'une mesure. En effet, il est difficile de savoir ce que contient exactement cette mesure et si des pics ponctuels ont été intégrés. Les résultats de l'étude menée par Stéphane Pawlak montraient que sur une longue distance parcourue en extérieur, une mesure unique intégrait en général à la fois un niveau de fond et un événement ponctuel. Il était nécessaire d'avoir un grand jeu de données pour lisser l'effet des événements ponctuels non représentatifs de l'état moyen.

Lors de nos mesures à Sense-City, nous avons remarqué que les capteurs e-PM ne font presque pas la distinction entre les particules PM₁₀ et PM_{2,5}. De plus, ces capteurs nécessitent un étalonnage régulier à partir d'instruments de référence en raison de leur dérive thermique au cours du temps. Les mesures de tels capteurs sont très bruitées, cependant Budde *et al.* (2013) ont montré qu'il était possible d'obtenir des résultats significatifs après étalonnage avec des instruments de référence.

⁶¹ http://www.wmo.int/pages/prog/arep/gaw/documents/Low_cost_sensors_post_review_final.pdf

Ces premières expériences mettent en évidence la difficulté de caractériser correctement la pollution avec de tels capteurs. Les mesures qu'ils fournissent sont très sensibles à la distance de la source et de nombreux facteurs de variabilité de la mesure ne peuvent pas être interprétés. Chaque mesure n'est pas exploitable car elle dépend d'éléments non reproductibles et aléatoires. Ce qui explique la nécessité d'intégrer une phase de répétabilité de la mesure et de déployer ces capteurs en grand nombre comme le mentionne le rapport WHO (2018).

- **Des contraintes pratiques et « humaines »**

Pour une utilisation quotidienne, il faudra s'assurer du bon positionnement du capteur pour que le filtre de celui-ci soit correctement placé de manière à être représentatif de l'exposition du sujet qui porte le capteur. Par exemple, éviter de mettre le capteur dans un endroit fermé, à l'intérieur d'un sac sans aération.

En cas de saturation du capteur, il faudrait prévoir une phase de dégrassement du filtre pour que celui-ci soit de nouveau opérationnel. Un voyant lumineux pourrait indiquer qu'il faut nettoyer ou changer le filtre par exemple.

À cela s'ajoutent des contraintes « humaines » dans le cadre de l'utilisation de ces capteurs où les participants adhèreraient à un observatoire participatif. Le porteur du capteur devra être proactif et s'investir dans sa mission. Pour que l'intégralité d'une ville soit couverte, des participants devront accepter de changer d'itinéraire et emprunter des itinéraires non recensés dans la base de données, et ce de façon répétée.

- **Des contraintes financières**

Et enfin, si l'on considère qu'un capteur doit rester à bas coût pour être déployé en grand nombre, nous ne pourrions pas équiper les micro-capteurs en cours de développement au LATMOS de capteurs de $PM_{2.5}$ qui sont trop chers. La contrainte budgétaire de 500 € par micro-capteur ne pourrait être respectée.

Comme cela a déjà été évoqué dans l'introduction de ce manuscrit, en raison du retard dans le développement des micro-capteurs – en phase de développement au LATMOS – nous n'avons pas équipé une centaine de volontaires comme prévu dans le projet initial. Nous n'avons donc pas davantage investigué sur ce sujet.

Cependant nous avons intégré le retour d'expérience provenant de ces campagnes exploratoires (connaissance de la variabilité spatio-temporelle de la pollution en région parisienne, contraintes des capteurs à bas coût) pour définir une autre stratégie avec le même objectif : caractériser la pollution le plus finement possible.

Nous avons utilisé les sorties d'un modèle urbain de dispersion haute résolution (PMSS) pour propager la mesure des stations AIRPARIF dans Paris à proximité de celles-ci (chapitre 4 et chapitre 5).

Chapitre 4. Représentativité spatiale des stations de mesures des NO_x et PM₁₀ à Paris /« On the spatial representativeness of NO_x and PM₁₀ monitoring-sites in Paris, France »

4.1 Notion de représentativité spatiale d'une mesure

Les quarante stations permanentes AIRPARIF de mesures des NO_x et les vingt-quatre stations pour les particules PM₁₀ ne sont pas suffisantes pour décrire l'hétérogénéité de la pollution en région parisienne avec sa superficie de 12 000 km².

En identifiant des zones où le niveau de concentration est très proche de celui de la station, on peut ainsi exploiter ces mesures sur une zone élargie autour de la station. Pour estimer la représentativité spatiale d'une station, les simulations de modèles numériques sont très souvent utilisées en raison de la couverture spatiale qu'ils fournissent. La caractérisation de ces zones dites «représentatives» peut se décliner en trois objectifs :

- Le premier objectif consiste à vérifier si le déploiement des stations de mesure est optimal pour mieux caractériser la complexité du paysage urbain (parcs, squares, avenues, grands boulevards, ronds-points).
- Le deuxième objectif est d'évaluer la performance des modèles numériques en les comparant aux observations.
- Et enfin, le troisième objectif est d'estimer l'exposition des individus.

La Commission Européenne a rappelé l'utilité de définir des zones de représentativité autour des stations de mesures sans préciser la méthodologie à suivre. Ce qui explique les nombreuses techniques proposées dans la littérature pour traiter de ce sujet dont la plupart sont recensées dans l'exercice d'inter-comparaison réalisé par FAIRMODE et AQUILA (rapport ⁶², 2017).

4.2 Résumé de l'article

Nous avons développé une nouvelle méthodologie pour estimer les aires de représentativité journalières des NO_x et PM₁₀ autour de dix stations trafic et de fond urbain d'AIRPARIF à Paris. Pour ce faire, nous avons utilisé les simulations d'un modèle de dispersion urbain PMSS (décrit section 2.2.3). Les simulations utilisées ont été fournies par la société ARIA Technologies. Elles avaient été réalisées dans le cadre d'un précédent projet pour une période

⁶² http://fairmode.jrc.ec.europa.eu/document/fairmode/WG1/IE_SR_final_report%202017.pdf

de dix jours en mars 2016. Cette période d'étude intègre un épisode hivernal de pollution aux PM_{10} associé à un régime anticyclonique. Ces simulations ont été configurées pour une résolution horizontale de 3 mètres couvrant l'ensemble du domaine parisien ($12 \text{ km} \times 10 \text{ km}$, ou encore 120 « tuiles ») sur une couche de surface de 2 mètres d'épaisseur.

Ici, nous nous intéressons à la variabilité locale de la pollution liée au trafic autour des stations AIRPARIF. Nous avons donc délimité des aires $1.2 \text{ km} \times 1.2 \text{ km}$ avec la station AIRPARIF au centre (« tuile »). Nous avons choisi de nous limiter à cette aire qui est suffisamment grande pour inclure des tronçons de trafic, des zones résidentielles et des zones à faibles émissions telles que les parcs ou des parties de la Seine.

Pour quatre des dix jours étudiés, la variabilité locale du signal des PM_{10} lié au trafic routier était recouverte par un signal de fond correspondant à un épisode de pollution aux particules provenant du nord de l'Europe. Nous avons normalisé le signal des PM_{10} pour faire ressortir le signal local que nous recherchions. Plusieurs tests ont été réalisés pour déterminer le fond à extraire dans la normalisation du signal : le fond du modèle CHIMERE, des mesures de stations de fond AIRPARIF, 10 % des « tuiles » de l'ensemble du domaine parisien avec la moyenne journalière la plus basse, la « tuile » avec la moyenne journalière la plus basse. Pour l'ensemble de ces tests, on retirait trop de signal y compris la variabilité locale liée au trafic.

Chaque « tuile » centrée autour des stations AIRPARIF a donc été traitée comme une image, en considérant le fond de l'image (« tuile ») comme étant la concentration minimale de chaque « tuile ».

L'aire de représentativité d'une station de mesures intègre tous les points autour de la station où la concentration est proche de la concentration de la station. Nous avons ici distingué les aires d'homogénéité et les aires de similarité. Dans la littérature existante, les zones d'homogénéité sont souvent définies comme respectant un seuil maximal à $\pm 20\%$ par rapport au niveau de pollution à l'endroit de la station (moyenne sur une longue période).

Ici, nous proposons une nouvelle approche en introduisant la notion de similarité. Une aire est dite « similaire » à la station de mesure si les concentrations dans la zone ont une variabilité temporelle journalière très corrélée avec la concentration au niveau de la station (fort coefficient de corrélation) et un faible écart en moyenne sur la journée par rapport à la station de référence (faible NRMSE, « Normalized Root Mean Square Error », l'erreur quadratique moyenne normalisée par la moyenne journalière).

Un algorithme sélectionne les pixels connexes qui satisfont le critère mentionné (R et NRMSE pour les zones de similarité et $\pm 20\%$ pour les zones d'homogénéité) et intègre le pixel de la station de mesure : « la pseudo-station » pour chaque jour dans la « tuile » de la station ($1.2 \text{ km} \times 1.2 \text{ km}$). Se limiter aux pixels connectés les uns aux autres permet de suivre la trajectoire des émissions à proximité de la station de mesure.

Une étude de sensibilité a été menée pour les stations trafic afin de déterminer les seuils du coefficient de corrélation R et du NRMSE à appliquer pour détecter les aires de similarité. Nous avons considéré que l'aire sélectionnée jour par jour, en fonction du seuil des deux critères, doit être comprise entre 300 m^2 et $400 \times 10^3 \text{ m}^2$ pour un maximum de jours et un maximum de sites. Les mêmes seuils ont ensuite été appliqués aux stations de fond urbain conduisant à des aires de représentativité réalistes.

Les seuils retenus pour les NO_x sont un coefficient de corrélation à 0.7 et un NRMSE à 0.45 et pour les PM_{10} un coefficient de corrélation à 0.75 et un NRMSE à 0.3. Les seuils des critères sélectionnés sont plus restrictifs pour les PM_{10} parce que les concentrations de NO_x ont une plus grande variabilité spatiale que les PM_{10} ce qui nécessite d'être moins restrictif pour obtenir une aire suffisamment grande ($> 300 \text{ m}^2$) autour de la station.

Pour caractériser les zones de représentativité on retiendra les zones les plus probables, c'est à dire les zones sélectionnées au moins sept jours sur dix. Pour les stations de fond urbain, les aires de similarité et d'homogénéité des NO_x et PM₁₀ ont le même ordre de grandeur. En revanche pour les stations trafic, les aires de similarité les plus probables sont plus grandes que les aires d'homogénéité.

Une aire de similarité est caractérisée à la fois par un seuil de coefficient de corrélation qui assure que les séries temporelles à l'intérieur de l'aire de similarité ont des variations corrélées aux variations de la station et un seuil de NRMSE qui assure que les aires de similarité ont des concentrations du même ordre de grandeur que celles de la station. La combinaison de ces deux critères intègre la condition d'homogénéité avec une contrainte supplémentaire qu'est le coefficient de corrélation. Le **Tableau 11** indique le rapport entre l'aire la plus probable sélectionnée uniquement à partir du critère NRMSE et l'aire d'homogénéité pour les NO_x et PM₁₀. Pour les stations de fond urbain, ces deux critères donnent des résultats très proches. Pour les stations trafic, l'aire la plus probable d'homogénéité est souvent plus restrictive que l'aire de NRMSE. L'aire d'homogénéité inclus donc l'aire avec le NRMSE.

Type station	Code station	Rapport de l'aire de NRMSE / l'aire d'homogénéité les plus probables	Rapport de l'aire de NRMSE / l'aire d'homogénéité les plus probables
		NO _x	PM ₁₀
Stations trafic	OPERA	13.5	6.9
	CELES_v	213	0.9
	BASCH_v	4.27	13.6
	ELYS	1.3	0.7
	HAUS	154	5.2
Stations de fond urbain	PA15L	1.1	1.0
	PA12	1.1	1.0
	PA13	1.1	1.0
	PA04C	1.6	1.0
	PA07	1.9	1.0

Tableau 11 : Les deux premières colonnes indiquent le type de station (colonne 1) et l'identifiant de la station (colonne 2). Les colonnes 3 et 4 indiquent le rapport entre l'aire la plus probable de NRMSE et l'aire d'homogénéité la plus probable pour les NO_x (colonne 3) et pour les PM₁₀ (colonne 4).

Une étude d'inter-comparaison a été menée par FAIRMODE (Forum for Air Quality Modelling in Europe) et AQUILA (Air Quality Reference Laboratories) pour une station trafic et deux stations de fond urbain sur un an de données pour les PM₁₀ (en Belgique). Onze méthodes ont été comparées et montrent des résultats très différents. Pour les aires de représentativité des stations trafic des PM₁₀, les aires de représentativité sont estimées entre 0.001 km² (10³ m²) et 160 km² alors que dans notre étude, les aires de similarité les plus probables sont comprises entre 0.001 km² (10⁶ m²) et 0.003 km² (38.4 m × 1000 m). Toujours dans le même rapport, les aires de représentativité des stations de fond urbain de PM₁₀ sont estimées entre 1 km² (10⁶ m²) et 700 km² alors que notre étude les aires de similarité et d'homogénéité sont comprises entre 0.4 km² et 0.9 km² en raison du domaine d'étude assez restreint autour de la station (1.2 km × 1.2 km).

Les aires de représentativité estimées dans notre étude montrent une large variabilité sur les dix jours. Cette variabilité s'explique très probablement par les variations des émissions de NO_x et PM₁₀ (1) liées au trafic (jours de semaines/week-end/jours fériés) et (2) liées aux conditions météorologiques (principalement le vent). Malgré le jeu de données contenant uniquement dix jours, nous avons pu observer l'influence du vent sur les aires de représentativité en comparant les jours avec vent et les jours sans vent, avec un seuil journalier à 3 m/s pour distinguer les deux cas. Pour étudier l'impact du vent, nous avons calculé la différence entre l'aire partagée quatre à six fois sur six pendant les jours sans vent (dans l'équation ci-dessous : « aire SANS VENT ») et l'aire partagée quatre fois sur quatre pour les jours avec vent (« aire AVEC VENT ») divisée par la moyenne des aires en utilisant l'équation ci-dessous (« aire »):

$$\text{rapport} = \frac{\text{aire SANS VENT} - \text{aire AVEC VENT}}{\langle \text{aire} \rangle}$$

Le **Tableau 12** regroupe ce rapport pour l'ensemble des stations pour les NO_x et PM₁₀. L'aire de similarité est toujours plus grande lorsqu'il n'y a pas de vent (cf. rapport toujours positif). L'écart entre les deux est d'autant plus marqué pour les stations trafic. Pour rendre ces résultats significatifs, il faudrait mener cette étude sur plus de jours.

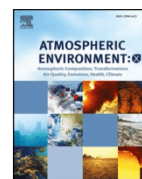
Type station	Code station	NO _x	PM ₁₀
Stations trafic	OPERA	1.9	1.9
	CELES_v	1.6	1.8
	BASCH_v	1.8	1.9
	ELYS	1.7	1.5
	HAUS	1.5	2
Stations de fond urbain	PA15L	0.3	0.6
	PA12	1.2	0.2
	PA13	0.8	0.5
	PA04C	1.4	0.7
	PA07	1.8	0.9

Tableau 12 : La première colonne indique le type de station et l'identifiant de la station (colonne 2) et les colonnes 3 et 4 indiquent le rapport de l'aire partagée les jours sans vent versus les jours avec pour l'ensemble des stations et les deux polluants NO_x (colonne 3) et PM₁₀ (colonne 4).

4.3 Article publié

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Atmospheric Environment: X

journal homepage: www.journals.elsevier.com/atmospheric-environment-x

On the spatial representativeness of NO_x and PM₁₀ monitoring-sites in Paris, France

Delphy Rodriguez^{a,*}, Myrto Valari^b, Sébastien Payan^a, Laurence Eymard^a^a LATMOS/IPSL, Sorbonne Université, UVSQ/CNRS, Paris, France^b LMD, Ecole Polytechnique, Palaiseau, France

HIGHLIGHTS

- A new method to quantify the spatial representativeness of urban monitoring-sites using high-resolution simulations.
- Measurements at urban background sites are representative of concentrations over the broad neighborhood around the monitor.
- Measurements at traffic sites are representative of concentrations over specific urban features: roads, pavements, etc.
- PM₁₀ representativeness areas are larger than NO_x representativeness areas.

ARTICLE INFO

Keywords:

Urban air pollution
Surface monitoring networks
High resolution modeling
Spatial representativeness

ABSTRACT

Ambient pollutant concentrations in Paris, France are routinely measured by the local surface-monitor network of the AIRPARIF agency. Such networks, however dense, have a limited spatial representativeness around the monitoring-site and are not capable to represent the strong horizontal gradients of pollutant concentrations over urban areas. High resolution models simulate 3D pollutant concentration fields at a spatial resolution as fine as a few meters over the urban area by integrating the underlying emission sources and accounting for the effect of buildings on the dispersion patterns. These models, provide a good spatial variability over the urban area but suffer from uncertainties related to the emission inventories, meteorological fields and parametrizations of the physical and chemical processes.

In this paper, simulations conducted by ARIA Technologies with the Parallel Micro-Swift-Spray (PMSS) model (<http://www.aria.fr/projets/aircity>) are used to assess NO_x and PM₁₀ representativeness areas around five urban background and five traffic-oriented monitoring-sites of the AIRPARIF network during ten days in March 2016. Commonly, the spatial representativeness of a monitor site is defined through homogeneity areas, namely the area around a monitoring-site where pollutant concentration is above 20% of the concentration at the location of the monitoring-site. Here, we propose a novel approach that uses similarity areas to define the spatial representation of monitor sites. Similarity areas integrate points that respect the additional condition to be highly correlated in time with the concentration at the monitor station. Thus, the criterion to select similarity areas is a combination of a high value of the correlation coefficient and a small value of the normalized root mean square error with regards to the concentration at the grid-cell corresponding to the location of the monitor. Criteria thresholds are determined through an iterative analysis and a representative area is defined through image processing that selects all the connected pixels that satisfy criteria thresholds and incorporate the grid-cell of the monitor.

Daily similarity areas estimated around each monitor are compared against homogeneity areas with regards to their shape, spatial extent, and urban specific characterization. Around urban background sites they are of the same order of magnitude, whereas around traffic sites similarity areas are generally larger than homogeneity areas. PM₁₀ representativeness areas are found to be 2.2 times larger than the NO_x ones. Urban background areas are representative of the broad neighborhood around the monitoring-site, whereas traffic-oriented monitoring-sites are representative of specific urban features such as sections of roads and sidewalks along the road. Averaged over the 10 days of the study and across all monitoring-sites, representativeness areas for urban background monitoring-sites are about 8 times larger than traffic representativeness areas (0.6 km² vs. 0.07 km²).

* Corresponding author.

E-mail address: delphy.rodriguez@latmos.ipsl.fr (D. Rodriguez).<https://doi.org/10.1016/j.aeaoa.2019.100010>

Received 20 January 2019; Accepted 20 January 2019

Available online 23 January 2019

2590-1621/ © 2019 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license

<http://creativecommons.org/licenses/by-nc-nd/4.0/>.

1. Introduction

The latest studies on the health impact of atmospheric pollution have pointed out that traffic-related atmospheric pollution is particularly dangerous for children asthma and cardiovascular diseases (APHEKHOM, 2012; Host, 2013). The health impacts due to exposure to particulate matter (PM), ozone (O₃) and nitrogen dioxide (NO₂) are presented in the technical report of the World Health Organization (WHO) « Review of Evidence on Health Aspects of Air pollution » (REVIHAAP Project, 2013). The International Agency for Research on Cancer (IARC), a specialized cancer agency of the WHO, has classified the outdoors atmospheric pollution as « carcinogenic to humans » (IARC gr.I, 2013). The annual financial and economic cost of the atmospheric pollution in France according to the Senate in July 2015 (Senate Report n°610, 2015) is 101.3 billion €. This amount is not only due to health impacts but also due to hidden costs such as the decrease of crop yield, deterioration of the buildings, etc.

This paper focuses on Nitrogen Oxides (NO_x = NO₂ + NO) and particulate matter with diameter lower than 10 μm (PM₁₀). NO_x are mostly emitted by fossil fuel combustion, industrial processes, and can be produced by chemical reactions involving NO and O₃. PM₁₀ compounds are characterized of a large variability of chemical composition, solid or liquid state and size distribution. PM₁₀ are directly emitted by urban heating, fossil fuel combustion, industrial and agriculture processes (CITEPA, French emissions inventory agency: <https://www.citepa.org>). Other, indirect sources of PM₁₀ are chemical reactions, large-scale transport of dust and sea-salt and the resuspension of particles already deposited on the ground due to the wind blowing or traffic circulation along the streets. In this paper, we will focus on the local emission sources to provide guidelines for local policies.

The highly heterogeneous emissions from various sources (traffic, residential, industrial ...) and the presence of buildings affect the dispersion of pollutants in the urban environment resulting to a high spatial and temporal variability of pollutant concentrations (Harrison, 2018). In Paris, France, ambient pollutant concentrations are routinely measured by the local network of the AIRPARIF agency (<https://www.airparif.asso.fr>). In the inner Paris area, this network has 15 and 8 permanent surface monitoring-sites, for NO_x and PM₁₀ respectively. The spatial representativeness is limited around each monitoring-site and therefore, such networks are not able to capture the strong horizontal gradients of pollutant concentrations over the urban area (Vardoulakis et al., 2005; Hagenbjörk et al., 2017; Hagemann et al., 2014; Blanchard et al., 2014).

Pollutant concentration fields are simulated with regional-scale chemistry transport models such as CHIMERE (<http://www.lmd.polytechnique.fr/chimere>) under air-quality forecasting platforms (e.g. Prev'Air <http://www.prevoir.org>, or Esmeralda <http://www.esmeralda-web.fr>) or research projects. Such models integrate emission fluxes, the horizontal and vertical transport of pollutants, turbulent mixing, chemical reactions and deposition through complex numerical systems to calculate the three-dimensional atmospheric concentration fields. However, these models have insufficient horizontal resolution (maximum of 1 km × 1 km) to resolve the intra-urban variability of the pollutant concentration.

High-resolution models such as the Parallel Micro-Swift-Spray (PMSS) model integrated in local scale air-quality modeling platforms such ARIA City (http://www.aria.fr/aria_city.php) simulate pollutant concentration fields at a spatial resolution of a few meters over the urban area. They take into account the advection from various emission sources, integrate buildings as obstacles to the wind flow by generating vortices downwind and upwind their location, calculate wet and dry deposition and in some cases deal with simple chemical reactions such as the Chapman cycle for ozone and NO_x interactions. Given that the street model used in this study implements a simple chemical mechanism we chose to focus on the less reactive NO_x pollutant. However, over the highly urbanized area of the study and close to traffic sources,

NO_x is a very good proxy of NO₂ since these pollutants are strongly correlated one with the other and with the traffic. Thus, chemistry is not expected to alter our results.

These models provide a good spatial variability within the urban environment but are still subjected to important biases due to the high uncertainties relating to the emission inventories, meteorological predictions and simplified chemistry.

Spatial representativeness assessment aims generally to improve the characterization of atmospheric pollution simulations by using surface measurements to test model performance, design and optimize monitoring-site deployment, better estimate population exposure to the atmospheric pollution, assess or report air quality, classify station, assimilate data for modeling.

In this paper, we focus on the first task, namely to define the spatial extent of the measurement around the monitoring-site. The European Commission stresses the need to assess the spatial representativeness around monitoring-sites without imposing a specific methodology (EC, 2011). An intercomparison exercise has been realized by FAIRMODE (Forum for Air Quality Modeling in Europe) and AQUILA (Air Quality Reference Laboratories) to harmonize a large set of 10 spatial representativeness approaches and provide a technical support to the implementation of Directive 2008/50/EC (EC, 2008) on ambient air quality and cleaner air for Europe. Study results obtained from the same common dataset reveal a strong variability between the different approaches (extent, shape, technical procedures). Most of these approaches use a common criterion to select representativeness areas: the pollutant concentration at each point of the representativeness area should not differ from the measurement at the monitoring-site location by more than a given threshold. Additionally, some approaches use also external parameters influencing air quality such emissions from different sources and/or climatic and topographic dispersion. Nguyen et al. (2012) calculate the means of Principal Component Analysis (PCA) to classify status station.

A brief overview of the most commonly used approaches found in the literature is presented here. These studies use a homogeneity criterion based on a certain threshold. Santiago et al. (2013) apply a methodology based on annual average simulated concentration maps obtained by means of weighted average of Computational Fluid Dynamics model for urban stations whereas Martin et al. (2014), use a Chemistry Transport Model (CHIMERE) for rural background stations. The applied threshold is ± 20% of the simulated concentration to the monitoring-site location. Vitali et al. (2016); Piersanti et al. (2015) used the Concentration Similarity Frequency (CSF) to assess representativeness areas from simulated concentrations. CSF approach is based on the definition given by Nappo et al. (1982): “a point measurement is representative of the average in a larger area (or volume) if the probability that the squared difference between point and area (volume) measurement is smaller than a certain threshold more than 90% of the time”. Representativeness area contains a grid-cell if the difference between concentrations at the monitoring-site and the selected grid-cell does not exceed 20%, 90% of the time. Spangl et al. (2007) compared modeled annual averaged concentration to the concentration to the monitoring-site location with similarity thresholds depending on pollutant. Additional criteria emissions are applied.

Beauchamp et al., (2016), Bobbia et al. (2008), assesses spatial representativeness areas on annual averages by estimating spatial concentration variability and uncertainties from passive sample measurements. Then those concentrations are interpolated from modeling output data applying a kriging approach. The applied threshold to select spatial representativeness points is 30%.

Janssen et al. (2012) assess representativeness of stations by calculating the relation between annual averaged levels at the location of the station and a land use characteristics.

These approaches require long datasets, much longer than the 10 days available in this study. Also, such homogeneity criteria provide an average estimate, not taking into account the daily temporal variability

due to specific meteorological conditions or traffic emissions which may involve pollution episodes. Moreover they are generally focusing on a small number of monitoring-sites. Another limitation of these approaches is that they do not filter out the variability due to regional scale transport (especially high for PM_{10}) and the long time averaged. They, therefore, fail to highlight the impact of the urban features (emissions, urban landscape) on the characterization of the representativeness area.

In this study, daily representativeness areas are derived from PMSS simulations with at 3 m horizontal resolution. The study focuses on NO_x and PM_{10} around five traffic and five urban-background monitoring-sites of the AIRPARIF network in Paris for ten non-successive days in March 2016.

Modeled and measured data and methodological aspects are presented in Section 2. The results of the study are presented in Section 3. Section 4 consists of a discussion on the general conclusions and future perspectives of the study.

2. Data and methodology

2.1. Monitoring-sites

Five traffic-oriented and five urban background sites have been selected for the study, covering a diversity of urban landscapes in Paris. Traffic monitoring-sites are located on sidewalks along roads, major crossroads or the border of the Seine River. Urban background monitoring-sites are located on low emission areas such as parks, pedestrian squares, schoolyards and leisure parks. NO_x are measured at all monitoring-sites whereas PM_{10} are measured at four traffic stations and two urban background stations (Table 1 and Fig. 1). Hourly averaged NO_x and PM_{10} surface concentrations are routinely measured and available on the website of the AIRPARIF air-quality agency (<http://www.airparif.fr>).

Table 1

Characteristic features of the study-areas. Locations marked with an asterisk indicate the location of the monitoring-site.

Type	ID station	Characteristic features	Pollutants
Traffic-oriented stations	CELES	Seine banks*, the Seine River, bridge, side roads	NO_x
	HAUS	Bld Haussmann*, side roads	NO_x , PM_{10}
	OPERA	Crossroad*, road traffic	NO_x , PM_{10}
	BASCH		
	ELYS	Champs-Élysées Avenue*: large road bordered with trees, side roads	NO_x , PM_{10}
Urban background stations	PA15L	Stade* and leisure park, "Boulevard Peripherique", traffic road	NO_x , PM_{10}
	PA12	Schoolyard*, railways, traffic road	NO_x
	PA13	Park*, traffic road, crossroad	NO_x
	PA04C	Place*, park, the Seine River, traffic road	NO_x , PM_{10}
	PA07	Square close to the Eiffel Tower*, Champ-de-Mars Garden, Seine banks, the Seine River, Bridge	NO_x

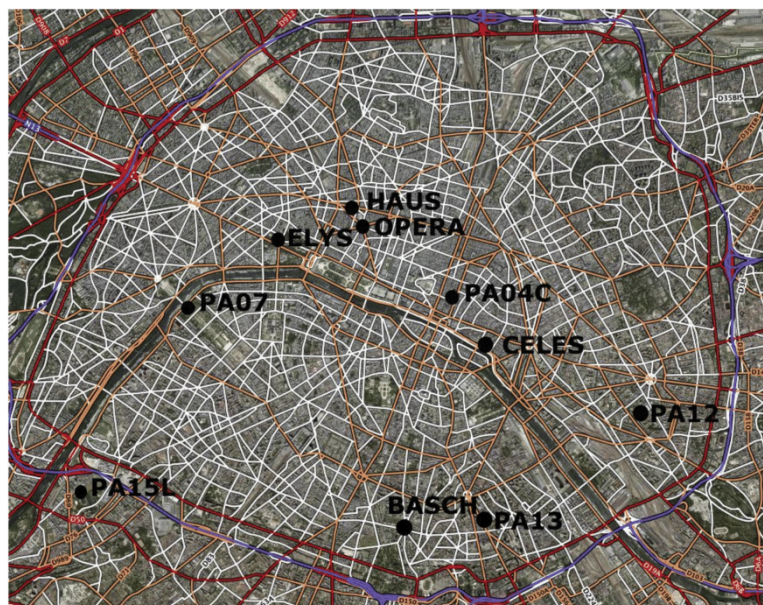


Fig. 1. Paris map with locations of studied Airparif stations, the Seine River and roads. The red thin line represents Paris city boundary "Boulevard Peripherique" and the white and orange line represents roads.

2.2. Model simulations

NO_x and PM₁₀ concentrations are simulated with the local-scale model PMSS, at a horizontal resolution of 3 m covering a 12 × 10 km² domain over the city of Paris.

The model has thirty-five vertical layers from the ground up to an altitude of 800 m. The vertical resolution decreases with height with a surface layer of 2 m thick (Moussafer et al., 2015). Emissions fluxes are taken from the AIRPARIF inventory. Meteorological conditions are simulated with the MM5 weather mesoscale model. Terrain elevation and buildings' contours and height are taken from the BDTOPO database developed by the French National Geographic Institute. Atmospheric dispersion is simulated with the PSPRAY lagrangian particles model. Traffic emissions are calculated with the HEAVEN chain (Healthier Environment through the Abatement of Vehicle Emissions and Noise, <http://www.airparif.asso.fr/etat-air/air-et-climat-emissions-heaven>). HEAVEN chain calculates in near real-time traffic situation and allows obtaining pollutant emissions on main traffic road of Parisian area from a "bottom-up" approach. Hourly traffic emissions are calculated from the combination of traffic model (car flow, mean speed, cold engines percentage) and emission factors (car speed, road type, temperature, car type, etc.) applied by loop-based counting systems.

PMSS simulations are available for ten non-successive days in March 2016. Domain-averaged hourly variations of NO_x and PM₁₀ simulated concentrations are shown in Fig. 2. A peak in NO_x concentration is observed on the 9th of March. A PM₁₀ pollution episode is simulated from the 11th to the 15th of March 2016 reaching a peak of 60 µg/m³ on the 12th of March. Another [PM₁₀] peak is observed on March 18.

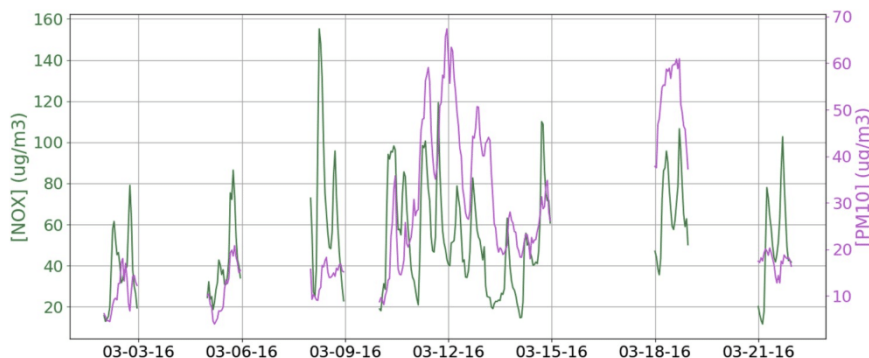


Fig. 2. Time-series of NO_x and PM₁₀ simulated concentrations (µg/m³) in Paris for 10 days in March 2016 (hourly data).

Table 2 summarizes daily and domain-averaged NO_x and PM₁₀ simulated concentrations, as well as mean daily surface pressure from the recent ECMWF model reanalyses ERA 5 (<https://www.ecmwf.int/en/forecasts/datasets/archive-datasets/reanalysis-datasets/era5>), wind direction and daily averaged velocity (superior to 3 m s⁻¹) simulated by the MM5 regional mesoscale weather forecasts model (<http://www2.mmm.ucar.edu/mm5/>). Wind patterns on these ten days change from West to North-East, accompanying the transition starting from a low-pressure system between 03-02 and 03-10 (994 hPa-1003 hPa) followed by a transition day 03-11-16 (1012 hPa), then the high-pressure situation took place on the 03-12-16 (1015 hPa). The beginning of a high-pressure situation is characterized by a strong atmospheric stability combined to a pollution episode due to agriculture activity from wide European area, as far as Eastern Europe, leading to an accumulation of PM₁₀ within the boundary layer. A sharp rise in PM₁₀ simulated concentration is observed on the 11th of March (more than twofold). These atmospheric conditions do not affect NO_x concentration, which is more driven by local traffic emissions. However, both PM₁₀ and NO_x concentrations are sensitive to rainfall which occurs on the 2nd and 5th of

March and results in pollutant deposition on the ground. During these days, NO_x concentration reaches its lowest levels.

NO_x concentration levels mostly depend on the density of the traffic

Table 2

The first column lists the days of March 2016, H designs Holiday. Rainy days are marked by an asterisk. In the second and the third columns are given the daily average simulated concentrations in Paris for NO_x and PM₁₀. The fourth column shows the surface pressure (hPa) from ERA reanalysis. The fifth column summarizes the mean wind direction and the range of daily-mean wind velocity modeled with the MM5 model.

March 2016	[NO _x] (µg/m ³)	[PM ₁₀] (µg/m ³)	Surface Pressure (hPa)	Wind Direction [Daily average velocity (m.s ⁻¹)]
02 Wednesday* H	38.2	10.5	994	W [5.4–5.8]
05 Saturday* H	41.0	11.6	983	NNW [3.2–3.3]
08 Tuesday	68.8	14.0	1003	–
10 Thursday	62.2	18.3	1001	–
11 Friday	63.5	46.2	1012	–
12 Saturday	54.5	44.3	1015	–
13 Sunday	31.1	28.5	1016	NE [3.9–4]
14 Monday	52.1	24.2	1014	NE [4–4.3]
18 Friday	68.0	52.1	1010	–
21 Monday	50.4	17.4	1008	–

network, with higher pollution levels during weekdays than during week-ends or holidays. Saturday the 5th of March, was a return-from-holiday and one would expect a peak in NO_x concentrations. However, this was attenuated by rainfall.

2.3. Model evaluation against local surface observations

PMSS simulations at the surface model layer are compared to measurements of the AIRPARIF network. A thorough work has been carried out to assign model grid-cells to each monitoring-site combining google street view and plans of the French National Geographic Institute. An error in the grid-cell selection would strongly affect the results at such high resolution (Duyzer et al., 2015). Two AIRPARIF stations have been omitted for the model evaluation process due to missing data in the emission inventory along street segments in the vicinity of these stations (BASCH and CELES).

PMSS simulations are in a good agreement with PM₁₀ measurements as shows the squared correlation coefficient (R²), equal to 0.83 and 0.92, and the root mean squared error (RMSE), equal to 9.57 µg/m³ and 6.21 µg/m³ for traffic and urban background sites respectively (Fig. 3). Model performance is not so satisfactory for NO_x measurements with a R² equal to 0.65 and 0.45 for traffic and urban background sites respectively. As shown in the scatter plots the model overestimates NO_x maxima, especially in traffic sites, resulting in large overall error (RMSE

equal to $81.57 \mu\text{g}/\text{m}^3$). Moreover, PMSS simulations fail to reproduce some nocturnal peaks of the NO_x concentration. On the contrary, PMSS underestimates NO_x concentrations at urban background sites. More specifically, at urban sites, PMSS simulations fail to reproduce the morning peak of NO_x concentration corresponding to the rush hour (between 7AM and 8AM). The large difference in the bias between NO_x and PM_{10} is probably due to errors in the emissions inventory and to the simple chemical mechanism of the PMSS model, both having a larger impact on NO_x concentrations than on PM_{10} .

Although the model evaluation has pointed out some limitations of the PMSS simulations, overall the model performance is good for PM_{10} . NO_x simulations are not so satisfactory for peaks due to traffic emissions input. The reason is most probably the simple chemical mechanism of the PMSS model that fails to represent the concentration of the highly reactive pollutants such as NO. Note that the error for NO_2 is smaller than for NO for traffic stations ($\text{RMSE}(\text{NO}_2) = 21.9 \mu\text{g}/\text{m}^3$ vs $\text{RMSE}(\text{NO}) = 72.3 \mu\text{g}/\text{m}^3$) and similar for urban background stations.

Despite of these model limitations, we assume that the model captures the actual spatial variability of pollution. In the following, these simulations are used to assess the spatial representativeness of each AIRPARIF station during the 10 days of March 2016.

daily variability through the definition of “similarity” area. It aims at being a complementary method, with respect to those used to evaluate air quality directive limits.

This area integrates points for which the daily temporal variability is highly correlated to the temporal variability of the concentration at the reference point (monitor), and the daily bias with respect to the concentration at the monitor location, remains below a chosen threshold.

The model grid-cell corresponding to the location of the monitor station, called “pseudo-station” hereafter, is used as reference value, instead of the monitoring stations. Indeed, due to the discrepancies between observations and simulations, using actual measurement would lead to inconsistent results when studying the spatial representativeness of stations using simulations.

In two cases, (BASCH and CELES stations), the lack of emission sources along a street segment in the vicinity of the station led us to slightly shift the “pseudo-station”. Those both virtual stations will be renamed “BASCH_v” and “CELES_v”. CELES has been shifted by 600 m from the real monitoring-site position in the northwest direction, on the same road. BASCH has been moved by 60 m to be placed in another avenue (exit of the crossroad “Avenue Jean Moulin” instead of “Avenue

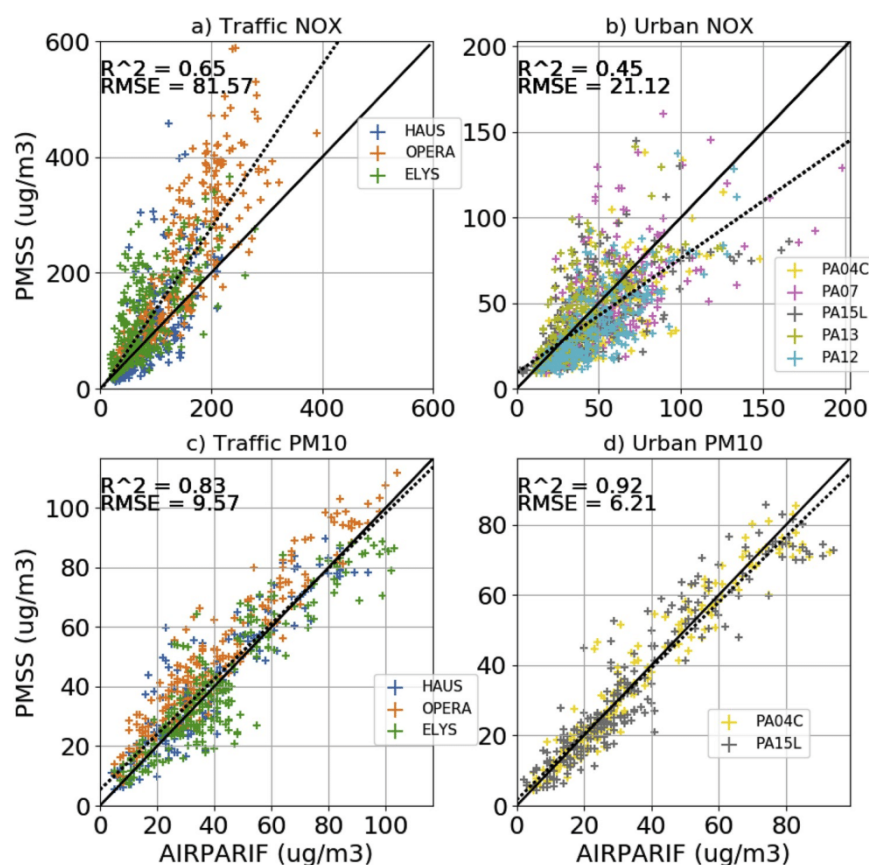


Fig. 3. Scatter plots of measured vs simulated hourly concentrations for NO_x (top) and PM_{10} (bottom) at traffic (left) and urban (right) monitoring sites. The black line represents $y = x$. The dotted line represents the linear regression between AIRPARIF measurements and PMSS simulations.

2.4. The spatial representativeness approach

The spatial representativeness area of a monitoring-site consists of all the points around the station where the concentration is linked to the concentration at the monitoring-site location, through a given criterion. Here we develop an alternative approach that focuses on the

du General Leclerc”). Those two transfers are made by respecting the same road and urban landscape conditions (subjective analysis). We indeed checked that there is no lack of emission sources along this new street segment in the vicinity of the virtual stations. Indeed, a further analysis with corrects simulations would be necessary for completing results around the actual station locations.

For evaluating the relevance of the similarity approach, with comparison to homogeneity optimal thresholds must be chosen, both for the correlation coefficient and the root mean square difference. We use the 20% threshold for homogeneity, to be consistent with previous studies. Criteria thresholds are applied over squared areas of $1.2 \times 1.2 \text{ km}^2$ centered around each pseudo-station. The size of these squares was chosen to be large enough to encompass both high (traffic network and dense residential areas) and low emission zones (parks or part of the Seine River).

Similarity areas are thus obtained by the combination of a (1) high correlation coefficient R (Eq (1)) and a (2) low normalized root mean square error NRMSE (Eq (2)).

$$R(ref, val_{(x,y)}) = \frac{\sum_{t=0}^{t=23} (ref_{(t)} - \langle ref_{(t)} \rangle) (val_{(t,x,y)} - \langle val_{(t,x,y)} \rangle)}{\sigma_{ref} \sigma_{val(x,y)}} \quad (1)$$

Where, $ref_{(t)}$ is the PM_{10} or NO_x hourly concentration at the pseudo-station, $val_{(t,x,y)}$ the PM_{10} or NO_x hourly concentration at each grid-cell, $\langle ref_{(t)} \rangle$ the PM_{10} or NO_x daily concentration at the pseudo-station, $\langle val_{(t,x,y)} \rangle$ the PM_{10} or NO_x daily concentration at each grid-cell, σ_{ref} the PM_{10} or NO_x daily standard deviation at the pseudo-station and $\sigma_{val(x,y)}$ the PM_{10} or NO_x daily standard deviation at each grid-cell.

$$NRMSE(ref, val_{(x,y)}) = \frac{\sqrt{\frac{1}{n} \left(\sum_{t=0}^{t=23} \left(ref_{(t)} - val_{(t,x,y)} \right)^2 \right)}}{\langle ref_{(t)} \rangle} \quad (2)$$

where $ref_{(t)}$ is the PM_{10} or NO_x hourly concentration at the pseudo-station, $val_{(t,x,y)}$ the PM_{10} or NO_x hourly concentration at each grid-cell and $\langle ref_{(t)} \rangle$ PM_{10} or NO_x the daily concentration at the pseudo-station.

Homogeneity areas are determined from daily-averaged simulated concentrations applying a threshold of 20% of the pseudo-measure, as recommended by the European Commission for uncertainty assessment in the Directive 2008/50/EC (EC, 2008) (from 15% to 25% depending on the measured pollutant) and used in literature (Vitali et al., 2016; Piersanti et al., 2015; Blanchard et al., 1999; Janssen et al., 2008). This criterion will be called *EC criterion* hereafter (Eq. (3)).

$$EC \text{ criterion} = \frac{abs(\langle val_{(t,x,y)} \rangle - \langle ref_{(t)} \rangle)}{\langle ref_{(t)} \rangle} * 100 \quad (3)$$

Where, $\langle ref_{(t)} \rangle$ is the PM_{10} or NO_x mean daily concentration at the pseudo-station and $\langle val_{(t,x,y)} \rangle$ the PM_{10} or NO_x mean daily concentration at each grid-cell.

PM_{10} concentrations are characterized by a large temporal variability due to dust episodes occurring on some of the days within the ten days period of the study (Fig. 1). To reduce this regional scale effect on the statistics and highlight local effects such as urban emission sources we normalized PM_{10} data on a daily basis (Eq. (4)).

$$[PM10]_{(t,x,y)normalized} = \frac{[PM10]_{(t,x,y)} - [PM10]_{min \text{ area}}}{[PM10]_{max \text{ area}} - [PM10]_{min \text{ area}}} \quad (4)$$

Where, $[PM10]_{(t,x,y)normalized}$ is the PM_{10} hourly normalized concentration at each grid-cell, $[PM10]_{(t,x,y)}$ is the PM_{10} hourly concentration simulated at each grid-cell, $[PM10]_{min \text{ area}}$ and $[PM10]_{max \text{ area}}$ are the 0.1st and 99.9th percentiles of the daily averaged PM_{10} concentrations over an area of $1.2 \times 1.2 \text{ km}^2$ around the station.

The previous criteria are applied on the simulated concentration fields around each traffic station. An algorithm selects the connected pixels that satisfy the aforementioned criteria and incorporate the « pseudo-station » pixel. Two pixels are considered « connected » when they are neighbors and they respect the threshold criteria. We thus focus on the representativeness area strongly connected to the station. In some cases with moderated traffic, both sidewalks are selected. However, traffic impact is not always symmetric over both sidewalks bordering the traffic road. It is difficult to fully assess this result, since it relies on PMSS simulations.

As an example, Fig. 4 shows the correlation coefficient (R) and NRMSE values estimated at each model grid-cell over an area of $1.2 \times 1.2 \text{ km}^2$ around the around Champs-Élysées Avenue monitoring-site (ELYS) on the 2nd of March 2016. The highest R values (purple color) are obtained in the vicinity of the station, along the avenue, and along the Seine bank. The minimal NRMSE values (from dark blue to green) are observed along a line inside the Champs-Élysées Avenue. The two criteria are thus complementary and their combination results in the definition of an area inside which concentration levels remain close to the concentration at the location of the station.

To choose thresholds for ρ and NRMSE, we consider two conditions: (i) that the size of the defined area lies within a reasonable range and (ii) that such an area is obtained most of the days of the study.

The European Air Quality Directive (EC/2008/50) specifies that « a sampling point must be sited in such a way that the air sampled is representative of air quality for a street segment no less than 100 m in length at traffic-oriented sites ». Indeed, the selected area must be large enough to get beyond the pseudo-station. Here we fix the lower limit at 300 m^2 and the upper limit at $400 \times 10^3 \text{ m}^2$ to avoid incorporating the entire domain centered around the monitoring-site (squares of $1.2 \times 1.2 \text{ km}^2$). For example, on the 2nd of March 2016, around the monitoring-site located on the avenue des Champs-Élysées, the selected similarity area with R coefficient equal to 0.7 and NRMSE value of 0.45 is considered as a success with a selected area of $8.9 \times 10^3 \text{ m}^2$, superior to 300 m^2 and inferior to $400 \times 10^3 \text{ m}^2$ (Fig. 5).

Criteria thresholds to select similarity areas have been determined by an iterative analysis from all traffic-oriented monitoring-site areas over the 10-days with a step of 0.05 for both correlation and NRMSE values. This analysis is carried out on traffic-sites, because between road and sidewalk there is the highest degree of variability so we can check if the retained thresholds allow the distinction between the road and sidewalk and evaluate the impact on distance to the road. Then the same threshold values are applied on the urban-background sites.

For each threshold, a score is estimated by counting the number of days on which the size of the selected area ranges within the chosen boundaries. The maximum score for each site is 10. Finally, a total score

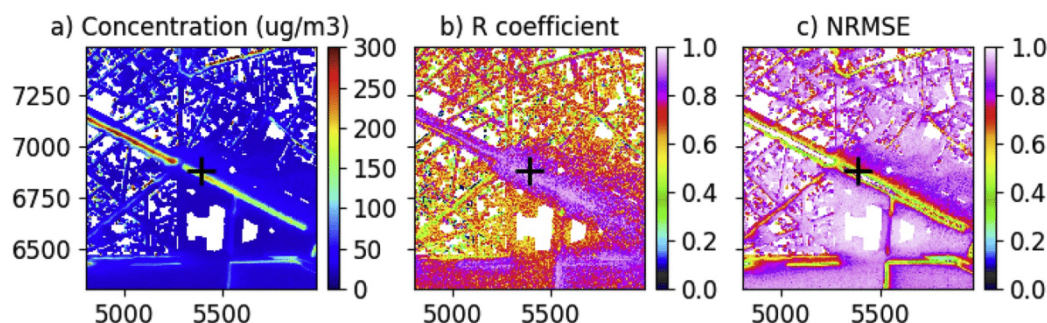


Fig. 4. Champs-Élysées Avenue area NO_x concentration (a), NO_x correlation coefficient (b), NO_x NRMSE on the 2nd of March 2016.

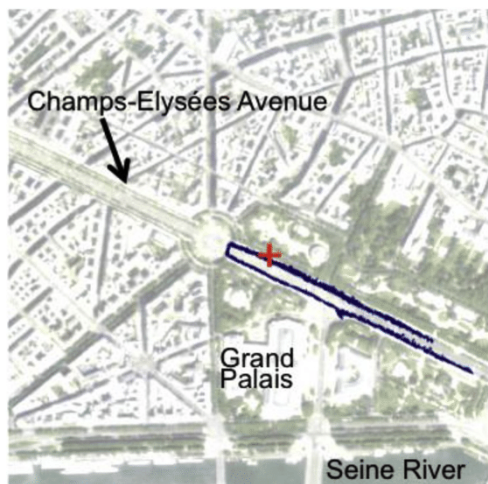


Fig. 5. NO_x similarity area around Champs-Élysées Avenue on the 2nd of March 2016. The red cross shows the pseudo-station location.

is calculated from all stations for the 10 days and reported as percentage of success.

The main difficulty is to find a combination of criteria thresholds that involves a good agreement for the maximum of days and the

maximum of traffic-oriented areas.

In the case where two different combinations of criteria thresholds give the same score (percentage of success), the retained combination is the more restrictive (i.e. the higher R correlation coefficient and/or the lowest NRMSE value). Table 3 summarizes the results of this iterative analysis for the 5 traffic-oriented monitoring-sites and for the 10 days of the study.

For NO_x similarity area, with a constant correlation coefficient of 0.7 and a NRMSE varying from 0.4 to 0.5 with a 0.05 step, the total success rate is comprised between 90 and 100%. Combination of R coefficient equal to 0.7 and NRMSE value of 0.4 selects too restrictive areas 5 times: 1 time (1 day) for 2 monitoring-site areas: BASCH_v, ELYS, CELES_v and 2 days for HAUS monitoring-site area. With correlation coefficients of 0.75 and 0.8 and NRMSE equal to 0.5, success rate is 98%. The combination of R coefficient equal to 0.7 and NRMSE equal to 0.45 and 0.5 has a 100% success rate for the 10 days and 5 traffic-oriented areas. In that case, the most restrictive criterion threshold is retained, namely R = 0.7 and NRMSE = 0.45.

Success rates in the selection of similarity areas are lower for PM₁₀ than for NO_x. This is probably due to the fact that part of the local variability is overshadowed by the homogeneous regional signal of PM₁₀ concentrations. Thus traffic has a smaller contribution for PM₁₀ than for NO_x resulting in similarity areas too large (> 400 × 10³ m²). For correlation coefficients of 0.7, 0.75 and 0.8 and NRMSE = 0.3, success rates are in the range of 76–78%. With a combination of a correlation coefficient of 0.75 and NRMSE values of 0.25, 0.3 and 0.35,

Table 3

Success rates in the selection of (a) similarity and (b) homogeneity areas for NO_x and PM₁₀ of all traffic monitoring-sites areas on the 10 days of the study. The retained criteria thresholds for similarity areas are shown in bold. In case of failure, the number of days and the selected area (m²) are mentioned in italics.

a)								
Pollutant	Criteria thresholds		Success rates for the ten days Traffic oriented stations					Total (%)
	R	NRMSE	BASCH_v	CELES_v	ELYS	HAUS	OPERA	
NO _x	0.7	0.4	9 <i>1 = 241</i>	9 <i>1 = 198</i>	9 <i>1 = 106</i>	8 <i>2 < 30</i>	10	90
	0.7	0.45	10	10	10	10	10	100
	0.7	0.5	10	10	10	10	10	100
	0.75	0.5	10	10	9 <i>1 = 193</i>	10	10	98
	0.8	0.5	10	10	9 <i>1 = 121</i>	10	10	98
PM ₁₀	0.7	0.3	9 <i>1 = 846x10³</i>	8 <i>1 = 148</i> <i>1 = 459x10³</i>	7 <i>3 > 833x10³</i>	6 <i>4 > 561x10³</i>	9 <i>1 = 161</i>	78
	0.75	0.3	9 <i>1 = 846x10³</i>	8 <i>1 = 148</i> <i>1 = 459x10³</i>	7 <i>3 > 826x10³</i>	6 <i>4 > 561x10³</i>	9 <i>1 = 161</i>	78
	0.8	0.3	9 <i>1 = 846x10³</i>	8 <i>1 = 148</i> <i>1 = 459x10³</i>	6 <i>3 > 833x10³</i> <i>1 = 133</i>	6 <i>4 > 561x10³</i>	9 <i>1 = 161</i>	76
	0.75	0.35	6 <i>4 > 876x10³</i>	8 <i>1 = 180</i> <i>1 = 459 x10³</i>	6 <i>4 > 833x10³</i>	4 <i>6 > 466x10³</i>	10	68
	0.75	0.25	10	7 <i>3 < 226</i>	8 <i>1 > 833x10³</i> <i>1 = 253</i>	4 <i>4 > 553x10³</i> <i>2 < 145</i>	7 <i>3 < 206</i>	72
b)								
Pollutant	Success rates for the ten days Traffic oriented stations					Total (%)		
	BASCH_v	CELES_v	ELYS	HAUS	OPERA			
NO _x	9 <i>1 = 97</i>	1 <i>9 < 96</i>	10	2 <i>8 < 30</i>	1 <i>9 < 152</i>	46		
PM ₁₀	10	10	8 <i>2 > 453x10³</i>	5 <i>5 > 506x10³</i>	7 <i>3 < 197</i>	80		

total success rates vary between 68% and 78%. However, NRMSE = 0.3 is not restrictive enough, resulting in the selection of vast, non-realistic similarity areas whereas 0.25 is too restrictive (OPERA area). The final similarity criteria thresholds resulting from the iterative analysis for similarity for the 5 traffic-oriented areas are given in bold letters in Table 3a.

Success rates for the selection of the daily homogeneity representativeness area using the fixed 20% “EC criterion” are shown in Table 3b. Total success rate for homogeneity areas are generally lower than for similarity areas. Detection of NO_x homogeneity areas has a success rate of 46% (100% for similarity areas). PM₁₀ success rates for homogeneity and similarity areas are of the same order of magnitude 80% and 78% respectively. NO_x homogeneity areas are particularly restrictive around Boulevard Haussmann monitoring-site (HAUS) and around the bank of the Seine River (CELES_v) due to the strong spatial variability of NO_x concentration. Respectively for HAUS and CELES_v, only 2 and 1 out of the 10 selected areas may be considered successful.

The retained thresholds for PM₁₀ are more restrictive than for NO_x. This is due to the higher spatial variability of NO_x concentrations, which requires less restrictive thresholds in order to retain areas large enough to get beyond the pseudo-station.

3. Results

3.1. Quantitative results

3.1.1. NO_x representative areas

3.1.1.1. Traffic-oriented areas. NO_x homogeneity and similarity areas around traffic sites for the 10 days in March 2016 are summarized in Fig. 6. NO_x traffic-oriented monitoring-sites have a large variability between stations, criteria (homogeneity vs similarity) and across the ten days period. NO_x traffic-oriented representativeness areas are comprised between 15 m² (minimum HAUS homogeneity area) and 51.5 × 10³ m² (maximum ELYS homogeneity area). A large variability is observed between the five traffic-oriented areas. ELYS has the largest similarity area, whereas OPERA has the smallest one, averaged over the 10 days, ELYS area is five times larger than the OPERA area (16.6 × 10³ m² vs 1.7 × 10³ m²). ELYS station is located on a sidewalk along the wide Champs-Élysées Avenue whereas OPERA station is located on a crossroad and integrates contributions from several streets. ELYS has the largest homogeneity area whereas CELES_v has the smallest one (18.3 × 10³ m² vs 74 m²).

Similarity areas are larger than homogeneity areas for CELES_v (ratio similarity (S)/homogeneity (H) = 164), HAUS (ratio S/H = 18) and BASCH_v (ratio S/H = 6). Similarity and homogeneity areas are of the same order of magnitude for ELYS. Homogeneity area is negligible for OPERA with a representativeness area averaged across the ten days of 74 m².

In order to get a synthesized result, the “most realistic” representativeness area is taken as the intersection of representativeness

areas obtained at least seven days out of ten (70%). Applying a more restrictive condition (e.g. 100% of the time) leads to very small areas, whereas using the mean representativeness area across the ten days leads to a non-realistic representativeness area, given the high variability of selected areas across the ten days. As shown in Table 4, CELES_v station has the smallest NO_x homogeneity area (18 m²) whereas BASCH_v has the smallest similarity area (846 m²). ELYS has the largest homogeneity area (2844 m²) and CELES_v has the largest similarity area (2745 m²). Similarity areas are larger than homogeneity areas with a ratio ranging between 3.6 (BASCH_v) and 152.5 (CELES_v), except the ELYS station, where both areas are of the same order of magnitude.

The most probable homogeneity area is in some cases not significant (18 m² and 27 m² for NO_x at CELES_v and HAUS monitoring-sites respectively). A possible explanation for this, is that the 20% threshold, stated for consistency with literature values, is not optimal for data issued from the PMSS simulation, contrary to the thresholds used for similarity areas.

3.1.1.2. Urban background areas. Fig. 7 summarizes the representativeness areas around NO_x urban background monitoring-sites for the 10 days. Across the study period and all stations, mean NO_x urban background areas are fifty-one times larger than NO_x traffic-oriented areas (411 × 10³ m² vs 8.2 × 10³ m²). NO_x urban background stations representativeness areas range between 9.6 × 10³ m² (minimum PA12 homogeneity area) and 932.0 × 10³ m² (maximum PA12, similarity area) whereas traffic-oriented stations representativeness areas range between 15 m² and 51.5 × 10³ m².

In most cases, the representativeness areas of urban background stations cover the neighborhood of the station. Representativeness areas of monitoring-sites located on schoolyards (PA12 and PA07) are in some cases very small but they still cover part of the neighborhood with a minimum value of 9.6 × 10³ m² (minimum PA12 homogeneity area) and 19.9 × 10³ m² (minimum PA07 similarity area). In fact, those monitoring-sites are partly influenced by traffic emissions due to their proximity to the roadside.

Large differences between the representativeness areas of the NO_x urban background stations are found. A ratio of 2.1 is estimated between the monitoring-site located on a pedestrian square close to the Centre Pompidou (PA04C) and the PA12 site (513.0 × 10³ m² vs. 242.7 × 10³ m² averaged over ten days for homogeneity and similarity criteria).

By averaging across the ten days and all urban background areas, we find that NO_x similarity areas are 1.3 times higher than NO_x homogeneity areas (468.7 × 10³ m² vs 353.5 × 10³ m²). Similarity area values span over a greater range across the ten days of the study than homogeneity areas, particularly for PA07 and PA12. Homogeneity area values at PA15L, a monitoring-site located within a stadium close to “Boulevard Périphérique” and another major road during the 10 days of the study have the smallest dispersion. This can be explained by the sharp spatial gradient of pollutant concentration at the proximity of

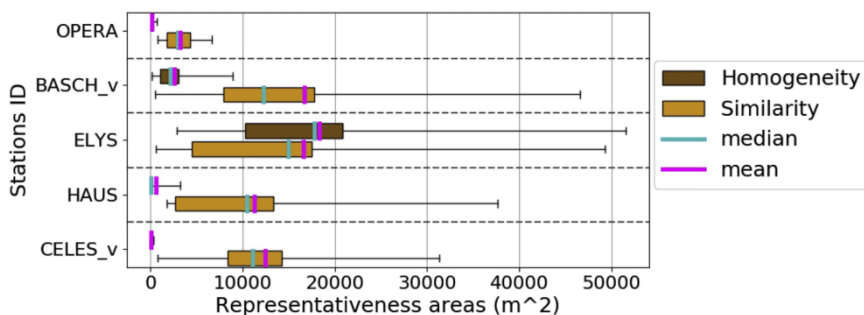


Fig. 6. Boxplots of NO_x homogeneity and similarity areas (m²) around traffic sites across the 10 days of the study. Blue and pink solid lines inside the boxes represent the median and mean values of the distribution respectively, lower and upper bounds of the boxes represent the 25th and 75th percentiles, and low and high whiskers represent the minimum and maximum values.

Table 4

NO_x traffic-oriented representativeness areas (m²) shared 70% of the time, average areas over the ten days of the study (m²) and standard deviation (m²).

ID STATIONS	Areas shared seven days out of ten		Statistics over ten days	
	Homogeneity (H) (m ²)	Similarity (S) (m ²)	Homogeneity: mean [standard deviation] (m ²)	Similarity: mean [standard deviation] (m ²)
OPERA	90	1125	165 [186]	3.2×10^3 [1.9×10^3]
BASCH_v	234	846	2.5×10^3 [2.4×10^3]	16.7×10^3 [13.6×10^3]
ELYS	2844	2529	18.3×10^3 [12.6×10^3]	16.6×10^3 [15.2×10^3]
HAUS	27	1728	615 [1189]	11.2×10^3 [10.3×10^3]
CELES_v	18	2745	74 [89]	12.4×10^3 [7.7×10^3]

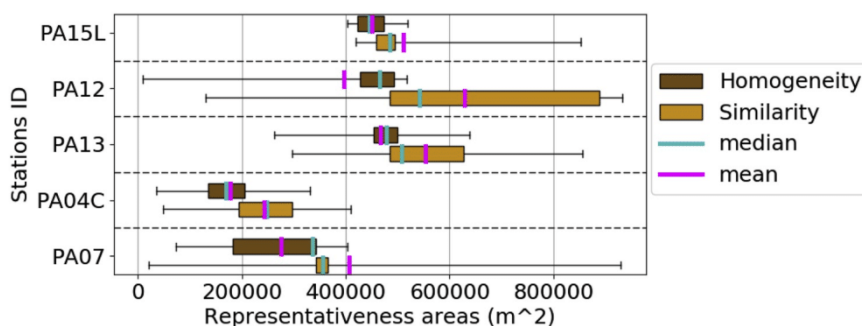


Fig. 7. Boxplots of NO_x homogeneity and similarity areas (m²) around urban background monitoring-sites during the ten days in March 2016. Blue and pink solid lines inside the boxes represent the median and mean values of the distribution respectively, the lower and upper bounds of the boxes represent the 25th and 75th percentiles and low and high whiskers represent the minimum and maximum values.

Table 5

NO_x background representativeness areas (m²) shared 70% of the time, average areas over the days (m²) and standard deviation of the areas over the ten days (m²).

ID STATIONS	Areas shared seven days out of ten		Statistics over ten days	
	Homogeneity (H) (m ²)	Similarity (S) (m ²)	Homogeneity: mean [standard deviation] (m ²)	Similarity: mean [standard deviation] (m ²)
PA15L	445.1×10^3	479.3×10^3	450.1×10^3 [34.7×10^3]	511.3×10^3 [117.5×10^3]
PA12	431.8×10^3	497.6×10^3	397.2×10^3 [161.7×10^3]	628.7×10^3 [252.9×10^3]
PA13	464.0×10^3	495.0×10^3	467.4×10^3 [95.3×10^3]	554.0×10^3 [154.9×10^3]
PA04C	121.7×10^3	193.8×10^3	176.8×10^3 [8.1×10^3]	242.7×10^3 [91.5×10^3]
PA07	186.0×10^3	350.7×10^3	275.9×10^3 [115.6×10^3]	406.9×10^3 [241.9×10^3]

these roads.

NO_x urban background representativeness areas obtained seven days out of ten (Table 5) are three hundred times wider than NO_x traffic representativeness areas (1.2×10^3 m² vs 366.5×10^3 m²). PA04C monitoring-site is the least representative (121.7×10^3 m² and 193.8×10^3 m² for the largest homogeneity and similarity areas respectively). PA13 (monitoring-site located in the Choisy Park) and PA12 in the schoolyard have the larger homogeneity and similarity areas (464.0×10^3 m² and 497.6×10^3 m² respectively). Homogeneity and similarity areas in most cases are of the same order of magnitude with the exception of PA07 station, where similarity area is two times larger than homogeneity area.

3.1.2. PM₁₀ representativeness areas

3.1.2.1. Traffic-oriented areas. PM₁₀ traffic-oriented areas are seventeen times larger than NO_x traffic-oriented areas. Fig. 8 summarizes PM₁₀ representativeness areas of traffic sites for the 10 days of the study. In some cases, PM₁₀ traffic-oriented areas cover a vast surface around the monitor (maximum of 846.2×10^3 m² for BASCH_v similarity area). A strong variability is observed between the monitoring-sites areas, with the highest ratio of twenty-eight being

found between the largest representative area around HAUS (279×10^3 m² averaged over ten days for both similarity and homogeneity) and the smallest one around OPERA (9.9×10^3 m²). PM₁₀ similarity areas are in general, larger than PM₁₀ homogeneity areas with a ratio between 1.5 (ELYS) and 5 (BASCH_v) except for HAUS areas where they both are of the same order of magnitude with very vast areas. In most of the cases, representativeness areas have a high variability across the ten days (see boxplots width of BASCH_v similarity area, ELYS, HAUS and CELES_v).

By considering areas shared 70% of the time (Table 6), PM₁₀ representativeness areas are nine times larger than NO_x representativeness area (11.0×10^3 m² vs 1.2×10^3 m²). A large variability is shown between stations with the OPERA site being the less representative monitor location (171 m² and 1.2×10^3 m² for homogeneity and similarity areas respectively). ELYS and BASCH_v have the largest homogeneity and similarity areas with 26.0×10^3 m² and 38.4×10^3 m² respectively. Similarity areas are larger than homogeneity areas for HAUS, OPERA and BASCH_v with a ratio of 4.8–13.6. Inversely, ELYS similarity areas are two times lower than the corresponding homogeneity areas. For CELES_v, similarity and homogeneity area are of the same order of magnitude.

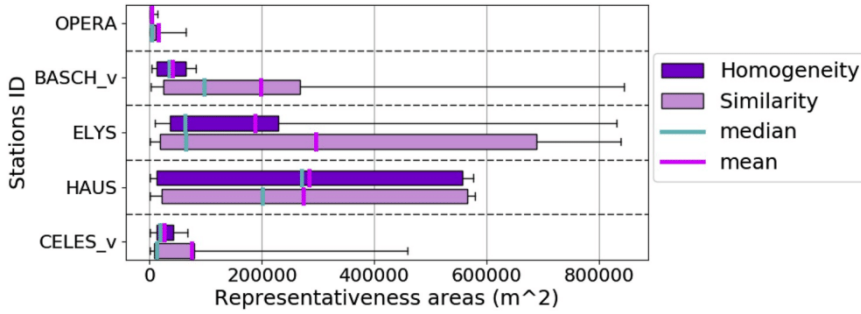


Fig. 8. Boxplots of PM₁₀ homogeneity and similarity areas (m²) around traffic station across the 10 days of the study. Blue pink solid lines inside the boxes represent the median and mean of the distributions respectively, lower and upper bounds of the boxes represent the 25th and 75th percentiles and low and high whiskers represent the minimum and maximum values.

Table 6

PM₁₀ traffic-oriented representativeness areas (m²) shared 70% of the time, average areas over the days (m²) and standard deviation of the areas over the ten days (m²).

ID STATIONS	Areas shared seven days out of ten		Statistics over ten days	
	Homogeneity (H) (m2)	Similarity (S) (m2)	Homogeneity: mean [standard deviation] (m2)	Similarity: mean [standard deviation] (m2)
OPERA	171	1.2×10^3	4.3×10^3 [4.4×10^3]	15.6×10^3 [23.6×10^3]
BASCH_v	2.8×10^3	38.4×10^3	40.5×10^3 [29.1×10^3]	197.7×10^3 [244.0×10^3]
ELYS	26.1×10^3	17.1×10^3	188.4×10^3 [252.4×10^3]	296.4×10^3 [360.0×10^3]
HAUS	2.6×10^3	12.3×10^3	283.8×10^3 [270.4×10^3]	274.4×10^3 [253.7×10^3]
CELES_v	5.2×10^3	4.6×10^3	25.7×10^3 [20.5×10^3]	74.4×10^3 [132.5×10^3]

3.1.2.2. Urban background areas. Fig. 9 summarizes PM₁₀ representativeness areas around urban background monitoring-sites for the 10 days in March 2016. PM₁₀ representativeness areas are generally larger than NO_x representativeness areas for urban background monitoring-site with a ratio of 1.9 across all days and urban background stations ($772 \times 10^3 \text{ m}^2$ vs $411 \times 10^3 \text{ m}^2$) and are comprised between $289 \times 10^3 \text{ m}^2$ and $988 \times 10^3 \text{ m}^2$.

Here again, we observe a large variability between representativeness areas of urban background monitoring-sites with a highest ratio of 1.8 between the most (PA12) and the less representative sites (PA04C). PM₁₀ homogeneity and similarity areas are of the same order of magnitude with mean values across all days and urban background stations equal to $791.5 \times 10^3 \text{ m}^2$ and $753.2 \times 10^3 \text{ m}^2$ respectively. PM₁₀ concentrations are more spatially homogeneous than NO_x concentrations

that have a high spatial gradient. For PA15L and PA12, the 25th percentile values of similarity areas are higher than the ones for homogeneity areas but the minimum and maximum values are very close.

By considering areas shared seven days out of ten (Table 7), PM₁₀ urban background representativeness areas are two times larger than NO_x urban background representativeness areas ($799.4 \times 10^3 \text{ m}^2$ vs $366.5 \times 10^3 \text{ m}^2$). PM₁₀ urban representativeness areas are also seventy-five times larger than PM₁₀ traffic representativeness areas ($799.4 \times 10^3 \text{ m}^2$ vs $10.6 \times 10^3 \text{ m}^2$). PA04C has the lowest representativeness area ($410.4 \times 10^3 \text{ m}^2$ and $428.2 \times 10^3 \text{ m}^2$ for homogeneity and similarity areas respectively). PA12 has the highest homogeneity ($931.7 \times 10^3 \text{ m}^2$) and similarity area ($955.9 \times 10^3 \text{ m}^2$).

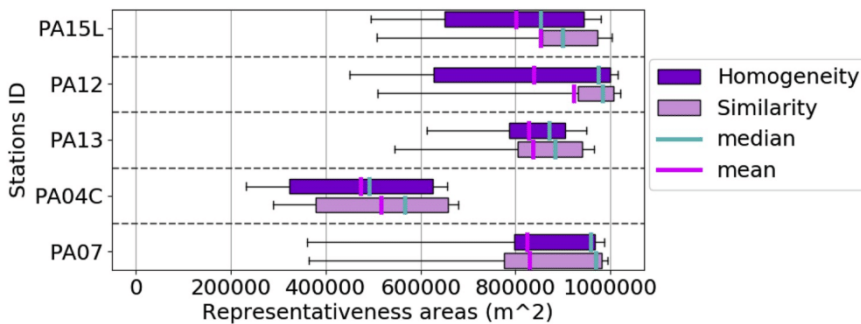


Fig. 9. Boxplots of PM₁₀ homogeneity and similarity areas (m²) on 10 days in March 2016 around urban background monitoring-sites. Blue and pink solid lines inside the boxes represent the median and mean of the distributions respectively, lower and upper bounds of boxes represent the 25th and 75th percentiles and low and high whiskers represent the minimum and maximum values.

Table 7

PM₁₀ background representativeness areas (m²) shared 70% of the time, average areas over the days (m²) and standard deviation of the areas over the ten days (m²).

ID STATIONS	Areas shared seven days on ten		Statistics over ten days	
	Homogeneity (H) (m ²)	Similarity (S) (m ²)	Homogeneity: mean [standard deviation] (m ²)	Similarity: mean [standard deviation] (m ²)
PA15L	833.0×10^3	881.9×10^3	801.7×10^3 [177.1×10^3]	852.4×10^3 [167.7×10^3]
PA12	931.7×10^3	955.9×10^3	838.3×10^3 [226.1×10^3]	922.8×10^3 [147.3×10^3]
PA13	849.4×10^3	855.7×10^3	828.0×10^3 [108.4×10^3]	836.5×10^3 [136.5×10^3]
PA04C	410.4×10^3	428.2×10^3	473.8×10^3 [156.9×10^3]	516.3×10^3 [146.6×10^3]
PA07	920.2×10^3	928.1×10^3	824.4×10^3 [238.1×10^3]	829×10^3 [242.8×10^3]

3.2. Characterization of urban areas

A large variability among stations is observed for the most probable traffic-oriented similarity areas (i.e. mean areas and areas shared seven days on ten), ranging from 18 m² (NO_x homogeneity area at CELES_v station) to 38.4×10^3 m² (PM₁₀ similarity area at BASCH_v station), see Tables 4 and 6. This large variability may be related to the monitoring-site location and its vicinity (see Table 1). Two traffic-oriented monitoring-sites are located on a crossroad (OPERA is in the middle of the crossroad and BASCH_v is at the exit of the crossroad). Three traffic-oriented monitoring-sites are located on a sidewalk along roadside (HAUS, ELYS and CELES_v).

In this section, we will investigate the possible impact of specific urban features around the station on the differences in similarity areas. The specific features characterizing the representativeness areas selected around traffic-oriented monitoring-sites are listed in Table 8. Their only common feature is the sidewalk in the immediate proximity to the road.

Table 8

Characteristic features selected around traffic-oriented monitoring-sites.

Sites	Characteristic features
BASCH _v – crossroad exit	Sidewalk and part of the road leading to the crossroad
OPERA – crossroad	Small area close to the monitoring-site
CELES _v – border of the Seine River	Sidewalk and sometimes Seine river and river bank
ELYS – Champs-Élysées Avenue	Sidewalk on one or both sides of the road
HAUS – Boulevard Haussmann	Sidewalk or all elements in the area (PM ₁₀)

3.2.1. Traffic-oriented monitoring-sites located on crossroad (BASCH_v and OPERA)

As shown in section 3.1.1 NO_x representativeness areas are smaller than the PM₁₀ representativeness areas for both monitoring-sites (0.6×10^3 m² vs 10.6×10^3 m² averaged across criteria and stations). NO_x and PM₁₀ similarity areas at the OPERA monitoring-site are of the same order of magnitude (1125 m² and 1179 m², see Tables 4 and 6) and both include a roadside segment of 140 m along the large crossroad. However, as shown in Fig. 10 the diurnal variability of the selected PM₁₀ similarity area is much higher than for NO_x (see also the standard deviations of similarity areas in Tables 4 and 6).

Two times out of ten, PM₁₀ similarity area around the OPERA monitoring-site includes main traffic-roads and secondary axes over a vast area around the station (Fig. 10d).

The most probable NO_x similarity area around the BASCH_v station is smaller than the OPERA one. It includes a small portion of the sidewalk and the street along a distance of about 100 m. The PM₁₀ similarity area around BASCH_v is larger than the one around the OPERA monitor and also, larger than the NO_x area around BASCH_v. It includes a part of the street side and the sidewalk where the monitoring-site is located (avenue Jean Moulin) over a length of 700 m but also two other large avenues, one leading to the crossroad, and another one further down Jean Moulin avenue on the south-east direction. Thus, traffic sources from different avenues more or less distant to the monitor site may contribute to the pollutant concentration measured at the BASCH_v monitoring-site. On the contrary, OPERA is isolated in the middle of the crossroad and does not integrate pollution sources from the nearby roadside except for a small road segment. These differences in the topography of the two stations explain the large discrepancy in PM₁₀ representativeness areas between the two sites (most probable PM₁₀ similarity area 38.4×10^3 m² (BASCH_v) vs 1179 m² (OPERA)).

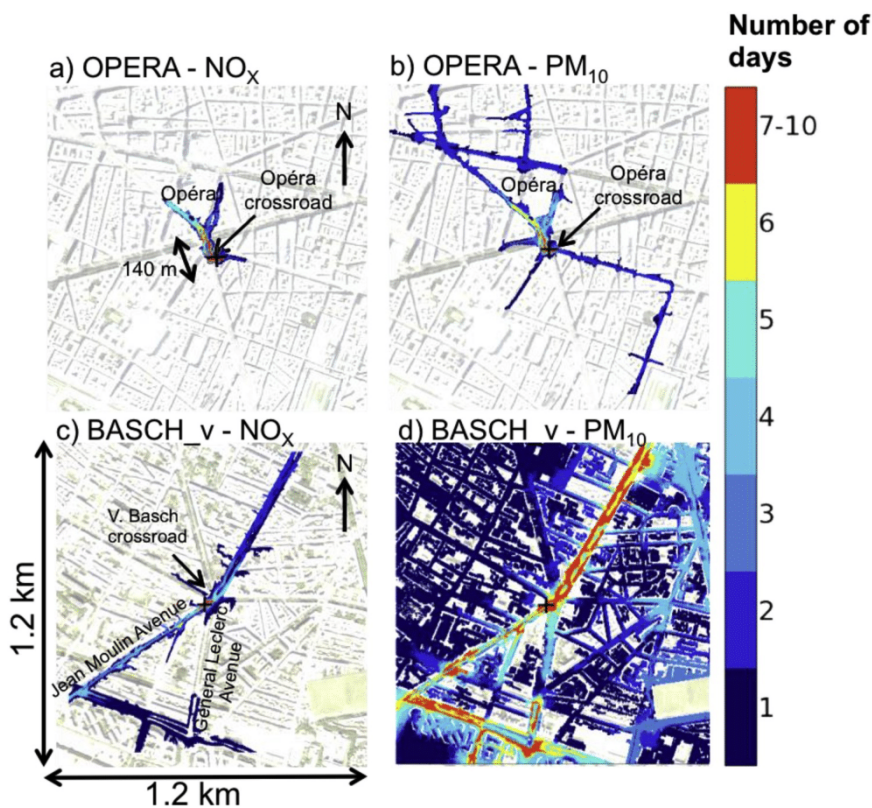


Fig. 10. NO_x (left) and PM₁₀ (right) similarity areas cumulated over the 10 days of the study for OPERA monitoring-site (top) and BASCH_v monitoring-site (bottom). The black cross indicates the location of the pseudo-station. The color bar indicates the number of days for which a street portion is selected. The most probable area (selected at least 7 days) is shown in red color in each case. Black arrow indicates the north wind direction.

3.2.2. Traffic-oriented monitoring-sites located along a roadside (ELYS, HAUS and CELES_v)

ELYS monitor station is located on the sidewalk of a wide urban boulevard. NO_x similarity area includes the sidewalk of the station along a distance of 500 m, while the much larger PM₁₀ area ($2.5 \times 10^3 \text{ m}^2$ vs $17.1 \times 10^3 \text{ m}^2$ for NO_x and PM₁₀ respectively) includes both sidewalks of the road along a distance of 700 m as well as a 400 m segment of a sidewalk from the station in northwest direction (Fig. 11). PM₁₀ similarity area at the ELYS monitor has a stronger variability during the ten days of the study than NO_x similarity area with a standard deviation over ten days of $360 \times 10^3 \text{ m}^2$ (PM₁₀) and $15.2 \times 10^3 \text{ m}^2$ (NO_x). Secondary arteries perpendicular to the primary axe are included in the selected area 10% of the time for NO_x and 50% of the time for PM₁₀. Four days out of ten, all the streets of the domain shown in Fig. 11b as well as part of the Seine River are included in the PM₁₀ similarity area.

The most probable PM₁₀ similarity area around Boulevard Haussmann monitoring-site (HAUS) is larger than the NO_x one ($12.3 \times 10^3 \text{ m}^2$ and $1.7 \times 10^3 \text{ m}^2$, respectively). NO_x similarity area includes the sidewalk of the station along a distance of 200 m from the station and a small street perpendicular to the Haussmann boulevard, in the northwest direction. PM₁₀ similarity area includes a portion of sidewalk over 400 m and secondary axes (Fig. 11). The PM₁₀ similarity area has a higher variability for these ten days than the NO_x area (see standard deviations on Tables 4 and 6). 20–30% of the time, NO_x

similarity areas include small segments of secondary axes, whereas 40 and 50% of the time, PM₁₀ similarity areas include all streets segments except from the busy roadside near the monitoring-site.

The extents of the most probable NO_x and PM₁₀ similarity areas around the Seine River Bank monitoring site (CELES_v) are close to each other ($2.7 \times 10^3 \text{ m}^2$ vs $4.6 \times 10^3 \text{ m}^2$ for NO_x and PM₁₀ respectively). They both include a segment of the sidewalk and roadside south-east from the monitoring-site along a distance of 640 m. Similarity areas extend along the bank of the Seine river four days out of ten for PM₁₀ whereas for NO_x selected similarity is not as widely spread. One day out of ten, PM₁₀ similarity area selects secondary axes and crosses the bridge. Note that the Southeast part of the road is not selected, due to the lack of emission data in the simulation.

The most probable NO_x similarity areas are of the same order of magnitude for the three monitoring sites (around $2 \times 10^3 \text{ m}^2$). However, the CELES_v PM₁₀ area is smaller than the two other ($4.6 \times 10^3 \text{ m}^2$ (CELES_v) vs. $17.1 \times 10^3 \text{ m}^2$ (ELYS) and $12.2 \times 10^3 \text{ m}^2$ (HAUS)). This difference may be attributed to the road configuration. ELYS is located at the edge on the 20 m wide sidewalk of a 30 m wide avenue, whereas HAUS is located at the edge on the 9 m wide sidewalk of a 18 m wide road. The CELES_v monitor has the particularity to be located on the sidewalk (6 m wide) of a road (12 m wide) from which it is separated with a 3 m wide bicycle path.

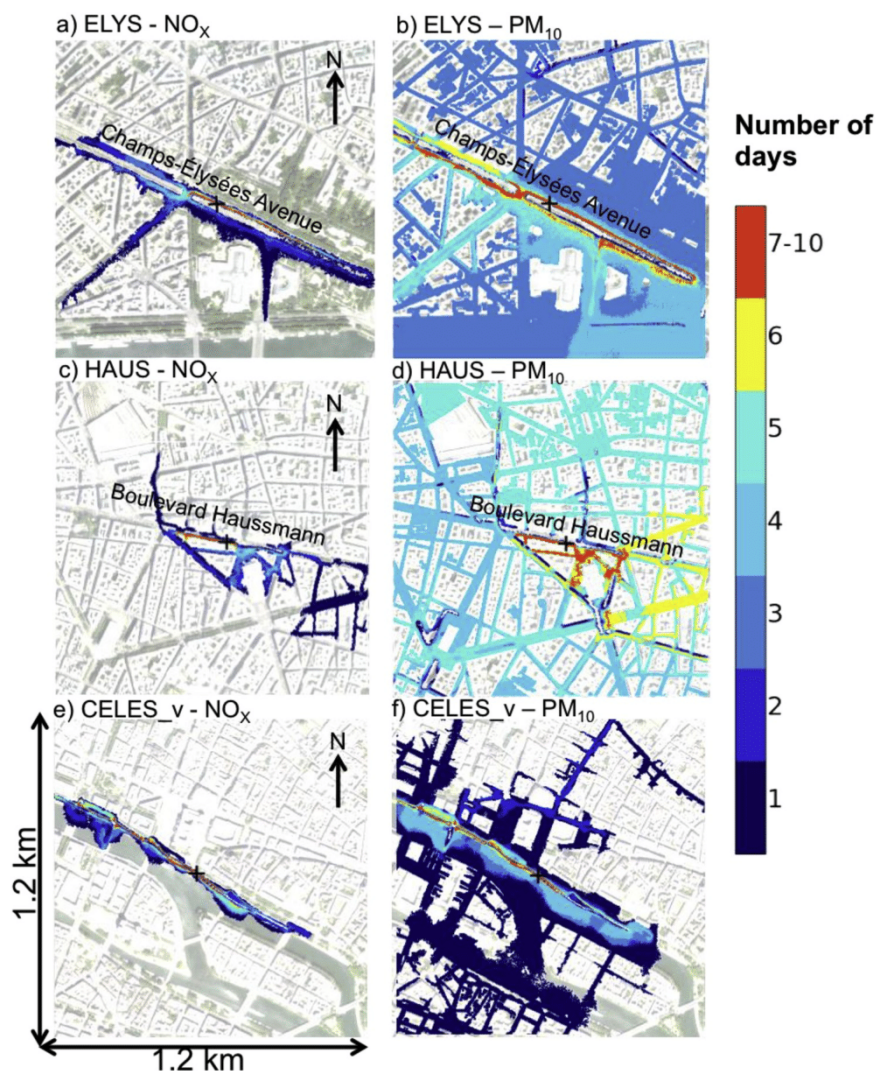


Fig. 11. Same as Fig. 10, but for three monitor sites, ELYS (top), HAUS (middle) and CELES_v (bottom).

4. Discussions and conclusions

A new methodology has been developed to assess daily NO_x and PM_{10} representativeness areas around traffic and urban-background monitoring-sites of the air-quality network of AIRPARIF in the city of Paris, France. It is based on a high-resolution model simulation (PMSS model) during ten days in March 2016, including a typical PM_{10} winter pollution episode due to high-pressure conditions. The PMSS modeling system provided pollutant concentration fields at a 3 m horizontal resolution with in a 2 m thick surface layer. Modeled concentrations were compared to measurements at 10 monitoring-sites, showing that simulations are in good agreement with the observations for PM_{10} and stay realistic for NO_x even if model has some difficulty to reproduce NO_x concentration peaks. To perform this study, we assume that the model properly reproduces the pollution spatial variability in Paris.

A representativeness area should be seen as an area where pollution variations are very close to those at the monitoring-site. Two kinds of such “resemblance” were investigated in this study: homogeneity and similarity. Homogeneity areas integrate points where daily concentrations do not differ more than 20% of the daily averaged value at the monitoring-site location. Similarity areas incorporate points satisfying

the additional condition to be strongly correlated with concentrations at the location of the monitoring-site. We first determined the optimal thresholds on correlation and normalized root mean square, considering that the obtained area surface must be confined within the $[300 \text{ and } 400 \times 10^3 \text{ m}^2]$ range, for a maximum number of days and a maximum number of sites. A sensitivity analysis was performed to choose the lowest NRMSE and the highest value of correlation coefficient specific for NO_x and PM_{10} traffic-oriented sites.

This new approach has been found to be robust despite some gaps in the traffic emissions and the small dataset (only ten days). The same criteria thresholds were applied for the different monitoring-sites of the AIRPARIF network leading to realistic representativeness areas for both urban background and traffic sites. The similarity criteria thresholds applied in the present study are most probably dependent of the study area and period as well as of the dispersion model. However the proposed methodology is generic and could be easily repeated under a different study setup.

Homogeneity and similarity areas are most of the time of the same order of magnitude for both NO_x and PM_{10} urban background monitoring-site. The difference between these two representativeness areas is more pronounced around traffic monitor-sites and in most cases, the

most probable similarity area is larger than the most probable homogeneity area. The definition of similarity areas is composed of two components, a threshold on the correlation coefficient, to ensure that the time series inside the similarity area have correlated variations, and a threshold on the NRMSE, to ensure that the time series variations have similar values. As a consequence, the daily average does not strongly differ as well. The overall criterion to select similarity areas integrates the condition of homogeneity with an additional constraint on the correlation coefficient. To check this hypothesis, we compared representativeness areas selected on the basis of a condition only on the NRMSE against homogeneity areas. For PM₁₀ of urban background stations and for NO_x of 3/5 urban background stations (PA15L, PA12, PA13), NRMSE and homogeneity criteria select an area (km²) similar with respectively an averaged ratio between the most probable NRMSE area and the most probable homogeneity area to 1 and 1.1. NRMSE most probable NO_x areas are wider than most probable NO_x homogeneity ones for 2/5 urban stations with a ratio to 1.6 (PA04C) and 1.9 (PA07), and much larger for traffic stations with an averaged ratio to 77.2 (range ratio between 1.3 (ELYS) and 213 (CELES_v)).

Most probable PM₁₀ NRMSE areas are larger than most probable PM₁₀ homogeneity area for OPERA, BASCH_v and HAUS with an averaged ratio between the most probable NRMSE area and the most probable homogeneity area of 8.6. For CELES_v they have the same value. For ELYS monitoring-site, most probable PM₁₀ homogeneity areas are slightly larger than most probable PM₁₀ NRMSE areas with a ratio of 0.7.

NRMSE thresholds were determined using a sensitivity analysis which integrates the daily variability and the diversity of urban elements landscape (by defining areas limits > 300 m² and < 400 × 10³ m²). Given the retained thresholds, this criteria is less stringent than homogeneity criteria. However, homogeneity criteria do not integrate any area specificities or daily peak.

In the FAIRMODE and AQUILA inter-comparison studies, representativeness areas of one traffic and two urban background monitoring-sites were estimated from annual year data (in Belgium). In this framework, results obtained with eleven methodologies have been compared, and show an important variability (FAIRMODE report, Kracht et al. 2017): PM₁₀ traffic-oriented representativeness areas are estimated between 0.001 km² (10³ m²) and 160 km² whereas here, most probable PM₁₀ traffic-oriented similarity area range between 0.001 km² (1.2 × 10³ m²) and 0.003 km² (38.4 × 10³ m²). In the FAIRMODE report, PM₁₀ urban background representativeness areas are estimated between 1 km² (10⁶ m²) and 700 km². In our study, most probable PM₁₀ similarity and homogeneity areas of urban background monitoring-sites range between 0.4 km² and 0.9 km² due to the limitation of the study domain fixed 1.2 km × 1.2 km around each station.

The representativeness areas estimated in this study show a large daily variability. Traffic variations between working days, week-ends and national holidays as well as meteorological conditions (mainly the wind) are probably the two principal causes of this variability for both PM₁₀ and NO_x. Unfortunately, the small dataset did not allow us to further investigate the daily traffic variation impact on representativeness areas.

Among the causes of this day-to-day variability, the wind can be suspected to play a role. The influence of the wind speed and wind direction has been investigated. Wind direction was not found to have a significant effect based on this small sample (only ten days and four windy days). The influence of the wind speed on the representativeness areas has been investigated by comparing similarity areas on windy days against non-windy days, we used 3 m/s as threshold to distinguish between the two. Preliminary results suggest that similarity areas are smaller in case of windy days and that the wind impact is stronger for traffic-oriented sites. The ratio between the difference of no windy area and windy area on the mean area for traffic stations is 1.7 and 1.8 respectively for NO_x and PM₁₀ and for urban background stations is 1.1 and 0.6 for NO_x and PM₁₀. However, such an analysis should be carried

out over longer simulation periods to provide statistically significant results.

The study showed that in general, urban background areas are representative of the neighborhood in the vicinity of the monitoring-site whereas traffic-oriented monitoring-sites are representative of specific urban features such as sections of roads and sidewalks along the road. Averaged over all monitoring-sites, most probable similarity areas for urban background monitoring-sites are about seventy times larger than traffic representativeness areas. Averaged over all stations, PM₁₀ most probable similarity areas are 2 times larger than NO_x areas around urban monitors and 8 times around traffic sites. Even if PM₁₀ concentrations were normalized to remove part of the regional signal (see Sect. 2.4) the remaining trace of large-scale effects seem to overshadow local effects such as the sharp concentration gradients over the road network.

As a next step, AIRPARIF monitoring-site measurements will be assimilated to the model simulations in order to correct some model artifacts and provide more reliable concentrations as input to the computation of the similarity areas. The present study should be expanded over a much greater number of simulated days, to provide more general conclusions and confirm our assumptions on the influence of parameters such as wind speed and traffic patterns. Then, measurements at monitor sites could be spread across similarity areas depending on the wind vector and provide an information particularly relevant for human exposure estimation.

Acknowledgements

Acknowledgments to ARIA Technologies for giving those simulations for ten days.

References

- APHEKOM, 2012. Improving Knowledge and Communication for Decision Making on Air Pollution and Health in Europe. <http://aphekom.org>.
- Beauchamp, M., Malherbe, L., 2016. Annexe technique au rapport intitulé estimation de l'exposition des populations aux dépassements de seuils réglementaires. In: Interpolation des sorties de modèles urbains par krigeage avec dérive polynomiale. note lcsqa. - <http://www.lcsqa.org>.
- Blanchard, C.L., Tanenbaum, S., Hidy, G.M., 2014. Spatial and temporal variability of air pollution in Birmingham, Alabama. *Atmos. Environ.* 89, 382–391. <https://doi.org/10.1016/j.atmosenv.2014.01.006>.
- Bobbia, M., Cori, A., de Fouquet, C., 2008. Représentativité spatiale d'une station de mesure de la pollution atmosphérique. *Pollut. Atmos.* <https://doi.org/10.4267/pollution-atmospherique>.
- CITEPA. Center Interprofessionnel Technique d'Etudes de la Pollution Atmosphérique <https://www.citepa.org/en/air-and-climate/pollutants-and-ghg>.
- Duyzer, J., van den Hout, D., Zandveld, P., van Ratingen, S., 2015. Representativeness of air quality monitoring networks. *Atmos. Environ.* 104, 88–101. <https://doi.org/10.1016/j.atmosenv.2014.12.067>.
- European Commission (EC), 2008. Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe. The framework directive. *Off. J. Eur. Union En. Ser.* L152/51.
- European Commission (EC), 2011/850/EU. Commission Implementing Decision of 12, December 2011. Laying down rules for Directives 2004/107/EC and 2008/50/EC of the European Parliament and of the Council as regards the reciprocal exchange of information and reporting on ambient air quality. *Off. J. Eur. Union En. Ser.* L 335/86.
- Hagemann, R., Corsmeier, U., Kottmeier, C., Rinke, R., Wieser, A., Vogel, B., 2014. Spatial variability of particle number concentrations and NO_x in the Karlsruhe (Germany) area obtained with the mobile laboratory "AERO-TRAM". *Atmos. Environ.* 94, 341–352. <https://doi.org/10.1016/j.atmosenv.2014.05.051>.
- Hagenbjörk, A., Malmqvist, E., Mattisson, K., Sommar, N.J., Modig, L., 2017. The spatial variation of O₃, NO, NO₂ and NO_x and the relation between them in two Swedish cities. *Environ. Monit. Assess.* 189. <https://doi.org/10.1007/s10661-017-5872-z>.
- Harrison, R.M., 2018. Urban atmospheric chemistry: a very special case for study. *Clim. Atmos. Sci.* 1, 5. <https://doi.org/10.1038/s41612-017-0010-8>.
- Host, S., 2013. « Exposition à la pollution atmosphérique liée au trafic routier et risques sanitaires », *Vertigo - la revue électronique en sciences de l'environnement*. <http://journals.openedition.org/vertigo/12816>.
- IARC report, 2013. https://www.iarc.fr/en/media-centre/iarcnews/pdf/pr221_E.pdf.
- Janssen, S., Dumont, G., Fierens, F., Deutsch, F., Maiheu, B., Celis, D., Trimpeneers, E., Mensink, C., 2012. Land use to characterize spatial representativeness of air quality monitoring stations and its relevance for model validation. *Atmos. Environ.* 59, 492–500. <https://doi.org/10.1016/j.atmosenv.2012.05.028>.

- Kracht, O., Santiago, J.-L., Martin, F., Piersanti, A., Cremona, G., Righini, G., Vitali, L., Delaney, K., Basu, B., Ghosh, B., Spangl, W., Brendle, C., Latikka, J., Kousa, A., Pärjälä, E., Meretoja, M., Malherbe, L., Letinois, L., Beauchamp, M., Lenartz, F., Hutsemekers, V., Nguyen, L., Hoogerbrugge, R., Eneroth, K., Silvergren, S., Hooyberghs, H., Viaene, P., Maiheu, B., Janssen, S., Roet, D., Gerboles, M., 2017. Spatial Representativeness of Air Quality Monitoring Sites - Outcomes of the FAIRMODE/AQUILA Intercomparison Exercise. EUR 28987 EN, Publications Office of the European Union, Luxembourg. <https://doi.org/10.2760/60611, JRC108791>. ISBN 978-92-79-77218-4.
- Martin, F., Fileni, L., Palomino, I., Vivanco, M.G., Garrido, J.L., 2014. Analysis of the spatial representativeness of rural background monitoring stations in Spain. *Atmos. Pollut. Res.* 5, 779–788. <https://doi.org/10.5094/APR.2014.087>.
- Moussafir, J., Olry, C., Nibart, M., Albergel, A., Armand, P., Duchenne, C., Mahé, F., Thobois, L., Loaëc, S., Oldrini, O., 2014. AIRCITY: A Very High Resolution Atmospheric Dispersion Modeling System for Paris, vol. 1 American Society of Mechanical Engineers, Fluids Engineering Division (Publication) FEDSM <https://doi.org/10.1115/FEDSM2014-21820>.
- Nappo, C.J., Caneill, J.Y., Furman, R.W., Gifford, F.A., Kaimal, J.C., Kramer, M.L., Lockhart, T.J., Pendergast, M.M., Pielke, R.A., Randerson, D., Shreffler, J.H., Wyngaard, J.C., 1982. The workshop on the representativeness of meteorological observations, June 1981. Boulder, Colorado. *Bull. Am. Meteorol. Soc.* 63, 761e764.
- Nguyen, P.L., Stefess, G., de Jonge, D., Snijder, A., Hermans, P.M.J.A., van Loon, P., Hoogerbrugge, R., 2012. Evaluation of the Representativeness of the Dutch Air Quality Monitoring Stations. The National, Amsterdam, Noord-Holland, Rijnmond-Area, Limburg and Noord-Brabant Networks. RIVM Report 680704021/2012.
- Piersanti, A., Vitali, L., Righini, G., Cremona, G., Ciancarella, L., 2015. Spatial representativeness of air quality monitoring stations: a grid model based approach. *Atmos. Pollut. Res.* 6, 953–960. <https://doi.org/10.1016/j.apr.2015.04.005>.
- REVIHAAP Project, 2013. Review of evidence on health aspects of air pollution. http://www.euro.who.int/_data/assets/pdf_file/0004/193108/REVIHAAP-Final-technical-report-final-version.pdf?ua=1.
- Santiago, J.L., Martin, F., Martilli, A., 2013. A computational fluid dynamic modelling approach to assess the representativeness of urban monitoring stations. *Sci. Total Environ.* 454–455, 61–72. Senat Report n°610, 2015. <https://doi.org/10.1016/j.scitotenv.2013.02.068>. http://www.senat.fr/fileadmin/Fichiers/Images/commission/enquete/pollution_air/Synthese_CE_Pollution_de_l_air.pdf.
- Spangl, W., Schneider, J., Moosmann, L., Nagl, C., 2007. Representativeness and Classification of Air Quality Monitoring Stations – Final Report. Service Contract to the European Commission - DG Environment Contract No. 07.0402/2005/419392/MAR/CI. Umweltbundesamt, Wien. Reports, Bd. REP-0121.
- Vardoulakis, S., Gonzalez-Flesca, N., Fisher, B.E.A., Pericleous, K., 2005. Spatial variability of air pollution in the vicinity of a permanent monitoring station in central Paris. *Atmos. Environ.* 2725–2736. Fourth International Conference on Urban Air Quality: Measurement, Modelling and Management, 25–28 March 2003 39. <https://doi.org/10.1016/j.atmosenv.2004.05.067>.
- Vitali, L., Morabito, A., Adani, M., Assennato, G., Ciancarella, L., Cremona, G., Giua, R., Pastore, T., Piersanti, A., Righini, G., Russo, F., Spagnolo, S., Tanzarella, A., Tinarelli, G., Zanini, G., 2016. A Lagrangian modelling approach to assess the representativeness area of an industrial air quality monitoring station. *Atmos. Pollut. Res.* 7, 990–1003. <https://doi.org/10.1016/j.apr.2016.06.002>.

Dans cette étude (chapitre 4 de ce manuscrit), nous avons déterminé les zones de représentativité des stations de mesures uniquement à partir des simulations du modèle PMSS dans des « tuiles » de 1.2 km × 1.2 km avec la station AIRPARIF au centre.

Nous avons considéré que l'approche de similarité - qui sélectionne des aires à la fois avec un fort coefficient de corrélation et aussi avec une faible erreur quadratique moyenne normalisée par la moyenne journalière (NRMSE) par rapport à la concentration simulée à l'endroit de la station - est plus adaptée pour définir la représentativité d'une mesure que l'approche d'homogénéité (écart moyen journalier : +/-20 % de la concentration moyenne simulée à l'endroit de la station). Cette définition de la représentativité est d'autant plus pertinente, si l'on souhaite étendre ce concept de représentativité à des capteurs mobiles à bas coût qui fourniront davantage une indication de la variabilité spatio-temporelle de la pollution que d'une valeur fixe.

Nous avons également considéré que l'aire de représentativité la plus probable est l'aire de représentativité sélectionnée au moins sept jours sur dix. En annexe de ce manuscrit, se trouvent de nouvelles figures présentées lors de la soutenance de cette thèse (slides 20/35 et 21/35), indiquant les aires de représentativité les plus probables sur la répartition des aires des dix jours sur l'ensemble des stations pour les NO_x et les PM₁₀. Et également une équivalence de ces aires de représentativité les plus probables en terme de (1) distance de représentativité en considérant une rue de six mètres de large pour les stations trafic et (2) un rayon de représentativité pour les stations de fond urbain. Pour ces approximations, nous avons supposé que la station était au centre de l'aire de représentativité.

Les stations trafic d'AIRPARIF à Paris, sont représentatives des trottoirs et parcelles de routes alors que les stations de fond sont représentatives d'un quartier.

Les aires de représentativité des PM₁₀ sont bien plus grandes que celles des NO_x. Ceci peut s'expliquer par l'homogénéité des PM₁₀ en Île-de-France dont la principale source d'émission est le secteur résidentiel et tertiaire et non le trafic routier (inventaire AIRPARIF de l'année 2015).

Nous avons également remarqué que le modèle PMSS est performant pour reproduire les mesures de PM₁₀ au niveau des stations AIRPARIF. Pour les NO_x, les erreurs entre le modèle et les maxima des mesures peuvent être importantes. Ce biais peut être attribué en partie au cadastre d'émissions de trafic en entrée du modèle PMSS.

Dans le chapitre 5 de ce manuscrit, nous avons développé un modèle bayésien qui combine les simulations numériques du modèle PMSS et les mesures passées et actuelles des stations AIRPARIF pour : (1) améliorer les simulations du modèle PMSS au niveau des stations et (2) propager la mesure des stations dans les zones de représentativité et ainsi rendre plus réalistes les champs de concentrations simulés dans ces zones.

Chapitre 5. Estimation bayésienne des concentrations de NO_x et PM₁₀ en utilisant des simulations numériques haute résolution et des mesures de surface à Paris /« NO_x and PM₁₀ Bayesian concentration estimates using high-resolution numerical simulations and ground measurements over Paris, France »

5.1 Concept bayésien

En statistique bayésienne, on évalue notre état des connaissances avant et après l'apprentissage de la donnée (O'Hagan, 2008). Avant l'apprentissage de la donnée, on parle de connaissance *a priori*. L'expert prédit des valeurs possibles du paramètre inconnu θ qu'il représente sous forme de distribution à partir d'informations disponibles (archives, etc.).

La connaissance *a priori* du paramètre inconnu θ est mise à jour grâce à l'apprentissage de la donnée pour obtenir une distribution *a posteriori*. La distribution *a posteriori* résume l'information qu'on a du paramètre θ après avoir observé la donnée.

Le théorème de Bayes s'écrit ainsi : $f(\theta|x) \propto f(\theta).f(x|\theta)$

La distribution *a posteriori* $f(\theta|x)$ est le produit du prior $f(\theta)$ et de la vraisemblance $f(x|\theta)$ à une constante près. La vraisemblance étant la distribution de probabilité des données conditionnées par le paramètre θ . La moyenne de la distribution *a posteriori* est un estimateur souvent utilisé pour indiquer la valeur la plus attendue pour le paramètre θ .

La distribution *a priori* est la plus grande différence entre l'approche bayésienne et une approche fréquentiste. Les probabilités bayésiennes changent selon l'évolution de l'état des connaissances et sont *subjectives* puisqu'elles reflètent un jugement probabiliste personnel.

Bien que *subjective*, la distribution *a priori* est un avantage car elle permet d'intégrer l'ensemble des informations disponibles pour analyser les données. La subjectivité se retrouve aussi dans la distribution *a posteriori* qui représente la croyance sur paramètre inconnu θ d'une personne en particulier : celle qui a construit le modèle bayésien.

En statistique classique, on interprète les données avec une fréquence de probabilité. Le paramètre inconnu θ est pré-assigné à une valeur fixe associée à une procédure statistique qui collecte aléatoirement les données dans les mêmes conditions. Toute l'information est

contenue dans la donnée. Cela suppose un grand jeu de données et une expérience reproductible. En bayésien, le paramètre inconnu θ est unique et ne peut pas être considéré comme une part d'une séquence répétable. On ne peut pas lui assigner une fréquence de probabilité. On utilise les données pour interpréter les paramètres ce qui revient à poser le problème dans le sens inverse.

Les statistiques fréquentistes n'échappent pas non plus à leur part de subjectivité, notamment par les différents choix qui sont faits à plusieurs niveaux (Collectif Bayes, 2015) :

- le choix de la fonction de vraisemblance (comme en bayésien)
- le choix d'une statistique de test et du niveau de significativité
- le choix d'une technique d'estimation : maximum de vraisemblance, interpolation optimale, assimilation variationnelle, etc.

La démarche bayésienne est cohérente. La loi *a posteriori* obtenue peut servir de loi *a priori* à une interprétation de nouvelles données. Cette approche est propice à l'accumulation de connaissances contrairement aux statistiques classiques où on repart de zéro.

Notons également que cette démarche est presque naturelle.

On peut donner comme exemple le diagnostic d'une maladie d'un patient par un médecin. Dans ce contexte, les résultats des analyses sanguines sont les données. Alors que l'expertise vis à vis du patient par le médecin se fait à partir de l'historique médical du patient, de son style de vie, de la survenue ou non d'une épidémie virale, etc. (prior). En combinant les résultats d'analyse médicale (données) et l'historique du patient (prior), le médecin met à jour sa connaissance *a priori* pour converger vers une connaissance *a posteriori* : le diagnostic de la maladie.

5.2 Résumé de l'article

5.2.1 Description du modèle bayésien

Depuis une quinzaine d'années, les statistiques bayésiennes sont très en vogue dans les sciences environnementales pour améliorer les prédictions spatiales des concentrations de polluants (Pirani *et al.*, 2018, Amin *et al.*, 2015, Millan *et al.*, 2009, Sahu *et al.*, 2005 entre autres) et le dépôt (Cowles *et al.*, 2003).

Au cours de ce travail de thèse, nous avons développé un modèle spatio-temporel bayésien qui combine les simulations haute résolution (3 m \times 3 m) du modèle PMSS et les mesures d'AIRPARIF pour dix jours de mars 2016. Cette méthode exploite de manière optimale ces plusieurs sources d'information : (1) la variabilité spatio-temporelle du modèle simulée à partir d'équations décrivant les processus physico-chimiques des polluants en milieu urbain, (2) la précision de la mesure au niveau des stations AIRPARIF (information locale), (3) la probabilité de représentativité des pixels inclus dans la zone de représentativité. L'objectif ici est d'améliorer l'estimation des concentrations de NO_x et PM_{10} au niveau des stations de mesures et de propager la mesure de la station dans la zone de représentativité préalablement définie (chapitre 4).

Dans ce modèle bayésien, les observations passées d'AIRPARIF des concentrations de NO_x et PM_{10} représentent la connaissance *a priori* du paramètre inconnu θ . Le paramètre inconnu θ

étant la concentration en polluant à un pixel donné dans l'espace et dans le temps. La donnée est la concentration simulée par le modèle PMSS.

Dans cette étude, nous avons choisi le logarithme népérien des concentrations et non les concentrations pour travailler uniquement avec des concentrations positives et également parce que la formule de la distribution *a posteriori* impose de travailler avec des distributions normales. La distribution *a priori* des concentrations étant une distribution proche de log-normale, en passant en logarithme de concentration nous obtenons ainsi une distribution normale.

Dans l'exemple (**Figure 35**, on applique le théorème de Bayes sur les logarithmes de concentrations (avec des concentrations exprimées en $\mu\text{g}/\text{m}^3$) au niveau d'une station de mesure AIRPARIF située Place de l'Opéra.

La connaissance *a priori*, logarithme népérien de la concentration des NO_x à 15 h le 2 mars, est comprise entre 4.8 et 6 avec un maximum à 5.4. La vraisemblance est comprise entre 4.4 et 5.3 avec un maximum à 4.9. La distribution *a posteriori* combine la distribution *a priori* et la vraisemblance. La vraisemblance décale la distribution du prior vers des valeurs plus faibles (vers la gauche) pour obtenir la distribution *a posteriori*. La distribution *a posteriori* est plus informative (distribution normale plus fine) que les deux autres distributions. Elle est centrée autour de 5, ce qui est un compromis entre la vraisemblance et le prior et on s'est donc rapproché de la concentration observée à la station AIRPARIF (vs concentration initiale simulée par le modèle PMSS : 5.2).

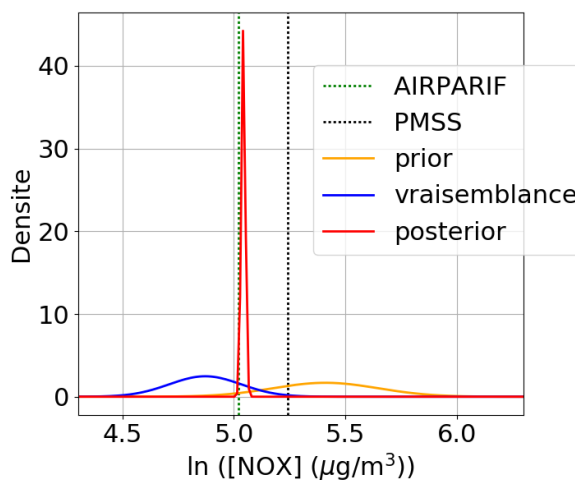


Figure 35 : Exemple de mise à jour bayésienne du logarithme de la concentration de NO_x au niveau de la station AIRPARIF située Place de l'Opéra à 15 h le 2 mars 2016 avec la distribution *a priori* (orange), la vraisemblance (bleu) et la distribution *a posteriori* (rouge), la mesure AIRPARIF (ligne verticale verte en pointillé) et la concentration simulée par le modèle PMSS (ligne verticale noire en pointillé).

Une des principales difficultés pour utiliser les statistiques bayésiennes est de définir (« elicitate » en anglais) la distribution *a priori* et la vraisemblance les plus adaptées.

Le choix du prior est la principale différence entre l'approche Bayésienne et les statistiques fréquentistes.

Dans notre cas, pour choisir le prior le plus approprié, on doit se poser la question suivante : quelle classification de données est la meilleure source d'information pour prédire les

concentrations horaires de NO_x et PM_{10} à l'endroit des stations de mesures ? Pour ces deux polluants, le trafic routier est la source principale d'émissions. Les conditions météorologiques peuvent également avoir un impact, en particulier pour les PM_{10} .

Deux priors différents ont ainsi été testés. Un basé sur la densité du trafic en classifiant les données par type de jours (jours de semaines/ samedi/ dimanche et jours fériés) et un autre basé sur les conditions météorologiques (jours anticycloniques/dépressionnaires/marais barométriques). En raison de la forte variabilité spatio-temporelle de la pollution en ville, pour chacun de ces priors nous utilisons les observations passées d'AIRPARIF à l'endroit de chaque station, pour chaque heure de la journée.

Le choix de la vraisemblance optimale joue également un rôle important dans le paramétrage du modèle bayésien. La vraisemblance étant la distribution de probabilité de la donnée (les simulations du modèle PMSS) conditionnée par le paramètre inconnu θ (les mesures d'AIRPARIF). Avec cette fonction, on met à jour ce que l'on sait à propos de la concentration de polluant (mesures AIRPARIF pendant les dix jours étudiés) basée sur une nouvelle preuve (les simulations du modèle PMSS). Le modèle probabiliste utilisé pour la vraisemblance est un modèle de régression linéaire simple : les concentrations simulées par le modèle PMSS vs les mesures d'AIRPARIF. Les régressions ont été calculées par paquet de données. Les données ont été regroupées par tranche horaire. La contrainte ici, est le faible jeu de données : des concentrations horaires pour dix jours de données. Or il ne faut pas moins de quinze points pour faire une régression linéaire.

5.2.2 À la recherche d'un modèle probabiliste optimal au niveau des stations de mesures

Une analyse de sensibilité a été réalisée pour déterminer le choix le plus approprié de la vraisemblance et du prior. Les concentrations de NO_x et PM_{10} simulées par le modèle PMSS ont été comparées par rapport aux mesures AIRPARIF avant et après la mise à jour bayésienne. Nous avons choisi comme indicateur le pourcentage de différence de RMSE avant et après mise à jour bayésienne sur les concentrations horaires pendant les dix jours de l'étude. Dans le cas où la mise à jour diminue l'erreur de la concentration estimée par rapport aux mesures AIRPARIF, ce pourcentage est négatif et la valeur est d'autant plus grande que l'erreur corrigée est forte.

Pour le cas le plus probant, la différence de l'écart relatif avant et après mise à jour bayésienne moyenné sur les dix jours a également été calculée.

Pour les NO_x , le prior « trafic » a donné des résultats bien plus probants que le prior « météo » dont la principale source d'émission est liée au trafic routier.

Les niveaux de PM_{10} sont très dépendants des deux facteurs testés : trafic routier et conditions météorologiques. Pour les PM_{10} , ces deux priors n'ont pas donné des résultats très différents (**Tableau 13**). Le modèle PMSS étant peu biaisé sur les valeurs de PM_{10} , les priors restent moins informatifs que la vraisemblance pour les PM_{10} . Il ressort tout de même une amélioration un peu plus forte pour le prior « météo » (mais pas sur l'ensemble des stations) que le prior « trafic ».

Stations		Prior	Vraisemblance : regroupement par paquets de données						
Type	Code		Jour	2h	3h	4h	6h	8h	12h
Trafic	HAUS	trafic	12%	-12%	-14%	-12%	-7%	-10%	-5%
		météo	7%	-14%	-17%	-15%	-10%	-12%	-8%
	OPERA	trafic	-30%	-38%	-36%	-36%	-37%	-34%	-32%
		météo	-33%	-42%	-40%	-41%	-40%	-38%	-36%
ELYS	trafic	12%	-10%	-8%	-8%	-5%	-7%	-4%	
	météo	+6%	-9%	-8%	-7%	-5%	-7%	-4%	
Fond urbain	PA04C	trafic	4%	-1%	+3%	+2%	+5%	+3%	+5%
		météo	+5%	-1%	+4%	+3%	+6%	+4%	+6%
	PA15L	trafic	19%	-8%	-3%	-4%	-4%	+1%	+3%
		météo	+14%	-9%	-4%	-4%	+2%	+1%	+3%

Tableau 13 : Différence de RMSE (en pourcentage %) sur les concentrations horaires des PM₁₀ pour les dix jours de l'étude, avec pour référence la mesure des stations AIRPARIF. Zone grisée et résultats en gras, pour la vraisemblance avec les données regroupées par tranches de 2 heures pour les priors « trafic » et « météo »

Nous avons fait le choix d'une méthode unifiée qui s'applique aussi bien pour les NO_x que pour les PM₁₀. Nous avons donc retenu un prior « trafic » en entrée de ce modèle pour les deux polluants.

La vraisemblance retenue regroupe le couple de données simulations/observations par tranche de 2 heures, tous les jours confondus pour chaque station. Avec ces choix de prior et vraisemblance, les concentrations estimées sont plus proches des mesures AIRPARIF que les concentrations initiales simulées par le modèle PMSS avec des valeurs moins dispersées pour les dix jours. En effet, pour les NO_x et PM₁₀ de l'ensemble des stations, l'erreur quadratique moyenne (RMSE) entre la moyenne de la concentration *a posteriori* et la mesure AIRPARIF est plus faible qu'avec la concentration initiale simulée par le modèle PMSS pendant cette période.

5.2.3 Propager la mesure locale dans la zone de représentativité

Le modèle bayésien paramétré de façon optimale améliore les estimations des concentrations de NO_x et PM₁₀ à l'endroit des stations AIRPARIF. L'autre enjeu de cette étude est de rendre les concentrations simulées par le modèle PMSS dans les zones de représentativité (définies dans le chapitre 4) plus réalistes en propageant la mesure de la station AIRPARIF dans toute la zone.

Nous utilisons le même prior et la même vraisemblance pour chaque point de la zone de représentativité. La différence entre chaque pixel de la mise à jour bayésienne est donc apportée par la concentration initiale simulée par PMSS à chaque point de grille.

Comment appliquer ce modèle bayésien à chaque point de la zone en intégrant la différence de représentativité (« ressemblance ») à la station de chaque pixel ?

Nous avons choisi d'appliquer une pondération entre les résultats issus de la mise à jour bayésienne (la moyenne de la concentration *a posteriori*) et la concentration initiale simulée par le modèle PMSS. Ainsi, un pixel représentatif à 100 % de la station (sélectionné comme étant représentatif dix jours sur dix) a pour concentration la plus probable la moyenne de la concentration *a posteriori*. Alors qu'un pixel représentatif à 10 % a pour concentration la plus probable la somme de 1/10^{ème} de la moyenne de la concentration *a posteriori* et 9/10^{ème} de la concentration initiale simulée par le modèle PMSS.

5.2.4 Principaux résultats

5.2.4.1 Estimation de la concentration des polluants au niveau de la station

Avec un prior « trafic » (par type de jours) et une vraisemblance définie pour chaque station en regroupant les concentrations horaires par tranche de deux heures pour les dix jours de données, on obtient une amélioration de l'estimation des concentrations au niveau des stations en réduisant la dispersion temporelle des valeurs par rapport aux mesures d'AIRPARIF.

L'erreur quadratique moyenne entre la moyenne de la concentration *a posteriori* et la mesure AIRPARIF est plus faible (RMSE et écart type moyen plus faibles calculés sur les dix jours) que l'erreur entre la concentration initiale de PMSS et la mesure AIRPARIF sur les concentrations horaires.

Pour les NO_x des stations trafic, la différence de RMSE est comprise entre -72 % (station OPERA) et -29 % (station ELYS). Pour les NO_x des stations de fond urbain, la différence de RMSE est comprise entre -31 % (station PA12) et -17 % (station PA07).

Pour les PM₁₀ des stations trafic, la différence de RMSE est comprise entre -38% (station OPERA) et -10 % (station ELYS). Pour les PM₁₀ des stations de fond urbain, la différence de RMSE est de -8 % (station PA15L) et pas de changement pour la station PA04C (-1 %)

Notons que pour les PM₁₀ de la station PA04C, le RMSE n'est pas diminué après mise à jour mais l'écart type moyenné sur les 10 jours est plus faible avec la moyenne de la concentration *a posteriori* qu'avec les concentrations initiales de PMSS (-8 %).

Les histogrammes de différence entre la moyenne des concentrations *a posteriori* et des mesures sont centrés autour de 0 (voir Fig.6 de l'article).

5.2.4.2 Estimation de la concentration des polluants pour chaque pixel de la zone de représentativité

Dans l'article, nous avons calculé la dispersion entre la concentration estimée avant et après mise à jour bayésienne et la concentration mesurée par AIRPARIF dans la zone de représentativité la plus probable (c'est à dire une zone sélectionnée au moins sept jours sur dix, zone où les pixels sont donc représentatifs à au moins 70 % de la station) pour des concentrations horaires pendant les dix jours de mars 2016. De plus, des cartes illustrent la différence entre les concentrations initiales simulées par le modèle PMSS et les concentrations résultantes de l'application bayésienne pondérée sur l'ensemble de la zone 1.2 km × 1.2 km (en pourcentage %).

Pour chaque station qui mesure les NO_x, en moyenne sur la zone de représentativité la plus probable, la mesure de la station AIRPARIF a été propagée sur l'ensemble de la zone : l'erreur entre la concentration estimée par la mise à jour bayésienne pondérée et la concentration mesurée à la station est plus faible qu'avec la concentration initiale simulée par le modèle PMSS (RMSE et écart type moyenné plus faibles).

Les modifications des concentrations pour les PM₁₀ dans les zones de représentativité les plus probables sont moins importantes que pour les NO_x. Les concentrations de NO_x étant moins bien représentées par le modèle PMSS que les PM₁₀.

Les cartes globales de différence de concentrations (1.2 km × 1.2 km) montrent également des modifications plus élevées pour les NO_x que pour les PM₁₀. Pour les NO_x, les concentrations

peuvent être modifiées jusqu'à 50 % par rapport à la concentration initiale simulée par le modèle PMSS alors que pour les PM_{10} le maximum de correction est autour de 15 %.

Les aires de représentativité des NO_x pour les stations trafic sont étroites et limitées à quelques mètres autour de la station. La correction apportée par notre méthode est donc limitée à une petite surface. Pour les stations de fond urbain, les aires de représentativité des NO_x sont plus larges. Des fortes corrections (de 35 % à 50 % par rapport à la concentration initiale) sont appliquées sur de grandes surfaces pour les stations PA12, PA04C et des corrections un peu plus faibles, mais toujours importantes (de 20 % à 30 %) sont appliquées sur des grandes surfaces pour les stations PA13, PA15L et PA07 (sur l'ensemble du jardin du Champs de Mars).

Pour les PM_{10} , les concentrations des stations trafic sont corrigées sur de plus grandes surfaces que pour les NO_x car les zones de représentativité des stations trafic sont plus étendues pour les PM_{10} que pour les NO_x . La concentration en PM_{10} est corrigée à 15 % sur une petite zone à proximité de la station pour les trois stations trafic et sur une large zone de la station PA15L.

5.2.5 Discussion

Nous proposons un modèle bayésien qui combine les simulations d'un modèle de dispersion haute résolution (3 m × 3 m) et les mesures de surface du réseau AIRPARIF à Paris.

Ce modèle spatio-temporel bayésien permet une meilleure estimation de la concentration de NO_x et PM_{10} à l'endroit des stations de mesures. Ce modèle modifie les concentrations de chaque pixel de la zone de représentativité et propage la mesure de la station à proximité de celle-ci, dans les zones préalablement définies. Pour estimer les concentrations dans la zone de représentativité de chaque station, nous utilisons la probabilité de représentativité de chaque pixel. Un pixel représentatif à 100 % de la station a pour concentration la plus probable la moyenne des concentrations *a posteriori*.

Le modèle a été paramétrisé au niveau de la station car nous ne disposons que des mesures de stations. Pour intégrer la différence de représentativité de chaque pixel, nous appliquons une pondération de la moyenne de la concentration *a posteriori* et de la concentration initiale simulée par le modèle PMSS. Le poids des résultats de la mise à jour bayésienne est d'autant plus important que le pixel est représentatif de la station. Ainsi, un pixel représentatif à 10 % de la station a pour concentration la plus probable la somme de $1/10^{\text{ème}}$ de la moyenne de la concentration *a posteriori* et $9/10^{\text{ème}}$ de la concentration initiale simulée par le modèle PMSS.

Les corrections appliquées aux concentrations initiales de PMSS sont d'autant plus importantes pour les NO_x que pour les PM_{10} car le modèle PMSS a plus de difficultés à reproduire les NO_x que les PM_{10} en raison de la forte réactivité de l'espèce NO avec les autres espèces chimiques et aux incertitudes de l'inventaire d'émissions liées au trafic routier (chaîne de calcul HEAVEN).

Bien que ces résultats contribuent à une meilleure estimation de l'exposition humaine dans les zones de représentativité, cette méthode a tout de même quelques limites. En effet, nous ne disposons que de dix jours de simulations ce qui nous a restreints dans le choix du modèle probabiliste (vraisemblance) et le classement du couple de données mesures/modèle par tranche d'heures. Et, l'épisode de pollution aux PM_{10} est étalé sur plusieurs jours ; ce qui peut également avoir un impact sur la paramétrisation du modèle bayésien.

De plus, nous ne disposons pas de mesures dans les zones de représentativité à proximité des stations AIRPARIF pour valider les concentrations les plus probables estimées par notre méthode de propagation spatiale du modèle bayésien.

Une piste intéressante qui pourrait être exploitée serait d'ajouter une incertitude à la concentration finale obtenue à partir de l'équation de pondération entre les concentrations initiales de PMSS et la moyenne des concentrations *a posteriori*.

$$C_{finale} = a \times C_{posterior} + (1 - a) \times C_{initiale} \text{ (eq.11)}$$

où C_{finale} est la nouvelle concentration estimée au niveau du pixel, a le coefficient de pondération, $C_{posterior}$ la moyenne des concentrations *a posteriori* au niveau du pixel et $C_{initiale}$ la concentration initiale simulée par le modèle PMSS au niveau du pixel.

On pourrait réécrire la formule de pondération (eq.11) et associer à $C_{posterior}$ et $C_{initiale}$ des incertitudes, respectivement $\delta C_{posterior}$ et $\delta C_{initiale}$.

$$C_{finale} = (w_1 \times C_{posterior} + w_2 \times C_{initiale}) / (w_1 + w_2)$$

où w_1 et w_2 sont les poids $w_1 = 1/\delta C_{posterior}^2$ et $w_2 = 1/\delta C_{initiale}^2$

La question serait de savoir si $C_{posterior}$ et $C_{initiale}$ sont compatibles compte tenu de leurs incertitudes. Pour cela, il faudrait ensuite calculer la différence : $diff = C_{posterior} - C_{initiale}$ et l'incertitude sur $diff$ qui est $\delta_{diff} = (\delta C_{posterior}^2 + \delta C_{initiale}^2)^{1/2}$

Si $|diff| > tol \times \delta_{diff}$, alors on pourrait dire qu'on a des valeurs aberrantes (de type "outlier") soit pour $C_{posterior}$ soit pour $C_{initiale}$ et il faudra choisir, après s'être fixé une tolérance $tol = 3.0$ par exemple ! Si les deux incertitudes sont compatibles, l'incertitude sur C_{finale} serait probablement la plus robuste.

Ce travail pourrait être enrichi avec un plus grand jeu de simulations PMSS. Ces données supplémentaires permettraient d'optimiser davantage le modèle probabiliste en construisant une nouvelle loi de vraisemblance ce qui probablement réduirait l'impact de l'épisode pollution sur les PM_{10} .

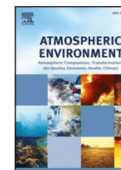
Des campagnes de mesures ponctuelles dans les zones de représentativité des stations permettraient d'évaluer la pertinence de la concentration finale obtenue avec ce modèle spatio-temporel bayésien.

5.3 Article publié



Contents lists available at ScienceDirect

Atmospheric Environment: X

journal homepage: www.journals.elsevier.com/atmospheric-environment-x

NO_x and PM₁₀ Bayesian concentration estimates using high-resolution numerical simulations and ground measurements over Paris, France

Delphy Rodriguez^{a,*}, Eric Parent^b, Laurence Eymard^a, Myrto Valari^c, Sébastien Payan^a^a LATMOS/IPSL, Sorbonne Université, UVSQ, CNRS, Paris, France^b AgroParisTech, 16 Rue Claude Bernard, Paris, France^c LMD, Ecole Polytechnique Palaiseau, France

HIGHLIGHTS

- A Bayesian scheme combining simulated pollutant concentrations with measured ones.
- Bias correction for high-resolution air quality modeling.
- Enhanced concentration estimates of air quality monitoring network.
- Spatially resolved concentration estimates around urban monitoring sites.

ARTICLE INFO

Keywords:
 Urban air pollution
 Surface monitoring networks
 High-resolution modeling
 Spatial representativeness
 Bayesian modeling

ABSTRACT

Air quality over cities is mainly monitored by in-situ surface measurements. However, these stations are too sparse to properly capture the inhomogeneity of pollutant concentrations over urban areas. The need for high-resolution concentration estimates has grown in recent years, together with the awareness of the harmful effects of air pollution. In this study, we develop a Bayesian scheme that combines the high-resolution (3 × 3 m) Particulate Micro SWIFT SPRAY numerical air quality simulations (PMSS) with operational surface measurements. The goal is to improve NO_x and PM₁₀ PMSS concentrations estimates over monitoring stations and with their vicinity. For this purpose, we simulate pollutant concentrations over the city of Paris for ten days over the period of March 2016. The Bayesian model provides an enhanced estimate of pollutant concentration in space and time. At the monitoring stations location, these estimates are characterized by lower temporal dispersion compared to the simulated data. Within the vicinity of the monitor stations, enhanced concentration estimates are closer to observations. For NO_x, the improvement is stronger and occurs in a larger area for urban background stations than for traffic stations. Overall, NO_x improvement is higher than PM₁₀ improvement. The initial PMSS model prediction is more biased for NO_x than for PM₁₀ due to large uncertainties in NO_x emissions over the traffic network.

1. Introduction

In large cities, air quality is often monitored by ground station networks measuring pollutant concentrations. Several types of areas are sampled: traffic, residential, etc. The spatial representativeness of these measurements is limited to a small area around the station, especially in dense urban areas: emissions from various sources, as well as air flow channeling in street canyons, result in sharp horizontal gradients of pollutant concentrations (Harrison, 2018). The need for high-resolution gridded estimates of pollutant concentrations over big cities has grown in recent years, together with the increasing awareness of the harmful

health effects of atmospheric pollution.

Various approaches have been developed to address the spatial distribution of pollutant concentration around monitoring sites. These approaches use real-time measurements from local networks and expand them in space and/or time based on additional relevant information such as land-use, weather conditions, and topography. They may be divided mainly into first-principle geophysical modeling and statistical ones. Atmospheric pollutant dispersion models describe the physical and chemical processes that govern the emission, transport, mixing, chemical reaction and deposition of pollutants. High-resolution models, such as the Particulate Micro Swift Spray (PMSS) air quality

* Corresponding author.

E-mail addresses: delphy.rodriguez@latmos.ipsl.fr, delphy.rodriguez@gmail.com (D. Rodriguez).

<https://doi.org/10.1016/j.aeoa.2019.100038>

Received 24 January 2019; Received in revised form 27 May 2019; Accepted 29 May 2019

Available online 17 June 2019

2590-1621/© 2019 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

model integrated into the ARIA City modeling platform (http://www.aria.fr/aria_city.php), provide realistic spatial distributions of pollutant concentrations. However, they suffer from large biases due to uncertainties on the emissions and on the parameterizations of the complex atmospheric processes governing urban air quality. Statistical approaches include land-use regression (Janssen et al., 2012; Son et al., 2018), kriging interpolation schemes (Beauchamp et al., 2017; Wu et al., 2018; Martin et al., 2012), methods for assimilation of measurements in models such as Kalman filtering (Hanea et al., 2004) or 4D variational assimilation algorithms (Elbern et al., 2001) and Bayesian models (Beloconi et al., 2018).

The scope of this study is thus to improve NO_x and PM_{10} pollutant concentrations simulated with the PMSS high-resolution air quality model over and in the vicinity of the monitoring stations of the AIRPARIF local air quality agency in Paris, France (<https://www.airparif.asso.fr>). To spatially expand the measurement around monitoring stations, we use the concept of “representativeness areas” defined in our previous work (Rodriguez et al., 2019). These areas consist of points where concentration levels are similar and time variations correlated to those at the monitoring site’s location. In practice, they are defined by applying criteria thresholds on the Normalized Root Mean Square Error (NRMSE) and on the correlation coefficient between simulated concentrations over all the pixels around the monitoring sites and simulated concentration at the monitoring site’s location.

Frequentist or Bayesian statistical approaches may be used. In both approaches, the purpose is to determine as accurately as possible an unknown parameter θ . Here, the unknown parameter θ is the concentration in the so-called “representativeness area” of an AIRPARIF monitoring site, i.e. at the vicinity of a station. In the frequentist setting, the unknown parameter θ is pre-assigned to a fixed single value. A statistical procedure is chosen to establish a confidence interval from any set of data collected in the same conditions, under the same model. It ensures the probability of this random interval to cover the fixed value θ , i.e. the reliability of the chosen statistical procedure. Then, the confidence interval is calculated from the available data as a random draw from this estimation procedure. All the information is contained in data. The frequentist procedure assumes a reproducible experience and a large data set. Frequentists claim that this is an *objective* way of reasoning.

Within the Bayesian paradigm, the unknown parameter θ is considered uncertain. What is known about θ is represented by a random variable conditioned upon the available information. The major issue is to quantify the uncertainty from all available data (Boreux et al., 2010). One of the most significant differences between frequentist and Baye-

and deposition (Cowles et al., 2003). In this study, we propose Bayesian model that updates the distribution of current and past ground station measurements with the high-resolution output of the PMSS model ($3 \times 3\text{m}^2$) within the city of Paris for ten days over the period of March 2016.

In Section 2, the AIRPARIF air quality network and PMSS modulations are presented. Then, representativeness areas, as obtained in Rodriguez et al. (2019), are summarized, and the Bayesian framework is described. Section 3 gathers results: first, the pollutant concentrations at the monitoring site locations (Sect. 3.1) and then, concentrations within representativeness areas (Sect. 3.2). Conclusions are given in section 4.

2. Data and methods

As in Rodriguez et al. (2019), a simulation set of ten non-continuous days in March 2016 are used. During this period (March 2–21), a high-pressure episode occurred, characterized by strong atmospheric stability and the accumulation of PM_{10} within the boundary layer. It resulted in a sharp rise in PM_{10} concentration on March 11. The effect of this synoptic-scale episode is less pronounced for NO_x than for PM_{10} . Indeed, in highly urbanized areas, the spatial and temporal variability of pollutant concentration is more driven by traffic patterns for NO_x than for PM_{10} . We note that the contribution of the traffic sector to the total emission concentrations is two times higher for NO_x than for PM_{10} in the Paris area (56% vs. 28% AIRPARIF, 2012). In the close proximity to roads, the NO_x emissions decrease much higher for NO_x (Pasquier et al., 2017). Data from eight local monitoring stations from the AIRPARIF network have been collected for the same period.

2.1. Surface measurements

The location of AIRPARIF monitoring sites have been chosen to sample various urban landscapes in Paris. They are classified as “traffic” or “urban background” depending on their proximity (or not) to the street network. A thorough description of the AIRPARIF network can be found on the agency’s website (<https://www.airparif.asso.fr/er-stations/index>). Data from three traffic-oriented and five urban background monitoring stations are used in this study as listed in Table 1. Traffic monitoring stations are located on sidewalks along roads and on major crossroads, whereas urban background monitors are located in low emission areas such as parks, pedestrian squares, schoolyards, and leisure parks. All traffic monitors measure both NO_x and PM_{10} , whereas PM_{10} is measured only at two out of five urban background sites.

Table 1

Characteristic features of the study areas. Locations marked with an asterisk indicate the location of the monitoring site.

Type	ID station	Characteristic features	Pollutants
Traffic-oriented stations	HAUS	Bld Haussmann*, side roads	NO_x , PM_{10}
	OPERA	Crossroad*, road traffic	NO_x , PM_{10}
	ELYS	Champs-Élysées Avenue*: large road bordered with trees, side roads	NO_x , PM_{10}
Urban background stations	PA15L	Stadium* and leisure park, “Boulevard Périphérique”, traffic road	NO_x , PM_{10}
	PA12	Schoolyard*, railways, traffic road	NO_x
	PA13	Park*, traffic road, crossroad	NO_x
	PA04C	Place*, park, the Seine River, traffic road	NO_x , PM_{10}
	PA07	Square close to the Eiffel Tower*, Champ-de-Mars Garden, Seine banks, the Seine River, Bridge	NO_x

sian approaches is the choice of a “prior” (O’Hagan, 2008). Prior characterizes our expert knowledge by giving suggestions on an unknown parameter θ (as a distribution of θ values) before observing data and is *subjective* since it reflects a personal probabilistic judgment.

In recent years, Bayesian statistics have become a standard methodology in environmental science, it has been used to improve spatial predictions of pollutant concentration by Pirani et al., 2014), Amin et al. (2015), McMillan et al., 2010, Sahu et al. (2005), among others,

These measurements are considered as reference, or “ground truth” of pollutant concentration, assuming that the measurement error is negligible.

2.2. Numerical air quality simulations

NO_x and PM_{10} concentrations were simulated using the high-resolution model PMSS, in a $12 \times 10\text{km}^2$ grid covering the city of Paris.

The model set-up is the same as in Rodriguez et al. (2019).

The model has thirty-five vertical layers from the ground up to an altitude of 800 m. The thickness of the surface layer is 2 m and increases with the altitude (Moussafir et al., 2014). The model simulates the turbulent flow around the buildings and pollutant dispersion through a 3-m horizontal resolution grid. Emission fluxes from urban sources are taken from the AIRPARIF inventory. Boundary conditions are taken from the CHIMERE chemistry-transport model (Mailler et al., 2017). Meteorological conditions are simulated with the MM5 weather mesoscale model (<http://www2.mmm.ucar.edu/mm5/>). Building contours and heights are also accounted for as obstacles for the air flow. They are taken from the BDTOPO database developed by the French National Geographic Institute. Atmospheric dispersion is simulated with the PSPRAY Lagrangian particles model. Traffic emissions are calculated with the HEAVEN chain (Healthier Environment through the Abatement of Vehicle Emissions and Noise, <http://www.airparif.asso.fr/etat-air/air-et-climat-emissions-heaven>) HEAVEN calculates in near real-time the traffic circulation and allows obtaining pollutant emissions on main traffic roads of the Paris area based on a “bottom-up” approach. Hourly traffic emissions are derived from the combination of traffic model (car flow, mean speed, cold engines percentage) and emission factors (car speed, road type, temperature, car type, etc.) applied by loop-based counting systems.

At each monitor, simulated concentrations are compared to measurements for both pollutants. PMSS simulations are in a very good agreement with PM₁₀ measurements, both for traffic and urban background sites, with a correlation coefficient (R²) equal to 0.83 and 0.92 respectively, and a root mean squared error (RMSE) equal to 9.57 µg/m³ and 6.21 µg/m³ respectively (Rodriguez et al., 2019). Model performance is not as good for NO_x with an R² value of 0.65 and 0.45, and RMSE of 81.57 µg/m³ and 21.12 µg/m³ for traffic and urban background sites, respectively.

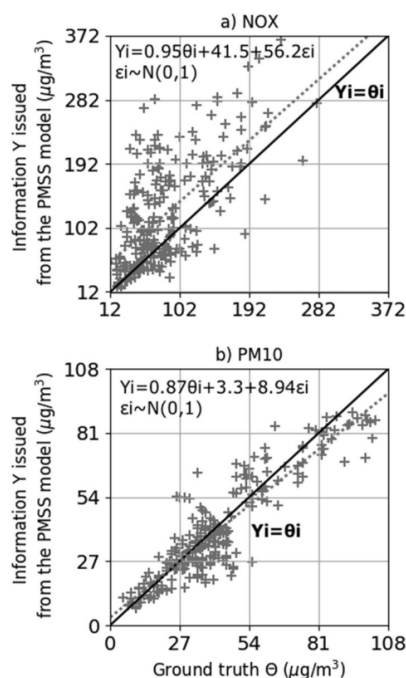


Fig. 1. Scatter plots of measured vs simulated hourly concentrations for NO_x (a) and PM₁₀ (b) of Champs-Élysées avenue monitoring station during the ten days of March 2016. The black line represents $Y = \theta$. The grey dotted line represents the linear regression between AIRPARIF measurements and PMSS simulations.

Fig. 1 shows scatter plots of concentrations at the ELYS monitoring station. The PMSS model fails to represent NO_x concentration maximum. As discussed in Rodriguez et al. (2019), this may be due to the simple chemical mechanism of the PMSS model that fails to represent the concentration of the highly reactive pollutants such as NO and/or the uncertainties in the NO_x emission derived by the HEAVEN traffic emission modeling chain. Note that the error for NO₂ is smaller than for NO for traffic stations (RMSE (NO₂) = 21.9 µg/m³ vs RMSE (NO) = 72.3 µg/m³). A similar result is obtained for urban background stations.

The model is not always able to represent the high concentrations levels during pollution episodes at the studied monitoring sites in Paris due to omissions in emission sources, errors in the meteorological predictions, or inaccuracy of the parameterization. However, it still provides realistic dispersion patterns of pollutant according to other studies during specific measurements campaigns in wind tunnel and real field experiment (Trini Castelli et al., 2018) and for academic project (Hanna et al., 2011). To perform this study, we assume that the model captures the actual spatial variability of pollutant in Paris (Rodriguez et al., 2019).

2.3. Spatial representativeness areas

PMSS simulations were used in Rodriguez et al. (2019) to determine representativeness areas around the AIRPARIF monitoring sites. The daily spatial representativeness area around a monitoring site was defined as the ensemble of adjacent pixels where the pollutant concentration is (1) strongly correlated with the concentration at the pixel of the station; (2) has a low statistical error (Normalized Root Mean Squared Error) with respect to the concentration at the pixel of the monitoring station. Correlation coefficient and NRMSE thresholds selection are based on an analysis over the 10-day with a step of 0.05 for both criteria. For each threshold, a score is estimated by counting the number of days on which the size of the selected area including the monitoring site is comprised between 300 m² and 400 × 10³ m².

The maximum score for each site is 10. Finally, a total score calculated from all stations for the 10 days and reported as the percentage of success. In the case where two different combinations of criteria thresholds give the same score (percentage of success), the retained combination is the more restrictive (i.e. the higher R correlation coefficient and/or the lowest NRMSE value).

Table 2 shows the applied criteria thresholds to select representativeness areas.

Table 2
Thresholds of representativeness criteria: correlation coefficient (ρ) and Normalized Root Mean Square Error (NRMSE) for NO_x and PM₁₀.

Pollutants	ρ	NRMSE
NO _x	0.7	0.45
PM ₁₀	0.75	0.3

The retained thresholds for PM₁₀ are more restrictive than for NO_x. This is due to the higher spatial variability of NO_x concentration which requires less restrictive thresholds in order to retain areas large enough to get beyond the pseudo-station.

This method is applied independently each day of the simulation to select the representative pixels. A representativeness probability is assigned to each pixel, depending on the frequency (i.e. number of days out of ten) at which each pixel is selected. An example of representativeness areas around the monitoring site of “Avenue des Champs-Élysées” is given in Fig. 2. This monitoring station is located on the sidewalk of a wide urban boulevard. NO_x most representative area includes the sidewalk where the instrument is located along a distance of 500 m (yellow to red color, indicating that this area is selected at least

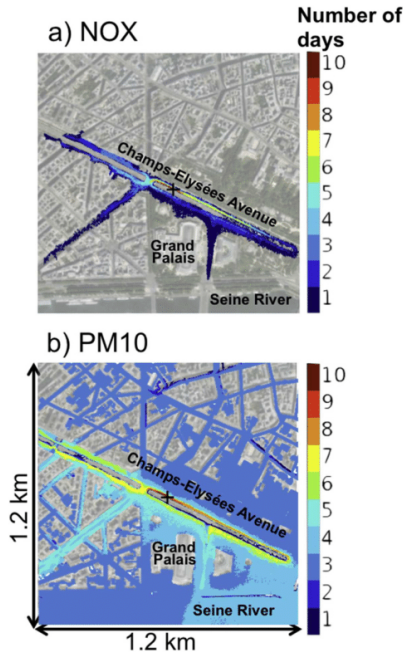


Fig. 2. NO_x (a) and PM₁₀ (b) representativeness areas cumulated over the 10 days of the study for ELYS monitoring site. The black cross indicates the location of the monitoring site.

days), whereas the dark blue area shows street portions which have been selected only one day. The much larger PM₁₀ area among the most probable representative ones (yellow to red, 7 days at least) includes both sidewalks along a distance of 700 m (boulevard portion including the station) as well as a 400-m segment of sidewalk in northwestern direction from the station.

2.4. The Bayesian framework

2.4.1. Model formulation

The goal of this study is to improve the PMSS pollutant simulations at, and around, each AIRPARIF station by using past and current station measurements. The scheme of the Bayesian model applied at the monitoring station location is represented with the Directed Acyclic Graph (DAG) of Fig. 3.

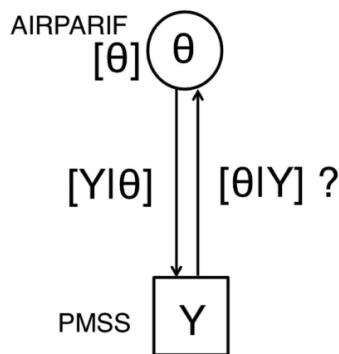


Fig. 3. Directed Acyclic Graph (DAG) describing our scientific question with θ : prior distribution, $[Y|\theta]$: likelihood, $[\theta|Y]$: posterior distribution.

Past AIRPARIF measurements represent the information on which the prior expert knowledge of the unknown parameter θ (i.e. pollutant concentration at a given point in space and time) is based (before getting the predicted data from the PMSS simulated concentration).

After integrating data information from the PMSS simulation, the prior information (past measurements from the station) is updated through the likelihood function to obtain the posterior knowledge. The likelihood function is the probability distribution of the data (PMSS simulations) conditional on the parameter θ (AIRPARIF measurements to the monitoring station), commonly used in other statistical approaches. Simple linear regression is used as a probabilistic model (PMSS simulated concentrations vs AIRPARIF measurements, Eq. (1))

$$Y_i = a\theta_i + b + \sigma \epsilon_i \tag{Eq. 1}$$

With error term $\epsilon_i \sim N(0,1)$. Where Y_i is the initial PMSS simulation ($\mu\text{g}/\text{m}^3$), θ_i is the AIRPARIF monitoring station measurements ($\mu\text{g}/\text{m}^3$) during the ten days in March 2016 and a , b and σ respectively the slope, intercept, and the standard deviation of the residuals of the linear regression. Parameter values (i.e. a high slope, a low intercept, an standard deviation) for which the likelihood is high, are those that have a high probability of producing the observed data.

Bayes'rule (Eq. (2)) updates the prior knowledge by learning from data (Y :PMSS simulated concentration) and provides a posterior distribution from which the *a posteriori* most probable concentration is estimated, associated with its uncertainty (the posterior standard deviation) (O'Hagan, 2008).

All three, prior, likelihood and posterior are distributions describing the probability that an event occurs. According to the Bayes'rule, the posterior distribution $[\theta|Y]$ is the product of the prior distribution $[\theta]$ and the likelihood function $[Y|\theta]$, except for a constant scaling factor (Eq. (2))

$$[\theta|Y] = \frac{[Y|\theta][\theta]}{[Y]} \tag{Eq. 2}$$

Where $[\theta|Y]$ is the posterior distribution, $[Y|\theta]$ is the likelihood, $[\theta]$ prior distribution, $[Y]$ a constant of normalization that we can omit to formulate Bayes'rule by saying that the posterior is proportional to the prior times the likelihood.

The distribution of NO_x and PM₁₀ concentrations are close to log-normal (Fig. 4a). Neperian logarithm concentrations (Fig. 4b) are used instead of concentrations in order to use normal distributions, which is convenient to get an explicit posterior distribution, following the *a posteriori* formula of the conjugate normal model (O'Hagan, 2008) (Eq. (3))

The posterior distribution is a normal distribution with a mean ($m_{y-\log}$) and a standard deviation ($s_{y-\log}$) obtained from the Bayesian updating formulae Eq. (3). Model precision is calculated as the inverse of the variance of the posterior distribution ($\frac{1}{s_{y-\log}^2}$). It reflects the degree of uncertainty related to the obtained mean of the estimated parameter

$$\frac{m_{y-\log}}{s_{y-\log}^2} = \left(\frac{m_{o-\log}}{s_{o-\log}^2} + \frac{(Y-b)}{a} \frac{a^2}{\sigma^2} \right) \tag{Eq. 3}$$

$$\frac{1}{s_{y-\log}^2} = \frac{1}{s_{o-\log}^2} + \frac{a^2}{\sigma^2}$$

Where $m_{o-\log}$ and $s_{o-\log}^2$ are respectively the mean and variance of the prior log concentration at the monitoring site's location, Y is the Neperian logarithm of the initial PMSS simulated concentration at given pixel, a , b and σ^2 are respectively the slope, the intercept, and the variance of the residuals of the linear regression model used as likelihood function (see Section 2.4.2).

At the end of the processing, the mean and standard deviation of the posterior distribution of NO_x and PM₁₀ concentration are obtained by applying a log-normal transformation to the model output.

Fig. 5 shows an example of the application of Bayes'rule for NC log-concentration with concentration expressed in $\mu\text{g}/\text{m}^3$ at the OPER

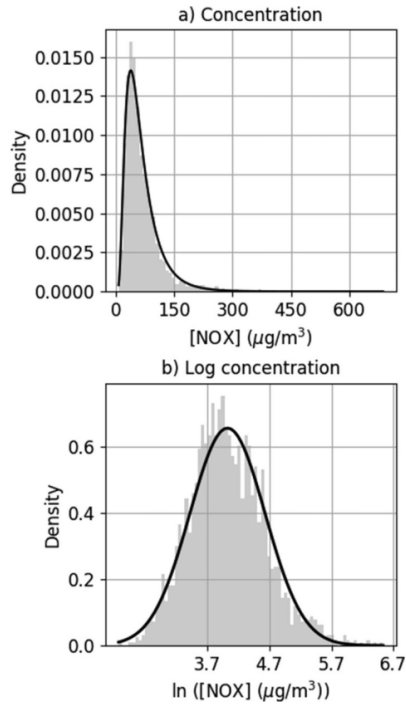


Fig. 4. NO_x prior mean concentration of Champs-Élysées avenue monitoring site over ten years (2006–2015) at 1 a.m. during Monday and Friday: (a) concentration and (b) Neperian logarithm concentration.

monitoring station at 3 p.m. on March 2nd. It is shown that the initial belief (prior) lies within a value ranging between 4.8 and 6 (Neperian logarithm concentration) with a maximum weight of 5.4. The likelihood values range between 4.4 and 5.3 with a maximal weight of 4.9. The posterior distribution is the combination between prior and likelihood. It is a peaky distribution centered around 5, which appears as a compromise between likelihood and prior information. In this case, data (likelihood) are more informative than the prior because the variance of the prior distribution is high. The posterior mean properly fits the AIRPARIF measurements whereas the PMSS mean initial simulated concentration was biased, at 5.2.

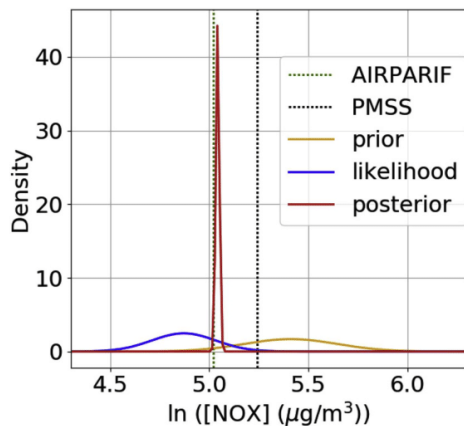


Fig. 5. Example of Bayesian update of the NO_x log-concentration at the Opera crossroad monitoring station at 3 p.m. on March 2, 2016.

2.4.2. Finding the optimal probabilistic model at the monitor location

A major challenge when using Bayesian statistics is to properly choose the prior distribution $[\theta]$ and the likelihood function $[Y|\theta]$.

Prior distribution $[\theta]$. The prior is a distribution of possible values of the unknown parameter θ . It describes the expert knowledge before the observation. In the present case, to build a prior, we must answer the question: which data classification is the best source of information, to predict hourly NO_x and PM₁₀ concentration during the ten days of March 2016 at any monitoring station location?

NO_x and PM₁₀ concentrations in the center of Paris are strongly correlated to traffic emissions. They also depend on meteorological conditions responsible for their transport and mixing. Consequently, two priors are tested, both based on data collected at the AIRPARIF station in past years: (1) the first one based on traffic variations, by classifying concentration data along three classes: working days (Saturdays, Sundays and public holidays); (2) the second one based on weather conditions, by distinguishing low and high-pressure days, as well as fast-flow situations. Because of the high spatial and temporal variability within the studied domain, different priors for each station and time of the day must be considered.

AIRPARIF monitoring stations have been operating starting from different years: 2006 for PA13, PA12, PA07 and ELYS stations, 2011 for OPERA and PA04C, 2010 for HAUS and 2014 for PA15L.

Fig. 6 shows daily mean profiles of NO_x concentration based on ten years of data (2006–2015) at the ELYS monitoring station. The diurnal cycle of mean concentrations is marked with a morning and afternoon peak corresponding to rush hours during working days. During Saturdays, Sundays and public holidays, the night peak is more marked than the morning peak which is also shifted later on. The prior weather classification (not shown) gives similar results for PM₁₀, but does not work properly for NO_x. The prior classification which best predicts NO_x and PM₁₀ hourly concentrations for the ten days in March 2016 is, therefore, a classification per day-of-week.

Likelihood function $[Y|\theta]$. The second challenge is the choice of the appropriate likelihood function. This is the probability distribution of the data (PMSS simulation) conditional to the unknown parameter (AIRPARIF measurements from the monitoring station), commonly used in most statistical approaches. Through this function, the knowledge about the pollutant concentrations (AIRPARIF measurements from the monitoring station during the studied period) is updated, based on new information (PMSS simulation).

A simple linear regression of PMSS simulated concentrations versus AIRPARIF measurements is used here as a probabilistic model. A strong constraint is imposed by the small size of our dataset (hourly concentrations during ten days) to define the appropriate classification

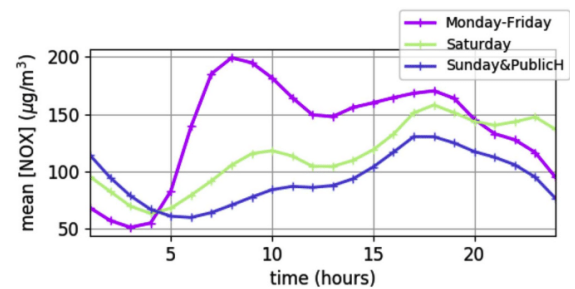


Fig. 6. Daily profiles of mean concentration for NO_x prior at the Champ-Élysées Avenue monitoring site. The magenta line represents working days data (Monday to Friday), the green line represents Saturday data and blue line Sunday and Public holidays data. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article)

Table 3

% RMSE and standard deviation (in brackets) difference between simulated and observed data before and after the Bayesian updating for several time frames classification of the hourly data (per day, by 2,3,4,6,8 and 12-h period for the ten days). [HAUS* - PM₁₀ results are averaged on 8/10 days, due to the lack of AIRPARIF measurements for two days].

a) NO _x									
Station			RMSE difference (%) [Standard deviation difference (%)]						
Type	ID	Initial RMSE [Initial standard deviation (%)]	Day	2h	3h	4h	6h	8h	12h
Traffic	HAUS	60.0 [42%]	-49%	-45% [-31%]	-43%	-43%	-43%	-41%	-40%
	OPERA	108.6 [50%]	-68%	-72% [-67%]	-71%	-70%	-69%	-69%	-67%
	ELYS	67.8 [75%]	-27%	-29% [-21%]	-27%	-27%	-27%	-25%	-22%
Urban background	PA12	20.1 [24%]	-19%	-31% [-35%]	-27%	-27%	-26%	-21%	-20%
	PA13	16.6 [33%]	-13%	-26% [-22%]	-21%	-22%	-18%	-15%	-14%
	PA04C	20.8 [35%]	-22%	-30% [-31%]	-25%	-25%	-28%	-21%	-21%
	PA07	25.5 [36%]	-27%	-17% [-12%]	-14%	-15%	-11%	-13%	-12%
	PA15L	23.3 [45%]	-17%	-26% [-24%]	-22%	-25%	-12%	-19%	-13%

b) PM ₁₀									
STATION			RMSE difference (%) [Standard deviation difference (%)]						
TYPE	ID	Initial RMSE [Initial standard deviation (%)]	Day	2h	3h	4h	6h	8h	12h
Traffic	HAUS*	8.6 [23%]	12%	-12% [-13%]	-14%	-12%	-7%	-10%	-5%
	OPERA	10.5 [30%]	-30%	-38% [-49%]	-36%	-36%	-37%	-34%	-32%
	ELYS	9.6 [19%]	12%	-10% [-8%]	-8%	-8%	-5%	-7%	-4%
Urban background	PA04C	5.1 [13%]	4%	-1% [-8%]	3%	2%	5%	3%	5%
	PA15L	7.1 [21%]	19%	-8% [-19%]	-3%	-4%	-4%	1%	3%

method. We couple simulated and observed concentration data over time-slots so that at least fifteen points are used for the linear regression (Eq. (1)). For each station, different time frames are tested to group the hourly data. Time windows of 2, 3, 4, 6, 8 and 12 consecutive hours, and whole day are tested as shown in Table 3.

A sensitivity analysis is performed for each type of likelihood and both proposed priors (traffic and weather conditions). NO_x and PM₁₀ PMSS simulated concentration at the monitoring station's location are compared with AIRPARIF monitoring station measurements before and after the Bayesian updating. To evaluate the best configuration, we calculate a percentage of Root Mean Square Error (RMSE) difference in each case (Eq. (4)). For the most relevant choice, to complete this first evaluation approach, we also calculate a percentage of standard deviation difference.

$$\% \text{ RMSE difference} = \frac{(RMSE_{\text{after bayesian updating}} - RMSE_{\text{initial}})}{RMSE_{\text{initial}}} * 100 \quad (\text{Eq.4})$$

where $RMSE(\text{ref}, \text{val}) = \sqrt{\frac{1}{n} \left(\sum_{t=0}^{240} (\text{ref}_{(t)} - \text{val}_{(t)})^2 \right)}$ is the root mean square error, $\text{ref}_{(t)}$ is the PM₁₀ or NO_x hourly concentration at the AIRPARIF monitoring site.

For $RMSE_{\text{initial}}$, $\text{val}_{(t)}$ is the PM₁₀ or NO_x hourly PMSS simulated concentration at the monitoring station location for the ten studied days. For $RMSE_{\text{after bayesian update}}$, $\text{val}_{(t)}$ is the PM₁₀ or NO_x hourly posterior mean concentration resulting from the Bayesian updating.

Table 3 summarizes results of the sensitivity analysis, associating the previously chosen prior and the time frames tested to construct the likelihood function. The best data classification corresponds to the largest decrease of the %RMSE (the highest negative value), meaning that the Bayesian updating provides less biased concentration estimates compared to the initial PMSS simulation. Note that we only focus on the evaluation of the most probable concentration (posterior distribution mean). The overall best model performance is obtained when gathering the hourly data in 2-h time frames. Data from all ten days corresponding to each 2-h time frame are gathered to obtain twenty points for calculating the linear regression.

2.4.3. Expanding the measured value within the representative area

We applied the previous Bayesian model, for each monitoring station and each hour of the day, to each pixel of the representative areas (see Section 2.3), by accounting for the probability of each pixel to belong to this area.

The Bayesian model update is performed on every pixel within the representativeness area, using the initial PMSS simulated concentration at the given pixel and with the selected prior and likelihood function at the station (Section 2.4.2). The fraction of the number of days, f_c which the pixel is selected to belong to the representativeness areas c the total number of days of the study, is used as weighting coefficient (Eq. (5)). Thus, for pixels within a representativeness area selected f_c days out of ten, we apply a weight equal to one. In this case, the concentration at the pixel is the posterior mean concentration at the given pixel (i.e. concentration after the Bayesian updating). For pixels selected only one day out of ten the final concentration will be calculated as the sum of 1/10th of the posterior mean concentration at the pixel and 9/10th of the initial PMSS simulated concentration at the pixel.

$$C_{\text{final}} = a * C_{\text{after bayesian updating}} + (1 - a) * C_{\text{initial}} \quad (\text{Eq.5})$$

where C_{final} is the new concentration estimate at the pixel, a the weighting coefficient, $C_{\text{after bayesian updating}}$ the enhanced concentration estimate at the pixel and C_{initial} the PMSS simulated concentration at the pixel.

We note here that the weighting coefficient used to expand the measured value in the representativeness areas depends on the representativeness criteria statistics (daily correlation coefficient and NRMSE) calculated from the ten-day period of the study.

Given the shortness of the time period and the random nature of meteorology, we expect to underestimate and/or misplace the daily variability of the concentration field. However, by taking the most probable representativeness area (representativeness area shared several days on ten) instead of daily representativeness area, the uncertainty of daily representative area and thresholds effects are minimized.

The methodology should be applied on longer study periods covering a larger variety of atmospheric conditions, to reduce uncertainties. Representativeness areas could then be defined based on

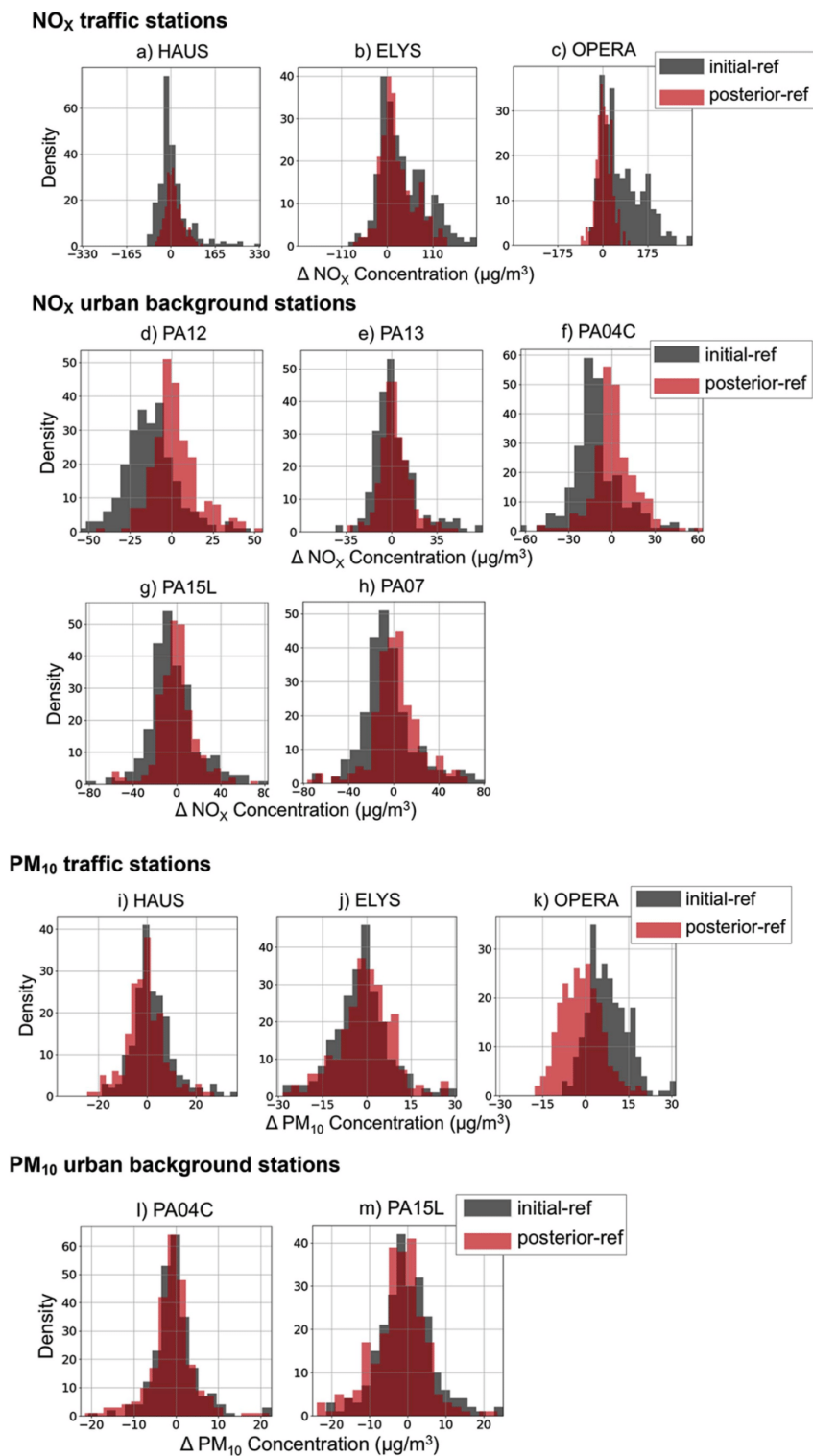


Fig. 7. Distribution of the bias of initial PMSS simulated concentrations (grey) and of the posterior mean concentration (red), with respect to the monitoring static measurements. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

dispersion pattern regimes. Alternative approaches include geostatistical kriging such as in [Beauchamp et al. \(2018c\)](#).

3. Results

3.1. Pollutant concentration estimates at the monitoring site location

As shown in [Table 3](#), concentration estimates by Bayesian updating present in almost all cases a lower RMSE than the initial PMSS simulated hourly concentrations. For both NO_x and PM_{10} pollutants, this improvement is higher for traffic stations (up to 72% decrease in RMSE at the OPERA monitor for NO_x) than for urban stations (31% at PA12 for NO_x). As shown in the histograms of [Fig. 7](#), the distribution of the PMSS simulation bias with respect to the monitoring station is not normal (grey histograms). This suggests the presence of systematic errors in the PMSS simulation. Especially for traffic stations, and for both pollutants, we find extreme values, suggesting high model over-estimation of the measured value. After the Bayesian updating, the bias distribution is closer to a normal distribution with errors centered around zero. This shows that the applied methodology is an efficient bias-correction method for the PMSS model.

3.2. Pollutant concentration estimates in the representativeness areas

The same Bayesian model (same prior and likelihood) is applied to the pixels of the representativeness areas in the vicinity of the monitoring station to update NO_x and PM_{10} PMSS simulated concentrations. To account for the probability of each pixel to belong in the representativeness area, the weighting function is applied to the posterior distributions to provide the final concentration estimates (see [Section 2.4.3](#)).

Maps of the differences between the updated concentration estimates and the initial PMSS simulation within the representativeness areas are shown in [Fig. 8](#) for NO_x and PM_{10} pollutants respectively. For NO_x , around traffic stations, a small number of pixels are modified, due to the small size of the representativeness area.

Changes can be seen in larger areas around the urban stations. The change in PM_{10} concentrations is smaller than in NO_x concentrations with a maximal correction of about 15% for PM_{10} vs. 50% for NO_x . As mentioned in [Section 2.2](#), this is most probably due to the better PMSS model performance for PM_{10} than for NO_x .

Moreover, for NO_x , around traffic stations, we only find small changes in the initial PMSS concentration (by 20–30% in average) whereas around urban background stations, up to 50% changes are observed on large areas for PA12 and PA04C. For PM_{10} , the highest modification is obtained for PA15L with 15% change and around the three traffic stations within a small-sized area surrounding the station.

[Table 4](#) shows results obtained within the most probable representativeness area around each monitoring station, i.e. the area selected at least seven days out of ten (see [Rodriguez et al., 2019](#)). The results are given, as in [Section 2.4.2](#), in terms of percentage decrease of the RMSE and standard deviation values between the updated concentration estimate and the initial PMSS simulated concentration with respect to the AIRPARIF measurement (Eq. (4)). RMSE and standard deviation shown in [Table 4](#) are spatially averaged across the pixels of the most probable representativeness areas.

NO_x concentrations after Bayesian updating are improved in all cases, with a maximal correction for the OPERA station (−46% for RMSE difference and -39% for the standard deviation difference).

PM_{10} concentrations are also closer to the measured values after Bayesian updating within the most probable representativeness area by considering RMSE and standard deviation ([Table 4b](#)).

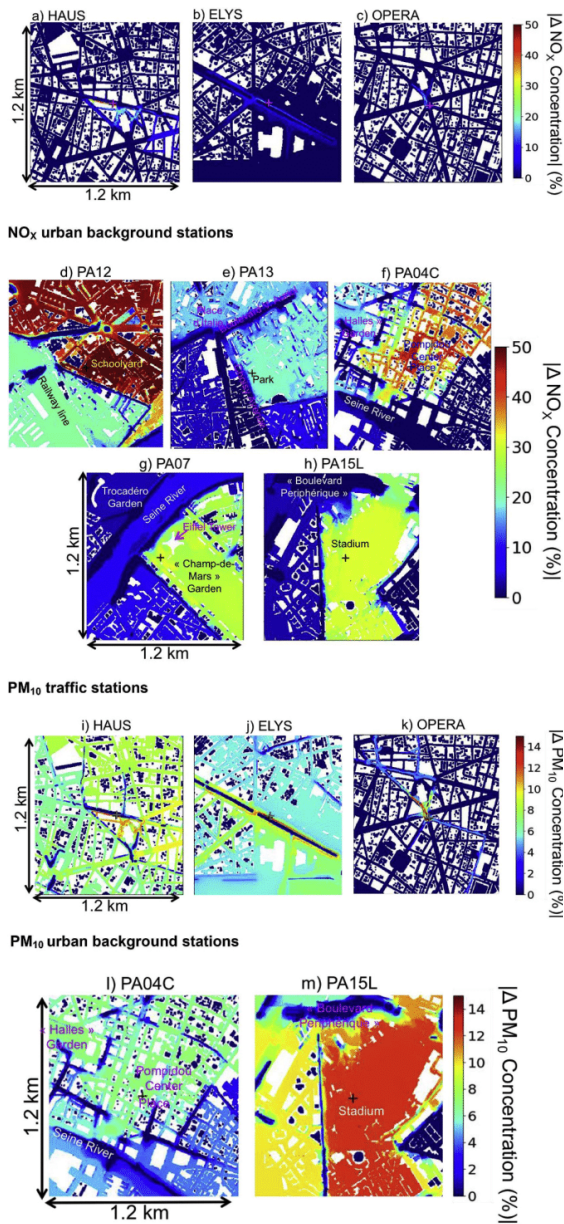


Fig. 8. Percentage difference in PM_{10} concentrations between the updated estimate and the initial PMSS simulation averaged over the ten days of the study. Crosses indicate the location of the monitoring sites.

4. Conclusion

In this study, our main goals are (1) to improve NO_x and PM_{10} simulated concentrations over AIRPARIF monitoring stations and (2) to extend this enhanced concentration estimate within representativeness areas at the vicinity of the station.

We show that the bias in the ten-day PMSS simulation with respect to the AIRPARIF measurements does not follow a normal distribution

Table 4

Average RMSE, standard deviation and % RMSE, and standard deviation differences between simulated and observed NO_x (a), and PM₁₀ (b) concentrations, before and after Bayesian updating, spatially averaged within the most probable representativeness area, for the ten days of the study.

a) NO _x			
TYPE STATION	ID STATION	Initial RMSE [Initial standard deviation (%)]	RMSE difference (%) [Standard deviation difference (%)]
Traffic	HAUS	72.0 [48%]	-37% [-33%]
	OPERA	105.5 [47%]	-46% [-39%]
	ELYS	64.0 [71%]	-20% [-14%]
Urban background	PA12	20.4 [34%]	-21% [-25%]
	PA13	18.5 [36%]	-21% [-19%]
	PA04C	25.3 [37%]	-17% [-22%]
	PA07	26.6 [36%]	-19% [-22%]
	PA15L	24.6 [48%]	-22% [-19%]

b) PM ₁₀			
TYPE STATION	ID STATION	Initial RMSE [Initial standard deviation (%)]	RMSE difference (%) [Standard deviation difference (%)]
Traffic	HAUS	25.3 [52%]	-8% [-12%]
	OPERA	10.0 [26%]	-18% [-24%]
	ELYS	10.5 [21%]	-12% [-13%]
Urban background	PA04C	5.5 [13%]	-5% [-8%]
	PA15L	7.2 [21%]	-10% [-18%]

Uncertainties due to the emission inventory, meteorological conditions and model parameterizations lead to systematic errors. Bayesian statistics is especially appropriate to handle model uncertainties and provide bias correction in such cases.

The proposed Bayesian model combines PMSS model simulations with current and past surface pollutant concentration measurements of the AIRPARIF stations (3 traffic-stations, 2 and 5 urban background stations respectively for PM₁₀ and NO_x). Combination of these two sources of information results in PMSS model error reduction at the station location and provides a spatially resolved concentration estimate in the vicinity of the monitoring site.

The most probable NO_x and PM₁₀ concentrations at the monitoring station location are given by the Bayesian posterior distribution. A sensitivity analysis is performed to find the optimal probabilistic model at each station, and determine the parameters of the prior and likelihood distributions. Past hourly AIRPARIF pollutant concentration measurements are used to establish the prior distribution for each monitoring station. Two different data classifications, based either on the intensity of traffic circulation or meteorological conditions, are tested. The classification per weekday, accounting for traffic circulation appears more appropriate. The hourly data were grouped in several time frames for the linear regression in order to test different likelihood functions. The 2-h period likelihood is found to give the best results by reducing RMSE and standard deviation between simulated concentrations and AIRPARIF measurements. For example, at the OPERA cross-road site, a -72% difference in the RMSE is obtained for NO_x concentrations.

We spatially extend the updated concentration estimates within the representativeness areas of each monitoring site, by applying a weighting function that considers the probability of each pixel to belong to the area. We propose to estimate the updated concentration at each pixel, by taking a weighted average between the posterior mean concentration and the initial PMSS simulation at the given pixel. The fraction of the number of days for which the pixel is selected to belong to the representativeness areas on the total number of days of the study is used as the weighting coefficient.

Results show that final NO_x and PM₁₀ concentration estimates within the most probable representative area (pixels selected seven

days among ten) are closer to AIRPARIF measurements than the initial PMSS simulation with a reduced error (decrease of % RMSE and of % standard deviation).

Bias correction is larger for NO_x concentrations than for PM₁₀ because the initial PMSS model error is smaller for PM₁₀. Modifications are observed over larger areas around the urban stations than around traffic ones due to the size of the representativeness area.

The Bayesian model developed in this study is an innovative and low computational cost method to spatially extend pollutant concentration measurements in the vicinity of the station. By providing low-bias high-resolution pollutant concentration estimates over urban areas, the method could contribute to a better assessment of human exposure to atmospheric pollution. This method should be further validated, by performing local measurements inside representativeness areas.

Finally, longer PMSS simulations would increase the available dataset for the linear regression, leading to more robust likelihood functions. In particular, the impact of the specific air pollution episode with a sharp PM₁₀ increase in the whole region would be attenuated in favor of more general statistics reflecting the baseline conditions.

Acknowledgments:

Acknowledgments to ARIA Technologies for providing the PMSS simulations for ten days.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.aeaoa.2019.100038>.

References

- Amin, N.A.M., Adam, M.B., Aris, A.Z., 2015. Bayesian extreme for modeling high PM10 concentration in johor. In: *Procedia Environmental Sciences, Environmental Forensics* 2015. 30. pp. 309–314. <https://doi.org/10.1016/j.proenv.2015.10.055>.
- Beauchamp, M., de Fouquet, C., Malherbe, L., 2017. Dealing with non-stationarity through explanatory variables in kriging-based air quality maps. *Spat. Statist.* 22, 18–46. <https://doi.org/10.1016/j.spa.2017.08.003>.
- Beauchamp, M., Malherbe, L., de Fouquet, C., Létinois, L., 2018. A necessary distinction between spatial representativeness of an air quality monitoring station and the delimitation of exceedance areas. *Environ. Monit. Assess.* 190 (7), 441. Jun 2018c. ISSN 1573-2959. <https://doi.org/10.1007/s10661-018-6788-y>.
- Beloconi, A., Chrysoulakis, N., Lyapustin, A., Utzinger, J., Vouontsou, P., 2018. Bayesian geostatistical modelling of PM10 and PM2.5 surface level concentrations in Europe using high-resolution satellite-derived products. *Environ. Int.* 121, 57–70. <https://doi.org/10.1016/j.envint.2018.08.041>.
- Boreux, J.J., Parent, E., Bernier, J., 2010. *Pratique du calcul bayésien, Statistique et probabilités appliquées*. Springer-Verlag.
- Cowles, M.K., Zimmerman, D.L., 2003. A Bayesian space-time analysis of acid deposition data combined from two monitoring networks. *J. Geophys. Res.: Atmosph.* 108. <https://doi.org/10.1029/2003JD004001>.
- Elbern, H., Schmidt, H., 2001. Ozone episode analysis by four-dimensional variational chemistry data assimilation. *J. Geophys. Res.* 106 (No. D4), 3569–3590. <https://doi.org/10.1029/2000JD900448>.
- Hanea, R.G., Velders, G.J., Heemink, A., 2004. Data assimilation of ground-level ozone in Europe with a Kalman filter and chemistry transport model. *J. Geophys. Res.* 109, D10302. <https://doi.org/10.1029/2003JD004283>.
- Hanna, S., White, J., Trolier, J., Vernot, R., Brown, M., Gowardhan, A., Kaplan, H., Alexander, Y., Moussafir, J., Wang, Y., Williamson, C., Hannan, J., Hendrick, E., 2011. Comparisons of JU2003 observations with four diagnostic urban wind flow and Lagrangian particle dispersion models. *Atmos. Environ.* 45, 4073–4081. <https://doi.org/10.1016/j.atmosenv.2011.03.058>.
- Harrison, R.M., 2018. Urban atmospheric chemistry: a very special case for study npj. *Clim. Atmosph. Sci.* 1–5. <https://doi.org/10.1038/s41612-017-0010-8>.
- Janssen, S., Dumont, G., Fierens, F., Deutsch, F., Maiheu, B., Celis, D., Trimpeneers, E., Mensink, C., 2012. Land use to characterize spatial representativeness of air quality monitoring stations and its relevance for model validation. *Atmos. Environ.* 59, 492–500. <https://doi.org/10.1016/j.atmosenv.2012.05.028>.
- Mailler, S., Menut, L., Khvorostyanov, D., Valari, M., Couvidat, F., Siour, G., Turquety, S., Briant, R., Tuccella, P., Bessagnet, B., Colette, A., Létinois, L., Markakis, K., Meleux, F., 2017. CHIMERE-2017: from urban to hemispheric chemistry-transport modeling. *Geosci. Model Dev. (GMD)* 10, 2397–2423. <https://doi.org/10.5194/gmd-10-2397-2017>.
- Martín, F., Palomino, I., Vivanco, M.G., 2012. Combination of measured and modelling

- data in air quality assessment in Spain. *Int. J. Environ. Pollut.* 49 (Nos. 1/2), 36–44. <https://doi.org/10.1504/IJEP.2012.049773>.
- McMillan, N.J., Holland, D.M., Morara, M., Feng, J., 2010. Combining numerical model output and particulate data using Bayesian space–time modeling. *Environmetrics* 21, 48–65. <https://doi.org/10.1002/env.984>.
- Moussafir, J., Olry, C., Nibart, M., Albergel, A., Armand, P., Duchenne, C., Mahé, F., Thobois, L., Loaec, S., Oldrini, O., 2014. AIRCITY: A Very High Resolution Atmospheric Dispersion Modeling System for Paris, vol 1 American Society of Mechanical Engineers, Fluids Engineering Division (Publication) FEDSM <https://doi.org/10.1115/FEDSM2014-21820>.
- O'Hagan, A., 2008. The Bayesian approach to statistics. In: Rudas, T. (Ed.), *Handbook of Probability: Theory and Applications*. SAGE Publications, Inc, Thousand Oaks, CA, pp. 85–100. <https://doi.org/10.4135/9781452226620>.
- Pasquier, A., Ré, M., 2017. Considering criteria related to spatial variabilities for the assessment of air pollution from traffic. In: *Transportation Research Procedia, World Conference on Transport Research - WCTR 2016 Shanghai*. 10–15 July 2016 25, pp. 3354–3369. <https://doi.org/10.1016/j.trpro.2017.05.210>.
- Pirani, M., Gulliver, J., Fuller, G.W., Blangiardo, M., 2014. Bayesian spatiotemporal modelling for the assessment of short-term exposure to particle pollution in urban areas. *J. Expo. Sci. Environ. Epidemiol.* 24, 319–327. <https://doi.org/10.1038/jes.2013.85>.
- Rodriguez, D., Valari, M., Payan, S., Eymard, L., 2019. On the spatial representativeness of NO_x and PM₁₀ monitoring-sites in Paris, France. *Atmospheric Environment: X*. <https://doi.org/10.1016/j.aea.2019.100010>.
- Sahu, Sujit K., Mardia, Kanti V., 2005. A bayesian kriged kalman model for short-term forecasting of air pollution levels. *J. Roy. Statist. Soc. Ser. C (Appl. Statist.)* 54 (1), 223–244. <http://www.jstor.org/stable/3592609>.
- Son, Y., Osornio-Vargas, Á.R., O'Neill, M.S., Hystad, P., Texcalac-Sangrador, J.L., Ohman-Strickland, P., Meng, Q., Schwander, S., 2018. Land use regression models to assess air pollution exposure in Mexico City using finer spatial and temporal input parameters. *Sci. Total Environ.* 639, 40–48. <https://doi.org/10.1016/j.scitotenv.2018.05.144>.
- Trini Castelli, S., Armand, P., Tinarelli, G., Duchenne, C., Nibart, M., 2018. Validation of a Lagrangian particle dispersion model with wind tunnel and field experiments in urban environment. *Atmos. Environ.* 193, 273–289. <https://doi.org/10.1016/j.atmosenv.2018.08.045>.
- Wu, C.-D., Zeng, Y.-T., Lung, S.-C.C., 2018. A hybrid kriging/land-use regression model to assess PM_{2.5} spatial-temporal variability. *Sci. Total Environ.* 645, 1456–1464. <https://doi.org/10.1016/j.scitotenv.2018.07.073>.

Further reading

- Fuentes, M., Raftery, A.E., 2005. Model evaluation and spatial interpolation by bayesian combination of observations with outputs from numerical models. *Biometrics* 61, 36–45. <https://doi.org/10.1111/j.0006-341X.2005.030821.x>.

Conclusion et perspectives

L'objectif principal de ce travail de thèse était de caractériser le plus finement possible l'air respiré par les franciliens.

Un premier axe de recherche a consisté en une approche expérimentale de l'exposition en réalisant des campagnes de mesures ponctuelles aux particules PM_{10} et $PM_{2.5}$ avec des instruments de référence et des capteurs mobiles à bas coût. Des contraintes techniques et humaines des capteurs à bas coût ont ainsi été identifiées en vue de l'élaboration de protocoles de mesures pour un prochain déploiement instrumental de plus grande ampleur. Les contraintes identifiées ne sont pas des moindres : dérive thermique dans le temps qui nécessite de réaliser des inter-étalonnages régulièrement, gamme de mesures limitée, fiabilité et reproductibilité de la mesure. Déployés en grand nombre, des capteurs à bas coût (<500€) - en cours de développement au LATMOS (avec des caractéristiques physiques similaires aux capteurs e-PM) - pourraient tout de même donner une indication de l'exposition de la pollution et permettre de cartographier plus finement la pollution de l'air sur l'ensemble de la région parisienne, et surtout renseigner sur sa variabilité spatiale et temporelle.

L'objectif serait d'équiper une centaine de volontaires, membres d'un observatoire participatif. Ces volontaires devraient être sélectionnés de manière à couvrir un périmètre géographique suffisamment large. Il faudrait également qu'ils acceptent de faire des détours afin d'utiliser des chemins peu empruntés de manière à caractériser plus finement l'ensemble du territoire. L'idéal serait que les volontaires passent – de temps en temps - à proximité des stations de mesures AIRPARIF pour vérifier la fiabilité de leurs données et qu'ils aillent au-delà des zones non-renseignées, à savoir au-delà des zones de représentativité des stations de mesures pour estimer la concentration dans ces zones non couvertes par notre approche.

Dans un deuxième temps, nous avons proposé une nouvelle méthodologie pour estimer les zones de représentativité les plus probables à partir des aires de représentativité journalières autour des stations de mesures AIRPARIF à Paris. Nous avons utilisé les champs de concentration horaires, à une résolution horizontale de 3 mètres, simulés par le modèle urbain de dispersion, le modèle PMSS. Les polluants étudiés sont les NO_x et les PM_{10} pour dix jours non consécutifs de mars 2016. Cette période inclut des épisodes de pollution aux particules PM_{10} . Pour cela, des aires de 1.2 km × 1.2 km ont été délimitées, avec la station au centre de la zone.

Nous avons distingué (1) les zones d'homogénéité où le biais journalier par rapport à la concentration simulée à l'endroit de la station ne diffère pas de plus de 20 % et (2) les zones de similarité où deux critères statistiques sont combinés : un fort coefficient de corrélation et une faible erreur (faible NRMSE) entre les pixels représentatifs de la station et le pixel de la station. Les zones de représentativité les plus probables ont été choisies comme les zones de similarité sélectionnées au moins sept jours sur dix. En effet, cette définition de la représentativité permet de conserver l'information sur la variabilité du signal dans la zone par rapport au signal à l'endroit de la station. Cette approche pourrait être transposable à la représentativité de mesures provenant de capteurs mobiles à bas coût. Ces capteurs, moins précis que les instruments de référence, peuvent donner une indication de la variabilité du signal.

Les stations de trafic sont représentatives d'éléments spécifiques du paysage urbain (trottoirs, section de route). Les stations de fond urbain sont représentatives d'un quartier et jusqu'à

soixante-dix fois plus grandes que les stations de trafic. Les zones de représentativité des PM_{10} sont environ deux fois plus grandes que pour les NO_x pour les stations de fond et huit fois plus grandes pour les stations trafic. Des tests de sensibilité ont été effectués pour extraire la variabilité locale liée au trafic du signal de fond urbain et régional des PM_{10} . Les résultats les plus probants sont obtenus avec la normalisation du signal des PM_{10} en retirant la concentration minimale de la « tuile » centrée autour de la station. Les autres options testées (fond CHIMERE, mesures de fond AIRPARIF, etc.) retireraient trop de signal, y compris la variabilité locale liée au trafic. Malgré ce traitement particulier pour le signal des PM_{10} , le signal reste homogène en raison des émissions du secteur résidentiel et tertiaire et un fond résiduel qui ne serait peut-être pas totalement retiré.

Dans un troisième volet, nous avons cherché à propager les mesures AIRPARIF pour rendre plus réalistes les sorties du modèle PMSS biaisées par rapport aux mesures de référence AIRPARIF. Pour ce faire, nous avons développé un modèle spatio-temporel bayésien à l'endroit de la station puis nous l'avons propagé dans les zones de représentativité. Nous avons mis à jour notre connaissance *a priori* au niveau de la station, à savoir les observations passées d'AIRPARIF, en les combinant avec les concentrations de NO_x et PM_{10} simulées par le modèle PMSS et les observations coïncidentes d'AIRPARIF. Nous avons ainsi obtenu une connaissance *a posteriori*. En statistique bayésienne, nous travaillons avec des distributions de concentration. La concentration la plus probable est en fait la moyenne des concentrations *a posteriori*.

Nous avons paramétrisé le modèle bayésien (choix de la vraisemblance et du prior) à l'endroit de la station. Le prior « météo » donnait également des résultats probants pour les PM_{10} . Nous avons fait le choix d'une méthode unifiée qui s'applique aussi bien pour les NO_x que pour les PM_{10} . Le prior de chaque pixel de la zone de représentativité est le même que celui obtenu à la position de la station et la donnée est la concentration simulée par le modèle PMSS au pixel considéré. Pour estimer la concentration la plus réaliste de tous les pixels dans la zone de représentativité on intègre la différence de représentativité de chaque pixel en appliquant une pondération pour les pixels sélectionnés de 9/10 jours (représentatif à 90 % de la station) à 1/10 jour (représentatif à 10 % de la station). Ainsi, un pixel sélectionné dix jours sur dix (représentatif à 100 % de la station) a pour concentration résultante la moyenne de la concentration *a posteriori* du pixel « station ». Un pixel représentatif 9/10 jours a pour concentration finale la somme de 9/10^{ème} de la moyenne de la concentration *a posteriori* et 1/10^{ème} de la concentration initiale simulée par le modèle PMSS.

En dépit du faible jeu de données et de l'épisode de pollution aux PM_{10} , nous avons construit une méthodologie robuste pour déterminer les zones de représentativité des stations AIRPARIF à partir des simulations du modèle PMSS.

Nous avons également proposé une approche novatrice en développant un modèle spatio-temporel bayésien qui combine les sorties du modèle PMSS et les mesures AIRPARIF qui intègre l'information sur la représentativité des stations.

Ce modèle permet d'estimer les concentrations les plus probables dans les zones de représentativité des stations AIRPARIF avec uniquement des mesures à l'endroit de la station. Avec des mesures de terrain à plusieurs endroits de la zone de représentativité, nous pourrions valider notre méthode de propagation du modèle bayésien par pondération.

En perspective de ce travail de thèse, il serait intéressant de mener une campagne de mesures dans la chambre climatique contrôlée de Sense-City en ayant préalablement étalonné les capteurs à bas coût. Des tests avec différents scénarios de la vie quotidienne pourraient être réalisés en intérieur et en extérieur en simulant des épisodes anticycloniques propices à la

pollution aux particules PM₁₀ ou bien des épisodes dépressionnaires favorables à la dispersion des polluants. Les objectifs de ces expériences seraient (1) de quantifier la variabilité spatio-temporelle de la pollution ordinaire (extérieure et intérieure) en milieu contrôlé (température, humidité, taux de pollution ambiante), (2) d'analyser finement le transfert de la pollution de l'extérieur vers l'intérieur, (3) de tester les capteurs (bruit instrumental, reproductibilité de la mesure, etc.) en tenant compte de leur gamme de mesure et en faisant varier la fréquence d'échantillonnage, (4) de travailler sur les protocoles de mesure en identifiant les sources de pollution. Ces nouveaux résultats nous permettraient de déterminer les corrections à appliquer sur les données des capteurs pour que celles-ci soient exploitables.

Les méthodes proposées utilisant les sorties du modèle PMSS pourraient être consolidées avec plus de jours de simulation :

- *pour la représentativité spatiale des stations :*

Dans notre étude, certaines stations ont été relocalisées (création de stations virtuelles) ou non traitées en raison de brins d'émissions manquants sur les simulations. Un jeu de données sans brins d'émissions manquants permettrait d'étudier l'ensemble des stations AIRPARIF à Paris. Un jeu de simulation plus grand, avec au minimum dix jours de chaque saison, permettrait : d'analyser un autre pas de temps, d'analyser la sensibilité de paramètres tels que : (1) les conditions météorologiques dont la vitesse du vent (confirmer les premiers résultats obtenus). En effet, le vent ne joue-t-il que sur la surface de représentativité, ou celle-ci se déplace-t-elle selon la direction du vent ? , (2) la typologie du site, (3) le trafic.

Le seuil des critères de représentativité pourrait être ajusté selon les besoins (un NRMSE plus restrictif si l'on s'intéresse à l'exposition).

- *pour l'application bayésienne :*

Un plus grand jeu de données permettrait également d'ajuster davantage les paramètres du modèle bayésien et de rendre plus robuste l'équation de pondération.

Nous pourrions également cartographier les incertitudes associées aux concentrations les plus probables.

Une étude de comparaison de modèles urbains de la qualité de l'air pourrait aussi être menée. Par exemple, nous pourrions étudier l'impact de la résolution horizontale du modèle et du type de modèle sur les résultats. Est-ce que les zones de représentativité avec les simulations du modèle ADMS-urban et sa résolution horizontale de 12 mètres sont identiques à celles obtenues avec le modèle PMSS ?

De plus, des campagnes de mesures dans les zones de représentativité pourraient être menées pour contrôler les concentrations les plus probables obtenues avec le modèle spatio-temporel bayésien. Ces campagnes de mesures pourraient être réalisées à la fois avec des capteurs mobiles à bas coût et des instruments de référence. Cela permettrait d'évaluer la variabilité spatio-temporelle prédite par le modèle.

Pour aller plus loin, nous pourrions également déterminer les zones de représentativité spatiale en chaque point du modèle pour (1) optimiser le déploiement des stations AIRPARIF, (2) étalonner un réseau de capteurs fixes (Rivano⁶³, Yahya *et al.*⁶⁴).

Une fois la méthode consolidée, les résultats permettraient de faire des cartographies de la pollution pour différentes typologies des sites urbains selon : le jour de la semaine (jours

⁶³http://sense-city.ifsttar.fr/fileadmin/contributeurs/Sens-City/Pose_de_la_premiere_pierre/11042016_-_RIVANO.pdf

⁶⁴ <http://lille2018.appa.asso.fr/wp-content/uploads/2018/12/rousseau.pdf>

ouvrés/samedi/dimanche et jours fériés), heures de pointes/week-end, épisode de pollution hivernal, conditions météorologiques.

Les données collectées par les capteurs mobiles lors de campagnes de mesures à venir pourraient ensuite être intégrées au modèle bayésien développé au cours de cette thèse. Une fois les données de capteurs corrigées par le modèle bayésien, elles pourraient servir de nouvelles entrées (en remplaçant le cadastre d'émission actuel) au modèle PMSS et ainsi améliorer les prévisions du modèle PMSS pour chaque zone de représentativité.

Depuis une quarantaine d'année, le "machine learning" a fait son apparition. Il est actuellement très utilisé dans les entreprises. Le "machine learning" propose un algorithme (régressions, algorithme de Naïve Bayes, clustering ou encore arbres de décisions) que l'expertise humaine adapte afin de conserver la meilleure approximation des données. Dans une étude, Berchet *et al.* (2017) utilisent les sorties d'un modèle lagrangien, comme le modèle PMSS, pour créer une base de données des concentrations et grâce au "machine learning", les concentrations à haute résolution sont estimées sur des longues périodes à partir de mesures météorologiques.

Depuis 2010, le "deep learning", une sous partie du "machine learning" est très utilisé chez les "data scientists" pour de nombreuses applications (reconnaissance faciale, véhicule autonome, etc.) y compris pour la prévision de la qualité de l'air (Cabaneros *et al.*, 2019).

Le "deep learning" ou "réseaux de neurones artificiels profonds" se passe de l'expertise humaine dans le tri des données (Lambert⁶⁵, 2018). Il s'inspire du fonctionnement de notre cerveau avec des réseaux de neurones, afin (1) d'extraire lui-même ses données et (2) d'ajuster son algorithme - en établissant lui-même les corrélations entre les données d'entrée et le résultat obtenu - pour que celui-ci soit le plus réaliste possible.

Le "deep learning" est une piste très intéressante qui mériterait d'être étudiée. Les deux approches : "deep learning" et statistiques bayésiennes sont conceptuellement différentes. Alors que les statistiques bayésiennes explicitent des hypothèses de structure, le "deep learning" ou "machine learning" se passe de modèle (Barber, 2012, « Bayesian Reasoning and Machine Learning », Chapitre 13). Avec les statistiques bayésiennes, les incertitudes sont explicitement quantifiées avec un intervalle de crédibilité probabiliste pour chaque prévision. Le "deep learning" est approprié lorsqu'on a un grand jeu de données à disposition, alors que les statistiques bayésiennes permettent de travailler avec peu de données si on formule les hypothèses et modèles appropriés. Ici, l'approche bayésienne était la plus appropriée car nous ne disposons que d'un petit jeu de données avec dix jours de simulation. Conserver l'expertise humaine pour concevoir le modèle bayésien permet l'encodage de la connaissance dans l'analyse des données. De plus, les statistiques bayésiennes utilisent un cadre formel unifié : le raisonnement probabiliste conditionnel pour résoudre les problèmes statistiques alors que le "deep learning" ou "machine learning" regroupe plusieurs méthodes.

Les perspectives de travail suggérées devraient permettre d'améliorer l'estimation de l'exposition à la pollution atmosphérique d'un francilien en généralisant les résultats obtenus lors de ces travaux de thèse et en couvrant une plus large zone géographique avec le déploiement de nombreux capteurs mobiles sur l'ensemble de la région Île-de-France et ceci sur une plus longue période. Ces résultats pourront ensuite être utilisés pour diminuer les incertitudes des études épidémiologiques qui évaluent l'impact de la pollution de l'air sur la santé d'un individu.

⁶⁵ <http://penseeartificielle.fr/difference-intelligence-artificielle-machine-learning-deep-learning/>

Bibliographie

Amin N.A.M., Adam M.B., Aris A.Z., 2015. Bayesian Extreme for Modeling High PM₁₀ Concentration in Johor. *Procedia Environmental Sciences, Environmental Forensics* 2015 30, 309–314. <https://doi.org/10.1016/j.proenv.2015.10.055>

APHEKOM, 2012. Improving Knowledge and Communication for Decision Making on Air Pollution and Health in Europe. <http://aphekom.org>

Ashmore MR., Dimitroulopoulou C., 2009. Personal exposure of children to air pollution, *Atmospheric Environment*, vol 43, no. 1, pp. 128–141. <https://doi.org/10.1016/j.atmosenv.2008.09.024>

Baklanov A., Lawrence M., Pandis S., Mahura A., Finardi S., Moussiopoulos N., Beekmann M., Laj P., Gomes L., Jaffrezo J.-L., Borbon A., Coll I., Gros V., Sciare J., Kukkonen J., Galmarini S., Giorgi F., Grimmond S., Esau I., Stohl A., Denby B., Wagner T., Butler T., Baltensperger U., Builtjes P., van der Hout D., van der Gon H. D., Collins B., Schluenzen H., Kulmala M., Zilitinkevich S., Sokhi R., Friedrich R., Theloke J., Hummer U., Jalkinen L., Halenka T., Wiedensohler A., Pyle J., and Rossow W.B., 2010: MEGAPOLI: concept of multiscale modeling of megacity impact on air quality and climate, *Adv. Sci. Res.*, 4, 115–120, <https://doi.org/10.5194/asr-4-115-2010>

Beauchamp M., de Fouquet C., Malherbe L., 2017. Dealing with non-stationarity through explanatory variables in kriging-based air quality maps. *Spatial Statistics* 22, 18–46. <https://doi.org/10.1016/j.spasta.2017.08.003>

Beauchamp M., Malherbe L., 2016. Annexe technique au rapport intitulé estimation de l'exposition des populations aux dépassements de seuils réglementaires. Interpolation des sorties de modèles urbains par krigeage avec dérive polynomiale. note lcsqa. <http://www.lcsqa.org>

Beauchamp M., Malherbe L., de Fouquet C., and Létinois L., 2018. A necessary distinction between spatial representativeness of an air quality monitoring station and the delimitation of exceedance areas. *Environmental Monitoring and Assessment*, 190(7):441, Jun 2018c. ISSN 1573-2959. <https://doi.org/10.1007/s10661-018-6788-y>.

Beekmann M., Prévôt A. S. H., Drewnick F., Sciare J., Pandis S.N., Denier van der Gon H. A. C., Crippa M., Freutel F., Poulain L., Ghersi V., Rodriguez E., Beirle S., Zotter P., von der Weiden-Reinmüller S.-L., Bressi M., Fountoukis C., Petetin H., Szidat S., Schneider J., Rosso A., El Haddad I., Megaritis A., Zhang Q. J., Michoud V., Slowik J. G., Moukhtar S., Kolmonen P., Stohl A., Eckhardt S., Borbon A., Gros V., Marchand N., Jaffrezo J.-L., [280] Schwarzenboeck A., Colomb A., Wiedensohler A., Borrmann S., Lawrence M., Baklanov A., and Baltensperger U., 2015: In situ, satellite measurement and model evidence on the dominant regional contribution to fine particulate matter levels in the Paris megacity, *Atmos. Chem. Phys.*, 15, 9577–9591, <https://doi.org/10.5194/acp-15-9577-2015>.

Beloconi A., Chrysoulakis N., Lyapustin A., Utzinger J., Vounatsou P., 2018. Bayesian geostatistical modelling of PM₁₀ and PM_{2.5} surface level concentrations in Europe using high-resolution satellite-derived products. *Environment International* 121, 57–70. <https://doi.org/10.1016/j.envint.2018.08.041>

Berchet A., Zink K., Muller C., Oettl D., Brunner J., Emmenegger L., Brunner D., 2017. A cost-effective method for simulating city-wide air flow and pollutant dispersion at building resolving scale. *Atmospheric Environment* 158, 181–196. <https://doi.org/10.1016/j.atmosenv.2017.03.030>

Blanchard C.L., Tanenbaum S., Hidy G.M., 2014. Spatial and temporal variability of air pollution in Birmingham, Alabama. *Atmospheric Environment* 89, 382–391. <https://doi.org/10.1016/j.atmosenv.2014.01.006>

Bobbia M., Cori A., de Fouquet C., 2008. Représentativité spatiale d'une station de mesure de la pollution atmosphérique. *Pollution atmosphérique*. <https://doi.org/10.4267/pollution-atmospherique>

Boreux J.J., Parent E., Bernier J., 2010. *Pratique du calcul bayésien, Statistique et probabilités appliquées*. Springer-Verlag.

Bressi M., Sciare J., Ghersi V., Bonnaire N., Nicolas J. B., Petit J.-E., Moukhtar S., Rosso A., Mihalopoulos N., and Féron A., 2013: A one-year comprehensive chemical characterisation of fine aerosol (PM_{2.5}) at urban, suburban and rural background sites in the region of Paris (France), *Atmos. Chem. Phys.*, 13, 7825-7844, <https://doi.org/10.5194/acp-13-7825-2013>, 2013.

Bressi M., Sciare J., Ghersi V., Mihalopoulos N., Petit J.-E., Nicolas J. B., Moukhtar S., Rosso A., Féron A., Bonnaire N., Poulakis E., and Theodosi C., 2014: Sources and geographical origins of fine aerosols in Paris (France), *Atmos. Chem. Phys.*, 14, 8813-8839, <https://doi.org/10.5194/acp-14-8813-2014>.

Budde M., El Masri R., Riedel T., Beigl M., 2013. Enabling low-cost particulate matter measurement for participatory sensing scenarios. *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia, MUM 2013*. <https://doi.org/10.1145/2541831.2541859>

Cabaneros S.M., Calautit J.K., Hughes B.R., 2019. A review of artificial neural network models for ambient air pollution prediction. *Environmental Modelling & Software* 119, 285–304. <https://doi.org/10.1016/j.envsoft.2019.06.014>

Camredon M., Aumont B., 2007. Modélisation chimique de l'ozone et des oxydants gazeux
Chemical modelling of ozone and gaseous oxidants I – L'ozone troposphérique :
production/consommation et régimes chimiques. *Pollution atmosphérique- N°193*, Janvier-
Mars 2007 <http://lodel.irevues.inist.fr/pollution-atmospherique/index.php?id=1404&format=print>

Castell N., Dauge F.R., Schneider P., Vogt M., Lerner U., Fishbain B., Broday D., Bartonova A., 2017. Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates?, *Environment International*, Volume 99, February 2017, Pages 293-302

<https://doi.org/10.1016/j.envint.2016.12.007>

CITEPA. Center Interprofessionnel Technique d'Etudes de la Pollution Atmosphérique

<https://www.citepa.org/en/air-and-climate/pollutants-and-ghg>.

Cowles M.K., Zimmerman D.L., 2003. A Bayesian space-time analysis of acid deposition data combined from two monitoring networks. *Journal of Geophysical Research: Atmospheres* 108. <https://doi.org/10.1029/2003JD004001>

Duyzer J., van den Hout D., Zandveld P., van Ratingen S., 2015. Representativeness of air quality monitoring networks. *Atmospheric Environment* 104, 88–101. <https://doi.org/10.1016/j.atmosenv.2014.12.067>

Elbern H., and Schmidt H., 2001. Ozone episode analysis by four-dimensional variational chemistry data assimilation, *J. Geophys. Res.*, 106, No. D4, 3569-3590. <https://doi.org/10.1029/2000JD900448>

European Commission (EC), 2008. Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe. The framework directive. *Off. J. Eur. Union En. Ser. L*152/51.

European Commission (EC), 2011/850/EU. Commission Implementing Decision of 12, December 2011. Laying down rules for Directives 2004/107/EC and 2008/50/EC of the European Parliament and of the Council as regards the reciprocal exchange of information and reporting on ambient air quality. *Off. J. Eur. Union En. Ser. L* 335/86.

Fuentes M., Raftery A.E., 2005. Model Evaluation and Spatial Interpolation by Bayesian Combination of Observations with Outputs from Numerical Models. *Biometrics* 61, 36–45. <https://doi.org/10.1111/j.0006-341X.2005.030821.x>

Gozzi F., Della Ventura G., Marcelli A., 2016. Mobile monitoring of particulate matter: State of art and perspectives, *Atmospheric Pollution Research*, Volume7, Issue 2, March 2016, Pages 228-234 <https://doi.org/10.1016/j.apr.2015.09.007>

Hagemann R., Corsmeier U., Kottmeier C., Rinke R., Wieser A., Vogel B., 2014. Spatial variability of particle number concentrations and NO_x in the Karlsruhe (Germany) area obtained with the mobile laboratory “AERO-TRAM.” *Atmospheric Environment* 94, 341–352. <https://doi.org/10.1016/j.atmosenv.2014.05.051>

Hagenbjörk A., Malmqvist E., Mattisson K., Sommar N.J., Modig L., 2017. The spatial variation of O₃, NO, NO₂ and NO_x and the relation between them in two Swedish cities. *Environmental Monitoring and Assessment* 189. <https://doi.org/10.1007/s10661-017-5872-z>

Hanea R.G., Velders G.J. and Heemink A., 2004. Data assimilation of ground-level ozone in Europe with a Kalman filter and chemistry transport model *J. Geophys. Res* 109, D10302. <https://doi.org/10.1029/2003JD004283>

Hanna S., White J., Trolier J., Vernot R., Brown M., Gowardhan A., Kaplan H., Alexander Y., Moussafir J., Wang Y., Williamson C., Hannan J., Hendrick E., 2011. Comparisons of JU2003 observations with four diagnostic urban wind flow and Lagrangian particle dispersion

models. Atmospheric Environment 45, 4073–4081.
<https://doi.org/10.1016/j.atmosenv.2011.03.058>

Harrison R.M., 2018. Urban atmospheric chemistry: a very special case for study. *Climate and Atmospheric Science* 1, 5. <https://doi.org/10.1038/s41612-017-0010-8>

Host S., 2013. « Exposition à la pollution atmosphérique liée au trafic routier et risques sanitaires », *VertigO - la revue électronique en sciences de l'environnement* <http://journals.openedition.org/vertigo/12816> ; DOI : 10.4000/vertigo.12816

IARC report. https://www.iarc.fr/en/media-centre/iarcnews/pdf/pr221_E.pdf

Janssen S., Dumont G., Fierens F., Deutsch F., Maiheu B., Celis D., Trimpeneers E., Mensink C., 2012. Land use to characterize spatial representativeness of air quality monitoring stations and its relevance for model validation. *Atmospheric Environment* 59, 492–500. <https://doi.org/10.1016/j.atmosenv.2012.05.028>

Jerrett M., Burnett R.T., Pope C.A., Ito K., Thurston G., Krewski D., Shi Y., Calle E., Thun M., 2009. Long-Term Ozone Exposure and Mortality. *New England Journal of Medicine* 360, 1085–1095. <https://doi.org/10.1056/NEJMoa0803894>

Koivisto A.J., Kling K.I., Hänninen O., Jayjock M., Löndahl J., Wierzbicka A., Fonseca A.S., Uhrbrand K., Boor B.E., Jiménez A.S., Hämeri K., Maso M.D., Arnold S.F., Jensen K.A., Viana M., Morawska L., Hussein T., 2019. Source specific exposure and risk assessment for indoor aerosols. *Science of The Total Environment* 668, 13–24. <https://doi.org/10.1016/j.scitotenv.2019.02.398>

Kracht O., Santiago J.-L., Martin F., Piersanti A., Cremona G., Righini G., Vitali L., Delaney K., Basu B., Ghosh B., Spangl W., Brendle C., Latikka J., Kousa A., Pärjälä E., Meretoja M., Malherbe L., Letinois L., Beauchamp M., Lenartz F., Hutsemekers V., Nguyen L., Hoogerbrugge R., Eneroth K., Silvergren S., Hooyberghs H., Viaene P., Maiheu B., Janssen S., Roet D. and Gerboles M., Spatial representativeness of air quality monitoring sites - Outcomes of the FAIRMODE/AQUILA intercomparison exercise, EUR 28987 EN, Publications Office of the European Union, Luxembourg, 2017, ISBN 978-92-79-77218-4, doi:10.2760/60611,JRC108791.

Kirchner S., Arènes J.F., Cochet C., Derbez M., Duboudin C., Elias P., Grégoire A., Jédor B., Lucas J.P., Pasquier N., Pignernet M., Ramalho O., 2007.- État de la qualité de l'air dans les logements français – *Environnement, Risques & Santé*, 6/4, juillet-août 2007, p. 250-269. doi:10.1684/ers.2007.0096.

Mailler S., Menut L., Khvorostyanov D., Valari M., Couvidat F., Siour G., Turquety S., Briant R., Tuccella P., Bessagnet B., Colette A., Létinois L., Markakis K., Meleux F., 2017. CHIMERE-2017: From urban to hemispheric chemistry-transport modeling. *Geoscientific Model Development*, 2017. <https://doi.org/10.5194/gmd-10-2397-2017>

Martin F., Fileni L., Palomino I., Vivanco M.G., Garrido J.L., 2014. Analysis of the spatial representativeness of rural background monitoring stations in Spain. *Atmospheric Pollution Research* 5, 779–788. <https://doi.org/10.5094/APR.2014.087>

Martín F., Palomino I. and Vivanco M.G., 2012. Combination of measured and modelling data in air quality assessment in Spain. *Int. J. Environment and Pollution*, Vol. 49, Nos. 1/2, 36-44. <https://doi.org/10.1504/IJEP.2012.049773>

Merbitz H., Fritz S., Schneider C., 2012. Mobile measurements and regression modeling of the spatial particulate matter variability in an urban area. *Sci. Total Environ.* <https://doi.org/10.1016/j.scitotenv.2012.08.049>

Monn C., Carabias V., Junker M., Waeber R., Karrer M., Wanner H.U., 1997. Smallscale spatial variability of particulate matter < 10 µm (PM₁₀) and nitrogen dioxide. *Atmos. Environ.* 31 (15), 2243-2247. https://ac.els-cdn.com/S1352231097000307/1-s2.0-S1352231097000307-main.pdf?_tid=5b784339-26e5-4dc7-a394270e138f6e3e&acdnat=1540990886_dca1442ff11c09f5155490e447af9d8e

Moussafir J., Olry C., Nibart M., Albergel A., Armand P., Duchenne C., Mahé F., Thobois L., Loaëc S., Oldrini O., 2014. AIRCITY: A Very High Resolution Atmospheric Dispersion Modeling System for Paris. American Society of Mechanical Engineers, Fluids Engineering Division (Publication) FEDSM. doi:10.1115/FEDSM2014-21820

Nappo C.J., Caneill J.Y., Furman R.W., Gifford F.A., Kaimal J.C., Kramer M.L., Lockhart T.J., Pendergast M.M., Pielke R.A., Randerson D., Shreffler J.H., Wyngaard J.C., 1982. The workshop on the representativeness of meteorological-observations, June 1981. Boulder, Colorado. *Bulletin Am. Meteorol. Soc.* 63, 761-764.

Nguyen P.L., Stefess G., de Jonge D., Snijder A., Hermans P.M.J.A., van Loon P., Hoogerbrugge R., 2012. Evaluation of the representativeness of the Dutch air quality monitoring stations. The National, Amsterdam, Noord-Holland, Rijnmond-area, Limburg and Noord-Brabant networks. RIVM Report 680704021/2012

O'Hagan A., 2008. The Bayesian approach to statistics. In Rudas, T. *Handbook of probability: Theory and applications* (pp. 85-100). Thousand Oaks, CA: SAGE Publications, Inc. doi:10.4135/9781452226620

Owen M.K., Ensor D.S., Sparks L.E., 1992. Airborne particle sizes and sources found in indoor air. *Atmospheric Environment. Part A. General Topics, Fifth International Conference on Indoor Air Quality and Climate Indoor Air '90: Characterization of Indoor Air* 26, 2149–2162. [https://doi.org/10.1016/0960-1686\(92\)90403-8](https://doi.org/10.1016/0960-1686(92)90403-8)

Pasquier A., and Ré M., 2017. Considering criteria related to spatial variabilities for the assessment of air pollution from traffic. *Transportation Research Procedia, World Conference on Transport Research - WCTR 2016 Shanghai*. 10-15 July 2016 25, 3354–3369. <https://doi.org/10.1016/j.trpro.2017.05.210>

Patton A.P., Perkins J., Zamore W., Levy J.I., Brugge D., Durant J.L., 2014. Spatial and temporal differences in traffic-related air pollution in three urban neighborhoods near an interstate highway. *Atmos. Environ.* 99, 309-321.

Pellizzari E., 1999. Particulate matter and manganese exposures in Toronto, Canada. *Atmospheric Environment* 33, 721–734. [https://doi.org/10.1016/S1352-2310\(98\)00229-5](https://doi.org/10.1016/S1352-2310(98)00229-5)

Petit C., Loubet B., Rémy E., Aubry C., Duguay F., Missonnier J., Cellier P., Feiz A.A., Blondeau C., Mauclair C., Durand B., 2013. Dépôt de polluants sur les espaces agricoles à proximité des voies de transport en Île-de-France. Quantification, perception et adaptation des acteurs. *Vertigo* - la revue électronique en sciences de l'environnement. <https://doi.org/10.4000/vertigo.12833>

Piersanti A., Vitali L., Righini G., Cremona G., Ciancarella L., 2015. Spatial representativeness of air quality monitoring stations: A grid model based approach. *Atmospheric Pollution Research* 6, 953–960. <https://doi.org/10.1016/j.apr.2015.04.005>

Pirani M., Gulliver J., Fuller G.W., Blangiardo M., 2014. Bayesian spatiotemporal modelling for the assessment of short-term exposure to particle pollution in urban areas. *J Expo Sci Environ Epidemiol* 24, 319–327. <https://doi.org/10.1038/jes.2013.85>

Ramalho O., Lucas J.-P., Mandin C., Derbez M., Kirchner S., 2012. Niveaux de particules dans les environnements intérieurs en France, *Revue Pollution Atmosphérique Climat, Santé, Société*, Novembre 2012, Pages 37-41, <http://www.appa.asso.fr/docs/1/fckeditor/file/Revue/PollutionAtmospherique/Hors-serie-particules-novembre-2012/P--Ramalho.pdf>

REVIHAAP Project, 2013. Review of evidence on health aspects of air pollution. http://www.euro.who.int/_data/assets/pdf_file/0004/193108/REVIHAAP-Final-technical-report-final-version.pdf?ua=1

Rouil L., 2012. Les modèles régionaux et les modèles urbains : nouveaux outils d'aide à la décision. *Pollution atmosphérique. Climat, santé, société. Numéro spécial particules – Novembre 2012.* <http://www.appa.asso.fr/docs/1/fckeditor/file/Revue/PollutionAtmospherique/Hors-serie-particules-novembre-2012/p--%20Rouil%281%29.pdf>

Sahu Sujit K., and Kanti V. Mardia, 2005. A Bayesian Kriged Kalman Model for Short-Term Forecasting of Air Pollution Levels. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 54, no. 1 (2005): 223-44. <http://www.jstor.org/stable/3592609>

Santiago J.L., Martín F., Martilli A., 2013. A computational fluid dynamic modeling approach to assess the representativeness of urban monitoring stations. *Science of The Total Environment* 454-455, 61–72. <https://doi.org/10.1016/j.scitotenv.2013.02.068>

Seinfeld J.H., Pandis S.N., 2016. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*. John Wiley & Sons.

Senat Report n°610, 2015. http://www.senat.fr/fileadmin/Fichiers/Images/commission/enquete/pollution_air/Synthese_C_E_Pollution_de_1_air.pdf

Son Y., Osornio-Vargas Á.R., O'Neill M.S., Hystad P., Texcalac-Sangrador J.L., Ohman-Strickland P., Meng Q., Schwander S., 2018. Land use regression models to assess air pollution exposure in Mexico City using finer spatial and temporal input parameters. *Science of The Total Environment* 639, 40–48. <https://doi.org/10.1016/j.scitotenv.2018.05.144>

Spangl W., Schneider J., Moosmann L., Nagl C., 2007: Representativeness and classification of air quality monitoring stations – final report. Service contract to the European Commission - DG Environment Contract No. 07.0402/2005/419392/MAR/C1. Umweltbundesamt, Wien, 2007, Reports, Bd. REP-0121.

Stein A.F., Isakov V., Godowitch J., Draxler R.R., 2007. A hybrid modeling approach to resolve pollutant concentrations in an urban area. *Atmospheric Environment* 41, 9410–9426. <https://doi.org/10.1016/j.atmosenv.2007.09.004>

Steinle S., Reis S., Sabel C.E., Semple S., Twigg M.M., Braban C.F., Leeson S.R., Heal M.R., Harrison D., Lin C., Wu H., 2015. Personal exposure monitoring of PM_{2.5} in indoor and outdoor microenvironments. *Science of The Total Environment* 508, 383–394. <https://doi.org/10.1016/j.scitotenv.2014.12.003>

Stetzenbach L.D., Buttner M.P., Cruz, P., 2004. Detection and enumeration of airborne biocontaminants. *Current Opinion in Biotechnology* 15, 170–174. <https://doi.org/10.1016/j.copbio.2004.04.009>

Trini Castelli S., Armand P., Tinarelli G., Duchenne C., Nibart M., 2018. Validation of a Lagrangian particle dispersion model with wind tunnel and field experiments in urban environment. *Atmospheric Environment* 193, 273–289. <https://doi.org/10.1016/j.atmosenv.2018.08.045>

Valari M., Menut L., Chatignoux E., 2011. Using a chemistry transport model to account for the spatial variability of exposure concentrations in epidemiologic air pollution studies. *J Air Waste Manag Assoc* 61, 164–179. <https://doi.org/10.3155/1047-3289.61.2.164>

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. *Atmospheric Environment* 105, 148–161. <https://doi.org/10.1016/j.atmosenv.2015.01.017>

Vardoulakis S., Gonzalez-Flesca N., Fisher B.E.A., Pericleous K., 2005. Spatial variability of air pollution in the vicinity of permanent monitoring station in central Paris. *Atmospheric Environment, Fourth International Conference on Urban Air Quality: Measurement, Modelling and Management*, 25-28 March 2003 39, 2725–2736. <https://doi.org/10.1016/j.atmosenv.2004.05.067>

Vitali L., Morabito A., Adani M., Assennato G., Ciancarella L., Cremona G., Giua R., Pastore T., Piersanti A., Righini G., Russo F., Spagnolo S., Tanzarella A., Tinarelli G., Zanini G., 2016. A Lagrangian modelling approach to assess the representativeness area of an industrial air quality monitoring station. *Atmospheric Pollution Research* 7, 990–1003. <https://doi.org/10.1016/j.apr.2016.06.002>

Wu C.-D., Zeng Y.-T., Lung S.-C.C., 2018. A hybrid kriging/land-use regression model to assess PM_{2.5} spatial-temporal variability. *Science of The Total Environment* 645, 1456–1464. <https://doi.org/10.1016/j.scitotenv.2018.07.073>

**Annexe : Support de la soutenance du 24 juin 2019, à
Paris (Campus Pierre et Marie Curie, Sorbonne
Université)**

Caractérisation de la pollution urbaine en Île-de-France par une synergie de mesures de surface et de modélisation fine échelle



Source : DUCLOS/SIPA

Delphy Rodriguez

Paris, Campus de Jussieu,
le 24 juin 2019

Étude de la pollution de l'air à l'échelle urbaine

Santé

- 48000 décès prématurés /an en France dus aux particules fines ($PM_{2.5}$)
- 87% de la population urbaine de l'UE respire des niveaux de particules fines ($PM_{2.5}$) >recommandations OMS
- Effets jusqu'à 300 à 500m des grands axes

Coût économique

- 101,3 milliard €/an en France (rapport Sénat, 2015)
- Santé
 - Environnement



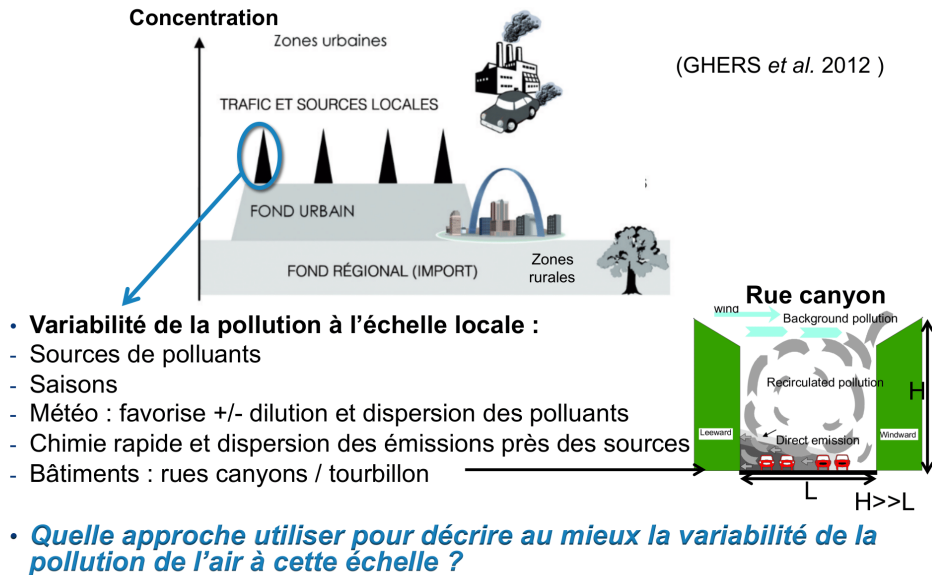
Mieux caractériser l'exposition de la population

État de l'art : exposition homogène vs. pollution air hétérogène, mobilité de la population, « effet cocktail »

Pollution de l'air à différentes échelles



Enjeux de la pollution à l'échelle locale



Cadre de la thèse

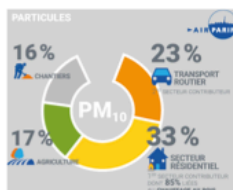
• Observatoire de surveillance de la qualité de l'air : AASQA, AIRPARIF (IDF)

• **Polluants étudiés :**

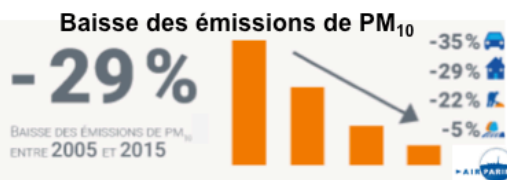
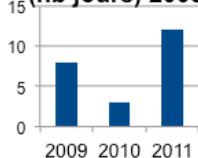
NO_x (NO₂+NO *monoxyde d'azote*) : sources anthropiques

PM₁₀ : sources naturelles et anthropiques

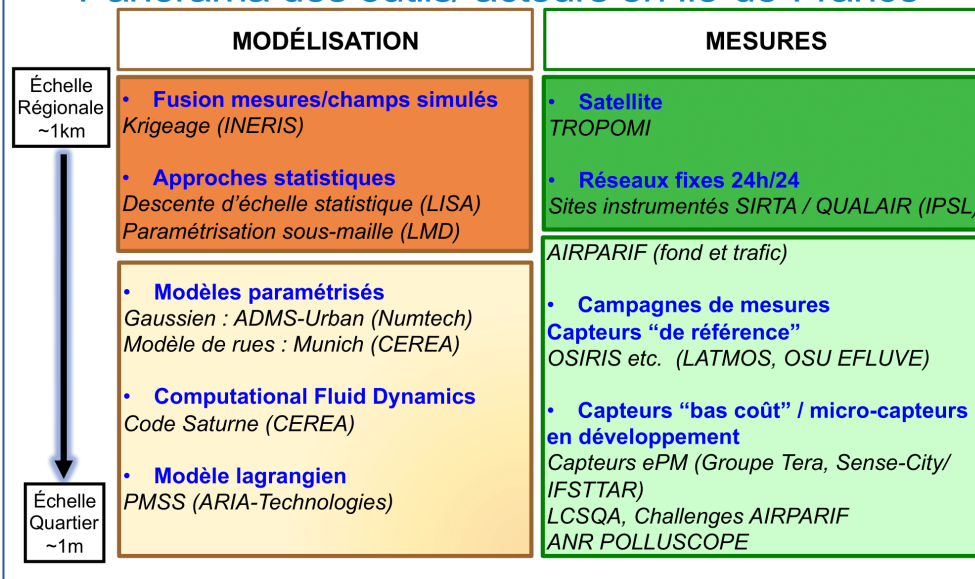
• **Répartition des sources d'émission anthropiques en IDF** *inventaire AIRPARIF 2015*



PM₁₀ : dépassements de seuils (nb jours) 2005-2015



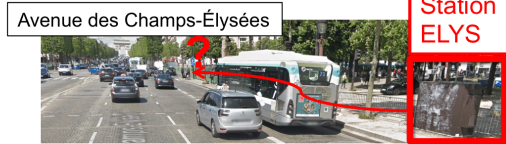
Pollution de proximité Panorama des outils/ acteurs en Île-de-France



Objectif principal et stratégies


Caractériser finement la pollution de l'air en Île-de-France

- Partie I : Quelle est la représentativité spatiale de chaque station AIRPARIF ?

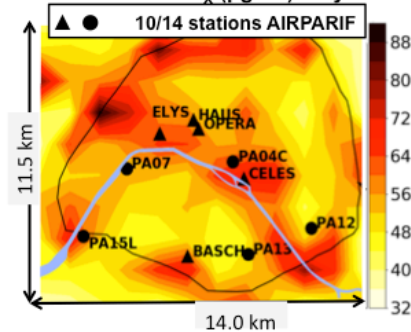


- Partie II : Peut-on améliorer les simulations d'un modèle avec les mesures des stations AIRPARIF ?

Outils utilisés

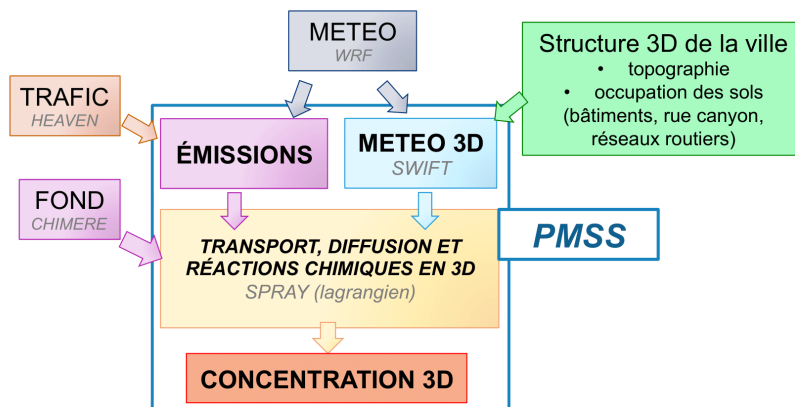
- Mesures de NO_x et PM_{10} de 10 stations AIRPARIF 
- Champs de concentration des NO_x et PM_{10} simulés par le modèle PMSS à 3 m de résolution, 10 jours en mars 2016

Concentrations de NO_x ($\mu\text{g}/\text{m}^3$) moyennées sur 10 jours de Mars 2016



Simulations fournies par
AriaTechnologies
dans le cadre du projet MoUVie
Mobilité et qualité de vie en milieu urbain

Modèle PMSS *Parallel Micro-SWIFT-SPRAY* à 3 m de résolution horizontale



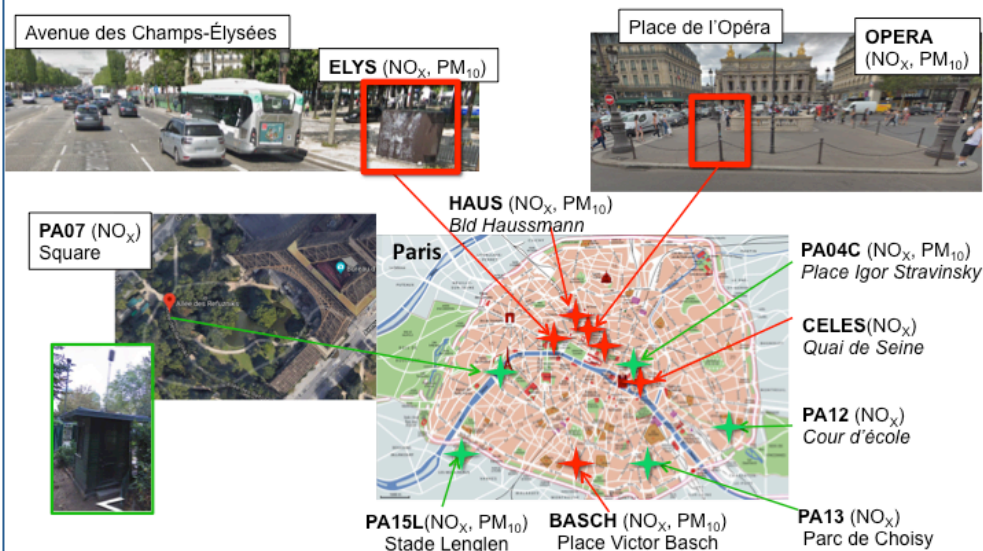
😊 **Atouts** : très haute résolution, processus fines échelles (tourbillons), géométrie des bâtiments

😞 **Limites** : mécanismes chimiques simplifiés, incertitudes (trafic et météo), accessibilité aux données 3D (bâti), temps de calcul

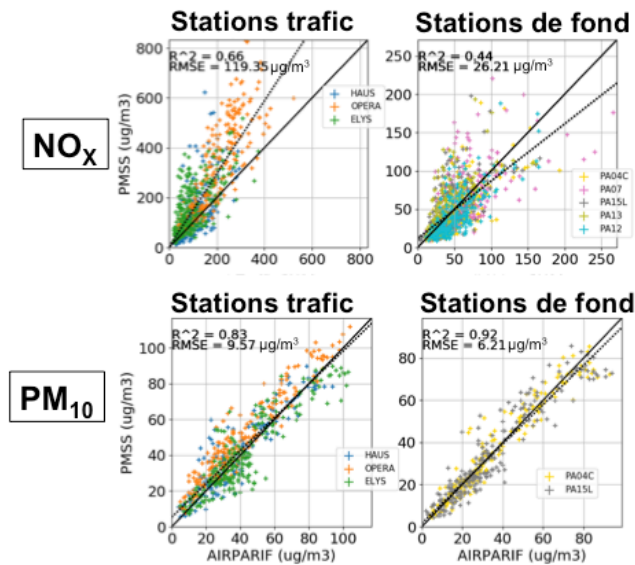
Mesures des stations AIRPARIF



★ TRAFIC ★ FOND URBAIN



Comparaison mesures AIRPARIF et simulations PMSS, 10 jours en mars 2016



Partie I. Représentativité spatiale des stations de mesures AIRPARIF de NO_x et de PM₁₀ à Paris



On the spatial representativeness of NO_x and PM₁₀ monitoring-sites in Paris, France

Delphy Rodriguez^{a,*}, Myrto Valari^b, Sébastien Payan^a, Laurence Eymard^a

<https://doi.org/10.1016/j.aeaoa.2019.100010>



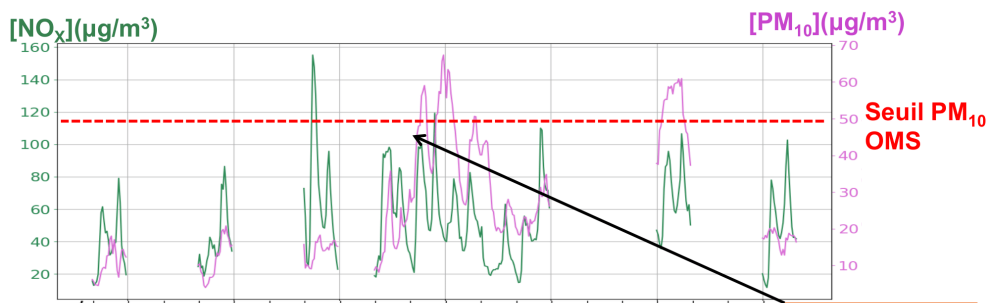
Domaine d'étude (1.2 km × 1.2 km) centré autour de chaque station AIRPARIF

« Tuile » centrée autour de la station ELYS



Caractéristiques des 10 jours de simulations PMSS

- Concentrations des NO_x et PM₁₀ simulées au-dessus de Paris en mars 2016



- Quel signal de PM₁₀ utilisé pour faire ressortir la variabilité locale des émissions liées au trafic ?

$$C(x,y)_{normalisée} = \frac{C(x,y)_{brute} - \min(C_{aire})}{\max(C_{aire}) - \min(C_{aire})}$$

Quel fond retiré ($\min(C_{aire})$) ?

Tests : CHIMERE, stations de fond AIRPARIF, 10% des « tuiles » avec la MJ la + basse, « tuile » avec la MJ la + basse de Paris => retrait de la variabilité locale due au trafic

Fond retenu : minimum de la « tuile » centrée autour de la station

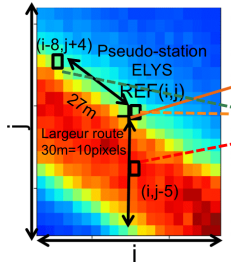
Début épisode de pollution aux PM₁₀
conditions anticycloniques : {vent faible + P_{atm} élevée}

Comment aborder la « représentativité spatiale » ?

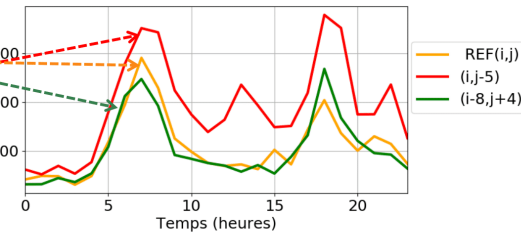
- REF=« pseudo-mesure » : travail méticuleux pour le choix du pixel



Concentration de NO_x (µg/m³), le 2 mars 2016, av. des Champs-Élysées



Concentration de NO_x (µg/m³), le 2 mars 2016

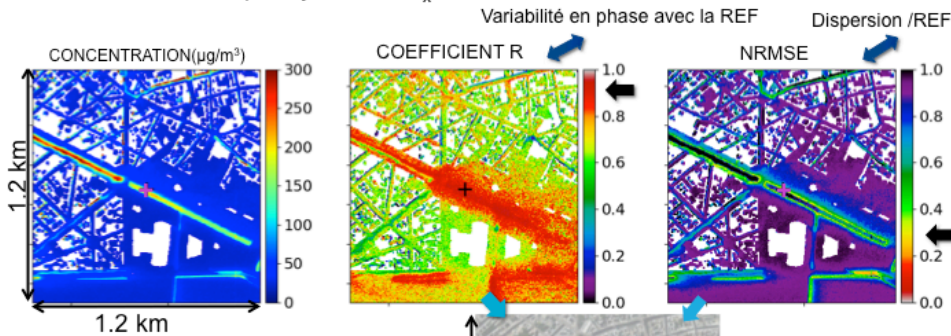


- Choix de critères qui intègrent la variabilité du signal

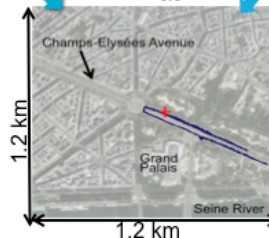
CRITÈRES : {Coefficient R & NRMSE}

- Avenue des Champs-Élysées : NO_x, le 2 mars 2016

REF=« pseudo-mesure »



- Détection de pixels connectés



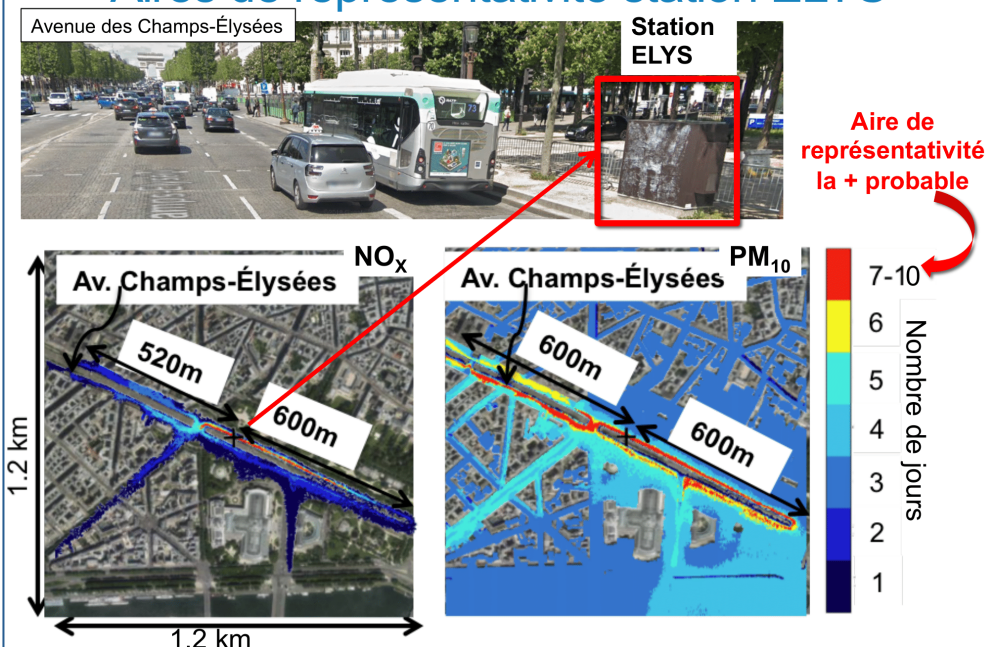
Analyse de sensibilité des seuils

- Étude menée sur stations trafics : vérifier distinction route/trottoir avec seuils sélectionnés
- Critères calculés jour par jour
- $300\text{m}^2 < \text{aire de représentativité} < 400 \times 10^3\text{m}^2$ (=succès)
- Seuils retenus : seuils avec un maximum de succès

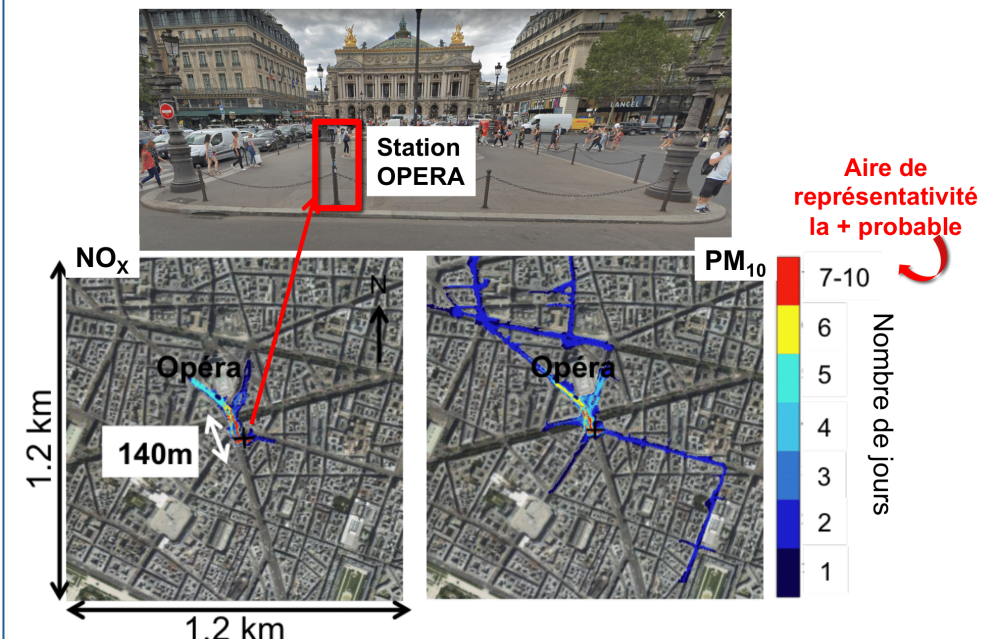
Polluants	R	NRMSE	Nombre de jours où aire de représentativité=succès					Succès total (%)
			BASCH_v	CELES_v	ELYS	HAUS	OPERA	
NO _x	0.7	0.45	10	10	10	10	10	100
PM ₁₀	0.75	0.3	9 <small>$1=846 \times 10^3$</small>	8 <small>$1=148$ $1=459 \times 10^3$</small>	7 <small>$3 < 826 \times 10^3$</small>	6 <small>$4 < 561 \times 10^3$</small>	9 <small>$1=161$</small>	78

- Seuils retenus ensuite appliqués aux stations de fond

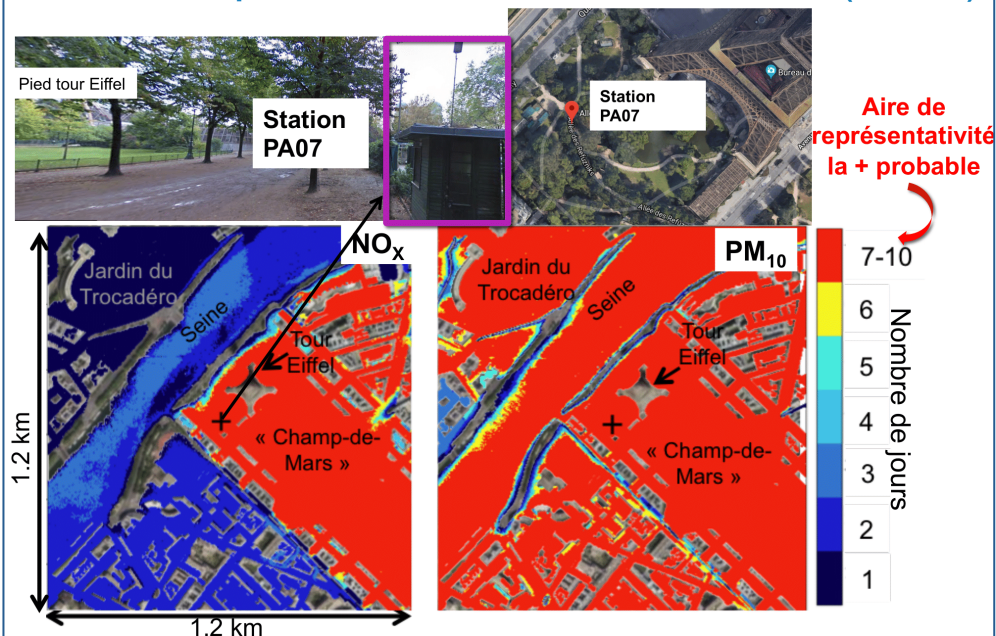
Aires de représentativité station ELYS



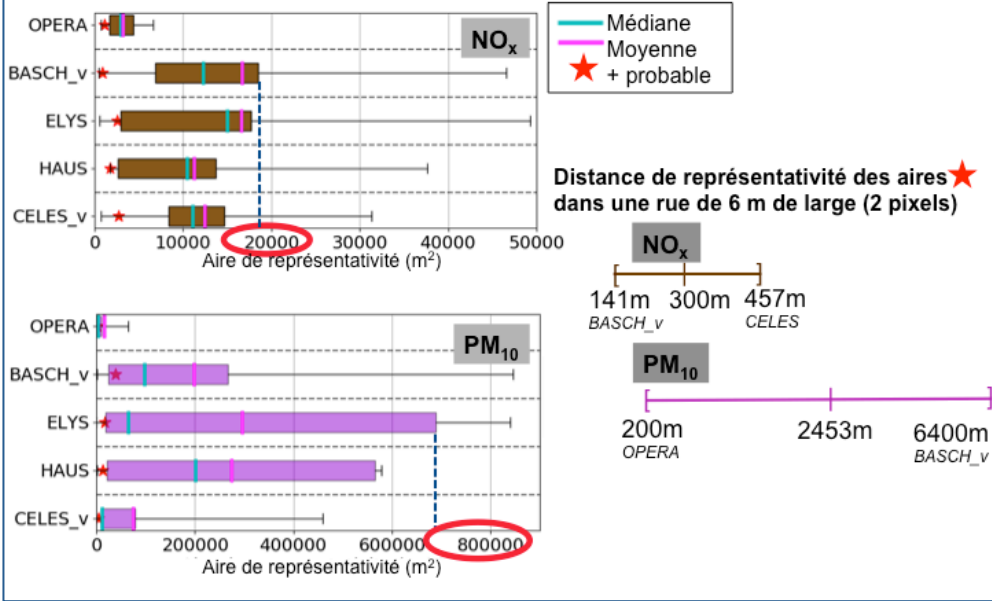
Aires de représentativité station OPERA



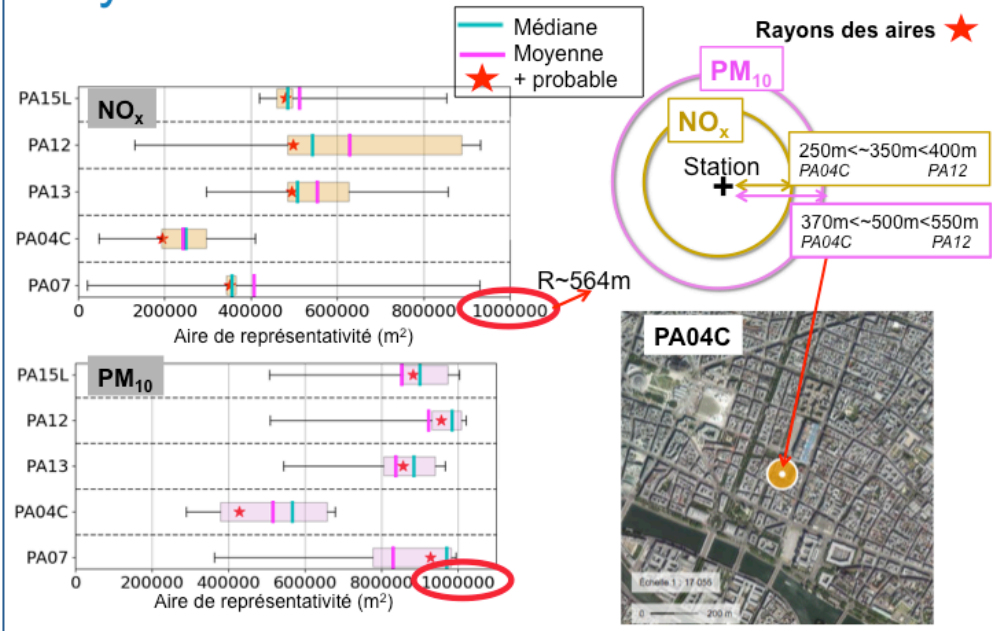
Aires de représentativité station de fond (PA07)



Synthèse sur les stations trafic



Synthèse sur les stations de fond



Bilan sur la représentativité spatiale des stations

- Aire de représentativité la + probable = aire sélectionnée de 7 à 10 jours/10
- Stations urbaines de fond : représentative d'un quartier
Stations de trafic : représentative d'un trottoir, parcelle de route

Comment exploiter ces aires les + probables pour propager au mieux la mesure de la station ?

Partie II. Estimation bayésienne des concentrations de NO_x et de PM_{10} à partir des simulations numériques haute résolution et des mesures AIRPARIF



Atmospheric Environment: X
Available online 17 June 2019, 100038
In Press, Accepted Manuscript



NO_x and PM_{10} Bayesian concentration estimates using high-resolution numerical simulations and ground measurements over Paris, France

Delphy Rodriguez ^{a, R, B}, Eric Parent ^b, Laurence Eymard ^a, Myrto Valari ^c, Sébastien Payan ^a

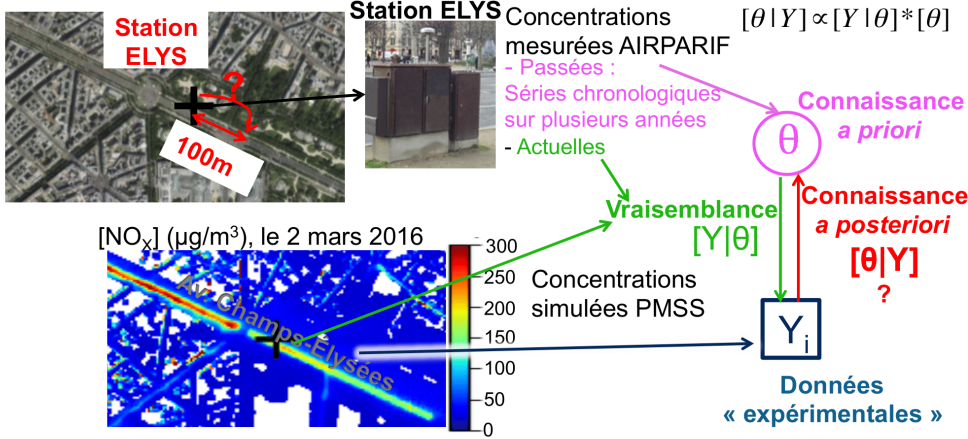
<https://doi.org/10.1016/j.aeaoa.2019.100038>

Approche bayésienne



Pasteur Thomas Bayes
1702-1761

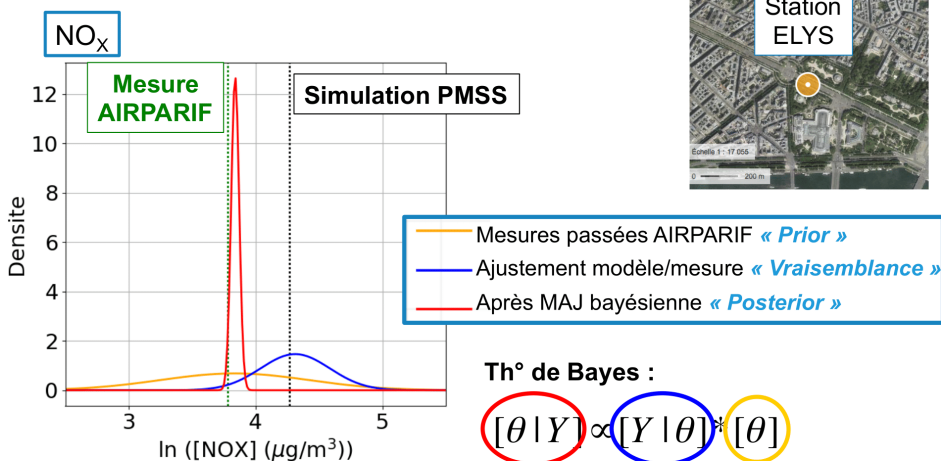
Objectif : propager la mesure de la station AIRPARIF dans sa zone de représentativité



Paramétrisation du modèle au niveau de la station (Prior, Vraisemblance)

Mise à jour bayésienne des concentrations au niveau de la station ELYS

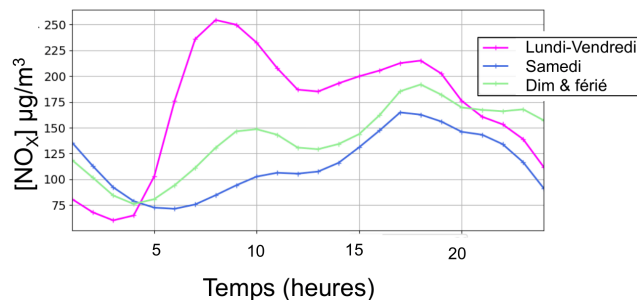
Station avenue des Champs-Élysées, le mardi 8 mars 2016 à 2h



Choix du prior

- Comment classer les mesures passées d’AIRPARIF pour prédire au mieux les concentrations à venir ? Que sait-on de la variabilité du polluant ?
 - tests : trafic et météo
 - **Prior retenu** : par station, heure, type de jours = prior « trafic »
- => Prior idem NO_x et PM₁₀ : méthode unifiée

Moyenne des concentrations a priori de NO_x, station avenue des Champs-Élysées

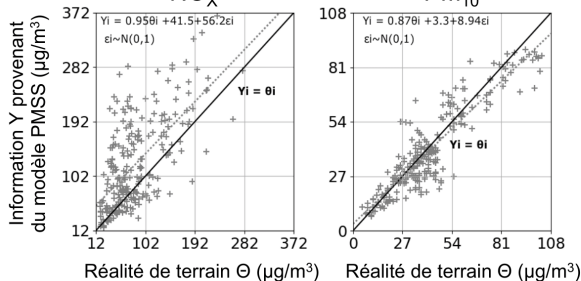


Choix de la vraisemblance

- Quelles sont les sources d’incertitudes du modèle PMSS ?
- Quel modèle probabiliste permet de reproduire au mieux les mesures AIRPARIF à partir des simulations PMSS pour les 10 jours de mars 2016 ?
- $Y_{PMSS} = a * \theta_{Airparif} + b + \text{stderr} * \epsilon$

moyenne écart type

Régressions linéaires sur l’ensemble des données, station avenue des Champs-Élysées

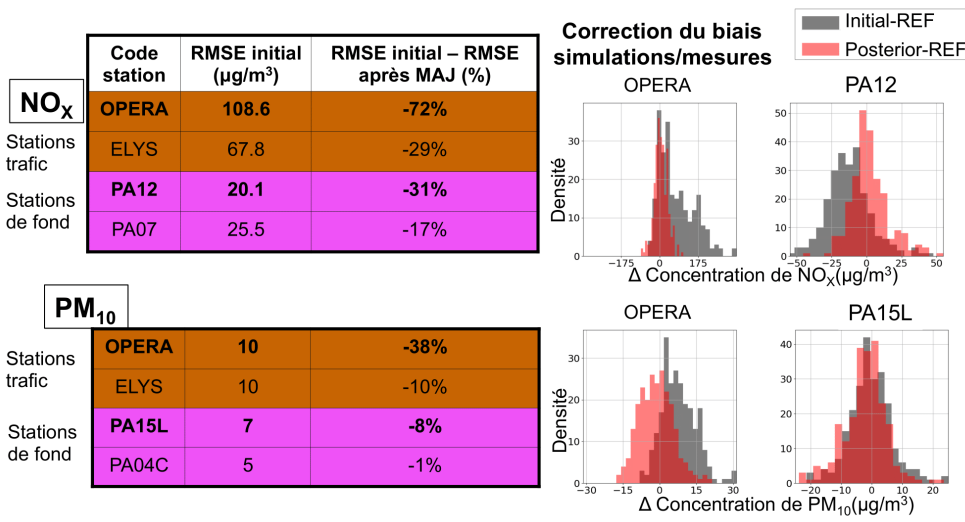


- tests : par jour, tranches de 2h, 3h, 4h, 6h, 8h, 12h pour les 10 jours

- **Vraisemblance retenue** : par station, tranches 2h pour les 10 jours

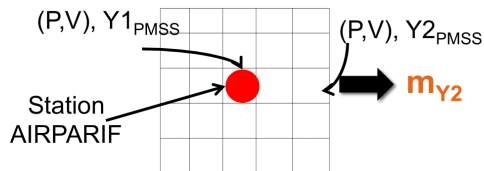
Amélioration des concentrations au niveau des stations AIRPARIF

Modèle optimal : prior « trafic » et vraisemblance par tranches de 2h

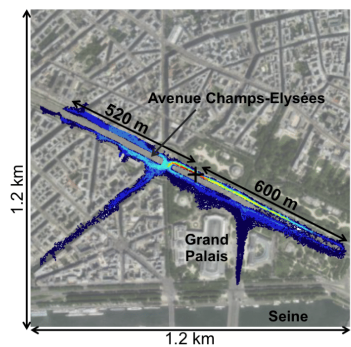


Propagation de la mesure dans les zones de représentativité

- Paramétrisation du modèle bayésien
- P : prior, V : vraisemblance



- Comment intégrer la différence de représentativité entre chaque pixel ?



Équation de pondération

$$C_{finale} = a * m_y + (1-a) * Y_{PMSS}$$

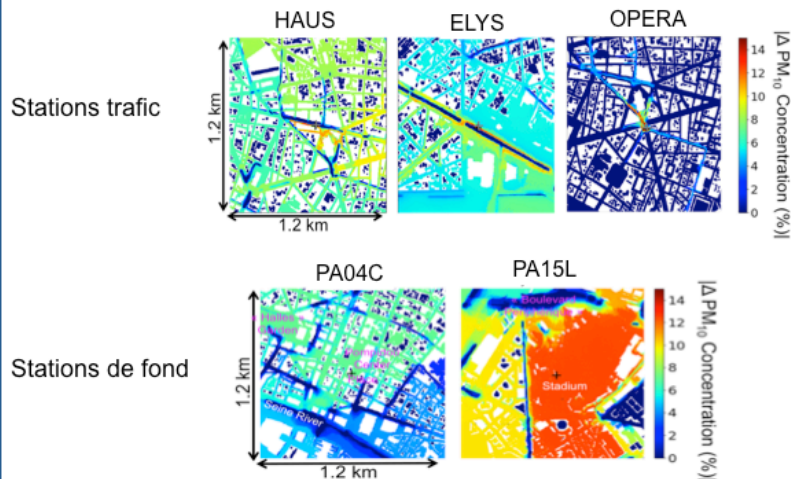
$a = 1$ (top of scale)
 $a = 0.1$ (bottom of scale)

C_{finale} : moyenne des concentrations a posteriori
 m_y : Station AIRPARIF
 Y_{PMSS} : concentration initiale PMSS

a : coefficient de pondération = nombre de jours de représentativité / 10 (=nombre total de jours)

Modification des concentrations avec la MAJ bayésienne – PM₁₀

$|Y_{PM_{10}} - \text{concentration finale}|_{\text{moyennée sur les 10 jours de mars 2016}} (\%)$



Propagation de la mesure dans zone de représentativité la + probable

	Code station	RMSE initial – RMSE après MAJ (%)
Stations trafic	NO_x OPERA	-46%
	ELYS	-20%
Stations de fond	PA15L	-22%
	PA04C	-17%
Stations trafic	PM₁₀ OPERA	-18%
	HAUS	-8%
Stations de fond	PA15L	-10%
	PA04C	-5%

Bilan sur l'estimation bayésienne des concentrations

- Combinaison mesures stations AIRPARIF & simulations PMSS
+ Intégration de la représentativité de la station
- Choix du modèle optimal : prior « trafic » et vraisemblance par tranches de 2h, adapté aux NO_x et PM₁₀
 - ✓ Diminution erreurs modèle/mesures au niveau des stations
 - ✓ Propagation de la mesure AIRPARIF dans la zone de représentativité
- Avantages et limites de l'approche bayésienne :
 - 😊 Encodage des connaissances avec incertitudes associées, travailler avec peu de données, modèle écrit ≠ boîte noire, calcul rapide
 - 😞 Optimisation du modèle prend du temps et nécessite expertise

Conclusion générale



Cartographier la pollution de l'air liée au trafic, le + finement possible (échelle de la rue)

simulations du modèle PMSS (résolution horizontale : 3m) + mesures des stations AIRPARIF



1/ Développement d'une méthode pour déterminer la représentativité des stations AIRPARIF

« similarité » : {varie en phase avec la REF et faible dispersion /REF}
 + zone la plus probable de représentativité

2/ Utilisation du théorème de Bayes pour ajuster le modèle autour de la station

Intégration de la représentativité (pondération) pour propager la mesure au-delà de la station

Travail méthodologique novateur qui nécessite d'être consolidé

Perspectives [1/2]



- **Consolider la méthode avec + de jours simulation**

- *représentativité des stations* :

- (1) tester un autre pas de temps

- (2) analyser la sensibilité de \neq paramètres : météo, typologie et trafic

- (3) ajustement des seuils (NRMSE +/- restrictif selon objectif)

- *application bayésienne* : confirmer équation de pondération et rendre + robuste vraisemblance

- Cartographier les incertitudes associées à ces concentrations

- Consolider la méthode avec un autre modèle urbain (influence du type de modèle et de la résolution /ADMS-Urban avec la collaboration d'Airparif ?)

Perspectives [2/2]




- Aller plus loin...**

- **Déterminer les zones de représentativité en chaque point du modèle**


- (1) optimiser le déploiement des stations AIRPARIF

- (2) étalonner un réseau de capteurs fixes

- **Classer la pollution selon le jour de la semaine et météo**

-  déplacement virtuel d'un individu sur chaque point du modèle

-  analyse + détaillée de l'exposition

-  impact de la pollution sur la santé d'un individu

Contact : delphy.rodriquez@gmail.com

