



HAL
open science

Caractérisation de la pollution de proximité et modélisation multi-agents de l'exposition à la pollution atmosphérique en Ile de France: situation actuelle, scénarios énergétiques et d'aménagement urbain

Inès Makni

► **To cite this version:**

Inès Makni. Caractérisation de la pollution de proximité et modélisation multi-agents de l'exposition à la pollution atmosphérique en Ile de France: situation actuelle, scénarios énergétiques et d'aménagement urbain. Océan, Atmosphère. Université Paris-Est Créteil Val-de-Marne - Paris 12, 2022. Français. NNT: 2022PA120059 . tel-04337129

HAL Id: tel-04337129

<https://theses.hal.science/tel-04337129>

Submitted on 12 Dec 2023

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.



Thèse de doctorat

Préparée à l'Université Paris-Est Créteil

Ecole Doctorale Sciences, Ingénierie et Environnement (ED SIE)

Laboratoire Interuniversitaire des Systèmes Atmosphériques (LISA)

Conseil Départemental du Val-de-Marne

Par **Inès Makni**

Caractérisation de la pollution de proximité et modélisation multi-agents de l'exposition à la pollution atmosphérique en Ile de France : Situation actuelle, scénarios énergétiques et d'aménagement urbain.

Membres du jury

Benjamin HANOUNE (Rapporteur)

Sébastien PAYAN (Rapporteur)

Karine SARTELET (Examinatrice)

Myrto VALARI (Examinatrice)

Patrice COLL (Examineur)

Isabelle COLL (Directrice de thèse)

Dorothee INGERT (Co-encadrante/Invitée)

Remerciements

Je tiens à remercier Isabelle Coll, professeure à l'Université et ma directrice thèse au LISA, qui m'a encadrée tout au long de ces quatre années et qui m'a fait bénéficier de sa grande expérience. Je salue ton professionnalisme et ta disponibilité malgré tes nombreuses responsabilités, mais aussi, ta capacité à traiter toute sorte de situation avec beaucoup de rigueur de pragmatisme et d'ouverture. Je te remercie également pour ta bonne humeur qui m'a toujours motivée. Travailler à tes côtés a été un réel plaisir pour moi.

Merci également à Dorothée Ingert, chargée de projet au Conseil Départemental du Val-de-Marne et co-encadrante de cette thèse, pour m'avoir chaleureusement accueillie et accompagnée au sein de l'équipe du STEG.

Merci à Patrice Coll, directeur du LISA, de m'avoir accueillie au sein de son laboratoire. Merci à Arthur Elessa Etuman pour m'avoir aidée à déboguer des scripts et pour m'avoir renseignée lorsque j'en avais besoin.

Merci à tous mes collègues, Taos, Thomas, Richard, Matthieu, Nicolas et Dilek pour leur gentillesse et leur bonne humeur au quotidien. Merci notamment à Taos, Thomas et Nicolas pour l'ambiance chaleureuse et conviviale qu'on a pu créer au sein du laboratoire. Vous êtes des collègues de bureaux incroyables ! C'était vraiment agréable de passer ces quatre années avec vous.

Merci à tous les permanents de l'équipe modélisation qui ont toujours été bienveillants et adorables tout le long de ma thèse. Un merci spécial à Benoit pour nous avoir permis d'avoir régulièrement une bonne ambiance dans le bureau ! C'était chouette de partager des snacks !

Merci à Arineh et Mathieu d'avoir été aussi de merveilleux collègues de bureaux pendant mon début de thèse. Merci Richard de m'avoir rappelée qu'il y avait Top Chef tous les mercredis soirs ! J'aurai dorénavant, le temps de rattraper tous les épisodes que j'ai ratés.

Merci aussi à tous les doctorants de la MSE que je n'avais, malheureusement pas l'occasion de voir assez souvent, mais avec qui j'ai beaucoup apprécié les pauses midi !

Merci également à toute l'équipe du STEG, au CD94, pour leur bonne humeur et pour les discussions matinales les vendredis matins autour de biscuits et d'un bon thé !

Merci à tous ceux qui m'ont soutenue au cours de cette thèse et que j'aurais oubliés.

Bien évidemment, un énorme merci à mes parents et à mes deux frères de m'avoir soutenue pendant les périodes les plus difficiles de cette thèse. Merci de m'avoir motivée et remontée le moral quand j'en avais besoin. Vous êtes les meilleurs.

Enfin, je ne remercie **pas** la COVID-19 de nous avoir empêché de travailler en présentiel. Les déjeuners dans la salle de convivialité m'ont énormément manqué pendant les périodes de confinements.

Résumé de la thèse

I. Français

Caractérisation de la pollution de proximité et modélisation multi-agents de l'exposition à la pollution atmosphérique en Ile de France : Situation actuelle, scénarios énergétiques et d'aménagement urbain.

Ce projet s'inscrit dans le cadre du développement au LISA d'une compétence transdisciplinaire en modélisation de la qualité de l'air urbain à fine échelle, l'objectif est de développer une chaîne de modélisation produisant et analysant des scénarios d'aménagement urbain et de mobilité des populations. Cette thèse vise à développer une méthodologie de calcul de l'exposition de la population francilienne aux polluants atmosphériques principalement émis par le trafic routier, qui sont connus pour leurs impacts néfastes sur la santé.

Les concentrations ambiantes de dioxyde d'azote et de particules sont modélisées avec le modèle CHIMERE. Ensuite, pour passer des concentrations à des valeurs d'expositions, une approche en deux temps est mise en œuvre : (i) Une approche statistique est développée pour affiner les concentrations simulées d'une échelle kilométrique à l'échelle de l'individu puis, (ii) une approche d'estimation de l'exposition qui prend en compte les habitudes et les mobilités des individus dans les micro-environnements urbains. L'approche statistique se divise en deux parties. D'abord, on procède à la correction de l'erreur observée entre la modélisation et la mesure par apprentissage sur un jeu de données. L'équation de correction est ensuite généralisée en s'appuyant sur les caractéristiques des sites de mesure proches du trafic. Ensuite, on valide l'approche en appliquant l'équation ainsi développée à un jeu de données de validation. La modélisation des expositions individuelles considère les habitudes et la mobilité des individus simulés par le modèle d'émission OLYMPUS. Les expositions sont estimées en fonction du mode de transport utilisé pour leurs déplacements ou les concentrations ambiantes des micro-environnements bâtis que les individus visitent, et selon le temps passé dans chaque micro-environnement.

L'objectif final est de pouvoir proposer des diagnostics sur l'exposition à la pollution, en fonction de l'organisation urbaine, de l'offre de transport et des pratiques de mobilité et de consommation d'énergie.

Mots-clés : Qualité de l'air, Evaluation statistique, Micro-environnement, Système multi-agent, Exposition, Exposition dynamique

II. Anglais

Characterization of near proximity pollution and multi-agent modeling of the atmospheric pollution in Ile de France area: current situation, energetic scenario and urban planning.

This project is part of the development at LISA of a transdisciplinary competence in urban air quality modelling at a fine scale. The objective is to develop a modelling chain producing and analyzing urban development and population mobility scenarios. This thesis aims to develop a methodology for calculating the exposure of the Ile-de-France population to air pollutants mainly emitted by road traffic, which are known to have harmful impacts on health.

Ambient concentrations of nitrogen dioxide and particulate matter are modeled with the CHIMERE model. Then, to pass from concentrations to exposure values, a two-step approach is implemented: (i) a statistical approach is developed to refine the simulated concentrations from a kilometer scale to the individual scale, then, (ii) an exposure estimation approach that takes into account the habits and mobilities of individuals in urban micro-environments. The statistical approach is divided into two parts. First, the observed error between modeling and measurement is corrected by learning on a data set. The correction equation is then generalized based on the characteristics of the measurement sites close to the traffic. Then, the approach is validated by applying the developed equation to a validation dataset. The modeling of individual exposures considers the habits and mobility of individuals simulated by the OLYMPUS emission model. Exposures are estimated according to the mode of transportation used for their travel or the ambient concentrations of the built micro-environments that individuals visit, and according to the time spent in each micro-environment.

The final objective is to be able to propose diagnoses of exposure to pollution as a function of urban organization, transportation supply and mobility and energy consumption practices.

Keywords: Air quality, Statistical evaluation, Micro-environment, Multi-agent system, Exposure, Dynamic exposure

Table des matières

Table des matières

Remerciements.....	3
Résumé de la thèse.....	5
Table des matières.....	7
Table des Figures et des Tableaux	8
Introduction	15
Chapitre I - Etat de l'art.....	17
Chapitre II - Augmentation de la résolution spatiale de simulations de la qualité de l'air : Développement d'une approche de descente en échelle statistique	31
Chapitre III - Estimation de l'exposition spatio-temporelle de la population : Prise en compte de la mobilité et des activités quotidienne de la population.....	101
Chapitre IV - Mise en place et analyse d'un scénario d'aménagement urbain – Etude d'impact sur la qualité de l'air de l'implantation de Grand Paris Express dans le quartier de Créteil l'Échat	141
Conclusion générale	156
Références	162
ANNEXE 1.....	180
ANNEXE 2.....	182
ANNEXE 3.....	186
ANNEXE 4.....	218
ANNEXE 5.....	235
ANNEXE 6.....	236
ANNEXE 7.....	240
ANNEXE 8.....	271
ANNEXE 9.....	296
ANNEXE 10.....	300

Table des Figures et des Tableaux

I. Table des illustrations

Figure 1 - Visualisation des 3 domaines imbriqués : en rouge le domaine EUROPE, en vert le domaine GRANDNORD3 et en bleu le domaine IDF1.	35
Figure 2 (a, b, c, d) - Comparaison des séries temporelles mesurées par le réseau Airparif (en rouge) et simulées par CHIMERE (en gris) sur des sites de mesures de fond urbain / périurbain, pour NO ₂ (en haut) et pour les PM ₁₀ (en bas). Les concentrations sont exprimées en µg/m ³	44
Figure 3 - Comparaison des moyennes glissantes (sur 12 et 24h) et des percentiles (20, 25 et 30) utilisées pour reproduire le bruit de fond du signal modélisé par CHIMERE au niveau de la station A1 pour NO ₂ et en 2009.	46
Figure 4 - Histogramme représentant les RMSE (en haut) et moyennes annuelles (en bas) de NO ₂ (en µg/m ³) sur chaque site telles que simulées par CHIMERE (beige hachuré), corrigées à partir des percentiles (en vert), corrigées à partir des moyennes glissantes (tons beiges) et issues de la mesure (rouge hachuré), en 2009.	49
Figure 5 - Histogramme représentant les RMSE (en haut) et moyennes annuelles (en bas) de PM ₁₀ (en µg/m ³) sur chaque site telles que simulées par CHIMERE (beige hachuré), corrigées à partir des percentiles (en vert), corrigées à partir des moyennes glissantes (tons beiges) et issues de la mesure (rouge hachuré), en 2009.	50
Figure 6 - Séries temporelles (i) CHIMERE en gris, (ii) <i>Cprox</i> en vert et (iii) issue de la mesure en rouge, sur la station trafic de l'Autoroute A1 pour NO ₂ en haut et pour la station de la place Victor BASCH des PM ₁₀ en bas, pour différentes parties de l'année 2009.	52
Figure 7 - Séries temporelles de concentration du NO ₂ (en µg/m ³) à la station de proximité au trafic des Champs-Élysées, sur l'année 2009. Les courbes sont celles de CHIMERE (en gris), de <i>Cprox</i> (en vert) et de la mesure (en rouge). Les parties surlignées en jaune montrent les périodes de très grande sous-estimation de <i>Cprox</i>	53
Figure 8 – Coïncidence entre les périodes de mauvaise restitution des concentrations (ici de NO ₂ , en µg/m ³) par CHIMERE et par <i>Cprox</i> , et la vitesse du vent issue de WRF, sur le site de trafic des Champs-Élysées pendant le mois de mars 2009.	54
Figure 9 - Représentation graphique des valeurs prises par le facteur CV en fonction des vitesses de vent (m/s).	56
Figure 10 - Comparaison des séries temporelles <i>Cprox</i> sans CV (en haut) et avec CV (en bas) par rapport à CHIMERE (en gris) et à la mesure (en rouge), près de l'autoroute A1 pour NO ₂ (à gauche) et sur le site Victor Basch pour les PM ₁₀ (à droite).	57
Figure 11.a - Comparaison des séries temporelles de CHIMERE (en gris), de <i>Cprox</i> (en vert) et de la mesure (en rouge) pour NO ₂ , en µg/m ³ , sur des périodes d'environ 1 mois durant l'année 2009. A droite des séries temporelles, des boîtes à moustache (même code couleur) montrent les minima, maxima, percentiles 25 et 75, moyenne et médiane de chaque série de données sur l'ensemble de l'année 2009.	60
Figure 11.b - Comparaison des séries temporelles de CHIMERE (en gris), de <i>Cprox</i> (en vert) et de la mesure (en rouge) pour NO ₂ , en µg/m ³ , sur des périodes d'environ 1 mois durant l'année 2009. A droite des séries temporelles, des boîtes à moustache (même code couleur) montrent les percentiles 5 et 95	

ainsi que les percentiles 25 et 75, les moyenne et médiane de chaque série de données sur l'ensemble de l'année 2009.	61
Figure 12 - Comparaison des variances pour les sites des jeux d'apprentissage de NO ₂ (en haut) et PM ₁₀ (en bas) pour l'année 2009. Les variances issues des mesures sont représentées en rouge, celles issues de <i>Cprox</i> fité sont en vert et celles issues de CHIMERE sont en gris.	62
Figure 13 - Rapport des coefficients α et β en pourcentage pour les stations de trafic et de fond des jeux d'apprentissage du NO ₂ et des PM ₁₀ pour l'année 2009.	64
Figure 14 - Impact sur la courbe de <i>Cprox</i> de la partition alpha, beta, sur le site de trafic d'Auteuil, pour NO ₂ sur une partie de l'année 2009. La courbe <i>Cprox</i> est obtenue avec alpha = 0,77 et la courbe <i>Cprox</i> moyen avec alpha = 0,57.	64
Figure 15 - Valeurs de k issues de l'apprentissage, sur les différents sous-groupes de mesure pour le dioxyde d'azote (couleurs claires) et pour les particules PM ₁₀ (couleurs foncées) : en fond (péri-)urbain (en vert), et en proximité au trafic (en orange).	65
Figure 16 - Impact de la moyenne des valeurs de k issues de l'apprentissage, sur les courbes de <i>Cprox</i> de particules fines en fond périurbain et urbain. En haut, courbes de CHIMERE en gris, de la mesure en rouge et de <i>Cprox</i> (issu de l'apprentissage en vert et moyen en bleu) sur deux sites d'environnement différents : Melun en haut et Vitry-sur-Seine en bas.	66
Figure 17 - Impact de la moyenne des valeurs de k issues de l'apprentissage, sur les courbes de <i>Cprox</i> de dioxyde d'azote en proximité au trafic : courbes de CHIMERE en gris, de la mesure en rouge et de <i>Cprox</i> (issu de l'apprentissage en vert et moyen en bleu) sur le site Célestins (quais de la Seine).	66
Figure 18 - Représentation cartographique des LCZ de la région Île-de-France.	69
Figure 19 - Corrélations entre le paramètre HE et les coefficients k des stations de trafic de NO ₂ (en haut) et de PM ₁₀ (en bas) pour 2009.	73
Figure 20.a - Boîtes à moustache de NO ₂ pour la mesure (rouge), CHIMERE (gris) et <i>Cprox</i> obtenue par l'approche automatisée (vert) - Elles donnent les percentiles 5 et 95, les percentiles 25 et 75 ainsi que les moyenne et médiane de chaque série de données sur l'ensemble de l'année 2009.	76
Figure 20.b - Boîtes à moustache de PM ₁₀ pour la mesure (rouge), CHIMERE (gris) et <i>Cprox</i> obtenue par la correction automatisée (vert) - Elles donnent les percentiles 5 et 95, les percentiles 25 et 75 ainsi que les moyenne et médiane de chaque série de données sur l'ensemble de l'année 2009.	76
Figure 21 - Comparaison des variances pour les sites des jeux de validation de NO ₂ (en haut) et PM ₁₀ (en bas) pour l'année 2009. RN2 est le site de trafic de validation, les autres sites sont des sites de fond urbain. Les variances issues des mesures sont représentées en rouge, celles issues de <i>Cprox</i> automatisée sont en vert et celles issues de CHIMERE sont en rose.	77
Figure 22 - Séries temporelles de CHIMERE (en gris), de la mesure (en rouge) et de <i>Cprox</i> (en vert) sur la station de trafic RN2 du jeu de validation RN2, sur une période représentative de l'année 2009. En haut, résultats pour NO ₂ et mise en évidence des périodes de l'année surestimées par l'approche de correction <i>Cprox</i> . En bas, résultats pour les PM ₁₀	78
Figure 23 - Comparaison des niveaux moyens corrigés (en vert) par l'étape d'universalisation de <i>Cprox</i> avec les niveaux mesurés (en rouge) et ceux de CHIMERE (en gris), pour NO ₂ et sur les sites périurbains de Gonesse et de Melun.	80
Figure 24 - Séries temporelles de CHIMERE (en gris), de la mesure (en rouge) et de <i>Cprox</i> (en vert) sur différentes stations de mesure des PM ₁₀ du jeu de validation, sur l'année 2009.	81
Figure 25 - Comparaison des cartographies annuelles des sorties de <i>Cprox</i> (en haut) et des bilans annuels d'Airparif (en bas) pour NO ₂ (à gauche) et pour les PM ₁₀ (à droite) pour l'année 2009.	83

Figure 26 - Gains obtenus sur la RMSE (RMSE CHIMERE - RMSE <i>Cprox</i>) pour 2017 (en bleu foncé) et 2009 (en bleu clair) pour NO ₂ (à gauche) et les PM ₁₀ (à droite).	88
Figure 27 - Comparaison des variances pour les sites d'apprentissage de NO ₂ (à gauche) et PM ₁₀ (à droite) pour l'année 2017. Les variances issues des mesures sont représentées en rouge, celles issues de <i>Cprox</i> sont en vert et celles issues de CHIMERE sont en beige.	88
Figure 28- Comparaison des moyennes annuelles calculées sur les sites d'apprentissage de trafic et de fond urbain, pour NO ₂ et pour les PM ₁₀ pour l'année 2017. Les moyennes issues des mesures sont représentées en rouge, celles issues de <i>Cprox</i> sont en vert et celles issues de CHIMERE sont en gris.89	89
Figure 29 - Séries temporelles sur le site de trafic Porte d'Auteuil pour NO ₂ (en haut) et pour les PM ₁₀ (en bas) pour l'année 2017. Les deux graphiques du centre sont des zooms issus du graphique du haut, sur des périodes plus courtes. Les courbes de la mesure sont représentées en rouge, celles issues de <i>Cprox</i> sont en vert et celles issues de CHIMERE sont en gris.	90
Figure 30 - Contribution des coefficients α et β à k , en pourcentage, pour les stations de trafic (à gauche) et de fond (à droite) des jeux d'apprentissage du NO ₂ (en haut) et des PM ₁₀ (en bas) pour l'année 2017.	92
Figure 31 - Corrélations obtenues entre le paramètre <i>HE</i> et les coefficients k des stations de trafic de NO ₂ (en haut) et de PM ₁₀ (en bas) pour 2017.....	93
Figure 32 - Comparaison des moyennes annuelles calculées sur les sites d'apprentissage et de validation (encadrés en bleu) en proximité au trafic, pour NO ₂ (à gauche) et pour les PM ₁₀ (à droite) pour l'année 2017. Les moyennes issues des mesures sont représentées en rouge, celles issues de <i>Cprox</i> fité sont en vert plein, celles issues de <i>Cprox</i> automatisé sont en vert pointillé et celles issues de CHIMERE sont en gris.	94
Figure 33 - Séries temporelles sur deux sites de validation de fond urbain, pour NO ₂ pour l'année 2017. Les courbes de la mesure sont représentées en rouge, celles issues de <i>Cprox</i> sont en vert et celles issues de CHIMERE sont en gris.	95
Figure 34 - Confrontation des moyennes annuelles obtenues avec les sorties du modèle HOR'AIR d'Airparif pour NO ₂ (en haut) et pour les PM ₁₀ (en bas). Les valeurs de CHIMERE sont représentées en gris, celles de la correction issue de l'apprentissage en vert, celles de la correction automatisée en vert pointillé, celles de la mesure en rouge et celles d'HOR'AIR en bleu.	96
Figure 35 - Comparaison des cartographies annuelles des sorties de <i>Cprox</i> (en haut) et des bilans annuels d'Airparif (en bas) pour NO ₂ (à gauche) et les PM ₁₀ (à droite) pour l'année 2017.	98
Figure 36 - Représentation schématique de la plateforme de modélisation du projet POLL-EXPO. Les noms des modèles sont indiqués dans les encadrés gris. En bleu figure le modèle auquel j'ai apporté des modifications et en orange figurent les parties de la plateforme que j'ai développées.	103
Figure 37 - Représentation schématique de l'organisation et des rôles des modules d'OLYMPUS. L'encadré orange indique la partie du code sur laquelle je suis intervenue. Source : Elessa Etuman & Coll, 2018.....	105
Figure 38 - Organisation des fichiers GTFS disponibles pour le réseau de transport en commun de l'Île-de-France.....	107
Figure 39 - Représentation schématique d'une correspondance entre deux arrêts de deux lignes de transport (entourés en rouge). La zone empruntée à pied par les agents est représentée par des pointillés.	108
Figure 40 - Schéma des paramètres considérés pour les calculs des temps de déplacement entre deux stations d'une même ligne de transport.....	110

Figure 41 - Représentation du réseau de transport en commun ajouté au modèle OLYMPUS. Les stations de transports en commun sont illustrées sous forme de point rose et les brins reliant les stations d'une même ligne sont illustrés sous forme de lignes grises.	111
Figure 42 (a, b) - Confrontation visuelle des itinéraires trouvés pour le réseau GTFS (en gris) et Google Maps (en couleur).....	113
Figure 42 (c, d) - Confrontation visuelle des itinéraires trouvés pour le réseau GTFS (en gris) et Google Maps (en couleur).....	114
Figure 43 - Organigramme du script d'estimation des expositions.....	116
Figure 44 - Comparaison des expositions*temps, estimées pour les trajets en bus en considérant une concentration ambiante de 160 $\mu\text{g}/\text{m}^3$ (à gauche) et de 26 $\mu\text{g}/\text{m}^3$ (à droite) pour NO_2	120
Figure 45 - Distribution des horaires de déplacement des individus de l'échantillon de population, tout mode de transport confondu.....	125
Figure 46 - Distribution des choix modaux en fonction des durées de déplacement pour les individus de l'échantillon de population.	125
Figure 47 - Boxplots des expositions totales sur les durées de déplacements (à gauche) et d'activités (à droite) pour NO_2 (en haut) et les PM_{10} (en bas).....	127
Figure 48 - Comparaison des distributions des durées de « Visite » (à gauche) et de « Résidence » (à droite).	128
Figure 49 - En haut, cartographie des expositions dynamiques cumulées moyennes sur la journée des individus au niveau de leur lieu de résidence pour le NO_2 . En bas, comparaison des cumuls d'exposition des individus de Créteil et d'Etampes avec les concentrations CHIMERE simulées au niveau des deux communes pour NO_2	130
Figure 50 - En haut, cartographie des expositions dynamiques cumulées moyennes sur la journée des individus au niveau de leur lieu de résidence pour les PM_{10} . En bas, comparaison des cumuls d'exposition des individus de Créteil et d'Etampes avec les concentrations CHIMERE simulées au niveau des deux communes pour PM_{10} . Tout en bas, illustration du décalage de journée d'un individu.	131
Figure 51 - Cumuls d'expositions dynamiques sur la journée pour 5 individus possédant de longs temps de transport. Le cumul d'exposition à NO_2 est présenté en vert pointillé, celui aux PM_{10} en tirets rouges. Les activités et déplacements sont présentés en couleur, en plein pour la mobilité et en hachuré pour les activités sur un site donné.....	132
Figure 52 - Réseau routier généré par le LVMT sur le domaine Ile-de-France (A), zoom sur un exemple de traversée des zones d'habitation par les tronçons de route (B) et réseau routier généré grâce à OSMNX (C). En dessous, zoom sur une partie de la ville de Paris où le bâti est représenté en marron, les tronçons de route du réseau LVMT sont représentés en rouge et ceux du réseau OSMnx sont représentés en noir.	135
Figure 53 - Représentation schématique de la simplification du réseau OSMnx.....	135
Figure 54 - Réseau OSMnx avec des VMAs manquantes au niveau de la ville de Paris (A) et complétées sur l'ensemble du réseau (B).	138
Figure 55 - À gauche, vue satellite du quartier de Créteil l'Échat. À droite, délimitation du quartier de gare de Créteil l'Échat. Source : Apur (SGP).....	143
Figure 56 - Cartographie des surfaces du quartier de gare (marron), de sa voirie (issue d'OSMnx en orange) et de l'emplacement de la gare (étoile rouge). Les mailles CHIMERE chevauchées par le quartier de gare sont représentées en hachuré dans le fond.	145

Figure 57 - À gauche, ajout des trois nouveaux tronçons de route au réseau LVMT (en rouge). À droite, ajout des trois nouveaux tronçons (en rouge plein) et suppression de deux tronçons du réseau OSMnx (en rouge pointillé).	147
Figure 58 - Différences entre les niveaux de fond simulés pour la situation de référence et pour le scénario Val-de-Marne pour NO ₂ (en haut) et les PM ₁₀ (en bas), et calculées dans le sens [Référence - Scénario]. Elles sont présentées pour les sorties brutes de CHIMERE (en gris) et pour les sorties corrigées par l'approche statistique <i>Cprox</i> (en vert).....	150
Figure 59 - Cartographies des différences de concentrations de NO ₂ (en haut) et des PM ₁₀ (en bas) entre la situation de référence et le scénario le long des routes du quartier de Créteil l'Échat.	151

II. Liste des tableaux

Tableau 1 - Détails des domaines CHIMERE - NX et NY représentent les nombres de mailles en longitude et latitude, DX et DY la dimension (en degrés) des mailles dans ces deux directions, et XMIN et YMIN les coordonnées (WGS84) du point sud-ouest de chaque maillage.....	35
Tableau 2 - Table des correspondances entre les catégories d'émetteurs de CHIMERE (colonne de gauche) et celles de l'inventaire francilien (colonne de droite).	38
Tableau 3 - Détails des domaines WRF - NX et NY décrivent le nombre de mailles en ouest-est et sud-nord ; DX et DY décrivent la résolution spatiale ouest-est et sud-nord (en m) des mailles ; Ref_lat et Ref_lon donnent les coordonnées du point central du domaine considéré ; Proj indique le système de projection considéré.....	39
Tableau 4 - Répartition des stations de mesure de NO ₂ et de PM ₁₀ en deux jeux d'apprentissage et de validation. Les noms des sites de mesure ayant été installés entre les années 2009 et 2017 sont écrits en bleu.	47
Tableau 5 – Résumé de la caractérisation des IMU en LCZ.....	69
Tableau 6 - Caractérisation des stations d'apprentissage de trafic en fonction des LCZ dans lesquelles elles se trouvent.	70
Tableau 7 - Données relatives aux comptages routiers (mesurés et modélisés par OLYMPUS) ainsi que les indicateurs d'émissions calculés au niveau des tronçons (TMJA pondérés), indicateurs moyens d'émission dans la maille (TMJA pondérés mailles), les vitesses de déplacement par tronçons (Vitesse) et les valeurs obtenues pour les Ratio HE de NO ₂ et des PM ₁₀	72
Tableau 8 - Résumé des données statistiques de la station de validation de trafic de NO ₂ et des PM ₁₀ (automatisation de l'approche et comparaison avec les données idéales « fit » de la station).....	79
<i>Tableau 9 - Évolution des émissions de NOx, PM₁₀, PM_{2,5}, SO₂ et COVNM entre 2008 et 2015 pour chaque secteur d'activité.</i>	<i>87</i>
Tableau 10 - Confrontation des durées d'itinéraires calculées via mon réseau GTFS avec les informations données par Google Maps et le nombre de correspondances indiqué entre parenthèses pour chacun des deux.....	112
Tableau 11 – Valeurs moyennes de concentrations associées à chaque type de mode de transport collectif.	121
Tableau 12 – Valeurs moyennes de concentrations associées à chaque type d'activité en µg/m ³ avec l'information sur les écarts-types entre parenthèses (quand celle-ci a été calculée).....	124

Tableau 13 - Récapitulatif des valeurs limites utilisées pour compléter les données du réseau OSMNX en fonction du type de route des tronçons.	136
Tableau 14 - <i>Moyenne des VMA pour les tronçons correspondant à plusieurs types de route.</i>	137
Tableau 15 – Parts modales décrites dans le Plan de Déplacement du Val-de-Marne (en %).....	146
Tableau 16 – Répartition modale (en %) de la population simulée pour la situation de référence et pour le scénario Val-de-Marne sur l’ensemble de l’Île-de-France et sur le Val-de-Marne.....	148

Introduction

La question de la qualité de l'air urbaine a commencé à être considérée sous le second empire avec Eugène Haussmann, préfet de Seine. Il œuvra le long de sa carrière pour l'amélioration et la modernisation des infrastructures routières afin de faciliter les déplacements de la population et des marchandises, mais il considéra aussi la circulation de l'air et de l'eau en ville pour faire de Paris une capitale moderne et propre. Il planifia, pour cela, l'aménagement de plusieurs parcs et jardins dans le but d'améliorer la qualité de l'air et remodela la ville pour élargir les rues, jusque-là, très étroites.

Aujourd'hui, après plusieurs décennies centrées sur la réduction des émissions de polluants atmosphériques par l'amélioration des technologies (de combustion notamment), les pistes pour continuer à améliorer la qualité de l'air se réorientent en grande partie vers la modification des comportements de consommation d'énergie, et les changements de mode de vie, des individus habitant l'espace urbain. En effet, depuis les années 2000, le constat que la maîtrise des émissions ne peut plus uniquement s'appuyer sur les avancées technologiques est de plus en plus partagé ([Batkin & Godard, 2019](#)).

La principale difficulté de la mise en place de telles politiques publiques, basées sur le réaménagement du territoire ou l'incitation à la modification des pratiques, est de pouvoir anticiper la manière dont les individus vont s'approprier ou non les changements proposés, et le signe ou l'ampleur des impacts que l'on peut espérer, sur les émissions de polluants et la qualité de l'air. Pour appréhender l'ensemble de manière systémique, il est nécessaire de croiser les différentes disciplines entrant en jeu dans l'élaboration d'une ville harmonieuse et durable : sciences de l'atmosphère, sciences et géographie de la santé, urbanisme et sociologie ainsi que les expertises menées sur les habitudes de mobilité, de consommation énergétique et concernant les politiques publiques ([IAU & ORS, 2014](#)). À la fin des années 2010, il est devenu clair qu'il fallait créer de nouveaux outils qui puissent des réflexions et des diagnostics plus transversaux et, donc, moins spécifiques à une communauté d'acteurs ou à une discipline. Les outils de modélisation existant jusqu'à présent, et dédiés à des domaines d'études précis (modélisation du trafic, des besoins en énergies, de la qualité de l'air, etc.) doivent communiquer autour d'une même question afin que le résultat puisse prendre en compte les connexions et les impacts de chacun des domaines d'études sur les autres. C'est donc un vrai défi que de pouvoir prendre en compte la complexité de la réalité lors des simulations surtout lorsque le facteur humain (choix et pratiques des individus face à une palette de possibilités) prend une part importante. Cette part influe sur l'émission de polluants via la réalisation d'activités et la mobilité des personnes, elle interagit avec la structure de l'espace urbain qui va contraindre l'offre et les choix de mobilité, sur la consommation énergétique via les choix de chauffage et/ou de climatisation des logements et bureaux.

L'objectif global de telles démarches est la production de diagnostics et bilans de la qualité de l'air dans des situations prospectives ou hypothétiques, afin d'accompagner l'action publique. Mon travail de thèse s'inscrit justement dans ce type de démarche et s'articule autour des travaux de mon équipe au laboratoire, qui consistent à développer une chaîne de modélisation transdisciplinaire et d'y apporter de nouvelles expertises et fonctionnalités.

Chapitre I

Etat de l'art

I. Lutte pour une meilleure maîtrise de la qualité de l'air : Passage en revue de ces dernières décennies

La problématique principale des politiques environnementales, jusqu'aux années 1980, concernait les émissions de fumées noires et de dioxyde de soufre car la pollution était essentiellement d'origine industrielle. Les années 1980 marquent un tournant avec l'émergence du trafic routier comme source majeure de la pollution du fait d'une utilisation plus massive de la voiture, et le déclin des émissions industrielles dû **aux efforts sur la qualité des combustibles et sur les méthodes de filtration et de captation des rejets**.

Il n'existait pas à cette période de mission publique d'information et de communication de l'état vers les populations sur la qualité de l'air, et les travaux sur ses impacts sanitaires étaient émergents. Pour ne pas compromettre l'accès de tous au véhicule particulier et ne pas brider le progrès social que représentait la possession d'un véhicule, les pouvoirs publics ont donc préféré miser sur la production de voitures de moins en moins polluantes sans poser de contraintes sur notre manière de l'utiliser (mise en place des normes de durcissement « Euro » des valeurs limites d'émissions pour les véhicules à moteur après la directive 91/441 en 1993 par exemple). La réglementation relative à la mobilité était ainsi liée à l'aménagement des territoires (régions/départements) et à la création de réseaux capables de faire fonctionner la ville (Loi d'Orientation des Transports Intérieurs, 1982)¹. Bien que la contribution du trafic routier à la pollution en ville soit un fait connu des scientifiques et des acteurs politiques, la communauté scientifique travaillant sur le sujet était assez réduite et l'action réglementaire s'appuyait – comme ce fut le cas pour l'industrie – uniquement sur des efforts technologiques d'amélioration de la qualité des combustions et de filtration des échappements (Bonnafoous & Brun, 1999). De cette démarche, essentielle, sont nées à partir des années 1990 les réglementations européennes cadrant les émissions de polluants par les véhicules à moteurs. Ces réglementations, appelées « normes Euro », imposaient un rythme d'amélioration des technologies et de la qualité des carburants en conditionnant les logiques de production des automobiles, pour chaque pays et avec des objectifs que chacun peut tenir. L'instauration de ces normes a montré son efficacité en inversant les tendances sur les émissions dès les années 1990 pour les oxydes d'azote et les particules (Serveau et al., 2015) malgré un trafic encore croissant, et contribue aujourd'hui encore, grâce à des véhicules de moins en moins polluants, à limiter les impacts sanitaires liés aux émissions du trafic routier urbain en France (Nedelec et al., 2010).

A partir de 1996, la LAURE² impose le développement de la surveillance de la qualité de l'air sur l'ensemble du territoire, par des approches normalisées et contraintes en termes d'incertitudes, qui permettent d'avoir un suivi précis de l'évolution de la composition de l'air dans les espaces habités. Elle permet l'instauration de nouveaux modes d'action régionaux, tels que les plans de protection de l'atmosphère et donne une dimension environnementale aux plans de déplacements urbains mis en place dans les grandes agglomérations avec la loi LOTI. L'évolution vers une approche plus compréhensive de l'action publique se fait notamment au travers de la directive sur les plafonds d'émissions nationaux, qui met en cohérence les normes de qualité de l'air et les objectifs de réduction des émissions. Dans ce contexte favorable à l'action locale, des mesures ponctuelles de lutte contre les

¹ <https://www.legifrance.gouv.fr/loda/id/LEGITEXT000006068730/>

² <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000381337/>

épisodes de pollution atmosphérique ont pu pour la première fois être mis en place. L'objectif était de renforcer localement le contrôle des émissions routières ou industrielles pour prévenir ou limiter l'ampleur d'un épisode en cours. Mais l'expérience a montré qu'il était difficile d'instaurer des actions rapides et efficaces, même avec l'aide des outils de modélisation pour anticiper l'évolution de la qualité de l'air. Il est alors devenu impératif d'avoir une réflexion sur la manière dont on consomme l'énergie et émet au quotidien dans les zones urbanisées.

Car les efforts arrivent toutefois en butée sur plusieurs points. Les phénomènes de congestion urbaine par exemple, découlant de l'importance de la place prise par la voiture dans le quotidien de la plupart des européens, ont dans un premier temps mené à la construction de nouvelles infrastructures routières plus larges permettant de faciliter les déplacements. Ces grandes infrastructures, quotidiennement congestionnées dans les grandes métropoles, continuent à poser des problèmes de gestion du trafic mais aussi d'aggravation de la pollution car les conditions de circulation associées et la densité des véhicules accentuent localement les émissions. Même si les véhicules sont moins émetteurs en pollution, l'intensité de l'activité routière actuelle contrebalance donc les bénéfices apportés par les normes Euro. Aujourd'hui, les politiques publiques se sont redirigées vers des actions de régulation pérenne des émissions urbaines en instaurant, par exemple, des Zones à Faibles Émissions (ZFE) instaurant une circulation choisie dans les centres urbains. Comme il en existe plus de 200 en Europe aujourd'hui, la ville de Paris s'est dotée en 2015 d'une ZFE basée sur des vignettes « Crit'air » permettant à la fois la sélection au quotidien des véhicules sur leurs caractéristiques d'émission, et la restriction des accès en cas d'épisode de pollution. En parallèle, les régions, départements, collectivités travaillent à promouvoir, par la réorganisation des transports ou l'aménagement urbain, l'utilisation de modes de transport alternatifs à la voiture. C'est dans ce cadre que le réseau du Grand-Paris Express a été mis en place. Son objectif est de faciliter les mobilités entre les banlieues sans avoir à passer par le centre de Paris et tout en permettant, en parallèle, d'augmenter l'offre de logement de la petite couronne.

Pour tendre vers une efficacité maximale des actions menées sur les différents leviers de la qualité de l'air, il faut pouvoir en proposer une évaluation *ex-ante*, c'est-à-dire anticiper leurs impacts sur la qualité de l'air. C'est seulement dans cette démarche que les transformations urbaines et les changements des pratiques sur le long terme pourront être amorcés avec confiance. Et que les problématiques d'exposition ponctuelle et chronique - qui existent dans les villes même lorsque les seuils réglementaires de pollution atmosphériques sont respectés - pourront être mises au défi.

Les questions qu'on doit se poser concernent les ambitions de ce type de projet et des schémas directeurs régionaux sont liés à leur capacité à accélérer l'amélioration de la qualité de l'air en ville, notamment par une incitation réussie au changement des comportements de la population. Les paramètres qui doivent être évalués sont bien sûr la composition de l'atmosphère, mais aussi l'exposition des individus, car cela touche aux questions de santé publique. Il faut pour cela pouvoir comprendre ce qu'est « l'exposition en ville », et savoir la quantifier. La notion d'exposition est en effet plurielle dans les grands centres urbains. Les personnes habitant et travaillant en centre-ville ont une exposition à la pollution qui diffère de ceux habitant et travaillant en banlieue par exemple, mais la proximité entre les logements ou des activités quotidiennes et les zones à fort trafic peut aussi fortement différencier l'exposition des individus. En outre, le fait d'utiliser des modes de transports individuels ou collectifs est un facteur de différenciation de l'exposition. On ne peut en outre, aujourd'hui, plus ignorer le fait que des inégalités d'exposition significatives sont constatées dans les espaces urbains, et les actions engagées doivent pouvoir garantir qu'elles agissent en faveur de leur réduction. Cette question recèle de grands enjeux sociétaux, compte tenu du fait que la pollution atmosphérique est une problématique qui intéresse de plus en plus les populations urbaines. En Île-de-France, l'Ifop a récemment réalisé un sondage qui a montré que 85% des franciliens se sentaient concernés par la qualité de l'air et ses impacts sanitaires (Ifop, 2018). Sur l'ensemble de l'Europe également, la population s'est emparée progressivement des problématiques de qualité de l'air en ville, qui figure parmi leurs trois principales préoccupations (European Commission, 2010).

Les problèmes de pollution de l'air en ville résultent, pour beaucoup, de l'omniprésence de la voiture et, donc, des émissions de polluants qui en découlent (Host, Chatignoux, Leal, et al., 2012; Host & Chatignoux, 2009; Khaniabadi et al., 2017; Manisalidis et al., 2020). La question de l'usage des véhicules et de la proximité aux voies de circulation est donc un incontournable de cette problématique.

II. Les impacts sanitaires de la pollution urbaine

L'histoire de la prise de conscience de la toxicité des polluants est récente et démarre au 20^{ème} siècle, à la suite de forts événements ponctuels de pollution ayant conduit, les jours suivant, à des augmentations des taux de morbidité et de mortalités (Bell & Davis, 2001; Firket, 1936; Logan, 1953; Snyder, 1994). Le smog de Londres est l'événement de pollution le plus notable de cette histoire en raison du fort taux de mortalité qui s'en est suivi (Bell & Davis, 2001). Les communautés scientifiques internationales se sont par la suite emparées de la question, et ont développé des approches et des outils permettant de décrire et d'expliquer les impacts des polluants sur la santé (Bell et al., 2004; Goldsmith, 1969; Goldsmith & Lester Breslow, 1959; Martin, 1964).

Une part importante des travaux s'est intéressée à l'occurrence et aux déterminants des effets de la pollution sur la santé, au travers d'études épidémiologiques - parfois de grande ampleur - ce qui a permis d'approfondir notre connaissance et notre compréhension des risques associés à l'environnement urbain, et en particulier à ceux liés à la grande proximité qui existe entre citadins et le trafic routier. Ces travaux visent également à mieux décrire les inégalités environnementales d'expositions observées dans les grands centres urbains (Bowatte et al., 2017; Host, Chatignoux, & Saunal, 2012; Khreis & Nieuwenhuijsen, 2017; Perez et al., 2013). Il est en effet désormais admis que la ville est un lieu où la qualité de l'air affecte significativement la santé des individus. On peut citer l'étude récente de (Achakulwisut et al., 2019) qui a montré que sur 4 millions de nouveaux cas pédiatriques d'asthme, 64% étaient enregistrés en zones urbaines. Plus encore, les études épidémiologiques dénoncent le danger spécifique d'une exposition aux polluants issus du trafic routier - une des sources majeures de pollution urbaine (CITEPA, 2021; EEA, 2013; Urban Air Pollution – What Are the Main Sources across the World? | EU Science Hub). A l'échelle mondiale, il est été estimé qu'entre 3 à 4 millions de décès prématurés sont attribuables à des expositions à une mauvaise qualité de l'air extérieure (Cohen et al., 2017; WHO, 2016).

Les effets de la pollution sur la santé sont multiples. Ils dépendent de la fragilité de l'individu, mais surtout de l'intensité et de la durée de l'exposition. Globalement, les effets d'expositions ponctuelles à la pollution conduisent à des difficultés respiratoires (épisodes de toux, souffles courts, asthmes et sifflements lors de la respiration) et à des augmentations des visites médicales et hospitalisations. Les expositions chroniques, à des niveaux même faibles, engendrent l'exacerbation des cas d'asthme, des insuffisances pulmonaires, des maladies cardiovasculaires et une augmentation de la mortalité des populations (Raaschou-Nielsen et al., 2013; WHO, 2013). Récemment, des études ont mis en avant d'autres problèmes induits par les expositions chroniques à la pollution atmosphérique, comme certaines formes de diabète (Eze et al., 2014; Rajagopalan & Brook, 2012) ou encore l'émergence de pathologies chroniques chez les enfants lors de leur développement prénatal. Plus spécifiquement, j'ai reporté ci-après les principales connaissances des effets dus aux particules fines et au dioxyde d'azote.

1. Les particules

Les particules font partie des polluants d'intérêt pour les épidémiologistes à cause des effets qu'elles peuvent avoir, sur le court et sur le long terme, sur la santé des populations qui y sont exposées. Elles présentent le panel d'effet délétère le plus important des polluants suivis par les épidémiologistes et leur dangerosité croît inversement à leur diamètre. En effet, plus les particules sont fines, plus elles ont

la capacité de pénétrer profondément dans l'organisme via le système respiratoire (Manisalidis et al., 2020; Wilson & Suh, 1997) et d'induire des maladies cardio-vasculaires et respiratoires, des dysfonctionnements du système nerveux central, le développement de certains cancers (Andersen et al., 2007; Cesaroni et al., 2014; Khreis et al., 2017; Künzli et al., 2000; Pope III et al., 2002; Schulz et al., 2005) et, *in fine*, l'augmentation du taux de mortalité des populations (Künzli et al., 2000; Lelieveld et al., 2015). Des corrélations significatives avaient déjà été trouvées, il y a plus de 20 ans, entre le nombre d'hospitalisation dû à de l'asthme et les niveaux de PM₁₀ et de NO₂ (Lipsett et al., 1997). Des études suspectent que l'impact de l'exposition aux particules sur le système nerveux central peut générer des troubles mentaux (anxiété, dépression, trouble du déficit de l'attention avec/sans hyperactivité, autisme) (Bakolis et al., 2021; Buoli et al., 2018).

Les particules ont des effets néfastes sur la santé même lorsque les concentrations ambiantes sont inférieures aux seuils réglementaires (Bakolis et al., 2021; Olmo et al., 2011; Vedal et al., 2003).

2. Le dioxyde d'azote

Le dioxyde d'azote est connu pour irriter le système respiratoire. Étant un polluant gazeux, il a la capacité de s'introduire profondément dans les poumons et provoque des problèmes respiratoires allant de la gêne (toux) au développement et à l'exacerbation de l'asthme avec sifflements lors de la respiration. Une forte concentration de NO₂ inhalée peut causer des œdèmes pulmonaires (Achakulwisut et al., 2019; Bowatte et al., 2017; Faustini et al., 2014; Khreis et al., 2017; Naidoo, 2017). Il a pu être constaté que le NO₂ avait des effets sur la santé même lorsque les niveaux de concentrations extérieures sont inférieures aux seuils réglementaires (Achakulwisut et al., 2019; Freire et al., 2010; M. Johnson et al., 2020).

III. La représentation de l'exposition

Historiquement, les estimations de l'exposition se sont basées sur les valeurs de concentrations des polluants dans l'air extérieur, croisées avec les lieux de résidence des individus (Beelen et al., 2007; Brunekreef et al., 2009; Cesaroni et al., 2013; de Hoogh et al., 2014; Jerrett et al., 2005). Mais cette méthodologie est biaisée car elle ne considère pas le fait que les individus fréquentent une multitude de lieux au cours de leur journée, et qu'ils passent du temps en intérieur où les concentrations diffèrent de l'extérieur (Park & Kwan, 2017; Setton et al., 2010). Avec l'intérêt croissant pour la qualité de l'air à l'échelle de la rue et l'évolution des outils et des capacités de calculs, les chercheurs ont pointé du doigt la nécessité d'introduire dans les estimations de l'exposition la notion de gradient de concentration - et notamment de gradient lié au trafic routier - afin d'obtenir des indicateurs plus robustes de l'exposition (Özkaynak et al., 2013; Setton et al., 2010). Ce n'est qu'en considérant les activités et les mobilités des individus qu'il est possible d'apprécier les diversités d'expositions (Dons et al., 2011) et de mettre en avant les problématiques liées aux inégalités d'expositions environnementales (Tonne et al., 2018).

Depuis les années 2000, de premiers travaux ont mené au développement d'outils permettant de caractériser les habitudes et les déplacements de populations (Borrego et al., 2009; De Nazelle et al., 2013; Gulliver & Briggs, 2005; Valari et al., 2011), et l'intérêt pour ces questions s'est fortement amplifié depuis la fin des années 2010. En effet, le développement d'outils et de méthodologies pour estimer précisément les expositions à l'échelle de chaque individu, et au regard de ses activités et modes de déplacement, est aujourd'hui sollicité par les décideurs locaux qui souhaitent accompagner les populations qui s'interrogent sur leur l'exposition. Il est également encouragé par les agences nationales : Santé publique France a ainsi publié un guide mettant en avant les prérequis pour pouvoir mener une étude quantitative des risques sur la santé (EQIS) basée sur la modélisation de la pollution

atmosphérique, dans lequel ils expliquent la nécessité de se baser sur des données de qualité de l'air simulées qui soient résolues spatialement (surtout pour des polluants comme le NO₂) et temporellement (Pascal et al., 2019).

Les verrous à lever pour améliorer notre estimation de l'exposition sont bien connus : il faut notamment inclure dans les calculs la variabilité de la qualité de l'air dans les micro environnements visités ainsi que les temps passés dans chacun d'entre eux (Batterman et al., 2014, 2015; X. Zhang et al., 2017). Cela nécessite pour commencer, d'être en capacité de caractériser les niveaux de pollution de l'air dans tous les environnements extérieurs, avec une résolution assez fine pour qu'elle puisse être représentative de l'échelle individuelle. La mesure comme la modélisation peuvent permettre cette caractérisation.

1. Par la mesure

La mesure est bien sûr le moyen le plus réaliste d'obtenir des niveaux de concentrations ambiants à l'échelle des individus.

La mesure réglementaire est adaptée au recueil de données à haute résolution temporelle, et de nombreux sites de mesure permettent la caractérisation continue de la qualité de l'air en fond urbain ou à proximité de la voirie. La principale limitation de ces sites reste leur représentativité spatiale, qui est très limitée dans l'espace. Rodriguez et al., 2019 ont travaillé sur la représentativité de ces sites. Même s'il a pu être démontré que la représentativité spatiale des mesures de PM est plus de deux fois plus élevée que celle des oxydes d'azote, elle diminue très fortement d'un site de fond urbain à un site de proximité au trafic. Et la zone dans laquelle une mesure fixe peut être représentative de la qualité de l'air peut se limiter à une portion du trottoir d'une voie de circulation donnée. Pour décrire l'ensemble des gradients existant dans la ville, il faudrait donc une densité bien plus grande de points de mesure. Les réseaux de surveillance ne sont pas dédiés, et logiquement pas adaptés, à renseigner les gradients de concentration existant à très fine échelle.

Équiper des individus ou des véhicules en déplacement dans la ville peut permettre d'aborder la question de la variabilité spatiale fine échelle. Plusieurs équipes ont récemment investi ce terrain, et proposent de premières cartographie fine échelle de la qualité de l'air extérieure grâce à des instruments embarqués sur des véhicules (Apte et al., 2017) mais aussi des estimations de la qualité de l'air à l'intérieur des transports en commun routiers sur leur trajet (Molle et al., 2013). L'équipement de personnes avec des capteurs individuels, durant leurs activités quotidiennes, est particulièrement adapté aux études d'expositions puisque cela permet d'accéder à la diversité des micro-environnements visités, et d'explorer de nouveaux espaces, quand la mesure sur les stations fixes ne peut que caractériser les environnements extérieurs (De Nazelle et al., 2013; Liang et al., 2019; Yang et al., 2015). Si cette approche est prometteuse et les évolutions technologiques rapides, l'interprétation des résultats doit encore être menée avec prudence et expertise, car les microcapteurs individuels restent encore fortement soumis à des problèmes de spécificité, de biais ou de robustesse de la mesure dans les environnements très ou très peu pollués (Castell et al., 2017; Morawska et al., 2018).

2. Par la modélisation

La modélisation de la qualité de l'air est une approche aujourd'hui très répandue, tant pour des applications réglementaires de prévision et d'analyse des niveaux de concentrations, que pour des besoins de recherche. L'état de l'art des outils existant permet de produire des champs de concentration suffisamment résolus pour représenter les variations intra-urbaines de concentrations (Batterman et al., 2014; Schaap et al., 2015).

Parmi les principaux déterminants de la qualité de l'air urbain, les cycles d'émission liés aux fluctuations du trafic sont évidemment critiques. Et plus la résolution spatiale du calcul augmente, plus la résolution temporelle des émissions devient nécessaire à l'atteinte d'un résultat réaliste. La représentation de la

dispersion atmosphérique dans les modèles est bien sûr l'autre facteur déterminant de la modélisation à l'échelle de la rue. Leur capacité à prendre en compte l'effet de la structure urbaine sur l'accumulation des polluants ou sur la redirection des masses d'air, constitue forcément un atout dans les études sur la relation entre ville, émissions routières et qualité de l'air.

Il existe aujourd'hui une grande diversité de modèles et leur usage est déterminé par leur application. Les premiers modèles, développés autour des années 1940, étaient destinés à des usages industriels. D'approche gaussienne, ils permettaient la représentation de la dispersion des panaches des cheminées d'usine avec une visée d'identification et de prévention des risques liés aux accidents industriels. Ils considèrent toutefois des polluants chimiquement peu réactifs, qui sont dispersés en régime stationnaire autour d'une source d'émission ponctuelle (en surface ou plus en altitude) et sous un vent dont la direction et la vitesse sont uniformes sur tout le domaine considéré. Le panache simulé est ainsi distribué selon deux gaussiennes (horizontale et verticale). Aujourd'hui, ces modèles sont toujours utilisés pour évaluer les impacts de la pollution de l'air autour des sites industriels ou agricoles (Abu-Allaban & Abu-Qudais, 2011; Bajwa et al., 2012; Lee & Keener, 2008). S'ils ne sont pas adaptés pour simuler les champs de concentrations qui mèneront aux calculs d'exposition des populations, ils sont aujourd'hui devenus incontournables pour représenter dans les modèles urbains la dispersion des polluants émis le long des routes (paragraphe sur le couplage de modèles).

En parallèle, la modélisation s'est développée vers la simulation de la composition globale de l'atmosphère et le suivi des panaches, dans le but de pouvoir étudier des impacts à longue distance et des phénomènes tels que l'acidification des pluies ou la formation d'ozone (Daly & Zannetti, 2007). Or ces échelles de temps et les phénomènes concernés n'étaient pas accessibles avec les hypothèses nécessaires à l'approche gaussienne (Silveira et al., 2019). Les outils développés pour l'étude de la qualité de l'air sont des modèles Lagrangiens et Eulériens (voir encadré à la page suivante). Ces deux types de modèles permettent, respectivement, de suivre soit l'évolution de masses d'air le long de leur déplacement, soit le devenir de masses d'air réactives sur un territoire fixe maillé. Les modèles eulériens, aussi appelés Chemistry-Transport Models (CTMs) sont aujourd'hui très utilisés dans les domaines de la prévision opérationnelle de la qualité de l'air en fond urbain pour tous les polluants. Mais l'échelle des processus considérés dans ces outils n'est pas assez fine pour permettre de caractériser les gradients de concentrations urbains.

3. La restitution des gradients de concentration

Avec l'émergence des questions d'exposition individuelle, de nombreuses études se sont donc attachées à faire évoluer les modèles de qualité de l'air vers une meilleure représentation de la diversité de l'espace urbain à fine échelle. Trois principaux types d'approche ont été considérés : (i) La connexion à un modèle à plus fine échelle, très souvent Gaussien, soit directement implémenté dans chaque maille du modèle Eulérien (avec prise en compte des rétroactions entre les deux modèles), soit mis en œuvre en sortie du CTM - et dans ce cas le modèle Eulérien ne sert qu'à forcer la simulation du modèle Gaussien, (ii) Le développement d'approches statistiques appliquées en sortie de modèle Eulérien pour

Encadré

Modèles Lagrangiens

Les modèles Lagrangiens ont initialement été développés autour des années 1972 à 1975, pour étudier le transport sur de longues distances du SO_2 , en considérant des périodes de temps d'une à plusieurs années. Ils permettent de calculer la trajectoire de parcelles de « particules » en considérant les vitesses et directions des vents ainsi que les phénomènes de turbulences qui en découlent (Leelőssy et al., 2014). Contrairement aux modèles Gaussiens, la trajectoire de la bouffée ne suit pas une ligne droite mais est dirigée par le vent. Les valeurs finales de concentrations sont calculées en fonction des concentrations de chaque parcelle, le modèle a donc la capacité de gérer les changements de direction des vents tout en gardant des temps de calculs raisonnables.

Aujourd'hui, ils sont beaucoup utilisés pour étudier les conséquences d'événements accidentels, ou, s'ils sont utilisés en mode dit « backward », ils peuvent permettre de remonter à la source d'une pollution observée sur un domaine en aval. Les modèles Lagrangiens présentent l'intérêt d'avoir des coûts de calculs relativement faibles lorsqu'on considère un petit nombre de sources et présentent de pouvoir rapidement évaluer les impacts des cadastres d'émissions sur les concentrations calculées. La principale contrainte de ce type de modèle est qu'il faut réintroduire un maillage Eulérien pour pouvoir estimer les concentrations des bouffées en fonction de l'emplacement des particules (Leelőssy et al., 2014). Ils ne permettent pas non plus de considérer les interactions des polluants entre les parcelles et ont tendance à surestimer les concentrations (El-Harbawi, 2013). Enfin, ils ne permettent pas de considérer les réactions chimiques complexes des polluants (El-Harbawi, 2013).

Modèles Eulériens

Les modèles Eulériens diffèrent des modèles Lagrangiens par le fait qu'ils considèrent un référentiel fixe dans le temps. Ils ont initialement été développés dans le but d'étudier (i) les phénomènes de pollution urbain liés à l'ozone dès 1973, (ii) les phénomènes de pollution urbaine liés au dioxyde de soufre dès 1974 et (iii) les phénomènes régionaux liés aux dioxyde de soufre (1976) (Daly & Zannetti, 2007). Ces modèles ne considéraient initialement qu'une représentation très simplifiée de l'atmosphère, sous la forme d'une couche horizontale dans laquelle les dispersions étaient calculées dans une couche limite de hauteur constante.

Aujourd'hui, ces modèles ont la capacité de réaliser des simulations sur de larges domaines spatiaux et sur des périodes de temps pouvant être très longues (Colette et al., 2011; Laurent Menut et al., 2012; Terrenoire et al., 2015). Ils présentent l'avantage de considérer une chimie assez poussée des composés atmosphériques et sont devenus des outils performants dans le cadre de l'analyse, de la prévision et/ou des études de scénarios de la qualité de l'air régionale (par les AASQA de chaque région) ou nationale. À l'échelle d'une zone urbaine, ces modèles permettent de simuler de façon satisfaisante les niveaux de pollution en fond urbain mais ont tendance à les sous-estimer à proximité des sources d'émissions, comme les zones de trafic routier (Borge et al., 2016). Le principe de simulation de ces modèles est basé sur la discrétisation de l'espace de calcul : les concentrations calculées sont homogènes au sein de chacune des mailles divisant le domaine, ce qui ne permet pas de représenter les forts gradients de concentrations qu'on observe à proximité des voies routières (Chang & Hanna, 2004; Valari & Menut, 2008; Vilà-Guerau de Arellano et al., 2004). Enfin, ils ne prennent pas explicitement en compte les obstacles urbains dans leur représentation de la rugosité du sol. Mais le fort intérêt des simulations avec les CTM porte sur le fait qu'ils permettent de traiter un grand nombre de polluants, primaires comme secondaires, ainsi que leurs réactivités sur de larges domaines spatiaux et temporels.

recréer des gradients de concentrations à une échelle dite « sous-maille » et enfin (iii) La micro simulation s'appuyant sur des représentations dynamiques à très fine échelle comme les Large Eddy Simulation qui simulent la structure de la turbulence dans la couche limite. Cette dernière méthode n'est, néanmoins, adaptée qu'à la caractérisation de la qualité de l'air à l'échelle d'une rue et ne peut être conduite que sur des domaines spatiaux très réduits. Elle s'avère donc peu adaptée à des applications sur une ville ou une région toute entière à cause des limitations liées aux ressources de calcul que cela engendrerait et au besoin en données d'entrées.

Lorsque l'on souhaite mettre en œuvre une étude de modélisation de l'exposition, il faut donc savoir sélectionner l'approche la plus pertinente non seulement par rapport à la question que l'on se pose, mais aussi par rapport aux données et outils que l'on peut mobiliser. Il faut, notamment se demander jusqu'où le raffinement de la modélisation a du sens et est réellement pertinent pour notre étude. Si l'approche déterministe à ultra-fine échelle permet de distinguer des gradients de qualité de l'air entre les deux trottoirs d'une même voie de circulation, elle ne devient plus pertinente lorsqu'on cherche à modéliser l'exposition de toute une population - pour laquelle les chemins empruntés ne sont pas individuellement connus. Les approches destinées à estimer l'exposition d'un grand nombre d'individus pourront par exemple se contenter d'une représentation statistique de la qualité de l'air urbain, fonctionnant par environnement-type ou par rue-type, ou exploiter les sorties d'un modèle urbain pour en extraire des concentrations moyennes par rue fréquentée. Mes travaux de thèse s'intéressent à la représentation statistique de l'exposition d'une population à l'échelle régionale, et s'appuient sur ce dernier type d'approches.

Couplages de modèles

Une première manière d'accéder à la fine échelle est de procéder au couplage de deux modèles d'échelles différentes, souvent à partir d'une première simulation de CTM. Un des gros avantages de réaliser une étape de descente en échelle couplée à une simulation de CTM est en effet de pouvoir l'appliquer à de larges domaines spatiaux.

L'approche la plus classique consiste à réaliser des couplages en post-traitement entre un modèle Eulérien et un modèle à plus fine échelle (Hood et al., 2018; Kim et al., 2018; Silveira et al., 2019). Dans cette approche, un modèle de chimie transport simule d'abord les concentrations de fond urbain/régional puis un modèle plus fin se base sur ces informations pour calculer la dispersion des polluants. Seules les données simulées par le modèle à plus large échelle serviront de forçage pour le modèle plus fin, et les impacts des processus physiques/chimiques simulés à petite échelle n'auront pas d'influence sur les simulations aux échelles plus grandes. C'est une limitation forte de ce type de couplage pour modéliser la qualité de l'air des plus grandes villes (Silveira et al., 2019; Siour et al., 2013).

Les approches de descente en échelle mises en œuvre pendant les simulations (à chaque pas de temps de calcul) sont beaucoup moins nombreuses que celles appliquées en post-traitement du fait de l'augmentation importante des temps de calculs que le couplage en ligne exige. Plusieurs modèles de chimie transport ont intégré des modules permettant d'effectuer des simulations à des résolutions plus fines. Le modèle uEMEP (Rolstad Denby et al., 2020) est constitué du couplage en ligne d'un CTM avec un module gaussien implémenté dans son maillage, et conçu pour éviter l'un des principaux écueils du couplage CTM-Gaussien en post-traitement, qui est le double comptage des émissions lors de la mise en œuvre du modèle Gaussien. Un module de descente en échelle a également été ajouté au modèle CHIMERE, dans sa version 2017 : il permet de séparer au sein d'une même maille les émissions par fraction de type de sol, afin de retrouver la variabilité sous-maille perdue par l'homogénéisation des concentrations (Valari & Menut, 2010). La principale difficulté de cette approche réside dans le choix des paramètres de mélange des différents sous-environnements pour une représentation réaliste du mélange des masses d'air.

Approches statistiques de restitution des gradients de concentrations

Dans une approche tout à fait différente de la représentation de la proximité aux sources, des méthodes statistiques peuvent être développées pour interpoler spatialement des données mesurées ou pour rendre plus hétérogènes les sorties de modélisation (soit au cours de la simulation, soit en sortie de la simulation) et ainsi les corriger pour qu'elles ressemblent plus à la réalité. Dans les deux cas, ces approches vont devoir s'appuyer sur des paramètres explicatifs locaux ou environnementaux - appelés souvent proxys - qui vont permettre de prédire l'intensité de la correction ou de l'interpolation statistique à appliquer.

Contrairement aux couplages de modèles, les approches statistiques ne vont pas considérer les processus physiques et chimiques de l'atmosphère mais vont se baser sur des données mesurées ou modélisées déjà existantes à l'échelle souhaitée, pour « apprendre » la structure à reproduire et ensuite pouvoir la prescrire. Les données utilisées pour les étapes d'apprentissage peuvent inclure les émissions, la dynamique ou la météorologie, l'usage des sols, la démographie de la zone étudiée, etc. Les approches basées sur des étapes d'apprentissage sont communément appelées « Machine Learning » et ont commencé à être développées autour des années 2010 (Rybarczyk & Zalakeviciute, 2018). Ces approches permettent de révéler des relations dans les données en considérant la non-linéarité entre les données de forçage et leurs impacts sur les concentrations. Le choix d'une approche de Machine Learning va fortement dépendre du type d'application que l'on veut en faire. Généralement, les problématiques d'estimation de la pollution nécessitent de considérer une approche d'Ensemble Learning ou de Régression alors que les problématiques liées à la prévision des niveaux de pollution considèrent des réseaux neuronaux (Rybarczyk & Zalakeviciute, 2018).

Les approches de descente en échelle statistique ont démontré leur capacité à affiner la résolution de simulations (Reich et al., 2014; Silibello et al., 2021) jusqu'à une centaine de mètres (Bonafe et al., 2011; Gariazzo et al., 2020; Huang et al., 2021; Joly et al., 2021), mais elles peuvent aussi être utilisées pour interpoler spatialement des données issues de réseaux de mesures (soit par des approches de Machine Learning, soit via des régressions).

Réseaux de neurones

Les réseaux de neurones (Artificial Neural Network en anglais) sont une des approches de Machine Learning inspirées du système nerveux humain. Ils sont composés d'une à plusieurs couches de neurones qui vont transformer l'information donnée en entrée en information selon des règles précises déterminées lors de l'étape d'apprentissage (Kohonen, 1988). Le développement de ce type d'approche a débuté dans les années 1940 (McCulloch & Pitts, 1943) mais ce n'est qu'en 1993 que cette approche a été appliquée pour la première fois dans le domaine de la qualité de l'air (Boznar et al., 1993). Aujourd'hui, les réseaux de neurones sont développés pour prédire la qualité de l'air dans des zones non couvertes par des sites de mesures en mettant en relation des données prédictives telles que la mesure, les variables météorologiques et d'usage des sols, les émissions et/ou des sorties de modélisation (Alimissis et al., 2018; J. Chen et al., 2019; Hrust et al., 2009; Su et al., 2016; M. Xu et al., 2021).

Ils sont très performants pour établir des interdépendances entre des données d'entrées et des paramètres cibles, suivant des fonctions non-linéaires, ce qui n'est pas accessible par des régressions linéaires (Alimissis et al., 2018; Dreyfus et al., 2002; Hornik et al., 1989). Néanmoins, développer un modèle robuste de réseau de neurone nécessite qu'il soit entraîné sur un grand nombre de données d'entrées (de 2 à plus d'une dizaine d'années) pour maximiser les relations entre les variables (Rybarczyk & Zalakeviciute, 2018).

Random-Forest

Une des approches des méthodes d'Ensemble Learning les plus populaires dans la littérature est le Random-Forest (Rybarczyk & Zalakeviciute, 2018). C'est une approche similaire aux modèles de régression mais qui est favorisée pour les cas d'étude impliquant des données non-linéaires ne nécessitant pas d'extrapolations (Araki et al., 2018). Cette approche est basée sur un ensemble d'arbres de décision qui ont pour but de mettre en avant les relations existant entre différentes variables prédictives. Les variables considérées, autres que les valeurs de concentration simulées, peuvent concerner des données démographiques, d'utilisation des sols, de trafic routier, de conditions météorologiques et de période de l'année. Les Random Forest peuvent être mis en place pour estimer les niveaux de concentrations sur des zones non couvertes par la mesure (Araki et al., 2018; G. Chen et al., 2018). Sinon, la mise en relation des données simulées avec ces prédicteurs permet de les affiner spatialement pour obtenir des résultats ayant une résolution plus importante (Gariazzo et al., 2020; Silibello et al., 2021) ou d'étudier des tendances de niveaux de pollution (Grange et al., 2018).

Comme pour le reste des approches de Machine Learning, la construction d'un modèle de Random Forest nécessite une grande quantité de données d'entrée pour que l'approche soit robuste. L'étude d'Araki et al., 2018 a discuté les potentiels sous-estimations du modèle qu'ils ont développé pour des axes routiers très fréquentés, et au niveau des intersections routières. Il est aussi important de noter que les modèles de Random-Forest, contrairement aux modèles de régression, ne permettent pas d'extrapoler des résultats qu'ils obtiennent et vont, donc, très rarement être généralisables à des données n'ayant pas servi à l'apprentissage du modèle (Hengl et al., 2018).

Régressions

Une autre approche utilisée pour retrouver la variabilité sous-maille des sorties de simulations d'un CTM est le développement de modèles de régression. Ces approches vont chercher à déterminer les relations statistiques existantes entre les données modélisées et des données à plus fine échelle, généralement issues de la mesure. Ainsi, il est possible d'ajuster les sorties d'un modèle Eulerien pour diminuer les écarts entre le modèle et la mesure. On garde, de cette façon, les avantages des deux approches car les interactions physiques/chimiques des composés sont caractérisées lors des simulations avec le CTM et les impacts des éléments locaux (proxys) sont pris en compte lors de l'établissement de la régression - contrairement aux approches LUR, présentées ci-après, pour lesquels aucune prise en compte des phénomènes physico-chimiques n'est possible (Alkuwari et al., 2013). La littérature compte assez peu d'études de qualité de l'air ayant cherché à affiner la résolution de simulations via des régressions (Guillas et al., 2008; Joly et al., 2021), cet outil statistique étant plus souvent utilisé pour les études sur le climat (Ribalaygua et al., 2013; Schoof & Pryor, 2001).

La plupart des études d'estimations des expositions dans la littérature s'appuie en effet sur le développement d'approches dites « Land-Use Regressions » (LUR) dont le principe est l'interpolation spatiale, par régression, de mesures sur un domaine donné, sur la base de variables d'environnement qui pourraient servir de proxys aux gradients de concentrations. Cette approche a été développée pour la première fois en 1997 dans le but de cartographier la pollution à fine échelle (D. J. Briggs et al., 1997) et est, depuis, beaucoup utilisée dans le domaine de l'estimation des expositions à la pollution atmosphérique (Beelen et al., 2013; Katsoulis et al., 2014). La littérature montre que des LURs de structures différentes peuvent prédire les variabilités spatiales de la pollution correctement (Hoek et al., 2008) et que, une fois développés pour un domaine, leur application reste utilisable et stable dans le temps (R. Wang et al., 2013).

Les échelles spatiales d'application des LURs vont dépendre de la portée du réseau de mesure déployé. Des études ont appliqué des LURs sur des pays entiers (Beelen et al., 2007, 2009; Hoek et al., 2001; Schmitz et al., 2019; Vienneau et al., 2010) mais beaucoup vont plutôt s'intéresser aux zones urbaines (Basagaña et al., 2012; Beelen et al., 2013; de Hoogh et al., 2014; Eeftens et al., 2012; Larkin et al., 2017). Plus le nombre de sites de mesure déployés sur le domaine étudié sera élevé, plus la cartographie des concentrations sera précise et permettra de restituer les gradients de concentrations effectivement observés. Il a, cependant, été suggéré de déployer un minimum de 80 sites de mesures pour les zones urbaines à cause de la complexité des gradients de concentrations (Basagaña et al., 2012). Cette approche permet d'obtenir des résolutions de cartographie très fine avec, à titre d'exemple, l'étude de Schmitz et al., 2019 où ils ont cartographié les valeurs de NO₂, NO_x et de PM₁₀ sur l'ensemble des Pays-Bas à 5m de résolution.

Les performances des LURs ont été étudiées en comparaison avec la mesure ou avec les sorties de modèles de dispersion (de Hoogh et al., 2014; Hoek et al., 2008). La revue de (Hoek et al., 2008) met en avant des résultats robustes en termes de moyennes annuelles pour les NO_x, les particules, les suies et les COV. Les corrélations obtenues peuvent être moins bonnes pour les particules, en comparaison avec celles des NO_x, possiblement parce que leur variabilité spatiale est moins prononcée ou à cause du manque de caractérisation de leurs sources en zones urbaines qui se traduit souvent par un plus faible nombre de site de mesure (de Hoogh et al., 2014; Eeftens et al., 2012).

Toutefois, malgré leur simplicité de mise en œuvre et leur popularité, les LURs restent limités dans leurs estimations des expositions. Les périodes temporelles pouvant être étudiées grâce à cette méthode sont assez courtes à cause du coût de maintien du réseau de mesure, quelques semaines tout au plus. De plus, ils ne permettent pas de prendre en compte la dispersion des polluants, ni leur accumulation au sein des rues puisque ni la chimie des polluants, ni la topographie de la zone étudiée ne sont prises en compte. D'ailleurs, récemment, des approches de Land Use Random Forest ont été développées (LURF) pour prendre en compte les relations non-linéaires existantes entre les données mesurées et les variables prédictives et augmenter la robustesse de l'approche (Araki et al., 2018; Brokamp et al., 2017).

4. La représentation des environnements intérieurs

Les problématiques de caractérisation des niveaux de concentrations en fonction des déplacements et des activités de la population n'ont pas été discutées dans les sections précédentes et, pourtant, la littérature reconnaît leur caractère indispensable pour estimer correctement l'exposition journalière cumulée de la population (D. Briggs, 2005; Buonanno et al., 2014; Dias & Tchepel, 2018; Dons et al., 2011; McCreddin et al., 2014; Park, 2020; Poom et al., 2021). En effet, un individu passe entre 80 et 90% de son temps en environnement intérieur (Boulanger et al., 2017; Klepeis et al., 2001) et les deux polluants visés par ce travail de thèse ont aussi des sources d'émissions en intérieur (toutes les combustions au gaz, liées au tabagisme ou à l'utilisation de bougies et d'encens par exemple).

Des études ont permis d'explorer l'exposition d'individus au cours de leurs journées, par le biais de la mesure. Les individus sont suivis via des applications smartphones qui identifient leurs localisations GPS et ils doivent souvent mentionner les activités qu'ils réalisent (Buonanno et al., 2014; De Nazelle et al., 2013; Dewulf et al., 2016; Dons et al., 2011; Glasgow et al., 2014; Kang et al., 2012; Liang et al., 2019; Mastroianni et al., 2016; Roberts & Helbich, 2021; D. Wang & Sun, 2015). Le nombre de volontaires pour ce type d'étude est forcément limité par la logistique que leur suivi implique, mais certaines études incluent jusqu'à plusieurs centaines d'individus (400 pour Roberts & Helbich, 2021). De manière assez exceptionnelle, l'étude de Dewulf et al., 2016 a utilisé les données issues d'un opérateur mobile pour suivre l'exposition de près de 3 millions d'individus.

Pour extrapoler ces résultats à l'ensemble d'une population, la modélisation de leurs habitudes et des concentrations des environnements intérieurs qu'ils visitent est alors mise en œuvre. Pour cela, il faut

distinguer deux grands types d'environnements intérieurs que sont (i) le bâti et (ii) les habitacles des véhicules ou espaces de transports en commun.

Comme on le verra en détail dans les chapitres de cette thèse, déterminer les niveaux de qualité de l'air en intérieur est une tâche complexe.

La modélisation des concentrations dans les bâtiments implique d'avoir une connaissance du parc de bâtiment de la zone étudiée, pour pouvoir prendre en compte les systèmes chauffage, de ventilation/filtration de l'air, la qualité énergétique et la forme du bâti, sa distance aux sources d'émissions extérieures, mais aussi les activités et pratiques de leurs occupants (Vardoulakis et al., 2020). L'ANSES cite dans son rapport de 2019 que la caractérisation du parc du bâtiment français ne peut-être pertinente si la présence ou non des systèmes de ventilation, les matériaux de construction et les habitudes des occupants ne sont pas pris en compte (ANSES, 2019) tandis que l'étude de Héroux et al., 2010 montre que les activités/habitudes des occupants des pièces peuvent fortement influencer la variabilité des niveaux de pollution. Enfin, les niveaux de concentrations à l'extérieur des bâtiments sont suspectés d'influencer à hauteur de 60% les concentrations intérieures en particules (Meier et al., 2015). Certaines études ont engagé une simulation de la qualité de l'air des environnements intérieurs en utilisant des approches intégrées avec des modèles calculant l'écoulement des polluants entre extérieur et intérieur ou basés sur des régressions en sortie de modèle méso-échelles (Gaffin et al., 2016) mais ces démarches sont encore très exploratoires et émergentes. La mise en place de modélisation de la qualité de l'air du bâti urbain d'une zone urbaine entière nécessite donc non seulement un grand nombre de données d'entrées, mais également une grande puissance de calcul du fait de la nécessité d'utiliser des modèles de dynamique des fluides pour caractériser l'écoulement des polluants entre extérieur et intérieur. Les chercheurs s'orientent pour l'instant plutôt vers la détermination de rapport Intérieur/Extérieur (communément appelés « Rapport I/E » ou « Ratio I/O » dans la littérature) (Annesi-Maesano et al., n.d.; Blondeau et al., 2005; Challoner & Gill, 2014; Cyrus et al., 2004; Gerharz et al., 2014; Kotzias et al., 2009; Ni Riain et al., 2003; Park, 2020; Salonen et al., 2019).

Pour ce qui est des déplacements, plusieurs études ont mesuré les niveaux de concentrations retrouvés dans les modes de transport des zones urbaines pour les véhicules particuliers, les bus, les métro/RER, les trains, les tramways et les modes de déplacement doux (Cepeda et al., 2017; Choi et al., 2019; de Nazelle et al., 2017; Delaunay et al., 2012; Johansson & Johansson, 2003; Molle et al., 2013; Strasser et al., 2018; B. Xu & Hao, 2017; Zuurbier et al., 2010). Comme pour la caractérisation des concentrations intérieures dans les bâtiments, il semble plus intéressant de pouvoir déterminer des coefficients de transfert des polluants afin de pouvoir prendre en compte les variabilités horaires et spatiales de la pollution. de Nazelle et al., 2017 déterminent justement des rapports de concentrations I/O pour des bus et des véhicules particuliers, et Matthaios et al., 2020 ont mesuré ces rapports pour des véhicules particuliers utilisant ou non les systèmes de ventilation et de filtration du véhicule. Il n'y a toutefois, à ma connaissance, pas d'article qui détermine ces rapports pour les métros/trains (plateforme et cabine) mais certaines études vont comparer les niveaux de concentrations des deux micro-environnements (Mao et al., 2019; Passi et al., 2021; B. Xu & Hao, 2017).

5. Modélisation des pratiques

Plusieurs outils et méthodes de modélisation ont été développés pour estimer les pratiques des individus impactant leur exposition à la pollution. Qu'elles soient basées sur des mesures (Smolak et al., 2020), des modèles d'activités, des réseaux de neurones ou des simulations de type Monte-Carlo, de nombreuses approches ont été mises en place pour caractériser les expositions des individus de toute une population, en fonction de leurs habitudes quotidiennes (De Montjoye et al., 2013; Dewulf et al., 2016; Gulliver & Briggs, 2005; McCreddin et al., 2014).

Les approches de modélisation basées sur l'activité ont ainsi un fort potentiel pour prédire le lien entre des caractéristiques d'une population et sa probable exposition. L'étude de McCreddin et al., 2014, en

particulier, compare les expositions estimées grâce aux trois méthodes mentionnées ci-dessus avec les expositions mesurées des individus. Les résultats des simulations Monte-Carlo et du réseau de neurone ont de bonnes corrélations avec la mesure (Coefficient de Pearson > 0,59) et celles issues des modèles d'activités sont plus mitigées (Coefficient de Pearson = 0,55). D'autres études trouvent, cependant, de meilleures corrélations des expositions modélisées et simulées (Gulliver & Briggs, 2005; Pilla & Broderick, 2015). Ces approches n'ont toutefois pas été réellement explorées par la communauté de qualité de l'air, et n'intègrent pas de lien entre les activités des populations (notamment mobilité), leurs émissions, et leur exposition.

IV. Conclusion et organisation du travail de thèse

Mes travaux de thèse s'intègrent justement dans une démarche de mise en œuvre d'une approche de modélisation intégrée, créant le lien entre populations, activités, mobilité, qualité de l'air et exposition. Ils s'insèrent au cœur du développement d'une plateforme de modélisation faisant intervenir :

- La modélisation d'un territoire, d'une population et de ses activités et pratiques de mobilité.
- La modélisation des émissions issues des activités de la population.
- La modélisation de la qualité de l'air depuis l'échelle régionale (CTM) jusqu'à l'échelle de l'individu (descente en échelle).
- L'estimation statistique de l'exposition de toute la population, en prenant en compte à la fois les environnements fréquentés dans la journée et les boucles de mobilité réalisées par chaque individu.

Les 4 points de ce projet sont développés au LISA, dans l'équipe dans laquelle j'ai réalisé ma thèse. Les deux premiers points sont assurés par le modèle d'émission OLYMPUS, basé sur l'activité des agents modélisés. Mes travaux sont rattachés à la construction d'une méthode de descente en échelle nécessaire au point 3, et à la conception et au codage de l'ensemble des modules nécessaires au point 4.

Les ambitions de ce travail de thèse étaient donc tout d'abord de mettre en place une approche pouvant générer des données de qualité de l'air dont la résolution soit assez fine pour une application pertinente dans le cadre de la modélisation de l'exposition individuelle. Pour cela, j'ai choisi d'emprunter la voie de la descente en échelle, par une approche statistique, et à partir des sorties d'un modèle méso-échelle de qualité de l'air (CHIMERE). Mon approche sera appliquée au dioxyde d'azote et aux particules PM₁₀. C'est le thème développé dans le chapitre 2.

Je me suis ensuite attachée (chapitre 3) à construire un module de calcul de l'exposition de la population francilienne à la pollution, avec l'exigence de prendre en compte l'agenda et la mobilité de tous les individus. Ces développements m'ont notamment demandé de construire tout un module sur la mobilité des individus par les transports en commun, module qui n'existait pas au début de ma thèse.

Enfin, cette thèse a également été conçue pour accompagner le conseil départemental du Val-de-Marne sur un exercice prospectif d'évaluation de la qualité de l'air sur une gare du futur réseau de transport du Grand Paris Express. Ce travail, qui fait l'objet du chapitre 4, a été initié en proposant une configuration de fonctionnement pour notre chaîne de modélisation, et en façonnant un premier scénario d'aménagement urbain et routier au niveau du quartier de la gare de Créteil l'Échat. Le but de ce scénario est d'étudier l'impact du projet de gare sur la qualité de l'air du quartier en prenant en considération les modifications de voiries, de bâti et d'attractivité de la zone qui en découlent. Cet exercice n'a pu qu'être initié, le scénario n'ayant pu être complètement construit, du fait des impératifs auxquels le Val-de-Marne a dû faire face pendant toute la durée de la crise sanitaire ces deux dernières années.

Chapitre II

Augmentation de la résolution spatiale de simulations de la qualité de l'air : Développement d'une approche de descente en échelle statistique

Table des matières

I.	Présentation générale de l'approche retenue	32
II.	Configuration de l'étude	33
1.	Configuration du modèle de chimie-transport CHIMERE	34
2.	Données météorologiques	39
III.	Conception et mise en œuvre d'une méthodologie de descente en échelle	40
1.	Conception de l'approche	40
2.	Conception de la phase d'apprentissage	42
3.	Mise en oeuvre de l'apprentissage	47
4.	Etape d'automatisation de la correction	63
5.	Validation de l'approche	74
6.	Discussion	81
7.	Cartographie des résultats	83
IV.	Travail sur l'année 2017	84
1.	Choix de l'année 2017	84
2.	Méthodologie	86
3.	Retravail du cadastre d'émission de 2015	86
4.	Résultats du travail sur l'année 2017	88
5.	Cartographie des résultats	98
V.	Conclusion	99

I. Présentation générale de l'approche retenue

L'objectif de ce chapitre est de proposer, dans nos modélisations de l'environnement urbain, une représentation des variations de concentration des polluants atmosphériques à fine échelle, afin qu'elles puissent être spécifiques de l'exposition d'un individu se déplaçant dans la ville.

L'outil que nous utilisons au laboratoire pour produire des champs de concentration de polluants est un modèle de chimie-transport (CTM), qui fonctionne donc sur une approche eulérienne dans laquelle l'espace est discrétisé en mailles homogènes du point de vue de la dynamique, de la topographie, des émissions... En résolvant l'équation de conservation de la masse dans chaque maille de la grille d'étude, ce type de modèle prend en considération les termes de perte et production de chaque espèce chimique liés aux émissions, à la dispersion dans l'atmosphère, à la photochimie, aux équilibres gaz-particule et aux processus de dépôt au sol. Mais par nature ces modèles ne représentent pas la dispersion progressive des panaches d'émissions dans les mailles, ils ne tiennent pas non plus compte de la présence du bâti en tant qu'obstacle au mélange des différentes masses d'air, et ils ne considèrent pas les phénomènes dynamiques de très petite échelle qui impactent la dispersion des polluants dans les rues entourées d'immeubles. Si l'on veut pouvoir simuler toute la variabilité des concentrations de polluants dans les différents micro-environnements urbains, il est donc inutile de raffiner excessivement le maillage car c'est l'approche numérique même du modèle qui n'est pas adaptée. En revanche, il est possible de retravailler les sorties du modèle pour à la fois bénéficier de la richesse de l'information qu'elles contiennent, et leur ajouter une information nouvelle à plus fine échelle. On parle de « descente en échelle ».

La revue de la littérature que j'ai présentée dans le chapitre précédent m'a permis d'identifier les méthodes statistiques d'apprentissage comme des outils très bien adaptés à mes contraintes et ambitions de descente en échelle. L'apprentissage permet d'analyser des données (d'observation notamment) et d'en dégager des motifs ou des comportements types. Les statistiques obtenues lors de l'entraînement sont utilisées pour construire un algorithme qui permettra alors de prédire les données sans passer par l'observation. On appelle « modèle de machine learning » l'algorithme résultant de l'entraînement d'un algorithme d'apprentissage avec des observations.

Comme mentionné plus haut, mon idée n'était pas d'utiliser l'apprentissage pour prédire la qualité de l'air urbain à partir des seules observations, mais plutôt de retravailler les données de modélisation à ma disposition. Comme le jeu de données issu du CTM contient déjà une information complexe et pertinente que je souhaitais garder - la variabilité des émissions, de la dynamique locale, de la météorologie, des transports à grande échelle... - mon choix a été de faire l'apprentissage de l'erreur du CTM à fine échelle. Les motifs que j'ai donc cherché à analyser et prédire sont ceux de l'écart entre (i) les sorties du CTM dans une maille et (ii) les observations dans différents environnements urbains de cette même maille. L'idée était de trouver une équation unique pour chaque site permettant de formaliser cette erreur sur l'ensemble de l'année, afin de la réintroduire dans mes sorties de modèle. Le fait qu'elle soit valable sur l'ensemble d'une année - et donc qu'elle ait été éprouvée sur une large gamme de situations météorologiques - devait me permettre de l'appliquer sur des périodes non couvertes par l'apprentissage.

L'intérêt de développer une approche statistique pour la descente en échelle était notamment de pouvoir m'affranchir des lourdes données d'entrées nécessaires pour les approches de modélisation urbaine, que ce soit en termes de quantité de paramètres ou de besoin en résolution. La démarche statistique doit m'amener à des algorithmes et/ou des facteurs correctifs plus simples à mettre en œuvre, tout en garantissant une représentation améliorée de la qualité de l'air dans les différents environnements urbains d'échelle inférieure au kilomètre. Mon ambition était également de pouvoir construire une méthode qui soit extrapolable à des situations hypothétiques (scénarios d'émissions, de pratiques ou d'aménagement). Cela rendait les approches de modélisation urbaine déterministe assez peu intéressantes, car j'aurais eu à reconstruire pour chaque scénario toutes les bases de données

d'entrée concernées, en prenant garde de garder leur cohérence. Toutefois, il fallait que je m'assure que la correction statistique puisse prendre en compte le changement des conditions s'appliquant à mes sites (moins de trafic, plus de bâti...) : il fallait donc qu'elle soit qualitativement et/ou quantitativement reliée à la géographie et aux usages du site. Les paramètres de la correction statistique sur chaque site devront donc pouvoir être définis par rapport à des indicateurs locaux (aussi appelés proxys). Cela doit me garantir non seulement une capacité d'extrapolation à d'autres sites que ceux de l'apprentissage, mais aussi la capacité de pouvoir prédire la correction à appliquer au CTM dans des scénarios urbains.

Dans ce chapitre, après avoir décrit le CTM à ma disposition, ainsi que sa configuration sur les territoires et période d'étude, je présenterai l'approche de correction des données du CTM que j'ai conçue et construite. Elle se découpe en trois étapes principales :

- La méthodologie d'apprentissage de l'erreur (données supportant la méthode, choix de construction, choix d'approches statistiques) et la formalisation de la correction des données du modèle de chimie-transport,
- La construction d'une relation empirique entre les proxys de chaque site pour pouvoir extrapoler spatialement et à d'autres situations la correction statistique,
- L'évaluation et la validation de la démarche, réalisées en appliquant l'algorithme précédemment mis en place à un jeu de stations non utilisées pour l'apprentissage.

Je présenterai enfin l'application de ma méthode à une seconde période temporelle pour discuter la possibilité d'extrapolation de la formalisation de l'erreur à d'autres situations météorologiques et d'émissions.

II. Configuration de l'étude

Mon domaine d'étude est la région Île-de-France. C'est une région dans laquelle se jouent de nombreux défis de durabilité : rejets de polluants anthropiques, changement climatique ou encore expansion urbaine et gestion du lien entre ville et ruralité. La région parisienne s'organise en effet autour d'un centre très urbanisé, à haute densité d'habitation et entouré d'une zone urbaine périphérique en expansion, avec des problématiques fortes de mobilité et de congestion du trafic automobile. Comme toutes les villes denses, elle connaît des dépassements des valeurs seuils pour NO₂, et parfois pour les particules fines, près des grands axes de circulation, ce qui est susceptible de créer des hétérogénéités d'exposition assez importantes parmi la population (Airparif; Host et al., 2012; Host & Chatignoux, 2009). Les questions liées à la mobilité pendulaire et à la qualité de l'air impactent au quotidien la qualité de vie des habitants. Par ailleurs, cette région offre l'avantage d'être un cas d'étude « simple » en modélisation, car elle présente peu d'éléments difficiles à reproduire - que ce soit une topographie trop marquée ou une météorologie complexe, elle est très bien documentée du point de vue de la mesure des polluants, et de nombreuses données d'enquête (ménages, mobilité) sont disponibles. Enfin, les outils de modélisation que nous utilisons ont déjà été validés sur ce territoire.

Pour développer et tester le module de calcul de l'exposition, j'ai choisi de travailler sur l'année 2009. Ce choix se justifie tout d'abord par une qualité de l'air globalement médiocre sur de nombreuses périodes et sur l'ensemble du domaine. L'hiver a été caractérisé par plusieurs épisodes de froid sévère associés à des conditions météorologiques peu dispersives, durant lesquels de fortes concentrations de NO₂, mais aussi de PM₁₀ issues du trafic routier et du chauffage au bois ont été mesurées. Utiliser une telle année permettait de soumettre une large gamme de situations à l'apprentissage. Par ailleurs, le plus récent inventaire d'émissions dont nous disposions était celui de l'année 2008. Il n'aurait pas été pertinent de l'utiliser sur une période trop récente. Pour appréhender tout de même des situations plus récentes, j'ai en parallèle récupéré des données d'émissions plus récentes sur l'Île-de-France (2017) que j'ai retravaillées pour les implémenter dans le modèle de chimie-transport. Cela me permettait de tester

la transposabilité dans le temps de ma méthode de descente en échelle, en vérifiant notamment l'impact d'un changement de cadastre d'émissions sur le calcul des facteurs de correction, ainsi que la pertinence de la formalisation retenue pour la correction. L'intérêt était également de pouvoir bénéficier d'un jeu plus important de stations de mesure (plusieurs ayant été installées entre les deux années que j'ai choisies). Il était également intéressant de réaliser l'étude comparée de la qualité de l'air à proximité des routes sur ces deux années car, en 8 ans, il y a eu une évolution significative du parc roulant et des motorisations des véhicules ainsi qu'une décroissance des flux globaux de trafic (M. André et al., 2020; Michel André et al., 2014, 2019). Il y a eu également des modifications locales comme la fermeture des voies sur berges en 2016 qui ont redistribué le trafic dans Paris. En somme, cette seconde situation pouvait presque se rapprocher d'un scénario, par rapport au cas ayant permis de développer l'algorithme. Un dernier intérêt, majeur, qui m'a menée à modéliser l'année 2017, était de pouvoir comparer mes résultats aux sorties de la chaîne de modélisation fine échelle d'AIRPARIF, disponibles à partir de 2016.

1. Configuration du modèle de chimie-transport CHIMERE

Le modèle de qualité de l'air que j'ai utilisé au cours de ma thèse est le modèle CHIMERE, un modèle de chimie-transport méso-échelle développé par l'INERIS, le LMD et le LISA (<https://www.lmd.polytechnique.fr/chimere/>). CHIMERE est un modèle utilisé tant pour la recherche sur les processus physico-chimiques pilotant l'évolution des polluants atmosphériques et la compréhension des déterminants des épisodes photochimiques (Laurent Menut et al., 2012; Monks et al., 2009; Petetin et al., 2016; Q. J. Zhang et al., 2013), que pour la prévision opérationnelle de la qualité de l'air (Copernicus Atmosphere Monitoring Service; Honoré et al., 2008; Prevaire) ou encore l'analyse de scénarios prospectifs de la qualité de l'air - qu'ils soient menés dans un cadre institutionnel ou académique, pour la compréhension des leviers de la qualité de l'air ou l'accompagnement de l'action publique (Coll et al., 2009; de Meij et al., 2009; Thunis et al., 2007).

Les modèles de chimie-transport ont été conçus pour suivre l'évolution spatiale et temporelle des principaux polluants (gazeux ou particulaires) impliqués dans la dégradation de la qualité de l'air, sur des domaines spatiaux régionaux à continentaux. CHIMERE permet aujourd'hui dans ses versions les plus récentes de traiter des questions de qualité de l'air depuis l'échelle d'une métropole (avec une résolution kilométrique) jusqu'à l'échelle hémisphérique, avec la possibilité de mener de manière couplée les calculs du modèle météorologique WRF et les calculs de CHIMERE (Laurent Menut et al., 2021). La version que j'ai utilisée est la version 2013 dont je décris la configuration ci-dessous (voir L. Menut et al., 2013 pour plus d'information).

Domaines de simulation

Pour pouvoir alimenter notre domaine francilien cible en données de qualité de l'air à ses limites, trois simulations ont été réalisées en cascade (ou en « Nesting »). Elles permettent de construire avec une résolution spatiale croissante les champs de concentrations en polluants aux bornes du domaine d'intérêt.

Le grand domaine de simulation, nommé « EUROPE », couvre la France et tout ou partie des pays limitrophes (Angleterre, Benelux, Allemagne, Suisse, Espagne et Italie) avec une résolution d'environ 55 km. Le domaine intermédiaire, de résolution 15 km et nommé « GRANDNORD3 », couvre la partie nord de la France. Enfin, notre domaine d'intérêt, en bleu dans la Figure, est nommé « IDF1 » et englobe la région Île-de-France avec une résolution kilométrique. Les caractéristiques des trois domaines sont détaillées dans le [Tableau 1](#).



Figure 1 - Visualisation des 3 domaines imbriqués : en rouge le domaine EUROPE, en vert le domaine GRANDNORD3 et en bleu le domaine IDF1.

Les domaines sont tous composés de 8 niveaux verticaux, avec une pression de 997 mbar dans le niveau le plus bas et une pression de 500 mbar dans le niveau le plus élevé. La caractérisation de la nature des sols sur chacun des domaines est assurée par l'utilisation de la base de données GLCF (Global Land Cover Facility³ REF).

Tableau 1 - Détails des domaines CHIMERE - NX et NY représentent les nombres de mailles en longitude et latitude, DX et DY la dimension (en degrés) des mailles dans ces deux directions, et XMIN et YMIN les coordonnées (WGS84) du point sud-ouest de chaque maillage.

Domaine	NX	NY	DX	DY	XMIN	YMIN
EUROPE	30	24	0.68	0.46	-7.33	42.0
GRANDNORD3	43	38	0.205	0.135	-3.91128	45.3705
IDF1	159	129	0.0137	0.009	1.457017	48.11395

Représentation des processus chimique dans CHIMERE

Pour les espèces gazeuses - CHIMERE propose plusieurs schémas chimiques pour décrire la réactivité des composés gazeux lors des simulations. Celui que j'ai utilisé est une version réduite du schéma MELCHIOR (Lattuati, 1997), obtenue via la méthode des opérateurs chimiques et appelée MELCHIOR2 (Carter, 1990; Derognat et al., 2003). Il considère environ 50 espèces et 120 réactions chimiques, contre 109 et 350 dans sa version complète. Dans cette version, les voies réactionnelles minoritaires aux échelles régionales à continentales ne sont pas prises en compte, ce qui permet de réduire les temps de calculs nécessaires aux simulations. Ce schéma présente l'avantage d'être régulièrement utilisé depuis près de 20 ans pour la prévision opérationnelle de la qualité de l'air avec CHIMERE à l'échelle nationale, mais aussi dans les différentes régions françaises - et notamment en Île-de-France, et il est quotidiennement évalué par comparaison avec la mesure (<http://www2.prevoir.org/content/propos-de-prevoir>). Le choix de ce schéma assure donc une grande robustesse à mes simulations sur les 3 domaines. Une raison supplémentaire à ce choix est que je prévoyais également de mettre en œuvre le

³ <https://geog.umd.edu/feature/global-land-cover-facility-%28glcf%29>

modèle urbain MUNICH pour un cas d'étude sur un quartier du Val-de-Marne, et MELCHIOR2 est l'un des deux schémas chimiques proposés par ce modèle. Ainsi je me suis dirigée vers ce schéma à des fins de cohérence entre les simulations.

Pour les espèces particulières - CHIMERE contient un module d'aérosol qui caractérise :

- Les émissions totales de particules (dites particules « primaires », PPM). Il est possible de distinguer les fractions organique (OCAR) et élémentaire (BCAR) des PPM si l'information est disponible dans le cadastre d'émission : dans ce cas, les émissions totales de PPM correspondent à la somme de BCAR, de OCAR et de la partie non spécifiée des particules primaires. Je n'ai pas retenu cette option car elle était indisponible dans le cadastre mis à ma disposition (voir plus tard).
- Les émissions des espèces secondaires telles que le nitrate, le sulfate, l'ammonium et les aérosols organiques secondaires (AOS).
- Les émissions naturelles de poussières et de sels de mer.

Les aérosols sont représentés dans le modèle en fonction de leur distribution en taille et de leur composition chimique. Pour chaque espèce considérée, la répartition des aérosols est réalisée sur un nombre N de classes de taille choisies par l'utilisateur, allant du diamètre le plus petit au plus grand (Mailler et al., 2017) : dans le cadre de mes simulations, j'ai considéré 9 classes de taille de particules. Les espèces considérées sont réparties dans 3 groupes de tailles : Elles prennent le suffixe « _fin » lorsque leur diamètre aérodynamique est inférieur à 2,5 μm (PM_{2.5}), « _coa » pour l'anglais « coarse » lorsque leur diamètre est compris entre 2,5 et 10 μm (PM₁₀-PM_{2.5}) ou « _big » lorsqu'il est supérieur à 10 μm (L. Menut et al., 2013). Plus de détails sur les partitions des classes d'aérosols peuvent être trouvés dans (Pun et al., 2002).

Les phénomènes de transport

Dans CHIMERE, des interfaces permettent d'utiliser des champs météorologiques issus de différents modèles pour calculer dans chaque maille l'advection des masses d'air, et faire un diagnostic des paramètres de turbulence dans la couche limite. Les champs météorologiques que j'ai utilisés sont issus de simulations du modèle WRF (voir plus bas). Le mélange vertical turbulent est calculé suivant la paramétrisation de Troen & Mahrt, 1985. Pour les transports horizontaux et verticaux des polluants, mon choix s'est dirigé vers le schéma d'advection de Van Leer (Hourdin & Armengaud, 1998). Il est reconnu en météorologie pour sa bonne précision numérique et sa diffusion plus faible que des schémas d'ordre 1, et représente dans CHIMERE un bon compromis entre robustesse et temps de calcul.

Les sources et puits d'émissions

L'outil CHIMERE permet de prendre en compte une variété de sources d'émission, que ce soit pour les polluants gazeux ou particuliers et pour des espèces réactives ou inertes. Elles peuvent être situées au sol ou présenter un profil vertical d'émission, et sont principalement distinguées selon leur origine anthropique ou naturelle.

Emissions anthropiques

Les émissions anthropiques incluent notamment les secteurs d'activité routière, industrielle, domestique, agricole, manufacturière ou encore aéroportuaire, et sont renseignées par des inventaires cadastrés adaptés à l'échelle simulée. CHIMERE possède un préprocesseur d'émission qui lui permet d'exploiter directement les données d'inventaire du programme EMEP (European Monitoring and Evaluation Programme) accessibles sur l'ensemble du territoire européen à une résolution de 0.5°. Pour

des simulations à plus petite échelle, CHIMERE peut utiliser les cadastres d'émissions régionaux fournis par les Associations Agréées de Surveillance de la Qualité de l'Air (AASQA) - et notamment ceux d'AIRPARIF pour la région parisienne - dès lors qu'ils sont fournis au format attendu. Bien sûr, le préprocesseur peut être retravaillé pour lire n'importe quel format d'émissions en entrée. Le prétraitement est réalisé en 3 étapes : (i) Les émissions de l'inventaire d'émission considéré sont projetées sur le maillage CHIMERE grâce à des proxies d'utilisation des sols ou de densité de population. Cette étape de projection sur le maillage de simulation est toujours nécessaire d'un point de vue technique, mais lorsque la résolution de l'inventaire est plus grossière que celle du modèle, elle permet à la simulation de gagner en précision. En effet, la redistribution (qui s'appuie sur des indicateurs d'usage des sols tels que la localisation des zones urbanisées (https://www.lmd.polytechnique.fr/chimere/docs/CHIMEREdoc_v2020r3.pdf)) va alors créer de l'information à plus fine échelle que celle du cadastre, ce qui limite l'effet d'une faible résolution des émissions, (ii) des profils temporels sont appliqués aux émissions de l'inventaire pour les convertir en flux horaires et (iii) une spéciation des espèces fournies par l'inventaire est réalisée pour les convertir en espèces modèles du mécanisme chimique choisi dans CHIMERE (par exemple, les NOx du cadastre d'émission seront séparés en NO, NO₂ et HONO). Cela se fait à l'aide de tables de correspondance entre familles d'espèces, espèces réelles et espèces modèles, et à l'aide de profils de spéciation d'émission par secteur d'activité, notamment pour les COV.

J'ai utilisé les données d'émission EMEP pour les simulations des deux grands domaines (EUROPE et GRANDNORD3). Les émissions EMEP sont composées des totaux d'émissions nationaux issus de rapports nationaux, mis à disposition par secteur d'activité et maillés. Les données annuelles d'émissions sont disponibles depuis 1990 jusqu'à l'année n-2, et pour 4 catégories de polluants : polluants dits majoritaires (incluant le CO, l'ammoniac NH₃, les composés organiques volatils non méthaniques, les NOx et SO₂), les particules (PM₁₀, PM_{2.5} et black carbon), les métaux lourds et les polluants organiques persistants « POPs ».

Sur le domaine cible (IDF1), j'ai pu bénéficier de l'inventaire régional d'émission d'Airparif constitué pour l'année 2008, et de résolution kilométrique. L'ancrage du domaine IDF1 et sa résolution ont d'ailleurs été choisis de façon à ce que les deux maillages se superposent parfaitement. Les données d'émission, ainsi que les étapes de prétraitement qu'il a été nécessaire de mener avant de procéder aux simulations, sont décrites ci-dessous.

- Les données d'émission d'Airparif sont issues d'un calcul « bottom-up », c'est-à-dire d'un calcul basé sur les données sectorielles d'activité les plus fines possibles qui est ensuite agrégé à une échelle administrative plus large, ou sur un maillage de simulation. Les données peuvent être déclaratives ou issues de mesures d'activité comme les comptages routiers. Les données à ma disposition étaient disponibles sur un maillage kilométrique, et exprimées en masse annuelle de polluant par maille, pour différents secteurs d'activité. Elles étaient accompagnées de facteurs mensuels, journaliers et horaires permettant la redistribution temporelle des totaux par catégorie d'activité. Les espèces d'intérêt pour la qualité de l'air et qui sont renseignées sont les COV non méthaniques, les NOx, SO₂ et les particules fines des classes de taille < 2.5µm et < 10µm. Le monoxyde de carbone n'est pas renseigné mais il n'altère pas la représentation de la réactivité atmosphérique d'échelle urbaine.
- Pour ce qui concerne le découpage en secteurs d'activité, cet inventaire proposait 31 catégories d'activité correspondant à des sous-catégories de la nomenclature européenne SNAP (Selected Nomenclature for Air Pollution). La correspondance avec les 15 catégories d'émetteurs considérées par CHIMERE a été réalisée au laboratoire, lors d'un travail précédant ma thèse. Le tableau des correspondances est présenté ci-après ([Tableau 2](#)).
- Enfin, l'inventaire d'Airparif étant développé dans un cadre institutionnel, il ne couvre que le territoire administratif de la région Île-de-France mais pas les zones limitrophes. Or mon domaine possède une forme géométrique simple qui englobe l'Île-de-France et qui inclut donc quelques territoires limitrophes (voir [Figure 1](#)). J'ai donc dû compléter les émissions sur les mailles non

franciliennes en utilisant les données EMEP, qui présentaient l'avantage d'être cohérentes avec les deux grands domaines. J'ai pu bénéficier pour cela d'un module de combinaison des inventaires, déjà écrit et mis en place à cet effet dans la version 2013 de CHIMERE par des membres de mon équipe. L'Île-de-France étant beaucoup moins urbanisée sur ses frontières, le fait d'avoir un inventaire moins résolu sur les coins n'a pas été considéré comme pénalisant.

Tableau 2 - Table des correspondances entre les catégories d'émetteurs de CHIMERE (colonne de gauche) et celles de l'inventaire francilien (colonne de droite).

Numéro du secteur modèle	Secteur équivalent SNAP dans le cadastre francilien
Secteur 1	010000
Secteur 2	020100, 020200, 020300
Secteur 3	020104, 020106, 020107, 020110, 020120, 020140, 020150, 020200
Secteur 4	020300, 020301
Secteur 5	030100, 030200, 030300
Secteur 6	040000, 040100, 040200, 040300, 040400, 040500, 040600, 040700, 040800
Secteur 7	050000, 050200, 050503, 050201, 050302, 050402, 050501, 050502, 050602, 050603
Secteur 8	060000, 060101
Secteur 9	060100, 060102, 060103, 060104, 060105, 060106, 060107, 060108, 060109, 060201, 060300, 060301, 060302, 060303
Secteur 10	060202, 060203, 060204
Secteur 11	060304, 060305, 060306, 060307, 060308, 060309, 060311, 060312, 060314, 060400, 060402, 060403, 060404, 060405, 060406, 060408, 060409, 060411, 060412, 060500, 060502, 060506, 060507, 060508, 060601, 060602, 060603
Secteur 12	070101, 070102, 070103, 070201, 070202, 070301, 070401, 070502, 070601, 070602, 070603, 070701, 070702, 070800
Secteur 13	080200, 080300, 080500, 080601, 080602, 080801, 080802, 081000
Secteur 14	090000, 090100, 090200, 090400, 090900, 091000
Secteur 15	100000, 100100, 100101, 100102, 100104, 100105, 100200, 100205, 100206, 100300, 100400, 100408, 100409, 100410, 100500, 100900, 110000

On rappelle les 11 secteurs d'activités considérés dans la nomenclature SNAP :

(01 00 00) Les combustions dans les industries de l'énergie et de la transformation de l'énergie

(02 00 00) les combustions hors industries (incluant les combustions domestiques)

(03 00 00) les combustions dans les industries manufacturières

(04 00 00) les procédés de production

(05 00 00) les extractions et distributions de combustibles fossiles/énergies géothermiques

(06 00 00) l'utilisation de composés chimiques et de solvants

(07 00 00) le transport routier

(08 00 00) les autres activités de transport (marin, aéroporté, agricole, militaire...)

(09 00 00) les traitements et éliminations des déchets

(10 00 00) l'agriculture et la sylviculture

(11 00 00) une catégorie « autres » incluant émissions par les feux de végétation, volcans, éclairs...

Leur code est présenté entre parenthèses et les 4 numéros suivants précisent une éventuelle sous-catégorie. A titre d'exemple, le SNAP 7 (07 00 00) représente la somme totale des émissions du transport

routier tandis que le SNAP 07 01 00 est limité aux émissions des véhicules particuliers et le SNAP 07 01 01 aux émissions des véhicules particuliers sur des trajets autoroutiers.

Comme mentionné précédemment, les années 2009 et 2017 ne pouvaient pas être simulées avec le même cadastre d'émission du fait de l'évolution significative des émissions entre ces deux années. Pour les simulations de l'année 2017, j'ai pu accéder à un cadastre des émissions de l'année 2015 produit par Airparif, mais dans un format très différent et beaucoup moins détaillé que celui de 2008. J'ai dû mener un travail important de redistribution des données dans les catégories d'émetteurs de CHIMERE. Ce travail fait l'objet de la section IV. 3, dans laquelle les modifications apportées sont décrites, et la qualité, la représentativité et la pertinence des données obtenues sont discutées.

Emissions biogéniques

Les émissions biogéniques représentent les émissions naturelles de la végétation et sont calculées par le modèle MEGAN (Guenther et al., 2006). MEGAN utilise des données de couvert végétal (espèces, densité) pour produire des flux d'espèces chimiques émises qui seront projetées sur la grille de CHIMERE. Les flux étant dépendants notamment de la période de l'année (croissance de la végétation, feuillage des arbres), de l'ensoleillement et de la température, le modèle utilise en entrée des données météorologiques horaires sur toute la durée de la simulation.

2. Données météorologiques

Tant pour le calcul de l'advection et de la diffusion des masses d'air que pour le calcul du dépôt des espèces ou des émissions biogéniques, CHIMERE a besoin de données météorologiques en entrée de simulation. Celles-ci sont fournies par le modèle non-hydrostatique WRF (Weather and Research Forecasting, version 3.9.1.1, <http://www.wrf-model.org>, (Powers et al., 2017)), un modèle développé pour les besoins de la recherche et de la prévision. Il projette les données météorologiques de réanalyse NCEP (National Centers of Environmental Predictions) sur son maillage comme données d'entrée et aux limites du domaine. Il modélise ensuite les champs météorologiques par une approche Eulérienne. Je n'ai pas utilisé la paramétrisation de la canopée urbaine qui vise à mieux restituer les spécificités de la dynamique de la couche limite urbaine dans WRF, car l'apport de cette option n'a pas été démontré comme significatif (REF). À la place, comme cela a été le cas dans les travaux précédents de mon équipe, j'ai opté pour l'option de correction de la dynamique en zone urbaine disponible dans CHIMERE, qui limite la vitesse du vent et renforce les flux de chaleur sensible à la surface.

Les domaines de simulation de WRF doivent englober ceux de CHIMERE afin de les contraindre à leurs limites. Les caractéristiques des 3 domaines WRF que j'ai défini sont résumées dans le [Tableau 3](#) ci-dessous. Le domaine CITY60 est utilisé pour forcer le domaine « EUROPE », CITY20 le domaine « GRANDNORD3 » et CITY04 force le domaine « IDF1 ».

Tableau 3 - Détails des domaines WRF - NX et NY décrivent le nombre de mailles en ouest-est et sud-nord ; DX et DY décrivent la résolution spatiale ouest-est et sud-nord (en m) des mailles ; Ref_lat et Ref_lon donnent les coordonnées du point central du domaine considéré ; Proj indique le système de projection considéré.

Domaine	NX	NY	DX	DY	Ref_lat	Ref_lon
CITY60	55	55	60 000	60 000	46.12	7.88
CITY20	58	52	20 000	20 000	48.40	2.5
CITY04	161	131	4 000	4 000	49.5	3.2

III. Conception et mise en œuvre d'une méthodologie de descente en échelle

1. Conception de l'approche

Le choix du jeu de données d'apprentissage est déterminant, puisque c'est lui qui contient l'information que l'on créera à fine échelle. Les données de mesure de la qualité de l'air les plus régulières et les mieux distribuées sur les territoires sont celles des AASQA. Elles sont relatives à l'ensemble des polluants réglementés dont la surveillance incombe aux AASQA, notamment les oxydes d'azote ($\text{NO} + \text{NO}_2$), les particules fines (PM_{10} et $\text{PM}_{2,5}$), l'ozone, le monoxyde de carbone, certains composés organiques volatils aromatiques dits BTEX (benzène, toluène, éthylbenzène et xylènes), le dioxyde de soufre, les hydrocarbures aromatiques polycycliques (HAP) et les métaux lourds (plomb Pb, cadmium Cd, arsenic As, nickel Ni). Pour les principaux polluants gazeux, les données sont disponibles au pas de temps horaire sur chaque site de mesure.

Les associations de surveillance de la qualité de l'air du réseau Atmo France doivent déployer un réseau de mesure qui répond à un certain nombre de contraintes réglementaires. Le nombre de stations à déployer est par exemple contraint par la taille des communes, le découpage de la région en territoires administratifs de surveillance, l'historique local des dépassements de seuil ainsi que tout contexte météorologique ou d'émission pouvant justifier un ajustement des conditions de surveillance (<https://www.legifrance.gouv.fr/loda/id/LEGIARTI000043396087/2021-04-19/>). Airparif compte ainsi une cinquantaine de stations de mesures fixes réparties sur un rayon de 100 km autour de la ville de Paris. Les sites de mesure qui sont déployés ont vocation à distinguer 4 grands types d'environnement : rural, urbain ou périurbain de fond, proximité au trafic et industriel.

- Les sites ruraux permettent d'observer les niveaux de concentration au vent et sous le vent du panache de la métropole. Ils ont été positionnés à la fin des années 90 pour répondre essentiellement à des problématiques de pointes d'ozone, et mesurent assez rarement les oxydes d'azote (qui s'y trouvent souvent au-dessous de la limite de détection des appareils classiques).
- Les stations de trafic correspondent aux stations de mesure situées à proximité d'axes routiers majeurs. Ces stations ont pour but de fournir des informations sur la qualité de l'air sous l'influence directe du trafic routier. Le rapport NO/NO_2 moyen doit y être supérieur à 1,5, ce qui garantit que la masse d'air n'a pas trop évolué depuis l'émission.
- Les stations de fond sont situées dans l'espace urbain ou périurbain. Elles visent à représenter les niveaux moyens de pollution retrouvés dans ou en périphéries des centres urbains, hors de l'influence directe du trafic routier. Pour s'en assurer, l'implantation des stations doit vérifier une distance minimale par rapport à l'axe routier, qui dépend du Trafic Moyen Journalier Annuel (TMJA) de la voie. Cette distance est par exemple de 10 m pour une voie ayant un TMJA de 1 000 à 3 000 véhicules/jour, et de 100 m lorsque le TMJA de la voie atteint 40 000 à 70 000 véhicules/jour. Il n'y a pas de distance minimale pour les axes de moins de 1 000 véhicules/jour. Une seconde contrainte vient compléter cette typologie : le rapport NO/NO_2 doit être inférieur à 1,5, ce qui garantit une évolution minimale du panache d'émission (par réactivité ou dilution). La distinction entre site urbain et périurbain de fond est liée au fait que le site de mesure se situe soit dans l'aire urbaine principale - de densité de population supérieure à 40 000 hab/km², soit dans une commune périphérique à l'aire urbaine principale.
- Les sites industriels sont positionnés sous le vent des sites industriels, afin de surveiller l'impact des activités industrielles sur les zones d'habitation alentour.

Parmi cette offre de données, seuls les environnements de fond périurbain, de fond urbain et de proximité au trafic en zone urbaine ont du sens dans ma démarche. En effet, on sait que le CTM ne peut reproduire la variabilité de la qualité de l'air à des échelles de 10 ou 50 m près des voies de circulation : la mesure en proximité au trafic apportera des éléments de correction qualitative et quantitative des

concentrations, qui n'existent pas en sortie du CTM. En outre, la diversité des mesures de fond pourra éventuellement permettre d'apporter une correction sur la manière dont CHIMERE reproduit les différents degrés de fond urbain, ou l'intensité des pics de concentration en fond urbain. En revanche, les stations rurales sont supposées être éloignées de toute influence directe des sources anthropiques. Elles ont une représentativité spatiale plutôt étendue et la qualité de l'air à proximité ne présente pas de gradient marqué de concentration des polluants primaires ou secondaires. Un éventuel besoin d'amélioration de la représentation des données rurales ne relèverait donc pas d'une démarche de descente en échelle. Pour ce qui est des données obtenues sur les sites industriels : (i) le nombre de stations de surveillance en zone industrielle est trop faible pour être compatible avec une démarche statistique d'apprentissage de l'erreur du modèle et (ii) la représentation de l'évolution des panaches industriels dans l'atmosphère par les CTM est trop peu explicite pour constituer un *a priori* satisfaisant pour l'apprentissage de l'erreur du modèle. Je ne pourrai pas appréhender la question de la qualité de l'air à fine échelle, sous le vent des industries, dans mon approche. Ce n'est pas bloquant pour mon objectif de calcul de l'exposition, car la dégradation de la qualité de l'air en zone industrielle ne constitue plus un problème d'exposition de premier ordre en région parisienne - le dioxyde de soufre est ainsi la plupart du temps inférieur à sa limite de détection. Cela peut être problématique localement pour certains métaux, comme c'est le cas de l'arsenic à proximité d'industries très spécifiques ([Airparif, 2021](#)), mais cela reste spatialement limité.

En conclusion, ma principale préoccupation sera donc de veiller à ce que l'approche de descente en échelle appliquée à CHIMERE permette de mieux reproduire les spécificités de la qualité de l'air en fond périurbain et en fond urbain, et d'intégrer une variabilité à fine échelle, dite « sous-maille » et liée au trafic, afin de prédire la diversité de la qualité de l'air près des axes routiers franciliens. L'absence de données de mesure spécifiques à d'autres microenvironnements, comme les espaces verts ou les quartiers piétons, ou encore le fait qu'il n'existe pas de stations à proximité de sources spécifiques de COV comme les stations-services ou les garages, délimite la portée - et limite la richesse - de ma correction à fine échelle sur les sorties du modèle CHIMERE. Ce point manquera lorsqu'on procédera à l'estimation de l'exposition des individus fréquentant de tels environnements au cours de leur journée.

Le choix des polluants retenus pour l'apprentissage, sur les stations (péri-)urbaines et de trafic, est également structurant. Puisque mon ambition est de travailler sur le risque d'exposition aux polluants en zone urbaine, je dois cibler des polluants d'intérêt sanitaire majeur. Et puisque la typologie des données à ma disposition ne me permet de distinguer que la distance au trafic routier, les polluants sur lesquels je devais me concentrer doivent présenter des gradients connus à proximité des axes routiers. Enfin, il fallait que le jeu de données à ma disposition soit suffisamment conséquent (en termes de nombre de sites) pour que je puisse mener un apprentissage, et que le modèle soit suffisamment robuste dans sa simulation des espèces d'intérêt. Les seuls polluants répondant à l'ensemble de ces critères sont le dioxyde d'azote et les PM₁₀. En effet, le dioxyde de soufre n'est pas émis spécifiquement par le trafic, il répond à une logique de surveillance des environnements industriels que j'ai choisi de ne pas considérer. Concernant les HAP et les métaux lourds, leur surveillance est peu répandue sur le territoire et elle ne se fait pas au pas de temps horaire car elle fait appel à des prélèvements passifs qui nécessitent une analyse différée en laboratoire : ces données n'étaient donc pas non plus exploitables pour notre travail. L'utilisation des mesures de BTEX pour la comparaison aux sorties du modèle était quant à elle délicate du fait que ces espèces ne sont pas explicitement représentées dans le mécanisme chimique MELCHIOR : elles sont contenues dans deux espèces modèles dont les contours de spéciation sont plus larges que la famille BTEX. Nous ne les avons donc pas retenues. Enfin, même si on peut estimer que la variabilité des PM_{2,5} est un très bon indicateur de l'exposition au trafic routier, les conditions pour leur implémentation dans l'approche d'apprentissage n'étaient pas réunies sur l'année 2009, en termes de nombre de sites de mesure. Le choix de NO₂ et des PM₁₀ couvre toutefois déjà une part significative des enjeux de l'exposition.

2. Conception de la phase d'apprentissage

Comme mentionné en introduction de ce chapitre, le choix que j'ai fait est de trouver une formalisation de l'erreur entre les données de CHIMERE et la mesure, qui soit indépendante de l'échéance temporelle considérée, et qui puisse donc constituer une correction pour les sorties de CHIMERE sur n'importe quel pas de temps. La correction devra s'ajouter à la valeur simulée CHIMERE qui comporte déjà l'essentiel de la variabilité horaire des concentrations de polluants, que ce soit en fonction des émissions, du flux solaire, de la situation météorologique ou de la dynamique locale.

Proposition d'une formalisation de l'erreur mesure-modèle

J'ai dans un premier étudié les motifs systématiques de l'erreur entre le modèle et la mesure aux différents sites, afin de proposer une équation pour la représenter. A titre d'illustration, j'ai rapporté, sur la [Figure 2](#), les séries temporelles de NO₂ et de PM₁₀ issues de la mesure et de la simulation CHIMERE de référence - menée sur IDF1 pour l'ensemble de l'année 2009. Suite à cette comparaison, j'ai pu tirer les constats suivants :

- Les concentrations modélisées par CHIMERE possèdent une variabilité à l'échelle horaire en bon accord avec la mesure et constituent un très bon *a priori* pour notre étude.
- Quel que soit le site et le polluant, la variabilité de la ligne de base des concentrations à plus basse fréquence (d'un jour sur l'autre, d'une semaine sur l'autre) est également assez bien reproduite par le modèle, qui amène donc une bonne compréhension des différents systèmes météorologiques qui se succèdent et des éventuels apports de polluants depuis l'extérieur du domaine. L'allure globale des courbes de la [Figure 2](#) illustre bien ce phénomène.
- Sur les sites de fond urbain et périurbains ([Figure 2a, b, c, d](#)), le modèle simule des niveaux de concentrations dans l'ensemble proches de ceux obtenus par la mesure, avec une fréquence et une intensité des pointes qui apparaît très réaliste. On constate toutefois une tendance du modèle à la sous-estimation des concentrations, qui s'exprime inégalement selon les sites. L'analyse des boîtes à moustache (boxplots) ([Annexe 3](#)) permet d'apprécier la distribution des données. Elle montre que, sur la grande majorité des sites de fond, l'amplitude du signal entre le percentile 25 (P25) et le percentile 75 (P75) n'est que légèrement sous-estimée par le modèle, mais cette partie de la distribution des données est presque systématiquement tirée vers les faibles valeurs dans le modèle, et les plus hautes valeurs mesurées ne sont en général pas reproduites par le modèle. C'est le cas par exemple à Paris 18 ([Figure 2b](#)). Certains sites font exception. Ces exceptions concernent aussi bien des sites où la distribution des données est presque parfaitement reproduite par CHIMERE (comme à Garches, [Figure 2a](#)), que des sites pour lesquels le modèle sous-évalue largement la mesure. Dans ce dernier cas, cela peut être lié à l'influence directe d'une source proche du site de mesure, ou à un environnement hétérogène dans le kilomètre autour de la station (mélange urbain / boisé / rural). En effet, en cas d'hétérogénéité de l'usage des sols dans la maille, le CTM sous-estime structurellement les concentrations observées au point le plus émetteur. Pour les PM₁₀, l'amplitude du signal entre les indicateurs P25 et P75 est dans l'ensemble mieux reproduite par CHIMERE. Si on constate toujours une sous-estimation des valeurs absolues de P25, de la moyenne et de P75 plus marquée sur certains sites (comme dans la [Figure 2c](#)) que sur d'autres, les plus hautes valeurs de la distribution sont dans l'ensemble mieux reproduites par CHIMERE que pour NO₂. Sur certains sites, les pointes sont toutefois surestimées, comme on peut le voir dans la [Figure 2d](#). Pour ce qui concerne la moyenne annuelle en NO₂ - qui est un indicateur réglementaire pour l'exposition - la sous-estimation par le modèle s'échelonne entre 1 et 15 µg/m³ sur l'ensemble des sites, avec un écart moyen de près de 20% sur cette valeur. Pour les PM₁₀, CHIMERE produit des valeurs annuelles moyennes qui sont de 4 à près de 15 µg/m³ plus basses que la mesure, ce qui représente sur de nombreux sites des biais de 30% à 4 % sur cet indicateur.

- L'écart mesure-modèle, que ce soit sur les valeurs moyennes ou sur l'amplitude du signal journalier, est beaucoup plus marqué sur les sites situés en proximité du trafic routier. J'ai présenté sur les [Figure 2e à 2h](#), la comparaison des données mesurées et celles produites par CHIMERE sur les sites de mesure situés Place Victor Basch dans Paris, et Porte d'Auteuil à l'entrée du périphérique parisien. Ici, la différence entre les niveaux moyens simulés et mesurés est telle qu'on n'a parfois pas de recouvrement des valeurs comprises entre P25 et P75. Les écarts sur la moyenne annuelle sont la plupart du temps conséquents. Ainsi, la moyenne annuelle en NO₂ mesurée à Porte d'Auteuil est de 97,2 µg/m³ tandis que le modèle prédit seulement 29,6 µg/m³ dans la maille concernée. L'écart sur NO₂ est presque aussi fort (94.6 contre 36.6 µg/m³) Place Victor Basch, qui n'est pourtant qu'un rond-point très fréquenté du centre de Paris. Pour les PM₁₀, le biais sur la moyenne la mesure indique une moyenne annuelle de 47,3 µg/m³ contre 22,3 µg/m³ dans le modèle.
- Par ailleurs, on observe un facteur 2 entre l'amplitude des pics simulés par CHIMERE et ceux mesurés par Airparif quel que soit le polluant. Il faut noter que, contrairement à ce qu'on a observé sur les sites de fond, l'ampleur absolue et relative des écarts est très dépendante du site considéré, et la station A1 est responsable des pires scores du modèle. Ce site permet toutefois de qualifier assez facilement la nature de l'erreur : comme pour les sites de fond elle est systématique et assez peu liée à la période et au système dynamique étudié, et elle s'exprime sous la forme d'un manque d'amplitude sur les pointes et les niveaux moyens qui pourrait se corriger par l'application d'un facteur de proportionnalité sur ces deux indicateurs. Il ressort de mon analyse que ce facteur correctif sera - possiblement au contraire des sites urbains - très dépendant du micro-environnement routier considéré.
- Enfin, on peut déduire de l'évolution de la ligne de base que la succession des régimes météorologiques (plus ou moins dispersifs) est bien restituée par CHIMERE.

L'erreur du modèle se présente donc plutôt comme une erreur systématique, induisant une sous-estimation des niveaux de fond - et de pointe pour NO₂ - et qui pourrait être corrigée par l'application d'un facteur correctif, dépendant du polluant et du site, sur les valeurs issues de CHIMERE. On peut ainsi espérer une amélioration sensible de notre estimation de l'exposition en fond urbain - tant moyenne que de pointe - en retravaillant les sorties de CHIMERE à partir de l'analyse de ses biais.

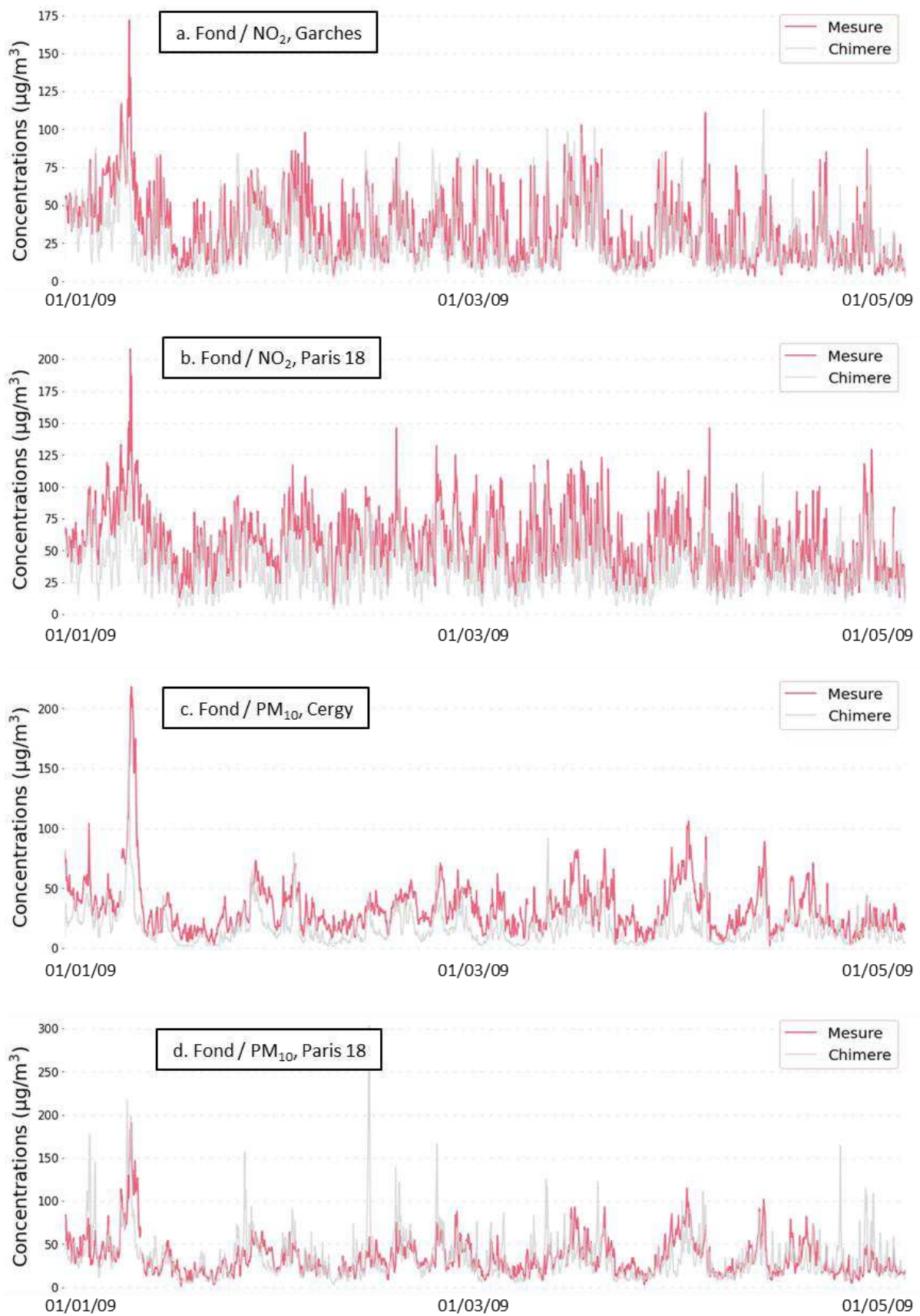


Figure 2 (a, b, c, d) - Comparaison des séries temporelles mesurées par le réseau Airparif (en rouge) et simulées par CHIMERE (en gris) sur des sites de mesures de fond urbain / périurbain, pour NO₂ (en haut) et pour les PM₁₀ (en bas). Les concentrations sont exprimées en $\mu\text{g}/\text{m}^3$.

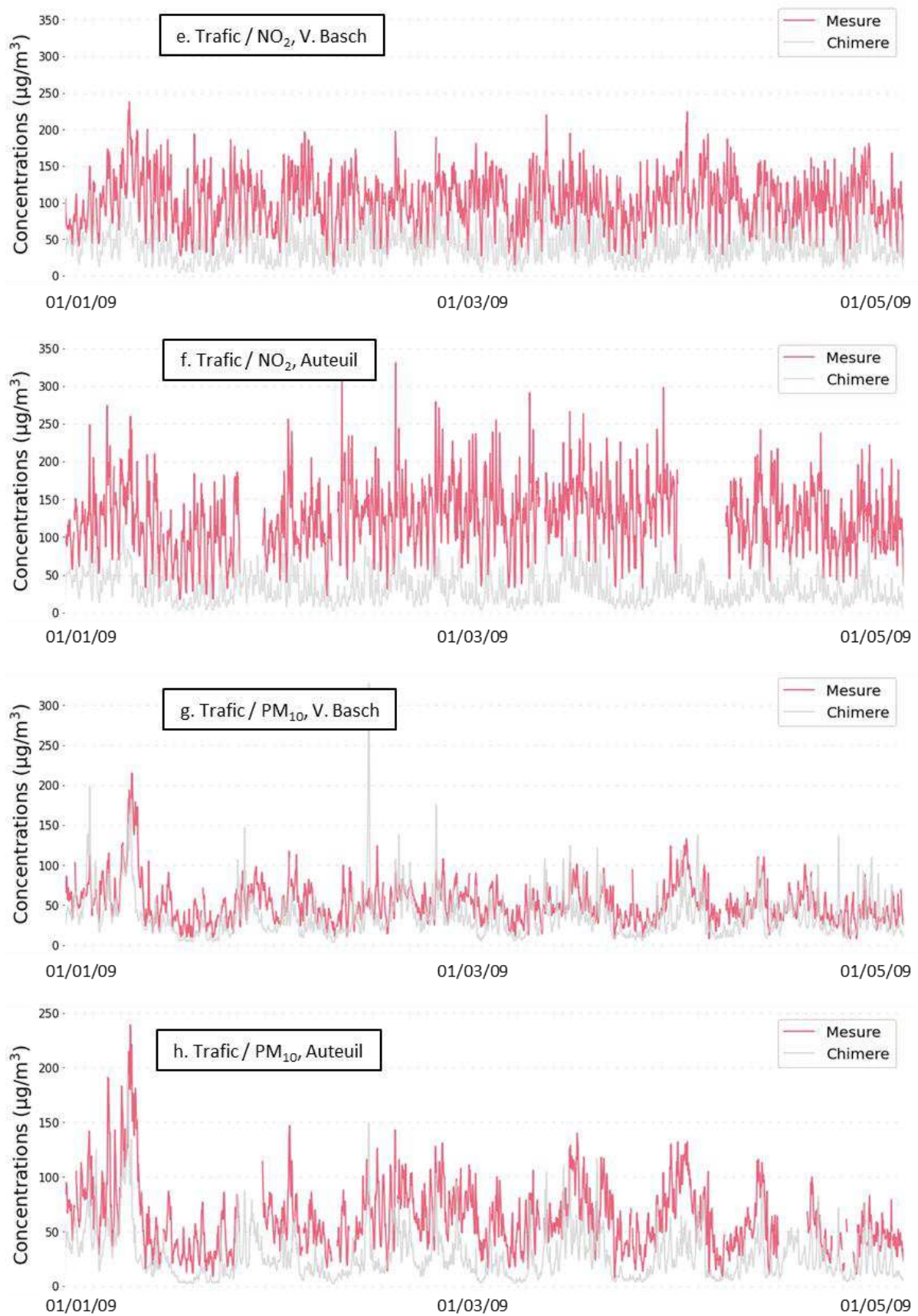


Figure 2 (e, f, g, h) - Comparaison des séries temporelles mesurées par le réseau Airparif (en rouge) et simulées par CHIMERE (en gris) sur des sites de mesures de proximité au trafic, pour NO₂ (en haut) et pour les PM₁₀ (en bas). Les concentrations sont exprimées en µg/m³.

Ce double constat m'a menée à proposer une formalisation pour la valeur CHIMERE corrigée (que j'appellerai *Cprox*), basée sur une amplification à la fois de la ligne de base et de l'amplitude des pointes de concentration :

$$C_{prox} = \alpha * \text{Ligne de base} + \beta * \text{Amplitude du signal} \quad \#(1)$$

Il fallait pour cela être en mesure de dériver, du signal simulé par CHIMERE, deux composantes représentant respectivement la ligne de base des concentrations, et l'amplitude horaire de variabilité des concentrations.

Ligne de base. Pour produire la ligne de base, j'ai considéré deux options :

- Utiliser la moyenne glissante des concentrations simulées dans chaque maille, qui permet de retirer au signal une part importante de sa variabilité horaire et de restituer l'évolution moyenne des concentrations. J'ai réalisé une étude visuelle afin de déterminer les fenêtres temporelles les plus pertinentes pour calculer la moyenne glissante. Après plusieurs tests, j'ai conclu que pour ne pas être trop influencée par les valeurs de pointe, la moyenne doit être calculée sur 12 à 24 heures. Les courbes correspondantes à ce calcul sont présentées en pointillés dans la Figure 3 et ont été retenues pour les tests d'apprentissage.
- Une seconde méthode que j'ai considérée est le calcul d'un percentile glissant des séries temporelles issues de CHIMERE. L'intérêt du percentile par rapport à la moyenne réside dans le fait qu'il relie les concentrations simulées les plus basses et n'est pas influencé par les valeurs absolues des concentrations de pointe. Comme pour la moyenne, j'ai fait plusieurs tests pour trouver la bonne fenêtre temporelle de calcul du percentile. J'ai retenu le calcul sur une période de 24h et j'ai testé plusieurs valeurs de percentiles (voir également la Figure 3). J'ai pu constater que les percentiles 20 et 25 étaient très proches et constituaient une bonne ligne de base du signal CHIMERE. A des valeurs de percentile plus élevées (courbe jaune), le signal commençait à intégrer des valeurs issues des pointes, ce qui n'était pas satisfaisant. Toujours dans le but de trouver le meilleur paramètre pour l'algorithme d'apprentissage, des tests ont été menés avec les courbes des percentiles glissants 20 et 25 sur 24 heures.

Les deux options ont été testées lors de la constitution de l'algorithme d'apprentissage. À chaque fois, le signal retenu est appelé « Bruit de fond » (*Bdf*) dans l'équation de calcul de *Cprox*.

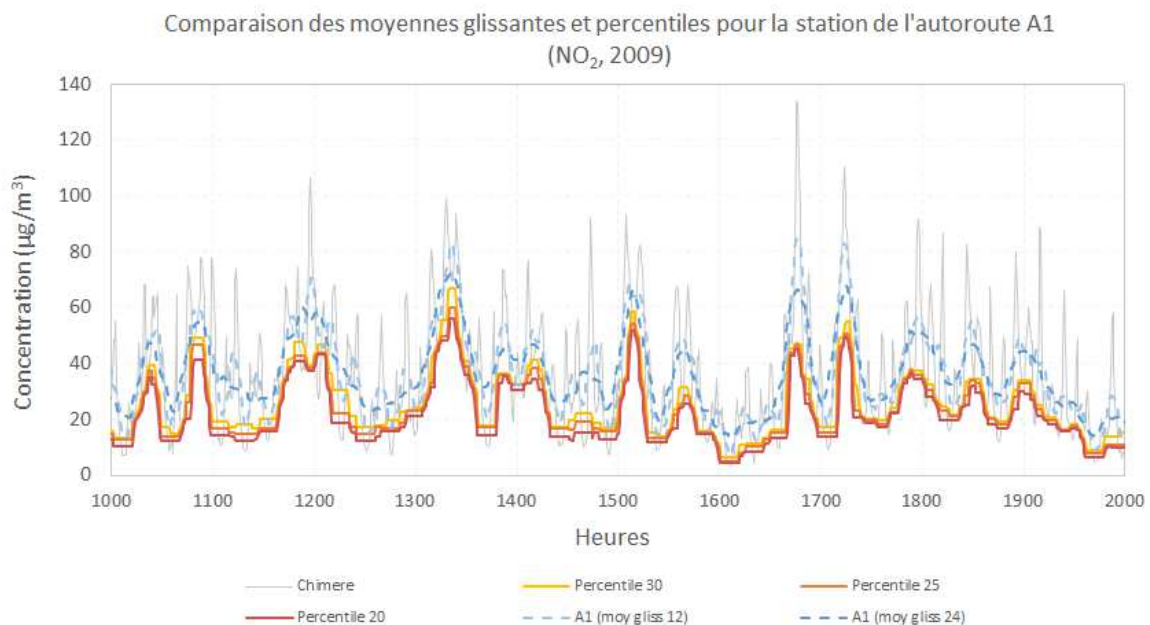


Figure 3 - Comparaison des moyennes glissantes (sur 12 et 24h) et des percentiles (20, 25 et 30) utilisées pour reproduire le bruit de fond du signal modélisé par CHIMERE au niveau de la station A1 pour NO₂ et en 2009.

Amplitude du signal. J'ai ensuite cherché quelle partie du signal CHIMERE pouvait être utilisée pour corriger l'amplitude des données simulées. Lorsque l'on considère les percentiles glissants comme bruit de fond, on peut utiliser l'écart au percentile à chaque heure comme un élément à amplifier puis à ajouter au bruit de fond. J'ai donc soustrait la valeur de *Bdf* au signal CHIMERE, en prenant garde de borner le résultat à zéro pour ne pas amplifier le signal vers le bas, voire vers des valeurs négatives. Avec la moyenne glissante comme bruit de fond, ce calcul n'était pas envisageable. J'ai décidé d'utiliser les données brutes issues de CHIMERE comme paramètre à corriger et à ajouter au bruit de fond. Quelle que soit l'approche choisie, le paramètre à amplifier est appelé ici « *Ampl* ». L'expression de *Cprox* s'écrira au final :

$$C_{prox} = \alpha * Bdf + \beta * Ampl \quad \#(2)$$

Séparation des données pour l'apprentissage et la validation

Conformément à la démarche décrite plus haut, je me suis appuyée sur les mesures aux différentes stations pour conduire la phase d'apprentissage. Il m'a tout d'abord fallu séparer les stations de mesure en un jeu d'apprentissage et un jeu de validation. Le découpage a été réalisé sur une sélection aléatoire des stations en activité avec la volonté de maintenir un surnombre léger mais systématique (plus de 50%) pour les stations d'apprentissage. Dès cette étape, j'ai rencontré une première limitation, relative au faible nombre de stations de mesures d'Airparif existant en 2009 à proximité du trafic (6 pour les PM₁₀ et 8 pour NO₂). Afin de garder un jeu d'apprentissage conséquent, je n'ai retenu qu'une seule station pour valider mon approche sur ces sites. En revanche, les stations de fond étant plus nombreuses (11 et 25 respectivement pour les PM₁₀ et NO₂), j'ai pu chaque fois garder une fraction significative (45%) des sites de mesure pour la validation de l'approche sans que cela appauvrisse de manière excessive mon jeu d'apprentissage. Le réseau de mesure étant en constante évolution, j'ai pu bénéficier de nouveaux sites pour le travail sur l'année 2017. L'assise de la validation de l'approche a ainsi pu être renforcée pour 2017, et s'appuyer chaque fois sur 40 à 50% des sites. Un résumé des noms de stations, de leurs typologies ainsi que de leur appartenance aux jeux d'apprentissage ou de validation est donné dans le [Tableau 4](#). Les noms de stations écrites en bleu correspondent à celles qui ont été installées après 2009.

Tableau 4 - Répartition des stations de mesure de NO₂ et de PM₁₀ en deux jeux d'apprentissage et de validation. Les noms des sites de mesure ayant été installés entre les années 2009 et 2017 sont écrits en bleu.

	Type	Jeu d'apprentissage	Jeu de validation
NO ₂	Trafic	A1, AUT, BASCH, BONAP, CELES, ELYS, RN6	RN2, BP_EST, HAUS, OPERA, SOULT
	Fond	MONTG, VERS, LOGNES, NEUIL, PA18, AUB, VILLEM, TREMB, MANT, CACH, CERGY, GARCH, NOGENT, PA13, PA04C	MELUN, VITRY, PA12, BOB, IVRY, CHAMP, STDEN, DEF, GON, EVRY, PA07
PM ₁₀	Trafic	A1, AUT, BASCH, ELYS, RN6	RN2, BP_EST, HAUS, OPERA
	Fond	LOGNES, VITRY, NOGENT, TREMB, GON, MELUN, RAMBO, PA15L, PA04C	GEN, DEF, BOB, PA18, CERGY

3. Mise en oeuvre de l'apprentissage

Les premiers tests d'apprentissage menés avec la formule #(2), ainsi que les choix pour la représentation du paramètre *Bdf*, et les aménagements que j'ai apportés à la formule, sont présentés et discutés ci-

dessous. Les résultats de ces tests seront pour l'essentiel illustrés par les résultats sur les sites de trafic, car ce sont ceux qui présentent les corrections les plus marquées. Mais les résultats obtenus sur tous les sites ont été pris en compte lors de ce travail. Les résultats de l'apprentissage dans sa version finalisée, conduits sur tous les sites, sont ensuite présentés et discutés.

Optimisation de la formule de C_{prox} : choix des paramètres d'entrée

L'objectif des tests d'apprentissage était de vérifier que je pouvais améliorer les données issues de CHIMERE à l'aide de coefficients de correction uniques pour chaque site. J'ai donc créé un programme de calcul des concentrations horaires C_{prox} sur toute une année, qui procédait par itération sur des gammes de valeurs prédéfinies pour α et β . Initialement, j'ai imposé une gamme de variation grossière pour ces coefficients (variation de 0 à 10 avec un pas de 1). Sur la base des premiers résultats, j'ai ramené cette gamme à l'intervalle [0;3] avec un pas de 0,1. J'ai ensuite dû définir quels indicateurs statistiques peuvent être utiles pour retenir les valeurs les plus pertinentes de α et β sur chaque station, c'est-à-dire pour produire la courbe C_{prox} la plus fidèle à la mesure à chaque site. J'ai considéré deux paramètres d'intérêt :

- la moyenne annuelle de chaque série de donnée C_{prox} , dont la comparaison à celle de la mesure permet d'identifier l'amplitude et le sens du biais éventuel du modèle. Ce paramètre est critique puisque la moyenne annuelle constitue le principal critère d'évaluation de l'exposition des individus aux particules et à NO_2 , et que cette valeur est fortement sous-estimée par CHIMERE près du trafic routier.
- l'erreur quadratique moyenne (ou RMSE) issue de la comparaison des données horaires C_{prox} avec la mesure. Ce paramètre mesure la somme des écarts, en valeur absolue, entre la courbe du modèle et la courbe de la mesure. Il est essentiel de retenir les plus basses valeurs de RMSE pour s'assurer que notre meilleure solution C_{prox} est réaliste à l'échelle horaire, et pas uniquement à l'échelle annuelle. La réduction de l'écart entre données CHIMERE et données d'observation, à l'échelle horaire, doit permettre de remédier au manque d'amplitude diurne des données modélisées précédemment constaté.

Pour chaque site, j'ai pu calculer le biais sur la moyenne annuelle ainsi que la RMSE pour l'ensemble des couples de coefficients α et β . J'ai mené un premier test en utilisant le biais sur la moyenne comme indicateur pour la sélection de α et β , et j'ai ensuite tracé les résultats obtenus. L'analyse visuelle a montré que les séries temporelles C_{prox} étaient trop peu bruitées, trop moyennes et avec une représentation insuffisante des concentrations élevées. En revanche, les jeux de coefficients obtenus par minimisation de la RMSE présentaient une évolution temporelle horaire beaucoup plus proche de celle constatée dans la mesure, avec une réduction du biais sur la moyenne qui s'est avérée significative. J'ai donc retenu ce mode de sélection des coefficients α et β .

Tous les tests ont été menés avec les différentes options pour le paramètre « bruit de fond ». Les scores obtenus avec les percentiles 20/25 et avec les moyennes glissantes 12/24h sur les sites de trafic sont présentés dans les [Figure 4](#) et [Figure 5](#) ci-après.

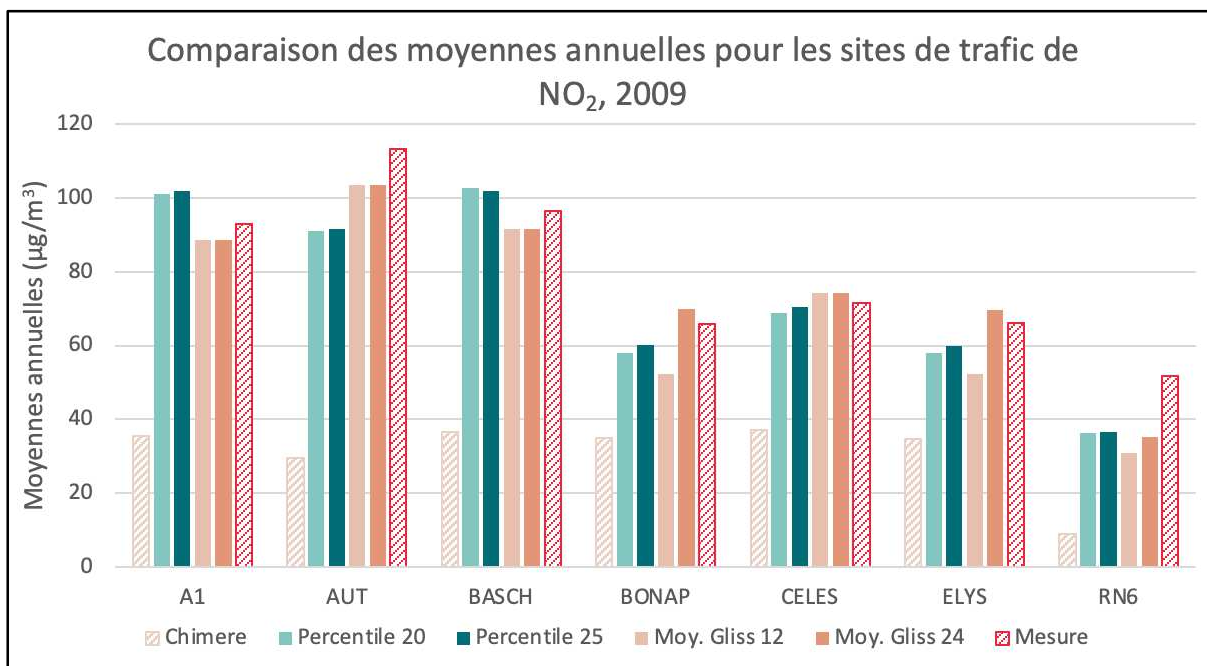
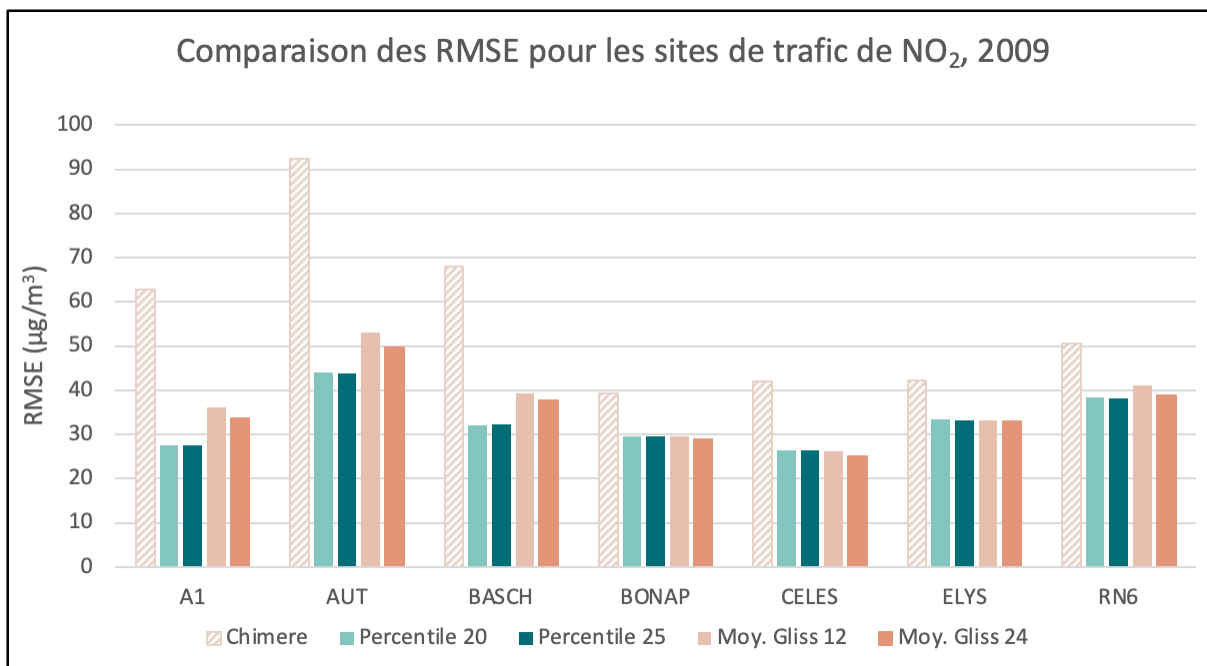


Figure 4 - Histogramme représentant les RMSE (en haut) et moyennes annuelles (en bas) de NO₂ (en µg/m³) sur chaque site telles que simulées par CHIMERE (beige hachuré), corrigées à partir des percentiles (en vert), corrigées à partir des moyennes glissantes (tons beiges) et issues de la mesure (rouge hachuré), en 2009.

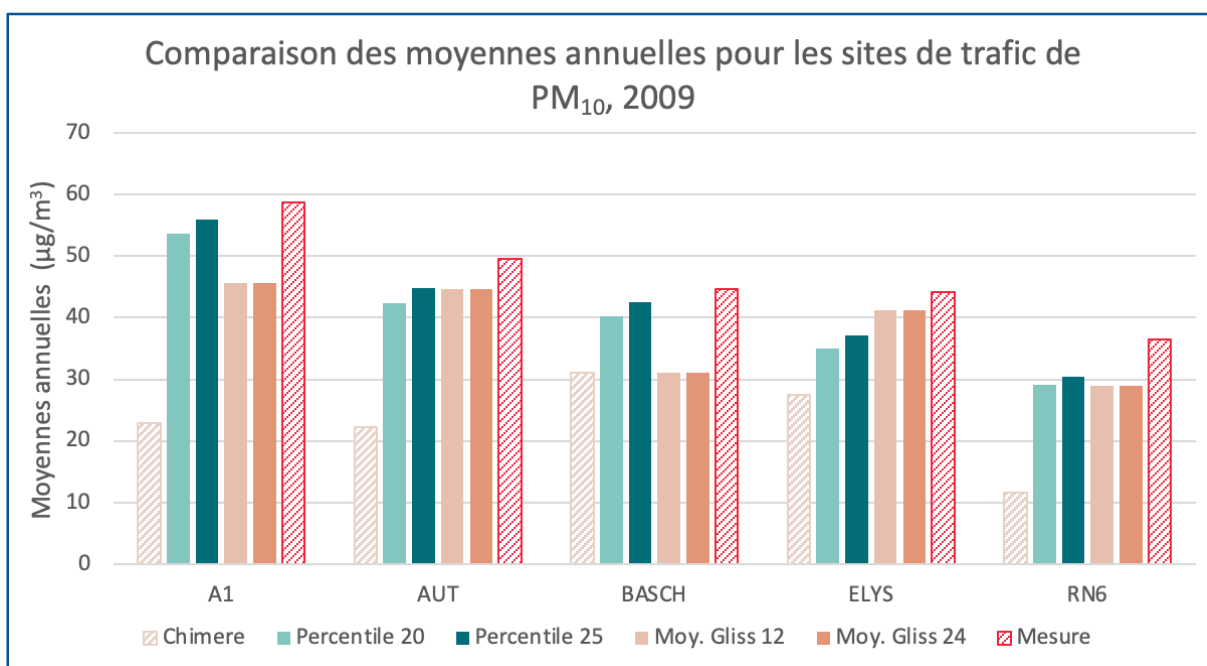
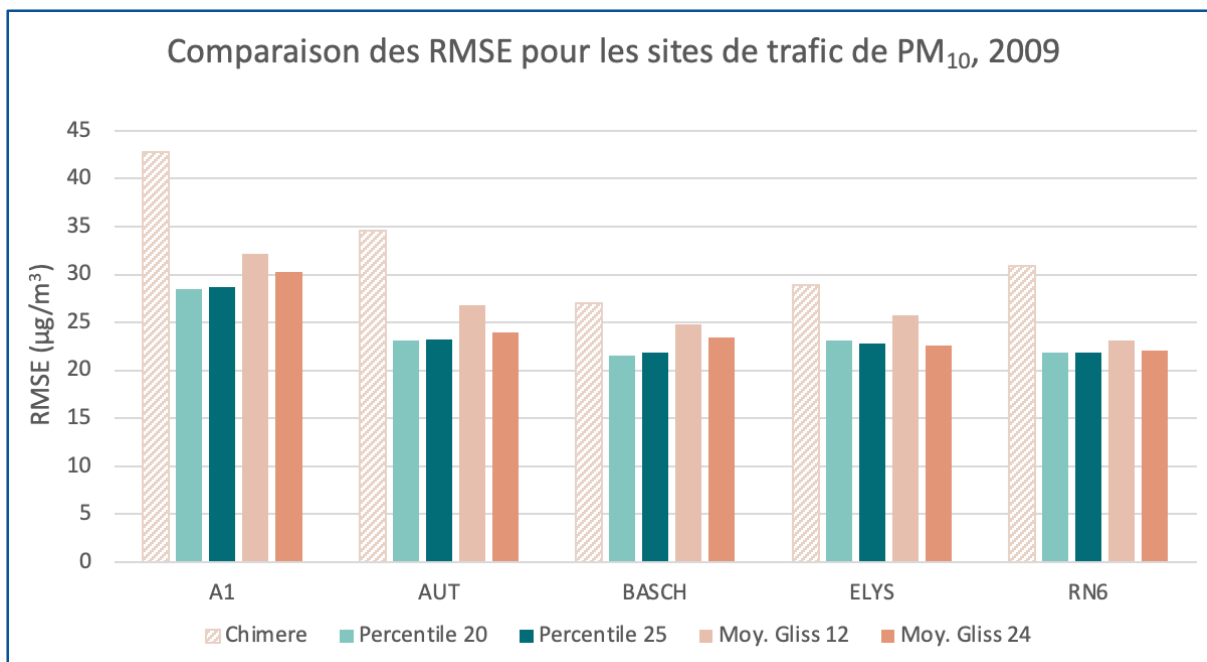


Figure 5 - Histogramme représentant les RMSE (en haut) et moyennes annuelles (en bas) de PM₁₀ (en µg/m³) sur chaque site telles que simulées par CHIMERE (beige hachuré), corrigées à partir des percentiles (en vert), corrigées à partir des moyennes glissantes (tons beiges) et issues de la mesure (rouge hachuré), en 2009.

La première information que délivrent ces figures c'est la forte diminution de la RMSE (histogrammes du haut) entre CHIMERE (à gauche en beige hachuré) et les différents calculs de *Cprox*. L'amélioration va de 25% à plus d'un facteur 2 pour NO₂, et elle atteint 20 à 35% pour les PM₁₀, sur la plupart de ces sites. En termes de moyenne annuelle simulée (histogrammes du bas) l'amélioration par rapport à CHIMERE est également majeure : les biais sur NO₂ diminuent d'au moins 20 µg/m³ et même de 50 µg/m³ sur certains sites. Pour les PM₁₀ le biais est réduit de 10 à 30 µg/m³ selon les sites.

Ces figures montrent par ailleurs qu'il y a peu de différences, et qu'elles ne sont pas systématiques, entre les résultats issus des différentes approches de calcul du bruit de fond.

Pour NO₂, il est difficile d'identifier des tendances pour les scores sur la moyenne et la RMSE, les meilleurs scores étant répartis entre les différentes approches. Pour choisir la meilleure approche, j'ai donc tracé les différentes séries temporelles *Cprox* ainsi que les données de mesure (non présenté ici). L'analyse visuelle de ces données a montré que l'utilisation des percentiles lisse trop le signal ce qui induit une trop faible variabilité horaire du signal corrigé. J'ai calculé la variance de tous les jeux de données *Cprox*, et les résultats confirment de mauvais scores avec l'utilisation des percentiles comme base de calcul de *Cprox*. Pour autant, si les scores de variance obtenus avec la moyenne glissante sont meilleurs - ce qui m'a amené à retenir cette approche - ils ne sont pas toujours satisfaisants et la variance de *Cprox* est souvent inférieure à celle de la mesure.

Pour les particules, aucune approche n'est réellement discriminante pour l'optimisation de la RMSE, mais on constate que celle basée sur les percentiles donne toujours des RMSE plus faibles. Par ailleurs, elle amène parfois de bien meilleurs résultats que la moyenne glissante sur la restitution de la moyenne annuelle mesurée, on le voit ici sur les sites A1 et BASCH. En outre, j'ai pu observer que - contrairement à NO₂ - l'utilisation de la moyenne glissante produit des séries temporelles avec une ligne de base trop variable, ce qui ressemble peu à la mesure (non présenté ici). L'inspection visuelle des séries temporelles m'a donc orientée vers l'utilisation des percentiles dans le calcul du bruit de fond pour les PM₁₀. Pour terminer, il m'a fallu choisir une valeur de percentile pour le calcul du bruit de fond. Les résultats obtenus avec P20 sont très proches de ceux obtenus avec P25 mais, même de très peu, ils sont presque toujours meilleurs. J'ai donc retenu le percentile 25 pour le calcul de *Cprox* pour les PM₁₀, mais l'impact de ce choix est probablement non significatif.

Les formules finales de l'équation de *Cprox* sont donc les suivantes :

$$Cprox(NO2) = \alpha * MG[CHIMERE](24h) + \beta * [CHIMERE] \quad \#(3a)$$

$$Cprox(PM10) = \alpha * P25[CHIMERE](24h) + \beta * DELTA(CHIMERE - P25) \quad \#(3b)$$

L'analyse visuelle des séries temporelles de mesure *versus* celles de *Cprox* a toutefois montré que *Cprox* possédait une variabilité insuffisante à l'échelle de la journée. J'en ai déduit que la sélection des paramètres α et β nécessitait de s'appuyer au moins partiellement sur un paramètre décrivant cette variabilité. J'ai choisi d'utiliser la variance, qui est constituée de la dispersion des points d'un échantillon. Et j'ai choisi de considérer, dans un même protocole de sélection de α et β , mes 3 paramètres d'intérêts (écart à la moyenne annuelle, RMSE et écart sur la variance) que j'appellerai ici « indicateurs » de qualité. Lors de mon itération sur α et β , j'ai donc stocké pour chaque couple de coefficients l'ensemble des valeurs des indicateurs de qualité du résultat dans une table. Puis j'ai classé successivement les valeurs de chacun de ces indicateurs par ordre croissant, et un rang leur a été attribué. Dans la table, chaque couple (α , β) était donc associé à 3 rangs. Il m'a ensuite suffi de moyenniser ces trois rangs et de sélectionner le couple amenant à la valeur de rang moyen la plus petite, afin de retenir celui offrant le meilleur compromis entre diminution des biais et amélioration de la RMSE. Les séries temporelles de *Cprox* obtenues via le tri des rangs présentent une amélioration de leur variance, tout en offrant toujours une réduction significative de la RMSE et du biais sur la moyenne, par rapport à CHIMERE. Par la suite, j'ai donc retenu ce protocole de sélection des coefficients. La [Figure 6](#) présente quelques résultats représentatifs de l'amélioration apportée par le calcul de *Cprox*, pour la restitution des concentrations de dioxyde d'azote et de PM₁₀ mesurées par AIRPARIF.

En haut, l'exemple donné pour NO₂ est celui du site de l'Autoroute A1 en début d'année 2009 (les heures de simulation correspondent à la période du 1er janvier au 11 février) : on peut voir que le bruit de fond *Cprox* (en vert), constitué par la ligne de base du signal, est surélevé par rapport à celui de CHIMERE (en gris) et qu'il est beaucoup plus proche de celui de la mesure. J'ai pu constater cette amélioration sur l'ensemble de l'année, et sur toutes les stations de proximité au trafic. Une autre caractéristique que

j'ai pu relever sur l'ensemble des sites de trafic est que *Cprox* reproduit assez bien la variabilité « basse fréquence » (*i.e.* à l'échelle de quelques jours) de la mesure : ce réalisme du signal *Cprox* est dû au fait que les sorties de CHIMERE prennent en compte le contexte d'émission, la dynamique et le transport à longue distance sur la région, ainsi que leur évolution d'un jour sur l'autre. *Cprox* apporte donc une réelle amélioration de la simulation de la qualité de l'air près du trafic même si la variabilité horaire de *Cprox* n'est pas encore pleinement satisfaisante. On peut en revanche distinguer des périodes pour lesquelles le signal de *Cprox* reste très faible par rapport à la mesure. Ce phénomène est par exemple constaté autour de la 400e heure de l'année sur cette figure, et il est observé à plusieurs reprises sur l'année. Il est d'ailleurs possible que ce phénomène survienne très régulièrement mais à des degrés divers, et que nous ne le constatons que lorsqu'il mène à une sous-estimation marquée des niveaux de concentration dans *Cprox*. Il semble donc que l'ensemble de l'année ne puisse pas être corrigé de manière identique, et qu'il faille introduire des paramètres supplémentaires dans l'équation #2.

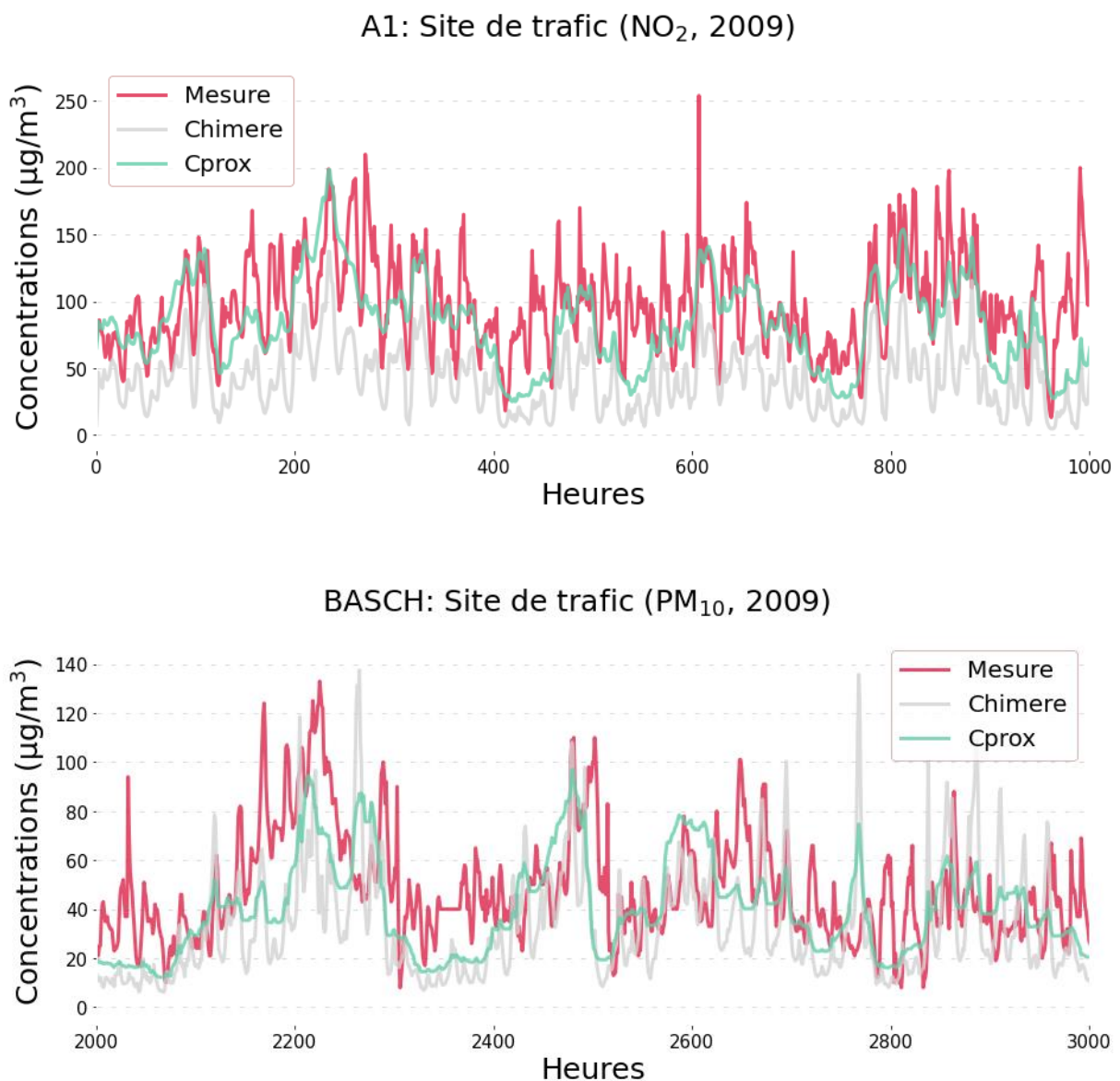


Figure 6 - Séries temporelles (i) CHIMERE en gris, (ii) *Cprox* en vert et (iii) issue de la mesure en rouge, sur la station trafic de l'Autoroute A1 pour NO₂ en haut et pour la station de la place Victor BASCH des PM₁₀ en bas, pour différentes parties de l'année 2009.

Pour les PM_{10} , j'ai représenté les données mesurées et simulées au rond-point de la place Victor Basch, pour la période allant du 25 mars au 6 mai 2009. La figure montre également une amélioration des données CHIMERE, sous la forme d'un rehaussement de la ligne de base du signal corrigé *Cprox* (en vert) par rapport au signal brut de CHIMERE (en gris). Cela a pour effet de réduire le biais et l'erreur quadratique du modèle. Toutefois, l'amélioration n'est pas aussi marquée que pour NO_2 , elle est même parfois de faible ampleur - comme on le voit sur les 100 premières heures de la période, ou entre les pas de temps 2300 et 2400. Un point remarquable de *Cprox* pour les PM_{10} en revanche, est la disparition des pics excessifs simulés par CHIMERE. L'exemple le plus parlant est celui du pic simulé vers l'heure 2800, et dont la concentration passe de $140 \mu g/m^3$ dans le CTM à seulement $75 \mu g/m^3$ dans *Cprox*, sachant que la plus forte valeur mesurée sur la journée est de $45 \mu g/m^3$. Toute la période qui suit, jusqu'au pas de temps 3000, illustre également parfaitement l'apport de *Cprox* pour l'élimination des pics non réalistes. L'utilisation du percentile glissant pour représenter le bruit de fond dans *Cprox* a toutefois tendance à fortement lisser le signal ce qui conduit - sur certains sites comme V. Basch - à une perte de variabilité horaire par rapport à ce qui était simulé par CHIMERE.

Optimisation de la formule de Cprox : prise en compte de l'effet du vent

Malgré ces résultats encourageants, la visualisation des séries temporelles *Cprox* montre des périodes durant lesquelles les concentrations corrigées restent très faibles par rapport à la mesure. Comme illustré en Figure 7, pour le site de mesure de NO_2 placé sur l'avenue des Champs-Élysées, ce phénomène est observé à plusieurs reprises au cours de l'année 2009 (périodes surlignées en jaune).

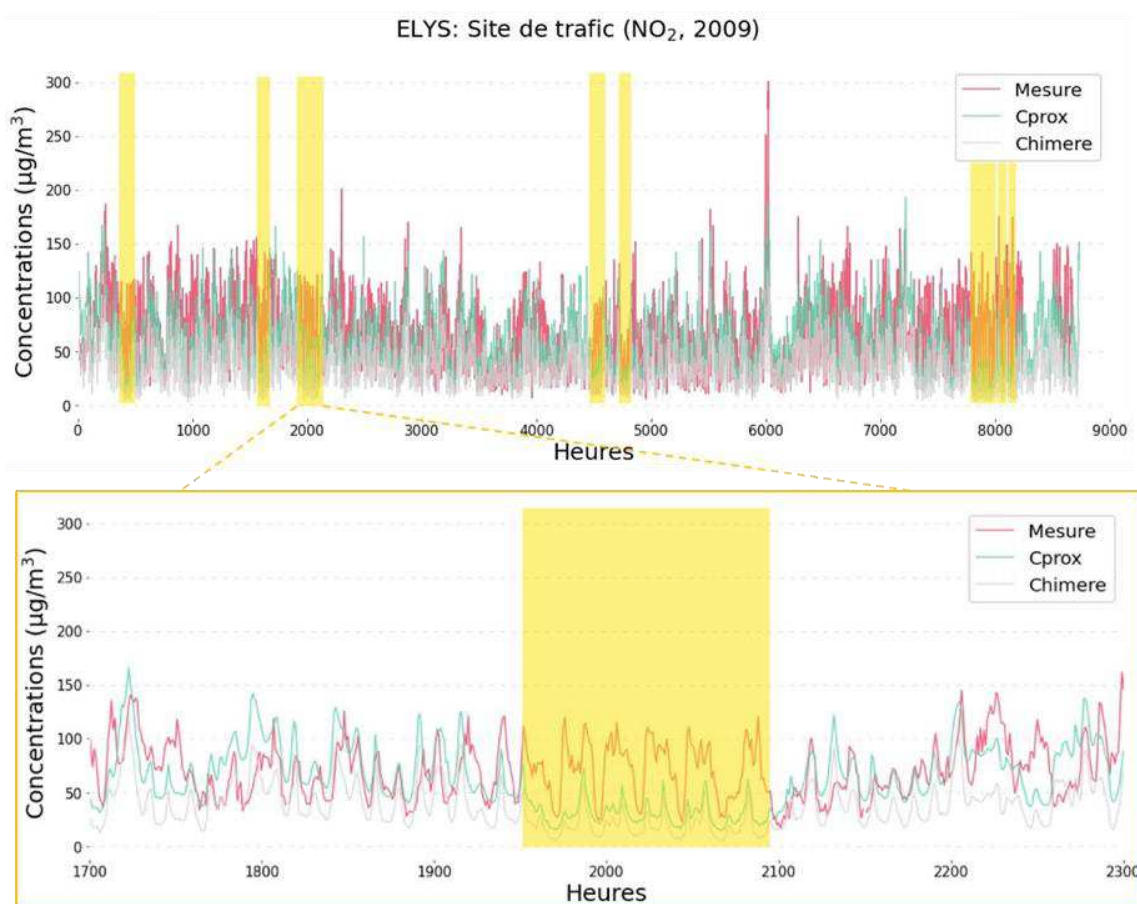


Figure 7 - Séries temporelles de concentration du NO_2 (en $\mu g/m^3$) à la station de proximité au trafic des Champs-Élysées, sur l'année 2009. Les courbes sont celles de CHIMERE (en gris), de *Cprox* (en vert) et de la mesure (en rouge). Les parties surlignées en jaune montrent les périodes de très grande sous-estimation de *Cprox*.

Sur cette figure, on voit d'une part que le signal de CHIMERE est plus faible sur la période en jaune que sur les jours environnants, et d'autre part que la valeur de *Cprox* reste particulièrement basse par rapport à la mesure malgré l'application des coefficients correcteurs. On remarque pourtant que durant ces événements, les niveaux de concentration mesurés sur les sites routiers restent élevés au cours de la journée. Ceci laisse penser que notre problématique est liée à un contexte - météorologique, de transport ou d'émission - mal restitué par CHIMERE. Ces événements sont aussi bien observés pour NO₂ que pour les particules. J'ai donc étudié les conditions dans lesquelles ces événements sont observés afin de voir s'il était possible de prévoir ces situations et de les prendre en compte dans l'équation de *Cprox*, afin de pouvoir moduler en conséquence ma correction.

Ma première piste de recherche a été celle des paramètres météorologiques, qui peuvent influencer sur la dispersion des polluants. Après analyse des données de WRF, j'ai pu constater que ces événements coïncidaient avec des épisodes de vent durablement plus élevé que les jours précédents ou suivants. Plus précisément, il s'agissait de vents avec des vitesses au-delà de 4 m/s et qui ne s'affaiblissent pas en journée, comme on le voit dans la Figure 8. L'effet observé est différent sur les sites de fond urbain. Durant ces périodes, les mesures en fond urbain sont beaucoup plus homogènes que les autres jours, avec des pointes beaucoup moins fortes mais un niveau de fond qui reste élevé. Et sur ces sites on n'observe pas de sous-estimation particulière dans les sorties de CHIMERE.



Figure 8 – Coïncidence entre les périodes de mauvaise restitution des concentrations (ici de NO₂, en µg/m³) par CHIMERE et par *Cprox*, et la vitesse du vent issue de WRF, sur le site de trafic des Champs-Élysées pendant le mois de mars 2009.

L'hypothèse que j'ai avancée pour la perte de réalisme des simulations est que, à partir d'un certain seuil, le modèle simule une dispersion horizontale excessive des polluants émis dans la rue. Les

obstacles urbains (bâtiments mais aussi végétation) n'étant pas représentés dans le modèle, il ne considère pas les effets d'accumulation ou d'écran qu'ils induisent localement. Sur les vitesses de vent concernées, qui restent tout de même modestes (environ 4-8 m/s), le modèle a ainsi tendance à prédire des niveaux de concentration assez faibles tandis que dans la réalité, le bâti urbain apparaît capable de freiner la dispersion des émissions. Pour expliquer les concentrations élevées toute la journée, on peut supposer que le vent permet de rabattre plus efficacement les émissions vers le trottoir où se fait la mesure, tandis que le bâti bordant la rue empêche l'évacuation de ces émissions, et favorise l'accumulation locale des polluants. Le module de correction des zones urbaines implémenté dans CHIMERE pour limiter ce problème - par une réduction des vitesses de vent au sol notamment - semble ici trouver certaines limites. Sur les sites de fond urbain, non soumis à l'influence directe du trafic routier et situés dans des environnements plus ouverts, il semble que ce soit l'effet de mélange avec l'atmosphère de fond qui prédomine - comme dans le modèle - et les concentrations résultantes mesurées et modélisées sont plus proches (Annexe 3).

Pour pallier ce problème, j'ai proposé d'ajouter une dépendance positive de la correction envers les conditions de vent, et ce uniquement sur les sites de proximité au trafic. Afin de ne pas faire apparaître d'effet de seuil sur C_{prox} en corrigeant plus fortement les concentrations au-delà d'une valeur limite de vent, j'ai opté pour une intensification progressive de la correction.

L'idée générale du facteur de correction sur le vent (CV) ajouté à C_{prox} (Équation 3) est donc d'avoir une correction s'amplifiant avec la vitesse du vent, qui puisse ainsi compenser progressivement le biais constaté précédemment. J'ai d'abord ajouté CV comme facteur multiplicatif de la correction totale, selon l'Équation 4 :

$$C_{prox} = [(\alpha * Bdf) + (\beta * Ampl)] * [1 + CV] \quad \#(4)$$

où Bdf et $Ampl$ sont définis conformément aux critères des équations #(3a) et #(3b).

Toutefois, au-delà d'un certain seuil de vent, il est probable que l'effet de dispersion des émissions l'emporte également dans la réalité, et renforcer continûment la correction de CHIMERE avec la vitesse du vent créerait des incohérences pour des vents très forts. Il fallait donc que CV ait une valeur significative uniquement sur une gamme de vitesses de vent modérées. J'ai donc conçu le paramètre CV comme une fonction de la vitesse du vent, avec une valeur nulle pour des vents inférieurs à 3 m/s, puis croissante jusqu'à une valeur (ou une gamme) seuil, et enfin rapidement décroissante au-delà de ce seuil. Pour fixer les limites de ces différents domaines, j'ai utilisé l'échelle de classification des vents de Beaufort. J'ai considéré *a priori* que la correction pourrait s'étendre sur l'ensemble des classes de « brise » et j'ai donc proposé (voir en Figure 9) que la valeur de CV augmente jusqu'à des situations de « brises modérées » - dont la limite supérieure est d'environ 7,5 m/s. Puis, entre 7,5 m/s et 10 m/s (ce qui correspond à la catégorie des brises les plus fortes), la correction est maximale mais ne dépend plus de la vitesse du vent afin de ne pas l'amplifier de manière excessive. Cette catégorie concerne à peine 3% des situations météorologiques rencontrées au cours de l'année 2009 sur Paris, mais j'ai pu constater que durant ces périodes, les teneurs en NO_2 étaient toujours fortement sous-estimées par C_{prox} , et nécessitaient une correction renforcée. Au-delà de 10 m/s, j'ai fait le choix de ramener rapidement la correction à zéro, car cela correspond à des situations de vent assez fort pour lesquelles il est envisageable que la dispersion des polluants soit efficace en ville. C'est une hypothèse que je n'ai pas pu évaluer car ces vitesses de vent sont observées à peine 1% du temps en 2009 sur la région parisienne, et constituent seulement la crête de quelques épisodes de vent. Pour ce qui est de la valeur absolue de CV , j'ai choisi de fixer le maximum de la fonction à 1 et de moduler son impact sur C_{prox} via des coefficients propres à Bdf et $Ampl$, comme on va le voir ci-dessous.

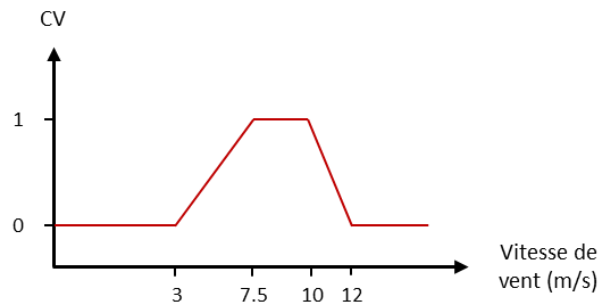


Figure 9 - Représentation graphique des valeurs prises par le facteur CV en fonction des vitesses de vent (m/s).

En effet, la question qui s'est posée est celle de la pertinence de l'application d'un coefficient identique de dépendance envers le vent, pour les deux termes de l'équation de *Cprox*. Il n'est en effet pas certain que le biais sur le vent dans le modèle ait également affecté la représentation des pointes et du bruit de fond urbain. J'ai donc réalisé une série de tests faisant varier - en plus de alpha et beta - et dans une gamme allant de 0 à 4 - deux coefficients entiers *x* et *y* intervenant dans le calcul de *Cprox* selon l'Équation #5 (où *Bdf* et *Ampl* sont définis selon les formules #3a) et #3b):

$$Cprox = [(\alpha * Bdf) * (1 + (x * CV))] + [(\beta * Ampl) * (1 + (y * CV))] \quad \#(5)$$

Les résultats de ces tests (illustrés en [Annexe 1](#)) montrent que, pour NO₂, on obtient les plus faibles valeurs de RMSE et de bons scores de variance avec le couple (1, 4), mais que les moyennes annuelles sont légèrement plus proches de la mesure avec le couple (1, 0). Pour les PM₁₀, le couple (1, 4) permet d'obtenir à la fois les meilleures RMSE et moyennes annuelles, ainsi qu'une variance ramenée à des valeurs très proches de celles de la mesure. C'est donc le couple (1, 4) qui a été retenu pour les deux polluants cible. Ce résultat signifie qu'en moyenne sur l'ensemble de l'année, la meilleure solution pour que *Cprox* soit proche des valeurs mesurées est d'appliquer, sur l'amplitude du signal, un facteur croissant avec le vent. L'application d'une correction uniquement sur le bruit de fond de NO₂ (couple (1,0)) permet en effet de rehausser la valeur moyenne du signal, mais sans réduire aussi efficacement l'erreur quadratique du modèle. Le travail itératif sur *y* a permis d'explorer des solutions de correction dans lesquelles un poids important est mis sur la correction du signal *Ampl*, et ces solutions, qui possèdent une variance plus réalistes, ont été retenues. Les améliorations apportées à *Cprox* par l'ajout de *CV* sont illustrées en [Figure 10](#).

Les séries temporelles de la [Figure 10](#) (en haut sans coefficient *CV*, en bas avec coefficient *CV*) montrent d'abord que l'ajout de *CV* est efficace pour corriger la très forte sous-estimation du NO₂ en période de vent (voir à gauche autour de l'heure 1600 sur A1) sans pour autant surestimer le signal le reste du temps. L'effet sur la restitution des concentrations de PM₁₀ est également visible sur de nombreuses périodes pourtant seulement modérément ventées : ce sont les heures 2000-2100, 2150-2200, 2300-2400 par exemple. Ce constat confirme que le biais lié au vent affecte le signal simulé tout au long de l'année, à des degrés variables. Au final, la prise en compte d'une correction croissante avec la vitesse du vent permet de trouver un meilleur accord avec la mesure tout au long de l'année : à la fois en termes de recouvrement des lignes de base des signaux, et en termes d'amélioration de la variabilité du signal - parce que le plus fort coefficient est appliqué à *Ampl*.

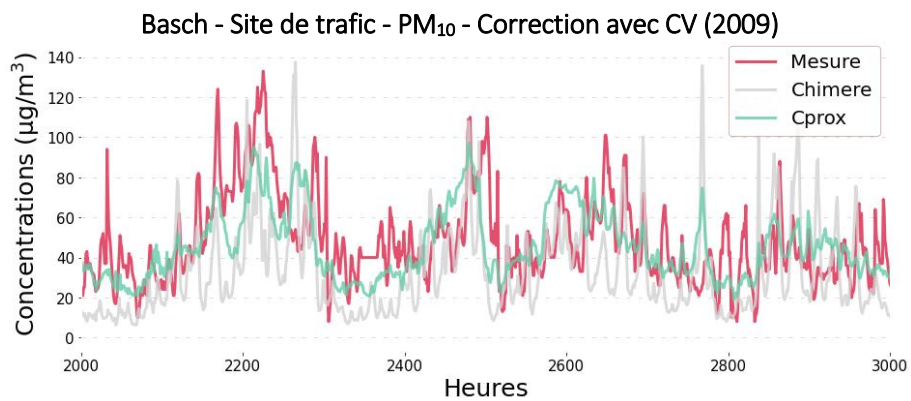
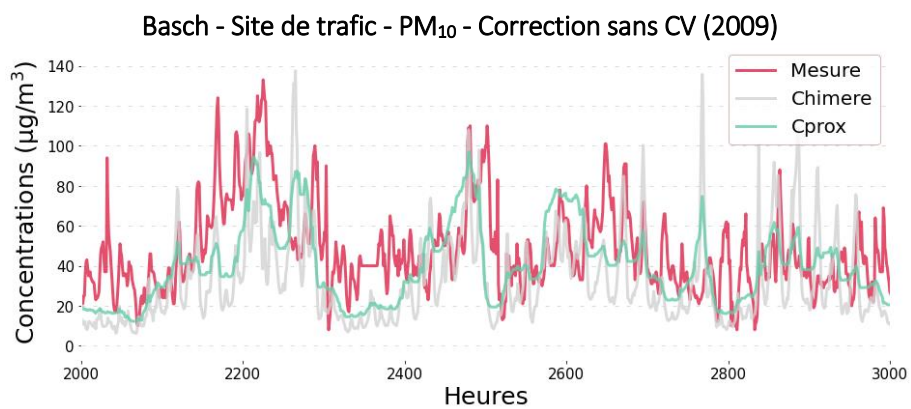
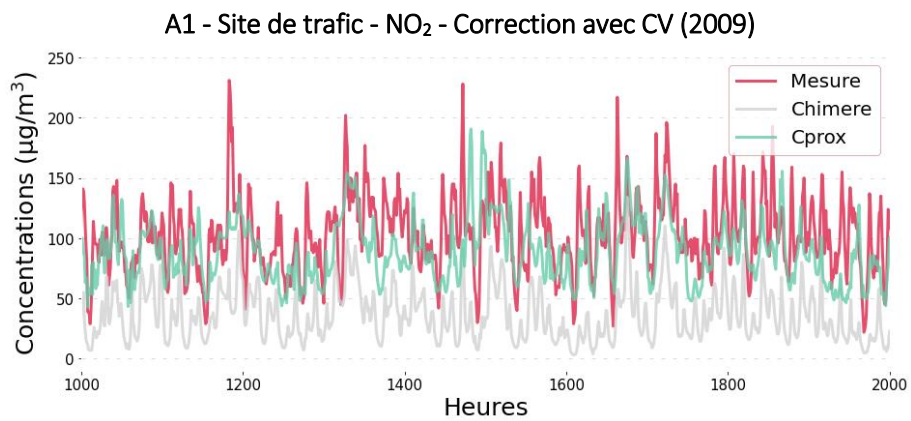
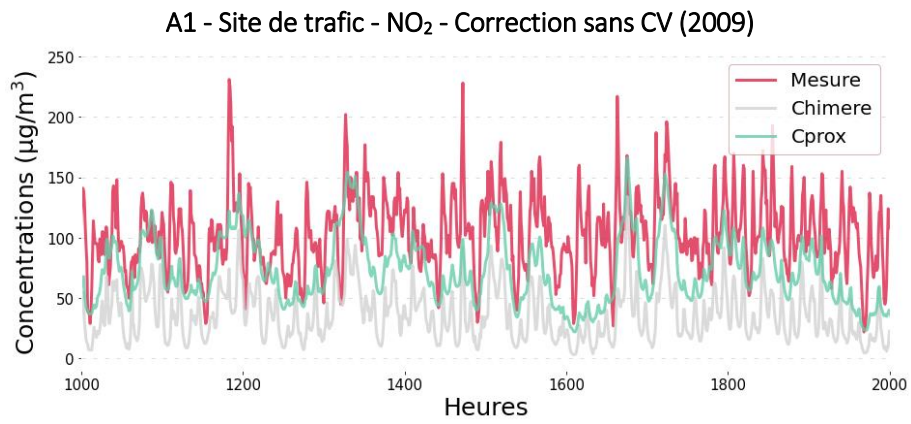


Figure 10 - Comparaison des séries temporelles *Cprox* sans CV (en haut) et avec CV (en bas) par rapport à CHIMERE (en gris) et à la mesure (en rouge), près de l'autoroute A1 pour NO₂ (à gauche) et sur le site Victor Basch pour les PM₁₀ (à droite).

Analyse des résultats de l'apprentissage

L'apprentissage de l'erreur du modèle a été mis en œuvre sur la base des formules #2) et #5), respectivement pour les sites de fond urbain et les sites de proximité au trafic de notre base de données d'apprentissage. Des boxplots pour la mesure, pour CHIMERE et pour *Cprox* ont été tracés pour chaque site (Annexe 3), afin d'apprécier l'amélioration apportée par la descente en échelle, et de comparer la distribution des valeurs de concentration de *Cprox* et de la mesure. Des résultats représentatifs de l'amélioration des séries temporelles sont présentés dans la Figure 11.a pour NO₂ et dans la Figure 11.b pour les PM₁₀. L'évolution de la variance des jeux de données simulés, par rapport à la mesure, est présentée dans la Figure 12.

Pour NO₂ en proximité au trafic (illustration sur 2 périodes de l'année, pour la station proche du périphérique à Auteuil), on constate que les niveaux moyens des séries temporelles *Cprox* se superposent avec les niveaux moyens mesurés. La sous-estimation de la mesure par le modèle passe en moyenne de 50 µg/m³ à moins de 3 µg/m³ après l'application de la correction aux sorties de CHIMERE. La RMSE est quant à elle réduite de 39% en moyenne. Il faut toutefois noter que la correction et les scores étant moyens sur l'année, on peut obtenir sur un même site des périodes de légère sous-estimation des pointes (comme on le voit entre les heures 3300 et 3400) dont l'effet est compensé par des périodes de surestimation des pointes (heures 7600 à 7800). On voit également que, grâce au contrôle de la variance lors de la phase d'apprentissage, on a pu redonner au signal *Cprox* une plage de variabilité horaire importante alors qu'elle était très faible par rapport à la mesure dans CHIMERE (Figure 12). Pour la plupart des stations de proximité au trafic, l'amplification attendue de la variance des données simulées est comprise entre +200 (µg/m³)² et +500 (µg/m³)². Et c'est sur le site d'Auteuil que ce phénomène est le plus marqué : la variance des données modélisées augmente de 366,8 à 1470,5 (µg/m³)² grâce à l'application de *Cprox*, en comparaison avec la mesure qui a une variance de 2094,1 (µg/m³)². Cette augmentation de variance est également visible sur le boxplot qui lui est associé (Figure 11.a), en particulier pour les valeurs situées entre les percentiles 25 et 75. On peut cependant remarquer sur ce boxplot que les valeurs les plus hautes de la distribution, comme les valeurs les plus basses, ne sont pas reproduites correctement par *Cprox*. On peut avancer deux raisons pour expliquer cela. Tout d'abord, la mise en œuvre d'une même équation sur l'ensemble de l'année impose que cette équation soit en mesure de réduire les écarts sur un maximum de données (effet de nombre), et la correction des valeurs extrêmes pèse trop peu sur les scores totaux. En outre, notre cible est la restitution de la moyenne annuelle (critère réglementaire pour l'exposition) et des valeurs les plus fréquentes en journée (contrôlées par la RMSE et la variance). Ni les indicateurs de qualité de notre apprentissage ni les paramètres de calcul de *Cprox* (moyenne glissante) ne ciblent les valeurs extrêmes : leur représentation dans *Cprox* n'est pas attendue.

Enfin, lorsque les données de CHIMERE sont très peu ressemblantes aux données de mesure, notre démarche peut échouer à représenter les données mesurées. Cela peut arriver pendant quelques jours sur un site, du fait d'une mauvaise restitution de la dynamique locale, ou d'un surplus d'émissions non prévu dans le cadastre, ou d'un import de pollution qui n'a pas été vu par le CTM. Ces situations peuvent difficilement être corrigées lors du calcul de *Cprox* car leur origine est très particulière et on ne peut pas cumuler les cas particuliers dans notre formule de correction. Mais cela peut aussi arriver lorsqu'une station est positionnée au sein d'une maille de calcul de CHIMERE très hétérogène. Il est alors difficile pour le CTM de restituer ce qui est observé au point de mesure. Notre approche n'est pas dimensionnée pour et n'a pas vocation à résoudre ces problématiques. Ainsi, les scores de *Cprox* sur le site de la route nationale 6 (« RN6 ») ne sont pas aussi bons que les autres (Annexe 2). Cela est lié au fait que les données de CHIMERE sont au départ excessivement faibles par rapport à la mesure. Il y a deux explications à cela. D'une part, le point de mesure est situé près d'une route à fort trafic, avec un passage important de camions. Or les spécificités du fret sur les voies ne sont pas incluses dans les cadastres qui, dans l'état de l'art, utilisent des pourcentages moyens de poids-lourds ou de véhicules utilitaires par type de voie. En outre, la maille du modèle dans laquelle cette station se trouve est une maille à fort caractère rural, peu émettrice en polluants routiers. Parce que le modèle calcul un flux moyen de

polluants par unité de surface sur chaque maille, il va alors sous-estimer les concentrations locales. La pertinence de notre correction dépendra donc de celle de CHIMERE au départ. Cela sera surtout problématique dans les mailles fortement hétérogènes, à la frontière entre urbain/routier et rural notamment.

Pour les stations de fond de NO₂. Les écarts entre les sorties de CHIMERE et la mesure sont beaucoup moins marqués pour les sites de fond que pour les sites de trafic et le chemin à faire pour corriger le modèle est dans l'ensemble beaucoup moins important qu'en proximité au trafic. En moyenne, les améliorations sur la moyenne annuelle sont de 4 µg/m³, et cela est suffisant pour obtenir un signal *Cprox* représentatif de la mesure. De même, la réduction sur la RMSE n'est que de 2 µg/m³ en moyenne pour l'ensemble des stations. Ces moyennes cachent pourtant des situations différentes. On notera ainsi que les données issues de CHIMERE pour la station du 18^e arrondissement parisien PA18 possèdent déjà un niveau moyen proche de celui de la mesure, avec des variations horaires déjà empreintes d'une forte variabilité, comme on l'observe dans la mesure. La courbe de *Cprox* est au final très fidèle aux observations, et presque superposée à celle de la mesure. Le boxplots des 3 séries montre toutefois que la correction amène une amélioration significative de la valeur moyenne annuelle. Inversement, pour des sites de fond périurbains comme celui de Mantes tout en bas de la [Figure 11.a](#), CHIMERE prédit localement un niveau de concentration en NO₂ et une variabilité de ces concentrations dans la maille qui sont très faibles par rapport à la mesure. La visualisation des séries temporelles avant/après correction montre que la courbe *Cprox* obtenue reste trop semblable à la ligne de base des concentrations mesurées et ne restitue pas les pointes de concentration liées au trafic. Comme précédemment discuté, on ne peut avoir trop d'attente lorsque le signal de départ est trop bas par rapport à la mesure. Toutefois, on doit noter les très bons scores de *Cprox* sur ce site, la correction permettant de restituer une valeur moyenne annuelle et une distribution annuelle des concentrations qui sont hautement représentatives de ce qui est mesuré. L'évolution des variances est plus mitigée dans le cas des sites de fond que pour les sites de trafic : 7 des stations présentent ainsi une augmentation de l'écart par rapport à la variance de la mesure, depuis CHIMERE jusqu'à *Cprox* ([Annexe 2](#)). Certaines de ces diminutions de la variance sont dues au fait que la correction va diminuer des pics trop forts dans le signal du CTM. La correction est alors bénéfique (exemple de la station d'Aubervilliers « AUB » en milieu d'année où le pic de CHIMERE en gris est diminué pour se rapprocher de la mesure). D'autres de ces diminutions sont, au contraire, un biais de la démarche qui cherche à restituer un signal satisfaisant, mais n'est pas contrainte par l'exigence de l'amélioration des scores lorsque l'on passe de CHIMERE à *Cprox*. Quand le signal original est très bon, la recherche d'un perfectionnement de la moyenne annuelle peut dégrader d'autres aspects du signal. Ce phénomène est notamment mis en avant par les boxplots des sites d'Aubervilliers, de Cachan, de Garches et de Paris 13^e pour lesquels on voit que CHIMERE représente très bien les niveaux de concentrations par rapport à la mesure, mais que l'application de *Cprox* a tendance à diminuer les valeurs des percentiles 5 et 95 des simulations ([Annexe 3](#)).



Figure 11.a - Comparaison des séries temporelles de CHIMERE (en gris), de *Cprox* (en vert) et de la mesure (en rouge) pour NO₂, en µg/m³, sur des périodes d'environ 1 mois durant l'année 2009. A droite des séries temporelles, des boîtes à moustache (même code couleur) montrent les minima, maxima, percentiles 25 et 75, moyenne et médiane de chaque série de données sur l'ensemble de l'année 2009.

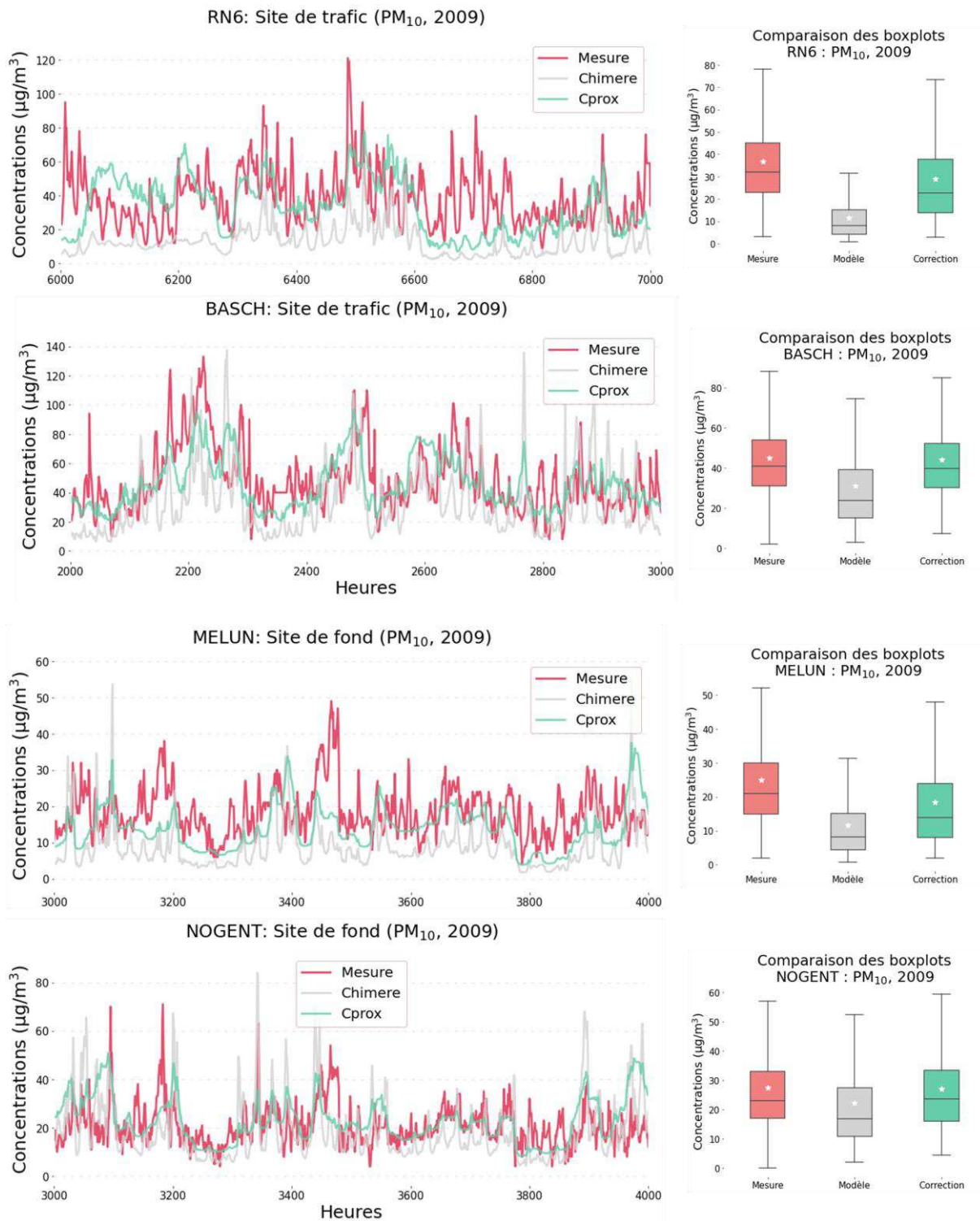


Figure 11.b - Comparaison des séries temporelles de CHIMERE (en gris), de Cprox (en vert) et de la mesure (en rouge) pour NO_2 , en $\mu g/m^3$, sur des périodes d'environ 1 mois durant l'année 2009. A droite des séries temporelles, des boîtes à moustache (même code couleur) montrent les percentiles 5 et 95 ainsi que les percentiles 25 et 75, les moyenne et médiane de chaque série de données sur l'ensemble de l'année 2009.

Pour les particules, sur les sites d'apprentissage, l'amélioration passe également par un rehaussement de la ligne de base des concentrations modélisées. Toutefois, la seconde problématique à résoudre n'était pas la faible variabilité des données CHIMERE, mais la prédiction de pointes trop fortes, avec une fréquence et une intensité variables selon les sites. La correction va donc chercher à réduire ces pointes,

ce qui est bénéfique pour le réalisme du calcul ultérieur d'exposition mais tend à réduire la variabilité intra-journalière des données modélisées. L'effet obtenu dépend de la fréquence initiale de ces pointes excessives. Toutefois, les scores de *Cprox* sont dans l'ensemble très fortement améliorés lors du passage de CHIMERE à *Cprox*.

Pour les stations de trafic de PM_{10} , on remarque très nettement la diminution des pics aberrants de CHIMERE pour se rapprocher de la mesure (comme on peut le voir sur le graphique, pour le site place Victor Basch) tout en assurant un rehaussement des concentrations moyennes le reste du temps. L'usage du percentile 25 comme ligne de base est un élément déterminant de cette réussite. Il contribue toutefois à aplatir la variabilité des données modélisées sur quelques heures, ce qui lui fait perdre un peu de sens à l'échelle de la journée, mais ce n'est pas pénalisant pour l'exploitation que l'on souhaite en faire. Les différences sur les moyennes annuelles entre la mesure et le modèle CHIMERE sont, en moyenne, de $23,9 \mu\text{g}/\text{m}^3$ et cette différence diminue à $4,8 \mu\text{g}/\text{m}^3$ avec *Cprox* (voir données en Annexe 2). De la même manière que pour le dioxyde d'azote, le biais sur la variance des concentrations de PM_{10} est intégré à l'ajustement de *Cprox*. Il est intéressant de noter que la correction permet autant de réduire que d'augmenter la variance selon le biais constaté au départ. Ainsi sur le site de BASCH où CHIMERE surestime de nombreux pics, la variance est réduite de près de $185 (\mu\text{g}/\text{m}^3)^2$ pour n'avoir, au final, que $10 (\mu\text{g}/\text{m}^3)^2$ d'écart avec la variance mesurée. A l'inverse, elle est augmentée sur le site porte d'Auteuil, et à peu près constante sur RN6. Les RMSE sont diminuées en moyenne de 32% après l'application de la correction aux signaux du CTM.

Sur les stations de fond pour les PM_{10} , les mêmes commentaires que pour les stations de fond de NO_2 peuvent être faits : la qualité de la correction d'un site à l'autre dépend en grande partie de la qualité de la simulation initiale. La correction est toutefois toujours pertinente, les RMSE diminuent de 25% et les différences entre les moyennes annuelles modélisées et mesurées passent de -7 à $-3 \mu\text{g}/\text{m}^3$ en moyenne. De même, un net gain au niveau de l'ajustement des variances est observé (Figure 12).

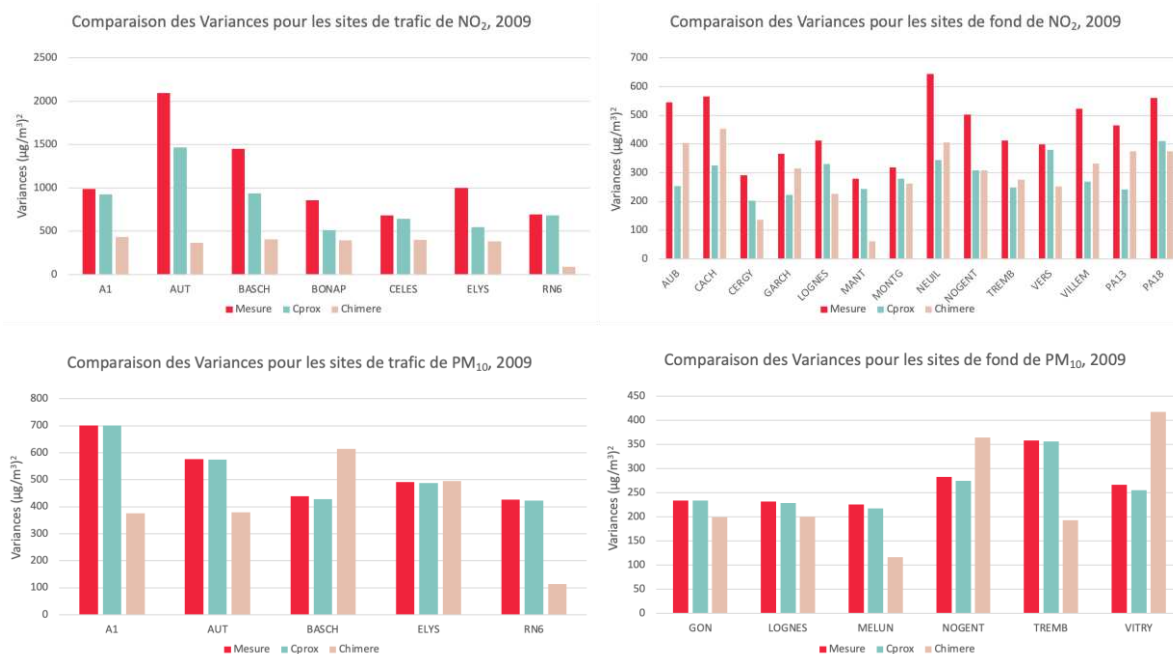


Figure 12 - Comparaison des variances pour les sites des jeux d'apprentissage de NO_2 (en haut) et PM_{10} (en bas) pour l'année 2009. Les variances issues des mesures sont représentées en rouge, celles issues de *Cprox* fité sont en vert et celles issues de CHIMERE sont en gris.

On peut conclure que l'application de *Cprox* aux sorties de CHIMERE permet de recréer les gradients de concentration entre stations de fond (péri-)urbain et de proximité aux voies, puisque la correction s'amplifie des sites de fond aux sites routiers, et que l'on recrée les principales caractéristique de niveau et d'amplitude des signaux mesurés. Au vu de l'amélioration globale des scores, j'ai considéré que les résultats des tests présentés ici étaient satisfaisants et que je pouvais me baser sur cette formulation de l'équation pour continuer le développement de l'approche.

La suite de la démarche consiste à universaliser les calculs pour prédire une correction sur les sites d'études non utilisés pour l'apprentissage. Il faut relier les corrections apportées par le modèle sur chacun des sites aux caractéristiques prédictibles du site, et définir une nouvelle formulation de la fonction *Cprox*. Cette étape m'amènera certainement à récupérer des bases de données supplémentaires contenant des données locales « invariables » (information sur le bâti ou sur les coefficients de confinement des rues par exemple) ou « variables » (flux de trafic, données liées aux conditions météorologiques par exemple).

4. Etape d'automatisation de la correction

Notre objectif est de proposer une correction qui soit statistiquement représentative de la qualité de l'air et des gradients de qualité de l'air dans l'espace urbain. La démarche de calcul de l'exposition que nous allons développer plus tard étant elle aussi de nature statistique, l'important est que nous retrouvions les caractéristiques, les tendances et la distribution des paramètres étudiés. Il ne s'agit donc pas de trouver la meilleure correction pour les différents quartiers du Grand Paris, et pour chacun des micro-environnements urbains routiers, mais plutôt de faire des compromis sur la qualité des résultats site par site, afin de privilégier une approche simple mais dans l'ensemble représentative, peu coûteuse en données et en calculs, et aisément applicable sur toute la région. Je devrai toutefois apprécier la qualité du résultat final.

Simplification des coefficients α , β et k

Tout d'abord, le fait d'avoir un couple de coefficient (α , β) par station dédouble le travail nécessaire pour leur extrapolation vers des sites non soumis à apprentissage. J'ai donc regardé si l'apprentissage faisait apparaître des tendances sur le lien entre les deux coefficients, par environnements type. J'ai pour cela tracé les rapports entre les coefficients et leur somme $k = \alpha + \beta$.

Les résultats sont reportés en [Figure 13](#). La figure met en avant plusieurs tendances dans la répartition des coefficients α et β . C'est sur les sites de PM₁₀ que l'on trouve la tendance la plus nette, avec une moyenne pour α supérieure à 80% dans les deux cas, et à peine 7% et 4% d'écart entre sa plus faible et sa plus forte fraction, respectivement dans les environnements routiers et de fond urbain. Dans le cas de NO₂, la variabilité des distributions est plus importante. En fond urbain, l'apprentissage propose pour la plupart des sites une part de correction sur le bruit de fond (alpha) qui s'échelonne entre 40 et 50%. En proximité au trafic, on observe plutôt deux tendances : une moitié des sites a une correction sur le bruit de fond également comprise entre 40 et 50%, et l'autre moitié entre 70 et 80%. J'ai considéré la possibilité d'imposer aux sites une partition moyenne de (α , β), soit (57,43) pour les sites de proximité routière et (48,52) pour les sites de proximité au trafic. Afin d'en évaluer l'impact, j'ai tracé la courbe de *Cprox* moyen pour le site de trafic d'Auteuil, dont la correction (77,23) était une des plus éloignées de la valeur moyenne. Les résultats, présentés sur la [Figure 13](#), montrent que - au sein d'une même formulation de *Cprox* - visuellement, le choix des coefficients n'altère pas beaucoup le réalisme de la courbe. La RMSE et la variance sont très légèrement dégradées n'est quasiment pas affectée, et la moyenne annuelle est améliorée de 7 $\mu\text{g}/\text{m}^3$. Ce résultat nous indique que les différentes valeurs du couple (α , β) donnent des résultats qui sont très proches les uns des autres, et que l'utilisation d'une moyenne pour tous les sites d'un même sous-groupe est une solution robuste. C'est ce que j'ai retenu pour tous les sous-groupes de mesure.

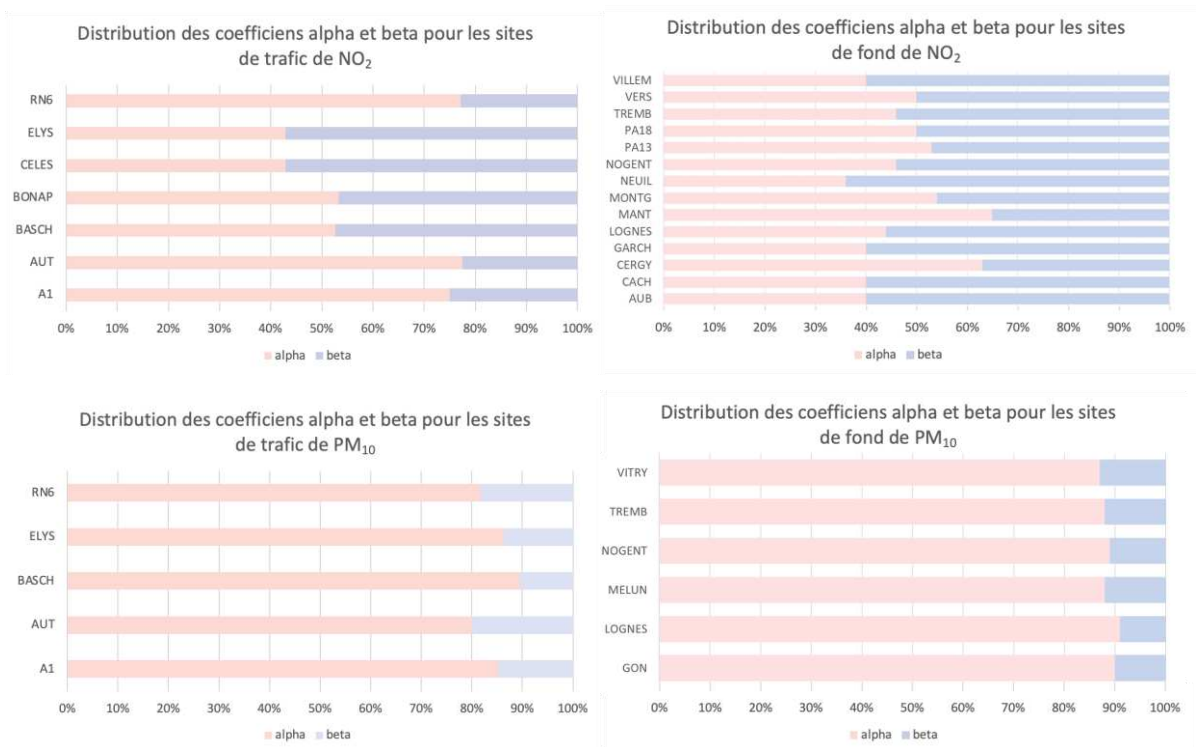


Figure 13 - Rapport des coefficients α et β en pourcentage pour les stations de trafic et de fond des jeux d'apprentissage du NO_2 et des PM_{10} pour l'année 2009.

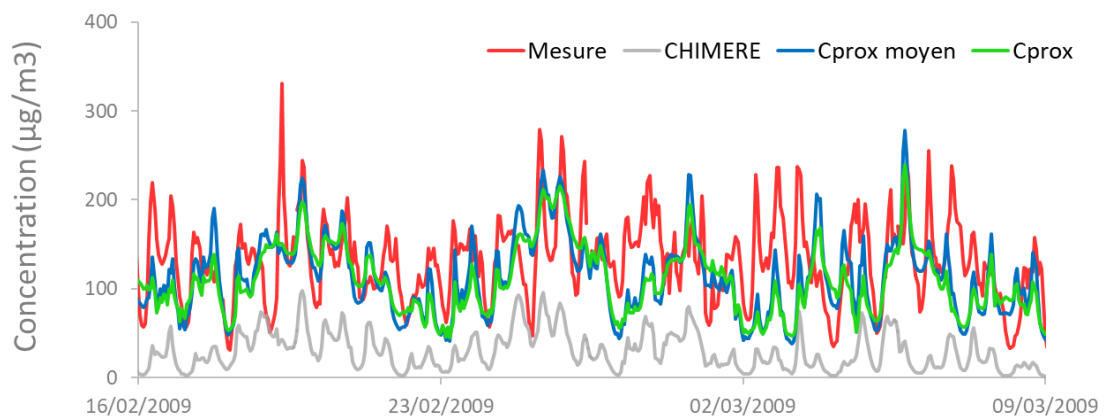


Figure 14 - Impact sur la courbe de C_{prox} de la partition alpha, beta, sur le site de trafic d'Auteuil, pour NO_2 sur une partie de l'année 2009. La courbe C_{prox} est obtenue avec $\alpha = 0,77$ et la courbe C_{prox} moyen avec $\alpha = 0,57$.

Par la suite, j'ai tracé en Figure 15 les valeurs de k pour chacun des sous-groupes de sites de mesure. Cette figure met en évidence une dispersion des valeurs beaucoup plus grande près des routes qu'en fond urbain, ce qui pouvait être attendu puisque les axes routiers diffèrent beaucoup de par leur dimension et leur fréquentation, alors que le fond urbain a vocation à représenter une teneur moyenne sur la zone urbaine. Le graphique montre ainsi la forte homogénéité des résultats obtenus pour le dioxyde d'azote en fond urbain, mis à part les résultats du site de Mantes-la-Jolie pour lequel nous avons identifié un biais particulièrement sévère dans les sorties de CHIMERE et dont la résolution est hors du cadre de mon travail - même si elle l'impacte. Au vu des résultats de ce sous-groupe, compris pour la plupart entre 1 et 1,5, il apparaît pertinent de considérer une valeur moyenne de correction k

ne tenant pas compte de l'environnement du site. J'ai donc retenu la valeur moyenne de 1,29 pour $k_{NO_2\ fond}$. Proposer une valeur moyenne de la correction k (qui serait de 1,95) pour les PM₁₀ en fond urbain est une hypothèse plus forte, et qui doit être discutée. En effet, les valeurs de k sont comprises entre 1,5 et 2,4, avec des valeurs plus fortes (autour de 2,2) pour les sites de grande couronne, et des valeurs plus faibles (autour de 1,7) pour les sites de petite couronne. Cet effet est lié au fait que CHIMERE sous-estime plus fortement les concentrations des polluants dans les mailles qui ne sont pas en majorité urbaines. Comme précédemment, j'ai tracé les différentes courbes en Figure 16, pour deux sites très éloignés dans la distribution de k . La figure montre que l'écart sur C_{prox} qu'induit l'utilisation d'un k moyen n'est pas significatif, que ce soit à la hausse ou à la baisse. Logiquement, on voit que les périodes de surestimation sur Vitry (vers le 20 février 2009 par exemple) pâtissent de cet effet de moyenne, mais la dégradation n'est pas majeure par rapport au biais déjà constaté sur cette période. Ces valeurs de k sont bien entendu liées à la qualité des sorties CHIMERE, donc à la qualité de l'inventaire de départ, et donc doivent être réappréciées lorsque l'on change d'inventaire et de terrain d'étude.

Les sites de trafic présentent quant à eux plus de dispersion (jusqu'à un facteur 2,3 pour NO₂) sur l'intensité de la correction à effectuer. Ceci, comme je l'ai mentionné plus haut, ceci peut être attendu. Par ailleurs, les concentrations mesurées sont élevées et cela rend encore plus critique le choix de k pour la prédiction de la qualité de l'air près des axes routiers. On voit ainsi sur la Figure 17 que l'utilisation d'un facteur moyen pour k affecte sévèrement la qualité (moyenne et valeurs de pointe) de la correction faite pour NO₂ sur le site des Célestins à Paris. Je n'ai donc pas considéré l'utilisation d'une valeur moyenne de k , mais je me suis attachée à identifier quel paramètre local présentait une co-variabilité avec la valeur de k et pouvait être inclus dans la formule de C_{prox} . Ce travail fait l'objet de la section suivante.

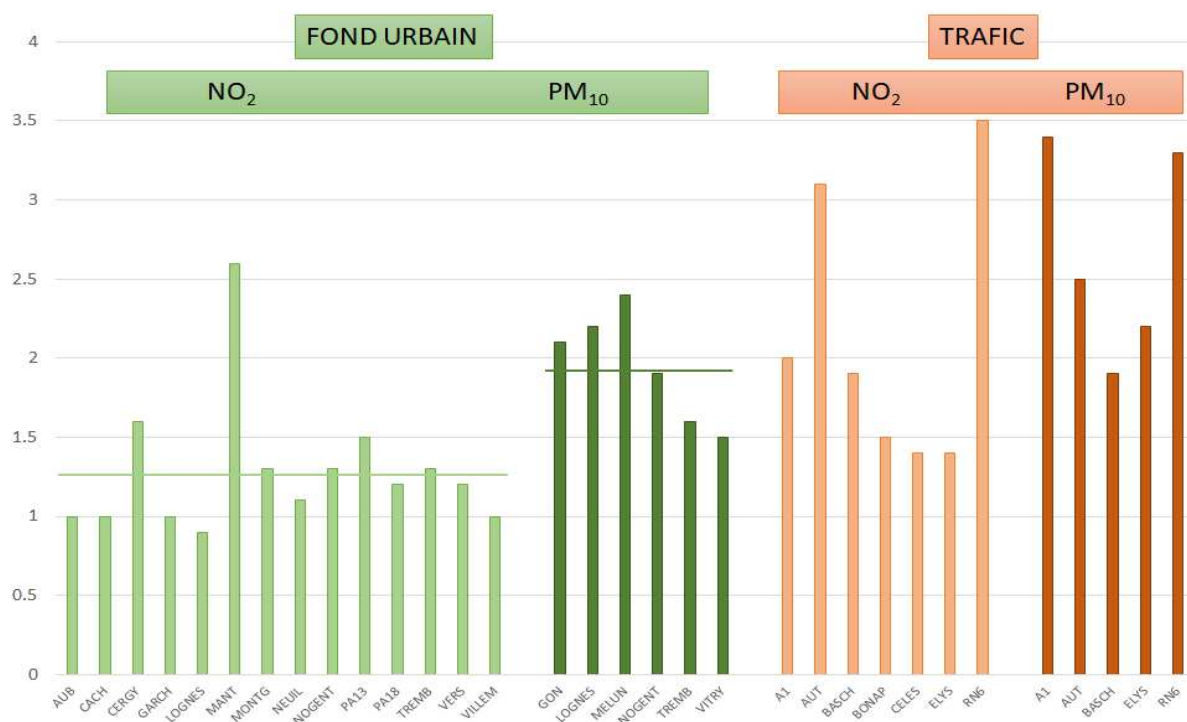


Figure 15 - Valeurs de k issues de l'apprentissage, sur les différents sous-groupes de mesure pour le dioxyde d'azote (couleurs claires) et pour les particules PM₁₀ (couleurs foncées) : en fond (péri-)urbain (en vert), et en proximité au trafic (en orange).

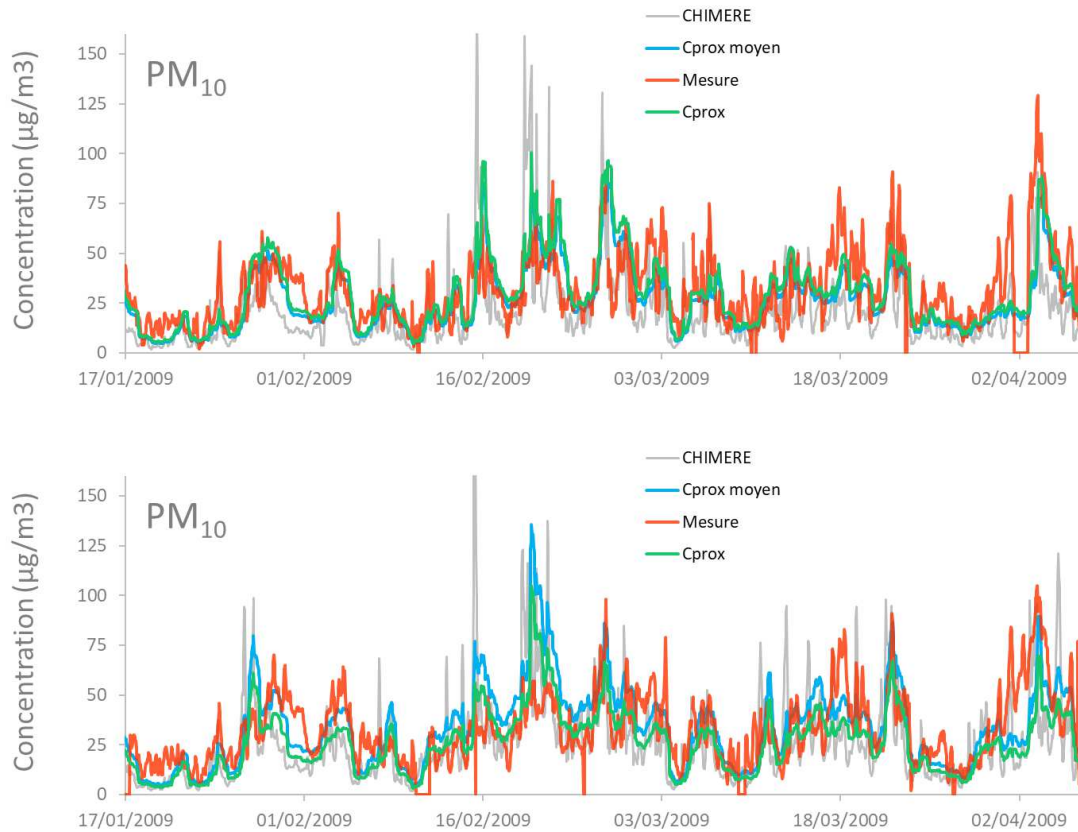


Figure 16 - Impact de la moyenne des valeurs de k issues de l'apprentissage, sur les courbes de C_{prox} de particules fines en fond périurbain et urbain. En haut, courbes de CHIMERE en gris, de la mesure en rouge et de C_{prox} (issu de l'apprentissage en vert et moyen en bleu) sur deux sites d'environnement différents : Melun en haut et Vitry-sur-Seine en bas.

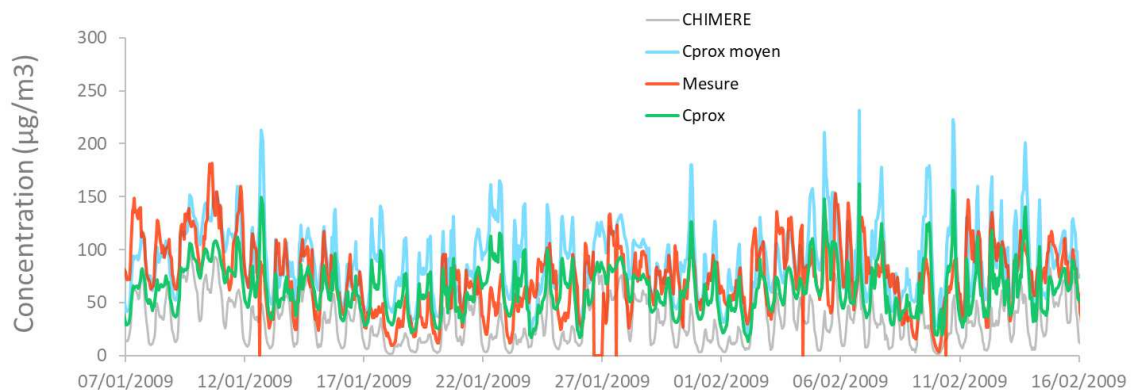


Figure 17 - Impact de la moyenne des valeurs de k issues de l'apprentissage, sur les courbes de C_{prox} de dioxyde d'azote en proximité au trafic : courbes de CHIMERE en gris, de la mesure en rouge et de C_{prox} (issu de l'apprentissage en vert et moyen en bleu) sur le site Célestins (quais de la Seine).

Au final, le signal corrigé C_{prox} s'écrira donc sous la forme de l'Equation 6a pour les stations de trafic et sous la forme de l'Equation 6b pour les stations de fond :

$$C_{prox} = k_{site} * \{[(Bdf * \% \alpha) * (1 + CV)] + [(Ampl * \% \beta) * (1 + (4 * CV))]\} \quad \#(6a)$$

$$C_{prox} = k_{fond} * \{[(Bdf * \% \alpha) * (1 + CV)] + [(Ampl * \% \beta) * (1 + (4 * CV))]\} \quad \#(6b)$$

où Bdf et $Ampl$ sont définis conformément aux critères des équations #3a) et #3b), où k_{fond} prend les valeurs de 1,29 et de 1,95 pour NO_2 et pour les PM_{10} respectivement et où k_{site} reste à prédire pour chaque site et chaque polluant.

Proposition de coefficients k site-dépendants en proximité au trafic

J'ai mené un travail de réflexion sur la méthodologie à adopter pour déterminer les coefficients k de toutes les zones d'Île-de-France situées à proximité de la voirie. J'ai identifié deux paramètres, en lien avec l'environnement des stations, susceptibles d'affecter localement l'erreur de CHIMERE lors de la simulation des espaces routiers, et qui pourraient donc être corrélés aux coefficients k sur chaque station d'apprentissage.

Le premier paramètre que j'ai jugé d'intérêt dans la démarche suivie est la morphologie urbaine dans l'environnement immédiat de chaque site de mesure. En effet, pour un même niveau d'émission, le fait qu'une rue soit étroite et enserrée entre des immeubles hauts va fortement influencer sur les niveaux de concentration qui s'y établissent. J'ai donc travaillé à la possibilité de corrélérer les degrés de confinement des rues avec les corrections nécessaires sur le modèle. Ce travail a été mené en collaboration avec une étudiante de Master. J'ai ensuite déterminé un second paramètre en me basant sur le fait que les erreurs d'un CTM sont en grande partie dues à une dilution horizontale forcée des polluants émis dans une maille, jusqu'aux limites de la maille en question. Si les émissions sont diluées dans un environnement riche en émission, l'effet de dilution sera faible. Si les émissions sont diluées dans un environnement peu émetteur, l'effet de dilution sera important. La capacité du modèle à restituer les gradients d'émissions observés pourrait dans ce cas dépendre du rapport entre (i) les émissions dans l'environnement immédiat du site de mesure et (ii) la totalité des émissions dans la maille. Ce paramètre a été étudié et mis en relation avec k .

Ces travaux sont présentés dans les deux sections qui suivent.

a. Environnement morphologique des rues

Comme expliqué plus haut, la morphologie urbaine environnant un site d'étude possède un rôle majeur sur l'accumulation ou la dispersion locale des polluants. Nos premières réflexions sur l'approche à suivre pour formaliser cette morphologie nous ont menées à considérer deux indicateurs, que j'ai appelés « proxys », exprimant le degré de confinement des rues.

Le premier proxy est l'effet de rue canyon. L'effet de rue-canyon est exprimé grâce au rapport entre la hauteur des bâtiments encadrant la rue et la largeur de la rue. Il est souvent exprimé dans la littérature par l'équation de l'« *aspect ratio* »:

$$\text{Aspect Ratio} = \frac{H}{W}$$

où H représente la hauteur du bâti bordant la rue canyon (en mètres) et W représente la largeur de la rue (en mètres). La rue est considérée comme « rue canyon » lorsque son aspect ratio est supérieur à 1 et comme rue ouverte lorsqu'il est inférieur à 1 (Vardoulakis et al., 2003). Pour l'Île-de-France, les données de hauteurs de bâtiment et de largeur ont pu être obtenues via la BDTPO qui est une base de données de l'IGN regroupant des informations topographiques d'un domaine donné.

L'objectif final était de cartographier la région en calculant l'*aspect ratio* moyen sur des mailles de $100 \times 100 m^2$ couvrant tout le domaine francilien, de façon à nous permettre de distinguer des gradients d'effet, depuis les zones les plus confinées vers les zones les plus ouvertes. Rapidement, la mise en œuvre de cette méthode s'est avérée très consommatrice en temps de calcul ce qui la rendait peu intéressante. Nous avons toutefois calculé les *aspect ratio* locaux sur chacun des sites de trafic routier, mais aucune relation ni tendance n'a pu être mise en avant entre ce paramètre et les coefficients k issus

de l'apprentissage, que ce soit pour le dioxyde d'azote ou pour les particules. L'hypothèse que l'on peut faire sur cette absence de lien est que ce paramètre H/W est trop spécifique et comporte une trop petite partie de l'explication de l'erreur modèle. Si on avait voulu intégrer une réflexion sur le confinement, il aurait été important de se poser la question de l'orientation des rues par rapport aux directions des vents dominants, considérer le fait que les hauteurs de bâtiment de part et d'autre des rues ne sont pas toujours homogènes, et pouvoir prendre en compte les intersections. Mais cela nous obligeait à rentrer dans une approche de type déterministe incompatible avec les ambitions de simplicité de la démarche entreprise.

Le second proxy est la notion de *Local Climate Zones*. Une autre approche permettant de caractériser les environnements urbains en fonction de leur capacité à accumuler ou disperser la pollution a été considérée. Nous nous sommes inspirées d'une méthodologie développée par [Stewart & Oke, 2012](#) nommée « **Local Climate Zones** » (LCZ). Plutôt que de caractériser les environnements urbains suivant des mailles régulières, cette méthode travaille plutôt sur des îlots tels que considérés dans les études de climatologie urbaine pour la détermination des îlots de chaleur urbains ([Bernard et al., 2018](#)). Les entités géographiques utilisées qui supportent ces LCZ sont les Îlots Morphologiques Urbains (IMU), définis au sein d'un référentiel géographique numérique par l'Institut Paris Région (IPR), anciennement Institut d'Aménagement et d'Urbanisme de la région Île-de-France ([IAU, 2016](#)). Les LCZ ont vocation à définir le taux d'urbanisation d'un îlot en distinguant les environnements végétalisés tels que les parcs et jardins. Mais de nombreux indicateurs urbains sont nécessaires pour définir avec précision à quelle classe de LCZ appartient un îlot. Nous avons utilisé les données de l'IPR relatives à chacun des IMU pour calculer les classes de LCZ auxquelles ils appartenaient. Nous avons ensuite utilisé une version simplifiée du tableau de détermination des zones de [Stewart & Oke, 2012](#) permettant de ne retenir que des grandes catégories jugées pertinentes pour notre démarche. Les quatre indicateurs retenus pour l'attribution des LCZ sur la région parisienne sont présentés ci-dessous, et résumés avec leurs valeurs dans le [Tableau 5](#) :

- Le mode d'occupation des sols (MOS) est utilisé pour suivre et analyser, à une certaine échelle, l'évolution du territoire francilien. Il permet d'identifier les IMU. La version simplifiée du MOS fourni par l'IPR contient 11 types d'occupation des sols appelés « postes ». Nous avons considéré 6 des 11 MOS pour cette approche. Il s'agit du bâti résidentiel (MOS 1) des autres types de bâti (MOS 2), des espaces ouverts artificialisés (MOS 3), des forêts et milieux semi-naturels (MOS 4), des espaces agricoles (MOS 5) et de l'eau (fleuve, lac... MOS 6).
- L'Aspect Ratio « AR » qui est, ici, définie de manière différente que dans la section précédente. $AR = \frac{(ES+DB)}{(SR+1)}$ où *ES* représente le coefficient d'emprise au sol, *DB* représente la densité du bâti et *SR* représente la surface des routes. Toutes ces valeurs sont tirées de couches cartographiques de l'IPR.
- La Hauteur de Rugosité « HR » (en mètre) indique la hauteur moyenne des bâtiments ou de la végétation de chaque îlot. Ce paramètre est déterminé selon le rapport entre le volume des masses construites et la surface au sol qu'elles occupent.
- La Surface Construite « SC » (en %) indique le pourcentage de bâtiments par îlot. Elle est déterminée selon la couche de surface au sol des bâtiments et la surface des IMU.

Tableau 5 – Résumé de la caractérisation des IMU en LCZ.

MOS	HR	AR	SC	LCZ
1 et 2	> 25	> 2		LCZ1
	> 25	< 2		LCZ4
	10 < HR < 25	> 0,75		LCZ2
	10 < HR < 25	> 0,75		LCZ5
	3 < HR < 10	0,75 > HR > 3		LCZ3
	3 < HR < 10			LCZ6
	< 3			LCZ7
			> 25	LCZ8
			< 25	LCZ9
3			LCZE	
4			LCZA-B-C	
5			LCZD	
6			LCZG	

La classification des LCZ a été réalisée sur l'ensemble du domaine francilien et est cartographiée dans la [Figure 18](#) ci-dessous. Cette classification du domaine ne nous permet pas de récupérer d'information sur le type d'environnement entourant chaque rue dans laquelle nous voulons prédire la qualité de l'air, mais sur le degré de confinement d'un îlot.

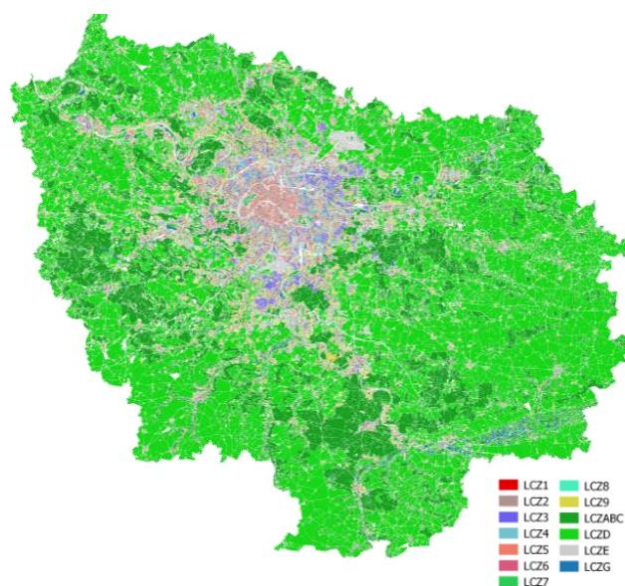


Figure 18 - Représentation cartographique des LCZ de la région Île-de-France.

On peut voir sur cette figure un gradient d'urbanisation qui s'intensifie lorsque l'on s'approche du centre du domaine. On note que toutes les LCZ déterminées par [Stewart & Oke, 2012](#) ne sont pas représentées sur la région.

Tableau 6 - Caractérisation des stations d'apprentissage de trafic en fonction des LCZ dans lesquelles elles se trouvent.

Station	A1	AUT	BASCH	BONAP	CELES	ELYS	RN6
LCZ	LCZ2	LCZE	LCZ2	LCZ2	LCZ2	LCZE	LCZ5
k_{NO_2}	2	3,1	1,9	1,5	1,4	1,4	3,5
$k_{PM_{10}}$	3,4	2,5	1,9			2,2	3,3

Pour pouvoir associer les concentrations de CHIMERE avec les LCZ, nous avons croisé les couches cartographiques les contenant aux coordonnées des sites sur lesquels nous avons mené l'apprentissage. L'objectif était de pouvoir mener cela sur chaque brin routier de notre domaine, et de prédire l'intensité k de la correction à apporter. Cette approche, quoique prometteuse, ne nous a pas permis de mettre en avant de relation entre les LCZ environnantes les rues ([Tableau 6](#)) et les coefficients k calculés plus haut. En effet, aucune relation logique n'est apparue entre ces deux paramètres, ni pour NO_2 ni pour les PM_{10} . Une première hypothèse que nous avons émise face à ce résultat est le fait qu'il y a un manque de diversité des LCZ au niveau des stations de trafic des jeux d'apprentissage. En effet, seules 3 LCZ sont représentées comme on le voit dans le [Tableau 6](#), et la diversité au sein de ces classes n'est pas considérée. La seconde hypothèse qui vient compléter notre diagnostic est que ces paramètres de confinement des rues ne sont pas des sources d'erreur du premier ordre du modèle CHIMERE, ou du moins, que les paramètres qui permettent de constituer ces classes ne sont pas discriminants sur les sites testés.

Nous avons donc considéré d'autres paramètres locaux pouvant être reliés à la variabilité de k .

b. Hétérogénéité des émissions (HE)

L'intensité du trafic d'une rue a un impact sur les niveaux de pollution mesurés au niveau des stations. Or, comme mentionné en introduction de cette section, les CTM travaillent sur un espace discrétisé en mailles qui mène à une homogénéisation des valeurs d'émissions dans chacune des mailles du domaine. Ceci induit une erreur sur l'intensité des taux d'émissions point de mesure. L'intensité de l'erreur du modèle est donc directement liée à la valeur du facteur de dilution des émissions, lors du calcul dans la maille.

Pour approfondir cette piste, nous avons choisi de représenter les gradients d'intensité des émissions dans les mailles, en construisant un rapport nommé HE entre (i) un indicateur des émissions émises au niveau du tronçon de route que l'on considère et (ii) un indicateur des émissions routières moyennes dans la maille.

$$HE = \frac{(FE_{poll} * TMJA_i * longueur_{i_{tronçon}})_{routières}}{\sum_{maille} (FE_{poll} * TMJA * longueurs_{tronçons})_{routières}}$$

Le flux d'émissions sur chaque brin routier se calcule comme le facteur d'émission (FE) moyen du polluant par les véhicules * l'activité routière (flux de trafic $TMJA$) * la longueur du tronçon de route. Un exemple de l'utilité de cette pondération par la longueur de la voie peut être donné en comparant deux tronçons entre eux : Plus un tronçon de route à l'intérieur d'une maille est long, plus les émissions

qui y sont émises seront importantes même si le nombre de véhicules totaux est faible. *A contrario*, un tronçon court aura un impact moins important en termes d'émissions pour un même nombre de véhicules par jour. La longueur du tronçon de route est notamment importante lorsque l'on a des vents dans l'axe de la rue, car ils peuvent ramener les polluants émis tout le long du brin vers le point de mesure. Le flux d'émissions routières sur le reste de la maille est calculé de la même manière, par addition sur l'ensemble des tronçons de la maille. C'est une hypothèse forte de ne pas considérer les autres émissions de la maille, car cela sous-entend que le gradient est intégralement formé par les effets spatiaux du trafic. Mais s'il est facile d'estimer les « autres émissions » dans la maille, il est très difficile d'identifier les émissions non routières le long du brin et donc corriger le dénominateur sans corriger le numérateur nous a paru peu robuste. C'est sans doute une piste à explorer.

Enfin, en faisant l'hypothèse d'un trafic homogène sur tout l'espace urbain, on peut s'affranchir du premier paramètre, le « facteur d'émission du polluant » sur la voie. En effet, celui-ci constitue un même facteur multiplicatif pour le numérateur et le dénominateur de notre rapport *HE*. Bien sûr, cette hypothèse d'homogénéité est discutable, mais elle constitue le mieux que l'on puisse produire comme donnée, puisque l'inventaire à notre disposition (et l'état de l'art des inventaires) ne considère qu'un seul parc roulant sur tout l'espace urbain.

Au final, le rapport *HE* s'écrit selon l'Équation 7 :

$$HE = \frac{(TMJA_i * longueur_{i_{tronçon\ i}})_{routes}}{\sum_{maille} (TMJA * longueurs_{tronçons})_{routes}} \quad \#(7)$$

Pour exprimer les flux de trafic au niveau de la voirie on utilise les données de Trafic Moyen Annuel Journalier (en nombre de véhicules par jour). Cette donnée se trouve être difficile à récupérer pour l'ensemble de la voirie francilienne étant donné que des entités administratives différentes sont chargées de réaliser ces comptages en fonction du département considéré, mais aussi que c'est essentiellement les routes les plus importantes qui font l'objet d'un suivi. Il en découle tout d'abord une inhomogénéité des bases de données due aux différents formats de données, à la séparation ou non des comptages pour les axes routiers à plusieurs voies, à la mise à disposition de moyennes annuelles ou mensuelles et aux différences d'années de référence pour les mesures.

Aussi, je me suis plutôt dirigée vers la modélisation des flux de trafic journalier afin d'avoir une information cohérente sur l'ensemble du réseau routier d'intérêt. J'ai utilisé pour cela les données produites par le modèle OLYMPUS développé dans notre équipe au LISA (<http://www.lisa.u-pec.fr/instruments/?id=1295>). Ce modèle sera décrit de manière détaillée pour les besoins du calcul d'exposition dans le chapitre suivant, mais on peut indiquer ici qu'il s'agit d'un modèle d'émissions qui calcule les flux de mobilité des individus sur un territoire à partir d'une représentation statistique de la demande en transport pour des déplacements quotidiens. Il prend en compte l'ensemble de motifs de mobilité (travail, loisir, achat, éducation) et tient compte dans son calcul des taux d'occupation des voies et de la congestion à tout instant. OLYMPUS produit ainsi des valeurs de TMJA sur chaque brin routier de la voirie francilienne. Il m'a simplement fallu extraire ces données et géolocaliser les stations de mesure afin de récupérer les TMJA des tronçons de route d'intérêt. Les TMJA réels ainsi que les TMJA modélisés des stations sont résumés dans le Tableau 7. Il est à noter que nous n'avons pas pu récupérer les valeurs des comptages au niveau des tous les tronçons de routes correspondant aux stations d'apprentissage, ce qui explique les cases grisées du tableau. Cette comparaison montre que la simulation d'OLYMPUS est très réaliste pour les stations « A1 » et « BASCH » avec seulement 5% de différence avec les observations, ce qui est très satisfaisant. « ELYS » est sous-estimé par rapport à la mesure à hauteur de 33% ce qui reste très satisfaisant. En revanche, le TMJA modélisé sur le tronçon lié à la station d'Auteuil est 40 fois plus élevé que les comptages. Il peut s'agir ici d'une erreur liée au fait que le comptage est lié à la bretelle d'accès Porte d'Auteuil alors qu'OLYMPUS considère le flux sur le

périphérique, qui ne sont évidemment pas du même ordre de grandeur. Dans l'ensemble, les données d'OLYMPUS présentent donc des valeurs cohérentes, et des travaux sont en cours dans l'équipe pour valider totalement les TMJA proposés par le modèle. Face à ce résultat encourageant, nous avons continué à travailler sur la généralisation de l'approche à l'aide des données d'OLYMPUS.

Tableau 7 - Données relatives aux comptages routiers (mesurés et modélisés par OLYMPUS) ainsi que les indicateurs d'émissions calculés au niveau des tronçons (TMJA pondérés), indicateurs moyens d'émission dans la maille (TMJA pondérés mailles), les vitesses de déplacement par tronçons (Vitesse) et les valeurs obtenues pour les Ratio HE de NO₂ et des PM₁₀.

Stations	Comptages (nombre véhicules/jours)		TMJA pondérés	TMJA pondérés mailles	Rapport HE (NO ₂)	Vitesse (km/h)	Ratio HE * Vitesse (PM ₁₀)
	OLYMPUS	Comptage	(*1E ⁰⁵)	(*1E ⁰⁸)			
A1	208 076	219 608	715	2,73	0,262	70	18,34
AUT	442 027	11 756	380	1,31	0,289	50	14,45
BASCH	103 735	109 499	548	2,73	0,2	30	6,01
BONAP	3 049		3,98	1,21	0,0033		
CELES	34 495		66,8	3,07	0,0218		
ELYS	57 675	77 000	23,4	2,05	0,0114	50	0,57
RN6	48 181		65	0,314	0,207	70	14,49

Les rapports *HE* ont été calculés grâce à l'Équation 7 pour chacune des stations et j'ai recherché leur corrélation avec les valeurs des coefficients *k*. Le résultat est présenté dans la Figure 19. On voit sur cette figure que les deux paramètres présentent une dépendance positive, mais que le point RN6 est totalement en dehors de cette tendance. Cela rejoint les commentaires formulés précédemment sur le fait que ce site est un site « particulier » dont la reproduction par le CTM est d'emblée mauvaise, et que cela augmente artificiellement le besoin de correction. Le fait de retirer la station de la corrélation permet d'obtenir une dépendance linéaire plus marquée, avec un coefficient de Pearson R² de **0,72**. Ceci nous permet de considérer le ratio *HE* comme un discriminant du besoin en correction du modèle. On peut déduire de cette régression une valeur de *k* pour les différents tronçons de la voirie d'Ile-de-France. On voit bien sûr que la variabilité de *HE* n'explique pas toute la variabilité de *k*, mais cela permet de réduire l'erreur qu'on aurait faite en appliquant une valeur moyenne pour *k*, et de prendre en compte la variation du trafic routier local. On doit en outre avoir en tête que les sites de trafic en zone peu urbanisée seront potentiellement mal reproduits et par CHIMERE, et par *Cprox* car nous ne pouvons pas apporter de correction pleinement satisfaisante sur cette base.

En revanche, aucune corrélation avec les rapports *HE* n'a été initialement trouvée pour les coefficients *k* des sites de trafic de PM₁₀ (R² = 0,12). À ce stade, j'ai émis deux hypothèses. Soit l'erreur du modèle n'est pas ou peu dépendante de la surémission locale (et donc il faut abandonner la piste de *HE*), soit le flux de trafic n'est pas un bon proxy pour représenter les émissions locales. Avant d'abandonner la piste de *HE*, j'ai exploré la seconde hypothèse. Aujourd'hui on sait en effet que, localement, les émissions routières non liées à la combustion, dites « non-exhaust » (NE) des particules ont un impact significatif sur les niveaux mesurés en proximité aux voies (OECD, 2020). Ces émissions « NE » sont reliées à la remise en suspension de particules se trouvant sur la voirie au passage des véhicules, à l'abrasion des pneus sur la route et au frottement des plaquettes de freins lors de la décélération. La quantification de ces émissions reste, aujourd'hui, encore complexe puisqu'elle va fortement dépendre des matériaux des freins et de la route (pour les phénomènes d'abrasion) et il est difficile de différencier la remise en suspension des particules liée au passage des véhicules avec celle liée aux vents (Gustafsson et al., 2009). On sait néanmoins que les vitesses de déplacement des véhicules sont reliées aux modes de conduite et qu'elles ont une influence sur les phénomènes de frottement et sur la remise en

suspension. Pour tenter d'intégrer cette dimension dans *HE*, j'ai introduit dans le [Tableau 7](#) la donnée relative à la vitesse de déplacement moyenne des véhicules sur chaque tronçon *HE* des PM_{10} et j'ai regardé si on pouvait améliorer l'émission locale en la multipliant par la vitesse. Cela constitue l'[Equation 8](#). Cet ajout permet d'augmenter la valeur du coefficient de Pearson à **0,66** ([Figure 19](#)). Comme précédemment, cette corrélation visant à prédire la valeur de *k* n'explique pas la totalité des différences entre sites. Mais l'amélioration du coefficient de corrélation permet d'avoir une proposition de formule pour le calcul de *k*, qui est plus pertinent que l'usage d'une valeur moyenne de *k*. Pour les PM_{10} on utilise donc :

$$HE = \frac{(Vitesse * TMJA_i * longueur_{i_{tronçon}})_{routes}}{\sum_{maille} (TMJA * longueurs_{tronçons})_{routes}} \quad \#(8)$$

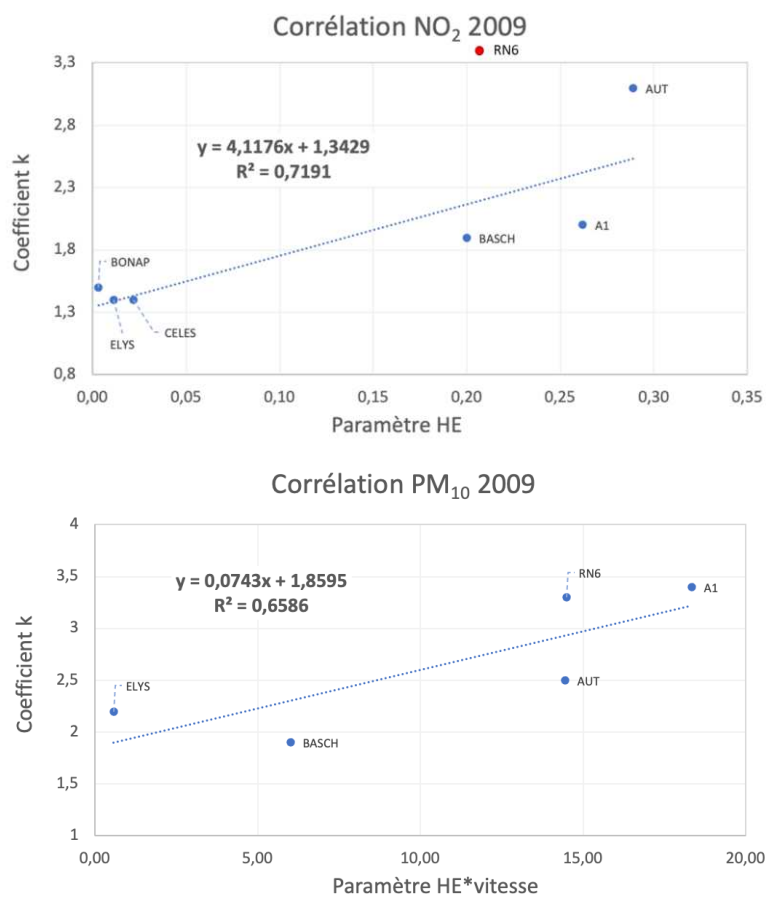


Figure 19 - Corrélations entre le paramètre *HE* et les coefficients *k* des stations de trafic de NO_2 (en haut) et de PM_{10} (en bas) pour 2009.

Bien sûr, il aurait été nécessaire d'avoir plus de points de contrainte pour réaliser nos tests statistiques sur la prédiction de la valeur de *k*, et le manque de données d'entrées ne nous permet pas de mener une analyse statistique détaillée. En particulier, plus de points auraient permis de savoir s'il faut donner plus de poids au point « AUT » ou pas.

Il est toutefois intéressant de noter que plusieurs méthodologies ont été développées dans la littérature pour prédire les niveaux de NO_2 et de PM_{10} à proximité des sources de trafic routier (LUR, Random Forest et Réseaux de neurone). Elles se basent sur une diversité d'indicateurs de pollution assez larges visant à reproduire les niveaux mais aussi les gradients de concentration de ces deux polluants entre

fond urbain et proximité routière. Les ordres de grandeurs des coefficients de corrélation que j'ai obtenue à la suite de cette étape de l'approche sont tout à fait cohérents avec ceux présentés dans la littérature, puisqu'ils sont compris dans l'intervalle des R^2 ainsi établis, quel que soit l'indicateur considéré et la méthode de corrélation utilisée. Soit entre 0,4 et 0,8 pour les particules et environ 0,75 pour la plupart des études portant sur le NO_2 . Bien que le nombre d'indicateurs que je considère dans mon approche soit faible par rapport à ces études, on peut donc considérer que les R^2 obtenus sont donc satisfaisants pour la prédiction des gradients de concentration des deux polluants. Grâce à ce travail de mise en corrélation de k avec les paramètres de l'environnement du site de mesure, des valeurs de k pourront être produites pour tous les tronçons de routes du domaine Île-de-France.

Enfin, il est important de noter que les valeurs du rapport HE présentées dans le [Tableau 7](#) et que j'ai utilisées pour construire la corrélation ont été soumises à un travail spécifique. Pour chaque site, j'ai analysé l'environnement routier afin de prendre en compte tous les brins d'intérêt pour l'exposition locale. Ainsi, certains environnements complexes comme la place Victor Basch (rond-point) ou la station sur l'autoroute A1 nécessitaient que l'on prenne en compte plusieurs brins. Ce travail, bien sûr, ne pourra être mené de façon automatique sur toutes les voies de la région, et l'attribution automatisée d'une valeur de HE à chaque site de validation ne pourra pas bénéficier d'un tel travail. C'est potentiellement une limitation de l'approche puisqu'on ne peut garantir la même robustesse à chaque point pour lequel on prédit le rapport HE . Toutefois cela ne concerne que les environnements routiers complexes et ce travail n'est pas nécessaire sur la plupart des rues. Pour limiter cette problématique, il faudra probablement mieux encadrer le calcul de HE dans OLYMPUS.

5. Validation de l'approche

Les résultats précédents m'ont permis de poursuivre avec l'étape de validation de la descente en échelle. Il s'agit d'appliquer les équations mises en place précédemment au niveau de sites d'études n'ayant pas été utilisés pour l'apprentissage de l'approche ([Tableau 4](#)).

Statistiques sur les corrections

Des séries temporelles de tous les sites ainsi que les boxplots des résultats de la phase de validation de l'approche sont présentés en [Annexe 4](#). Les boxplots de stations représentatives sont présentés en [Figure 20a](#) et [Figure 20b](#). L'impact de la correction sur la variance est présenté dans la [Figure 21](#).

La correction prévue sur les sites de trafic fait à la fois intervenir des paramètres fixes (répartition de la correction sur le signal lissé et sur les pointes, effet du vent) et des paramètres site-dépendants (valeur de k). Il est donc intéressant de regarder si la prédiction de correction s'avère bien dimensionnée. En 2009, seul un site de proximité au trafic est disponible pour la validation, le site situé sur la route nationale RN2. On voit en [Figure 20](#) que la correction permet de rehausser l'ensemble de la distribution des données, et pour NO_2 de gagner en amplitude sur cette distribution des données. On constate toutefois que la correction est trop forte pour NO_2 puisque C_{prox} surestime de près de $30 \mu\text{g}/\text{m}^3$ la moyenne annuelle mesurée. Elle est en revanche pertinente pour les PM_{10} . La rareté des sites de proximité au trafic a fortement limité le volant de données disponibles pour la validation et il est difficile de conclure sur la pertinence de notre approche avec un seul site de validation. Qui plus est, le site RN2 s'avère particulier comme on le verra plus tard, dans l'analyse détaillée de sa correction C_{prox} . Et, bien que ce site amène des conclusions d'intérêt, il reste intéressant de travailler sur une autre année pour pouvoir explorer plus en détails le potentiel de notre approche de descente en échelle en proximité au trafic.

La correction prévue sur les sites de fond est quant à elle fixe et de faible ampleur. Il ne peut pas y avoir de modulation basée sur la qualité des données initiales de CHIMERE. Il est donc intéressant de voir comment cela améliore, ou dégrade, les sorties du CTM. Pour NO_2 comme pour les PM_{10} , on constate

sur l'ensemble des sites une amélioration des valeurs annuelles moyennes ce qui reste l'objectif numéro 1 de mon travail. Toutefois, lorsque les sorties de CHIMERE sont très proches de la mesure, le fait d'imposer une correction vers de plus hautes valeurs et une plus forte amplitude du signal peut avoir tendance à générer une série de données un peu trop fortes, comme on le voit sur les boxplots du site de Gonesse pour NO₂, ou de Paris 18e pour les particules. Quel que soit le résultat sur la moyenne annuelle, on voit également sur la [Figure 20](#) que la distribution des données est plus réaliste sur la plupart des sites de fond urbain, avec un élargissement de la gamme de concentrations entre les percentiles 25 et 75. L'impact sur la RMSE est, aux exceptions précédentes près, modéré (2-3 µg/m³) et en faveur de *Cprox*. Les quelques exceptions, comme Gonesse et Ivry pour les PM₁₀, subissent une augmentation de la RMSE qui est aussi de +2-3 µg/m³. L'impact sur la variance est plus complexe. Il est en général négligeable pour NO₂ même s'il tend toujours à la réduire, ce qui n'est pas le sens espéré. Pour les PM₁₀, il est également faible mais systématiquement dirigé dans le bon sens.

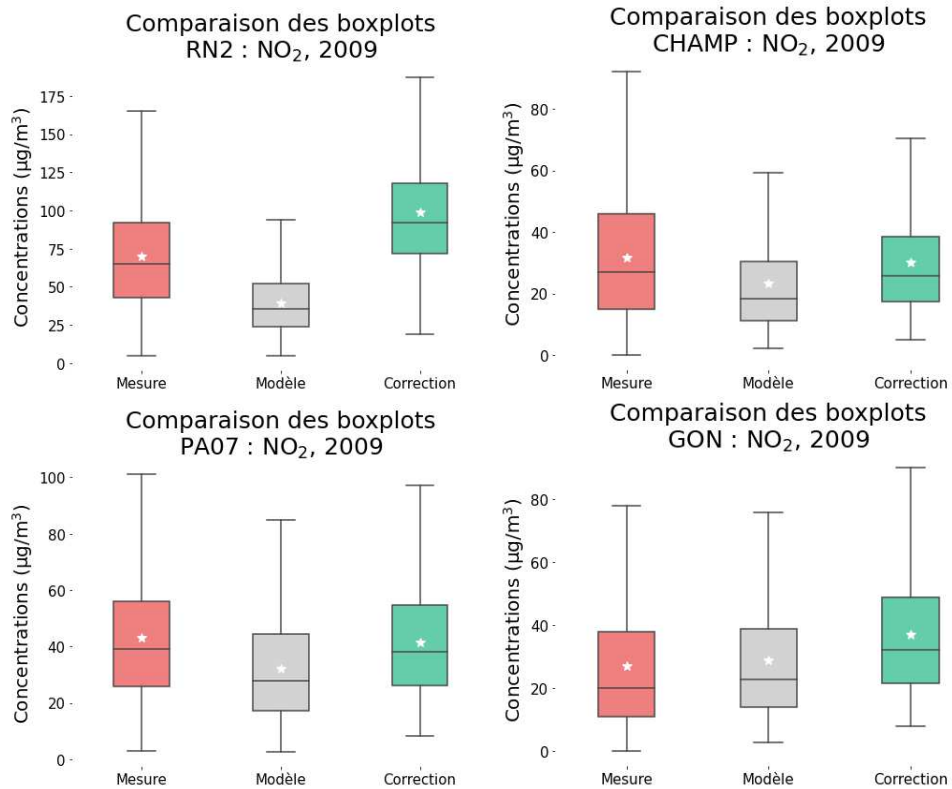


Figure 20.a - Boîtes à moustache de NO₂ pour la mesure (rouge), CHIMERE (gris) et *Cprox* obtenue par l'approche automatisée (vert) - Elles donnent les percentiles 5 et 95, les percentiles 25 et 75 ainsi que les moyenne et médiane de chaque série de données sur l'ensemble de l'année 2009.

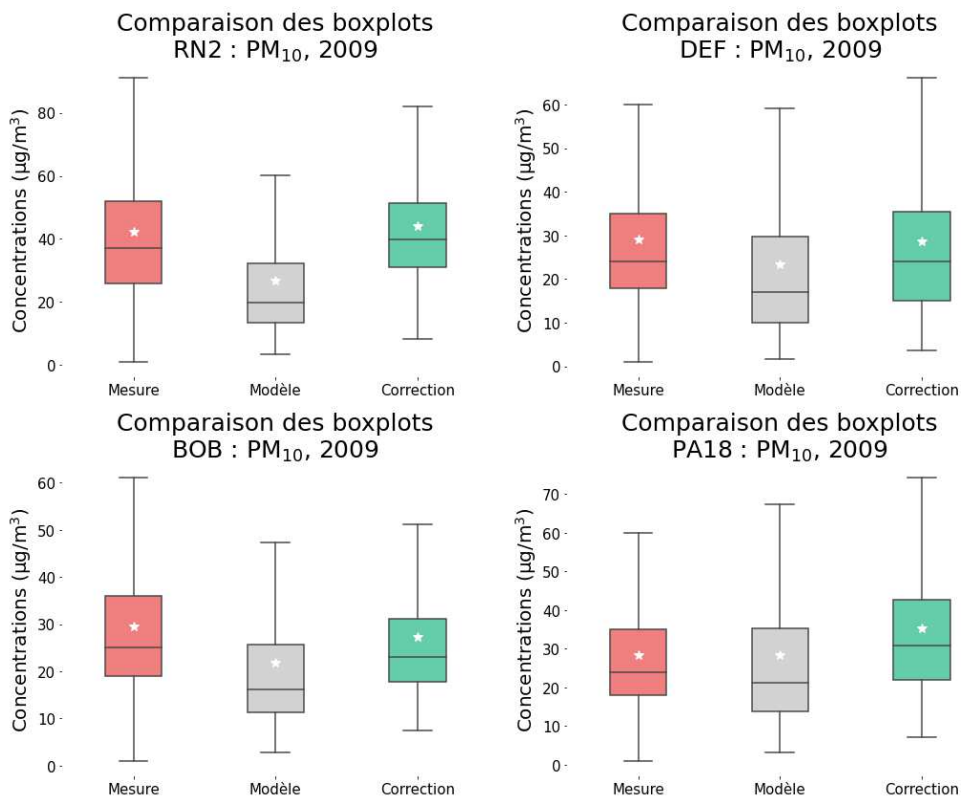


Figure 20.b - Boîtes à moustache de PM₁₀ pour la mesure (rouge), CHIMERE (gris) et *Cprox* obtenue par la correction automatisée (vert) - Elles donnent les percentiles 5 et 95, les percentiles 25 et 75 ainsi que les moyenne et médiane de chaque série de données sur l'ensemble de l'année 2009.

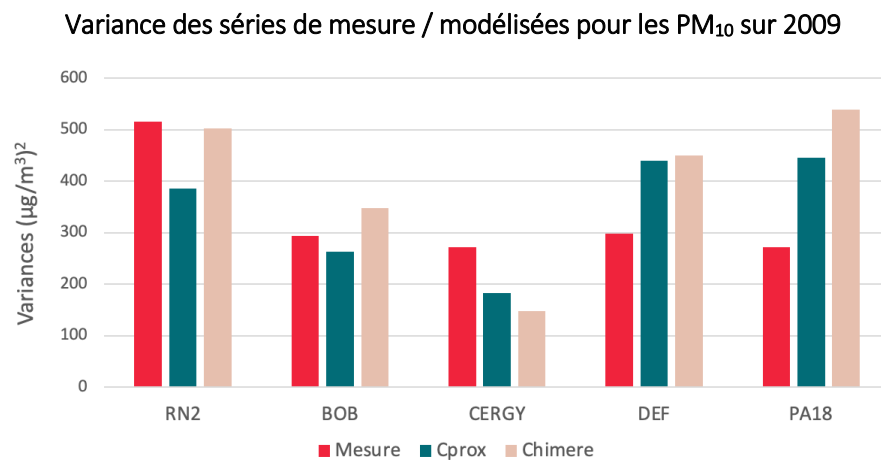
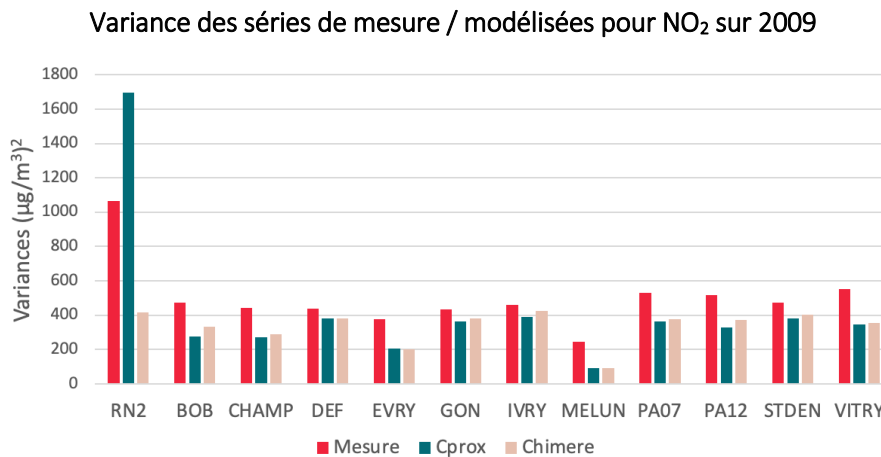


Figure 21 - Comparaison des variances pour les sites des jeux de validation de NO₂ (en haut) et PM₁₀ (en bas) pour l'année 2009. RN2 est le site de trafic de validation, les autres sites sont des sites de fond urbain. Les variances issues des mesures sont représentées en rouge, celles issues de *Cprox* automatisée sont en vert et celles issues de CHIMERE sont en rose.

Analyses des séries *Cprox* produites

Comme on l'a mentionné précédemment, il est intéressant de faire un point sur le site de trafic RN2. Les données statistiques de cette station sont données dans le [Tableau 8](#).

La donnée k_{auto} est la valeur de k prédite par le calcul automatisé de *Cprox*. Pour alimenter la discussion, nous pouvons comparer les résultats obtenus pour ce site de validation avec les niveaux obtenus pour les sites d'apprentissages. La donnée k_{auto} de la station RN2 est du même ordre que trois des sites d'apprentissage de trafic - BONAP, CELES et ELYS - avec des valeurs autour de 1,4. Par contre, une RMSE et des écarts plus élevés entre les moyennes annuelles et les variances sont calculées pour le site de RN2 après l'application de l'approche alors qu'on tendait vers une diminution de ces dernières pour les sites d'apprentissage.

On peut alors associer la surestimation des concentrations de NO₂ au fait que la valeur de *HE* est trop élevée, ce qui peut être lié, par exemple, à une surestimation du flux local de trafic par OLYMPUS, mais ce n'est pas vérifiable car aucune mesure de TMJA n'est disponible sur ce tronçon de route. On peut dans un deuxième temps se demander s'il n'y a pas une spécificité du site qui explique le phénomène de surestimation récurrente. Pour explorer cette hypothèse j'ai tracé les séries temporelles liées à cette station en [Figure 22](#). L'analyse visuelle des courbes de concentration de NO₂ montre un résultat à deux facettes. Sur certaines périodes de temps, on obtient comme espéré une amélioration de la courbe de concentration modélisée par rapport à la mesure, et la courbe de *Cprox* se superpose même assez

fidèlement à celle de la mesure. Mais sur d'autres périodes de temps, récurrentes, on observe une surestimation assez forte des niveaux mesurés par l'approche *Cprox*. J'ai choisi de représenter en Figure 22 une période où cela arrive particulièrement souvent, et j'ai surligné ces périodes en jaune.

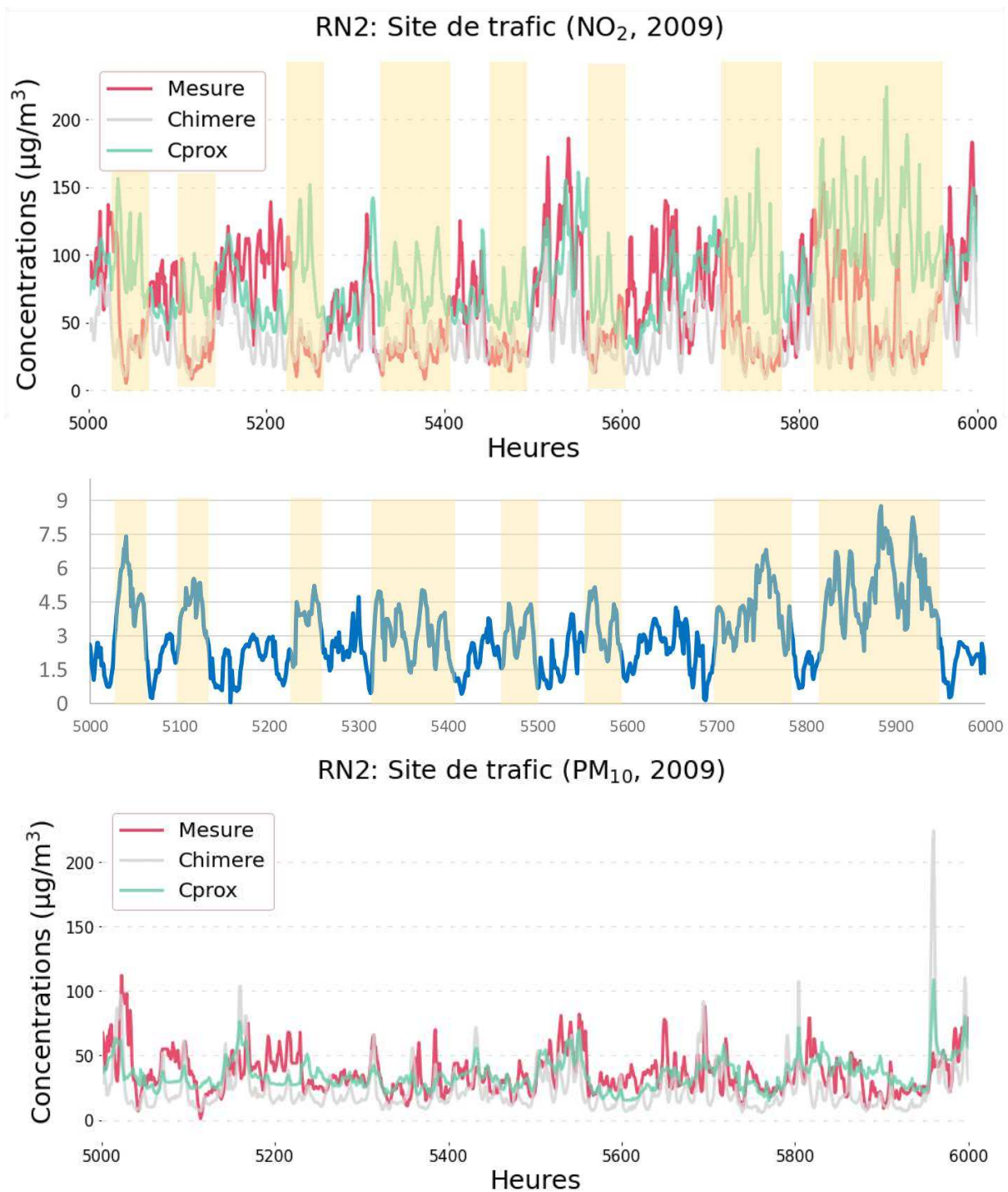


Figure 22 -Séries temporelles de CHIMERE (en gris), de la mesure (en rouge) et de *Cprox* (en vert) sur la station de trafic RN2 du jeu de validation RN2, sur une période représentative de l'année 2009. En haut, résultats pour NO₂ et mise en évidence des périodes de l'année surestimées par l'approche de correction *Cprox*. En bas, résultats pour les PM₁₀.

Les périodes de surestimation sont associées à des périodes pendant lesquelles les niveaux mesurés restent bas, et sont bien reproduits par CHIMERE. Ainsi, l'application de la correction va donc augmenter des valeurs qui étaient, initialement, correctes par rapport à la mesure, induisant des événements de surestimations. Ce comportement des mesures de NO₂ rappelle les phénomènes observés par vent fort en site de fond urbain. J'ai donc tracé en [Figure 22](#) le vent horaire sur le domaine, et j'ai également surligné en jaune les valeurs de vent supérieur à 3 m/s. On constate que les périodes de surestimation sont en fait des événements de vent fort pour lesquels la correction est amplifiée à tort. En effet, le site RN2 est très ouvert et se comporte face au vent plutôt comme un site de fond. Nous n'avions pas de telle station dans notre échantillon d'apprentissage. Il est donc probable qu'il faille dans le futur n'appliquer la correction sur le vent (1,4) qu'à des sites émetteurs pour lesquels il existe un environnement bâti permettant l'accumulation locale des polluants émis. C'est un des enseignements de cette étape de validation. L'effet de dégradation n'est en revanche pas observé sur les particules car, comme nous l'avons vu, la part de *Ampl* dans la correction est très minoritaire ($\beta = 10\%$ de *k*). Les résultats sur les particules sur ce même site (illustrés dans la [Figure 22](#), en bas) sont même à l'inverse améliorés sur toute la période. On retrouve ce bon score sur le boxplot de la [Figure 20b](#).

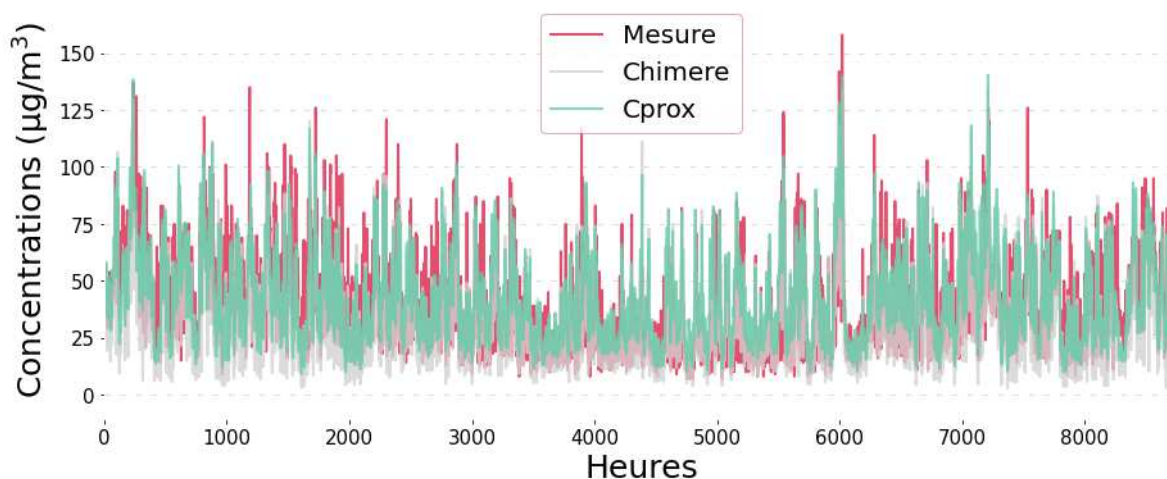
Tableau 8 - Résumé des données statistiques de la station de validation de trafic de NO₂ et des PM₁₀ (automatisation de l'approche et comparaison avec les données idéales « fit » de la station).

Station	<i>k</i>	RMSE CHIM (µg/m ³)	RMSE <i>Cprox</i> (µg/m ³)	Moy. CHIM (µg/m ³)	Moy. <i>Cprox</i> (µg/m ³)	Moy. Mesure (µg/m ³)	Variance CHIM (µg/m ³) ²	Variance <i>Cprox</i> (µg/m ³) ²	Variance Mesure (µg/m ³) ²
RN2 NO ₂ (auto)	1,6	41,0	54,6	39,4	100,4	69,9	415,3	1694,6	1065,6
RN2 PM ₁₀ (auto)	2,1	26,3	19,0	26,8	42,1	42,3	501,8	385,9	515,3

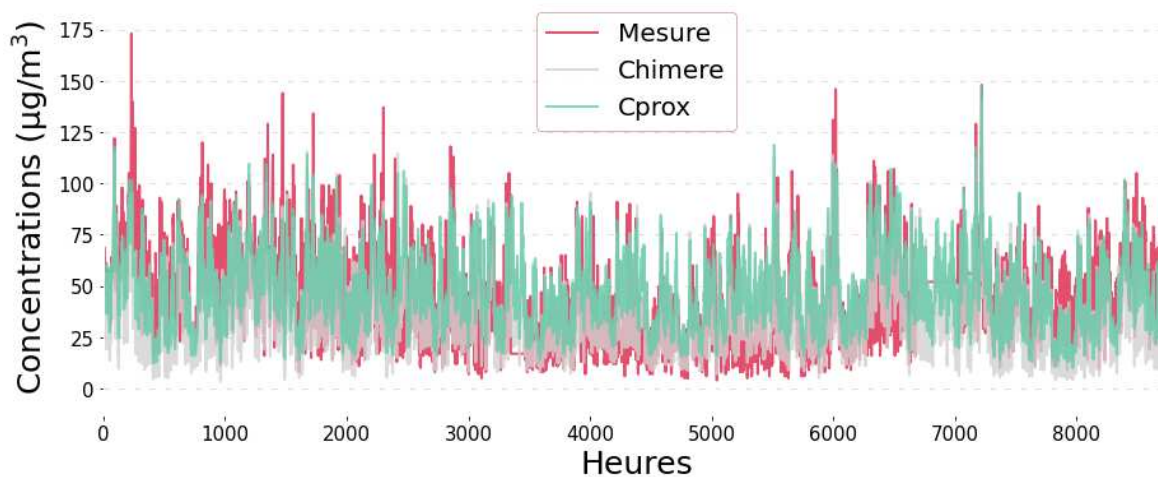
Les séries temporelles présentées ci-après en [Figure 23](#) (pour le dioxyde d'azote) et [Figure 24](#) (pour les particules) retracent les principaux atouts et défauts de l'approche fixe de calcul de *Cprox*, commentés sur la base des scores de *Cprox* ci-dessus. De haut en bas de la [Figure 23](#) on constate : (i) sur St Denis comme sur une majorité de sites, l'amélioration apportée par *Cprox* qui élimine les valeurs trop basses simulées par CHIMERE et permet d'obtenir une moyenne annuelle plus juste, (ii) le manque d'une flexibilité de la correction au cours du temps. Comme on le voit sur le site de Gonesse, la formulation fixe ne permet pas d'ajuster la correction en fonction des périodes de l'année, ce qui pourrait être intéressant sur plusieurs sites de fond. Et enfin (iii) le problème récurrent lié à CHIMERE qui sous-estime trop fortement les polluants routiers dans les espaces assez peu urbanisés et que la formule de *Cprox* n'inclut pas, par choix, au moins dans la version actuelle. C'est ce qu'on observe sur la commune de Melun en grande couronne. Les concentrations mesurées à cette station, ainsi que leurs variabilités horaires, sont beaucoup plus importantes que celles modélisées. La moyenne annuelle mesurée est presque 3 fois supérieure à celle prédite par CHIMERE (23,8 µg/m³ contre 9,5 µg/m³) et la variance des concentrations, qui est de 245,3 (µg/m³)² pour la mesure n'est que de 91,6 (µg/m³)² dans les données de CHIMERE. *Cprox* ne permet pas d'améliorer la variance de ce site, qui reste constante du fait du choix d'une correction modérée.

Pour les particules fines, à l'image de ce qu'on observe sur La Défense, la correction apportée en site de fond périurbain et urbain est plus systématiquement favorable et suffisante. Le site de Cergy est le pendant de ce qu'on a constaté sur Melun pour le dioxyde d'azote.

St Denis - Site de fond urbain - Concentrations de NO₂ sur l'année 2009



Paris 12e arrondissement - Site de fond urbain - Concentrations de NO₂ sur l'année 2009



Melun - Site de fond périurbain - Concentrations de NO₂ sur l'année 2009

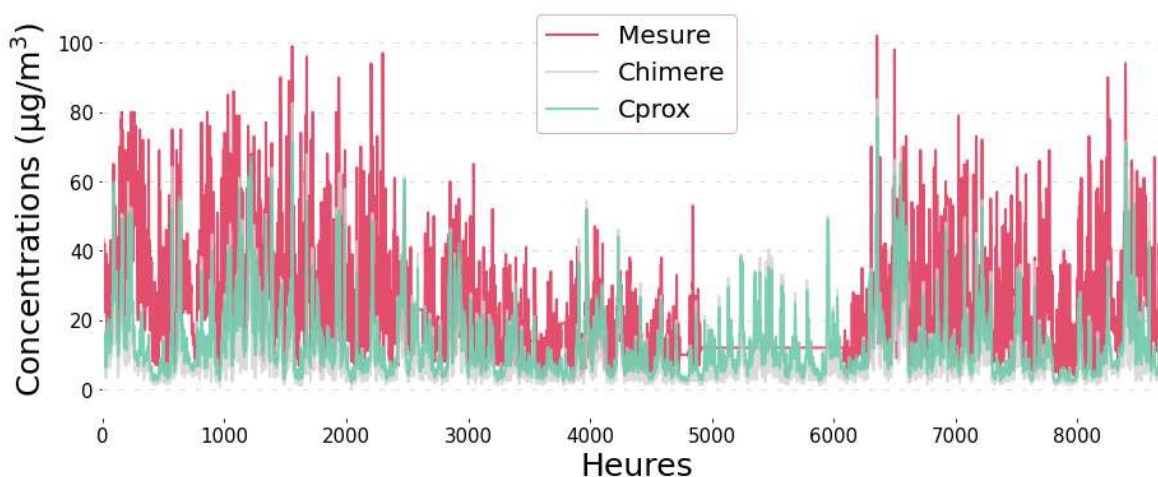
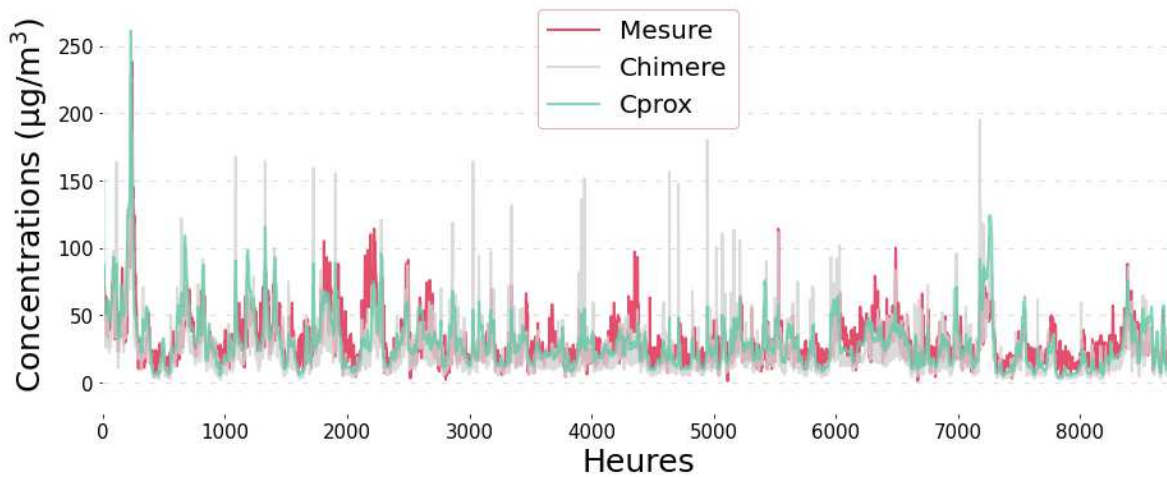


Figure 23 - Comparaison des niveaux moyens corrigés (en vert) par l'étape d'universalisation de *Cprox* avec les niveaux mesurés (en rouge) et ceux de CHIMERE (en gris), pour NO₂ et sur les sites périurbains de Gonesse et de Melun.

La Défense - Site de fond urbain - Concentrations des PM₁₀ sur l'année 2009



Cergy - Site de fond périurbain - Concentrations des PM₁₀ sur l'année 2009

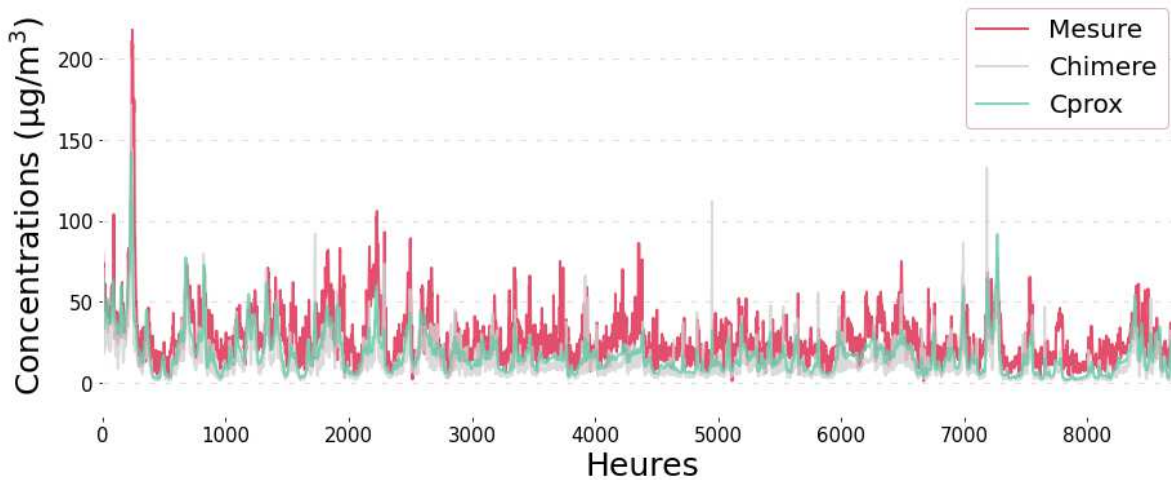


Figure 24 - Séries temporelles de CHIMERE (en gris), de la mesure (en rouge) et de *Cprox* (en vert) sur différentes stations de mesure des PM₁₀ du jeu de validation, sur l'année 2009.

6. Discussion

L'approche statistique développée ici nous permet d'obtenir des résultats permettant de différencier la correction en fonction de la typologie des sites étudiés, et du polluant concerné. Le fait qu'on ait développé *Cprox* en considérant un nombre minimal de données d'entrées impose une nature simplificatrice à la restitution des gradients de concentrations, par rapport à d'autres approches qui nécessitent d'avoir accès à un plus grand nombre ou à une plus grande précision des bases de données d'entrée.

Cela permet d'une part de recréer des gradients de concentration réalistes, statistiquement parlant, sans temps et coût de calcul excessif. Ainsi, même si certains sites précis sont légèrement sur- ou sous-estimés, l'obtention de *Cprox* sur un simple calcul de remise à l'échelle des données de CHIMERE permet d'avoir à une représentation réaliste de la palette des micro-environnements urbains de fond et en proximité au trafic. Et de pouvoir estimer sur une base statistique la complexité des risques d'exposition pour un individu se déplaçant en Île-de-France. Le fait de ne pas avoir une corrélation

parfaite pour deviner k sur les sites de trafic n'est à ce titre pas rédhitoire puisqu'elle permet de créer un relief réaliste dans les champs CHIMERE, même s'il n'est pas forcément très précis partout. Par ailleurs, le choix de partir des champs de concentration d'un CTM comme CHIMERE pour calculer *Cprox* permet d'intégrer un maximum de paramètres issus d'une démarche déterministe et compréhensive : cela inclut la prise en compte de la variabilité des systèmes météorologiques, de la variabilité des émissions au cours de la semaine ou de l'année, des événements d'import de pollution sur la région... La ligne de base de *Cprox* et son excellent accord avec celle de la mesure a ainsi été relevée à plusieurs reprises et sur une variété de sites. Finalement, d'un point de vue quantitatif, l'amélioration des scores de NO_2 et de PM_{10} sur les sites de trafic justifie pleinement l'intérêt de cette démarche de correction et constitue un élément robuste de sa validation. En outre, la disparition des pics excessifs issus du CTM sur les séries temporelles des PM_{10} constitue une amélioration sensible de nos champs de concentration simulés.

Le choix d'une approche simplifiée conduit en revanche à certaines limitations.

La première est liée à la nature de l'usage que l'on peut faire des résultats. Il ne s'agit pas de recréer parfaitement un environnement donné mais la gamme des environnements urbains d'une métropole. La démarche n'est donc pas adaptée à une étude sur un site ou un axe routier particulier, ou alors il faut mener un apprentissage sur le site d'intérêt.

L'absence de prise en compte des spécificités des sites de fond (notamment périurbains) a été pointée du doigt lors de l'étape de validation. Certains sites de trafic situés dans des communes éloignées de la zone urbaine principale (grande couronne parisienne) souffrent en effet d'une sous-estimation marquée dans les sorties de CHIMERE. Si l'on veut recréer les gradients forts qui existent dans ces communes, il faut identifier ces espaces émetteurs existant au sein de mailles à fort caractère rural, afin de prédire le besoin de les corriger plus intensément. L'usage des sols dans la maille, ou encore un facteur similaire à *HE* sont des pistes pour prédire les zones à risque de sous-estimation et y remédier. A l'inverse, la légère dégradation des scores de certaines stations urbaines de fond ne justifie pas la complexification de notre approche, car l'effet est minime. Plus largement, la question de la spécificité de certains sites, inhomogènes, doit être discutée. On constate des limitations de performances pour certains lieux dont les environnements sont hétérogènes mais aussi sujets à des phénomènes particuliers, d'émission ou de circulation. Pour NO_2 , on peut citer la station de la porte d'Auteuil « AUT », et celle de la route nationale 6 à Melun « RN6 ». Pour ces deux sites, c'est dès l'étape du CTM que le biais sur la représentation est particulièrement élevé. « AUT » et « RN6 » ont des séries temporelles simulées par CHIMERE qui sont parmi les moins performantes de nos sites d'apprentissage, et elles nécessitent une forte correction. Il existe en fait sur ces sites des particularités environnementales qui affectent le résultat du CTM. « AUT » est située dans une maille avec un fort taux de trafic routier mais avec beaucoup de verdure à proximité. La station est en outre située sous un échangeur routier donc de potentiels phénomènes de recirculation de l'air ou d'accumulation des polluants peuvent avoir lieu. Pour « RN6 », comme nous l'avons mentionné, il s'agit d'une route nationale beaucoup empruntée par les poids-lourds et la station de mesure est située proche d'un petit tunnel ce qui peut conduire à des mouvements d'air particuliers. Dans une approche statistique, ce n'est pas très pénalisant de ne pas reproduire fidèlement ces sites en particulier, mais ils sont des pièces essentielles de notre dispositif d'apprentissage et de l'automatisation de notre correction. Afin d'éviter d'être trop influencés par trop de cas particuliers, il faut aller vers de plus grands volumes de données d'apprentissage.

On a également pu constater que l'absence de temporalité dans notre correction peut être problématique. C'est essentiellement l'été que la correction tend à surestimer la mesure, sans doute parce que la diffusion verticale est plus intense et que CHIMERE sous-estime moins (ou pas du tout) la mesure. Cela me semble une évolution prioritaire de mon travail, et elle pourrait être menée comme je l'ai fait pour l'effet du vent, via la modération de la correction selon la valeur d'un indicateur de température moyenne diurne - par exemple.

Pour les particules, les données *Cprox* manquent souvent de variabilité horaire. Ceci peut être lié au fait qu'on utilise une courbe de percentile glissant pour exprimer le bruit de fond urbain en tant que donnée de départ pour *Cprox*. Cela n'affecte pas les valeurs annuelles moyennes par site qui sont ma cible principale, mais on peut s'interroger sur les données dont on a besoin pour le calcul de l'exposition. Faut-il simuler les pics matinaux ? Ont-ils un impact sur les scores de l'exposition ?

Enfin, de manière structurelle, le choix de baser notre prédiction de *k* sur les émissions routières dans la maille revient à considérer que ces émetteurs et le traitement qui en est fait dans CHIMERE sont la cause presque exclusive des gradients observés sur les sites de proximité au trafic. Et que d'autres sources présentes dans la maille n'interfèrent pas avec l'erreur du modèle. C'est sans doute un point pertinent pour les particules fines, dont le rapport entre sources et bruit de fond est moins en faveur des émissions routières que pour NO₂. Mais les bons scores obtenus sur les sites urbains ne pointent pas de manque significatif dans l'équation de correction.

7. Cartographie des résultats

J'ai mis en œuvre deux étapes pour cartographier les résultats obtenus par l'approche *Cprox* :

- J'ai tout d'abord appliqué l'équation de correction sur le fond (péri-)urbain aux valeurs maillées de CHIMERE, puis j'ai projeté ces valeurs maillées sur une grille plus fine (100 m²).
- Ensuite, j'ai effectué le même travail avec la correction de proximité au trafic en projetant les valeurs sur les lignes du réseau routier d'Ile-de-France avant de les projeter, une seconde fois, sur une grille de 100 m² qui ne couvre que les tronçons de routes.

J'ai ensuite fusionné les deux grilles résultantes et effectué une interpolation bilinéaire pour obtenir les cartographies annuelles présentées ci-dessous en [Figure 25](#) pour NO₂ (à gauche) et pour les PM₁₀ (à droite).

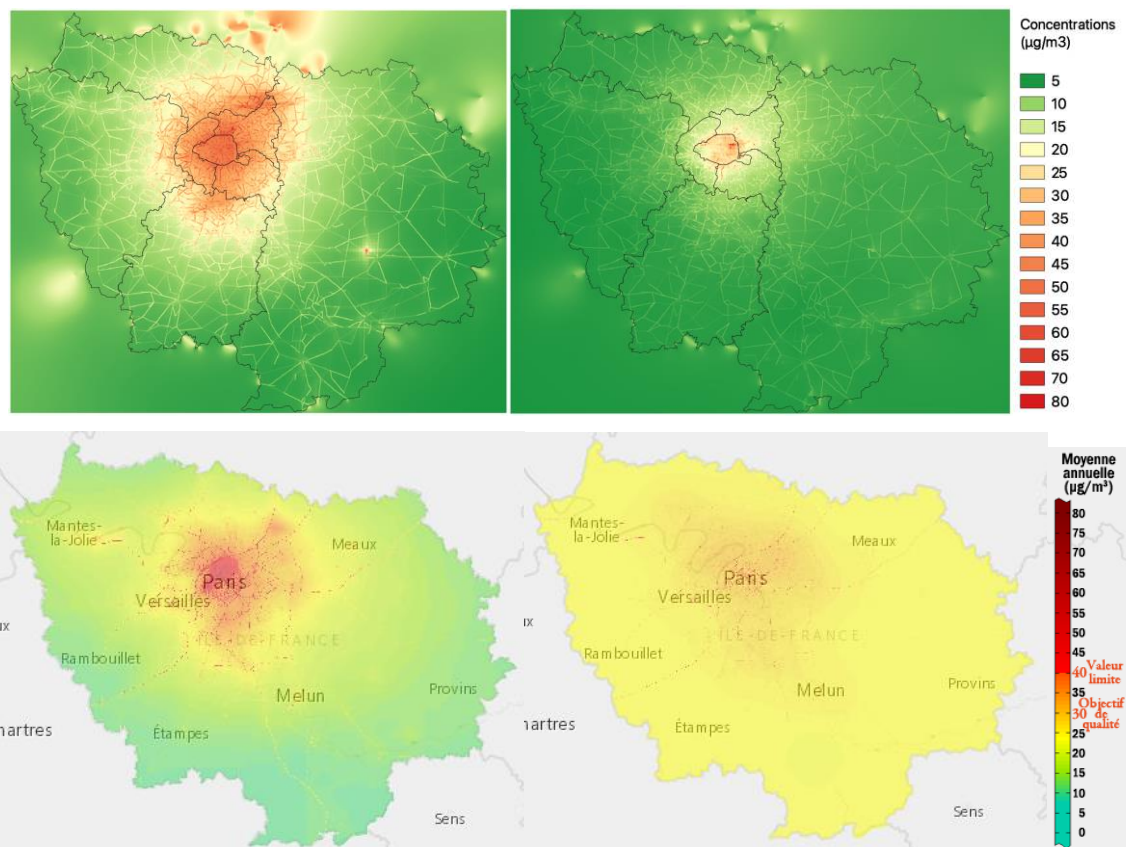


Figure 25 - Comparaison des cartographies annuelles des sorties de *Cprox* (en haut) et des bilans annuels d'Airparif (en bas) pour NO₂ (à gauche) et pour les PM₁₀ (à droite) pour l'année 2009.

Les cartes de qualité de l'air d'Airparif sont également présentées ici. Elles sont issues d'une réanalyse des valeurs simulées. Les simulations sont menées à l'aide d'un CTM (CHIMERE), cette simulation servant à forcer une modélisation urbaine fine échelle (résolution de 12,5 m à 50 m selon la proximité au centre de Paris) qui est ensuite retravaillée statistiquement pour réduire l'erreur par rapport à la mesure. On peut constater sur la figure la similitude des cartes que j'ai créées avec celles produites par Airparif pour le dioxyde d'azote. Il est intéressant de voir que, pour NO₂, on distingue sur les deux cartes l'impact de l'activité sur l'aéroport de Charles de Gaulle au Nord-Est de Paris.

Pour les PM₁₀, *Cprox* conduit à une valeur de concentration de fond urbain en général de 5 µg/m³ sur la périphérie du territoire, valeur qui augmente lorsqu'on s'approche du centre parisien et qui atteint 15 µg/m³ en petite couronne. Ces valeurs semblent sous-estimées puisque Airparif simule des teneurs sur ces zones comprises entre 20 et 25 µg/m³. Dans les deux cartes, les concentrations annuelles moyennes sur Paris sont pour l'essentiel comprises entre 20 et 35 µg/m³ mais *Cprox* excède par endroit 40 µg/m³.

Notre manière de cartographier, moins lissée que celle d'Airparif sur la grande couronne, fait apparaître des gradients plus marqués sur les routes pour NO₂ et les PM₁₀.

IV. Travail sur l'année 2017

Comme je l'ai mentionné précédemment, après ce travail d'extrapolation de la descente en échelle à des sites non utilisés pour l'apprentissage, j'ai décidé de tester la formulation de *Cprox* sur une autre période temporelle et donc un autre jeu de données météorologiques, afin de tester la robustesse du protocole de descente en échelle. L'idée était également d'élargir la base de données de travail, puisque le nombre de sites de proximité au trafic, et mesurant les PM₁₀, a augmenté depuis 2009. Bien sûr, l'erreur du modèle étant dépendante de la justesse de son inventaire, l'apprentissage doit être reconduit puisque j'ai utilisé un nouvel inventaire.

1. Choix de l'année 2017

Qualité de l'air

Airparif décrit dans son bilan de la qualité de l'air de 2017 que les niveaux de pollution de la ville de Paris se sont améliorés par rapport aux années précédentes. La valeur moyenne de concentration en NO₂ mesurée sur les stations de trafic avait diminué de 113 µg/m³ à 90 µg/m³ mais les niveaux sur les stations de fond en agglomération et hors agglomération étaient restés plutôt stables au cours du temps (les diminutions suivantes ont respectivement été observées : 45 → 39 µg/m³ et 12 → 10 µg/m³). Cependant, les tendances relatives à l'évolution du NO₂ diffèrent fortement entre les sites de trafic et les sites de fond. Le rapport fait ainsi état d'une diminution de 34% des niveaux de NO₂ en fond urbain parisien alors qu'ils restent relativement constants à proximité du trafic. *A contrario*, sur l'ensemble de la région, les concentrations en NO₂ sont stables avec un niveau de fond régional moyen proche de 10 µg/m³ pour les sites de fond tandis qu'on observe une légère baisse à proximité du trafic.

Pour les PM₁₀, polluant dont les gradients de concentration sont moins importants, une tendance à la baisse est observée dans Paris depuis une dizaine d'années avec une diminution de 25% des concentrations sur les sites de trafic et de 20% sur les sites de fond. Sur l'ensemble de l'Île-de-France, une légère diminution des niveaux est observée depuis 2016 grâce à des conditions météorologiques dispersives qui ont permis de maintenir une bonne qualité de l'air sur la globalité de l'année. De fait, le nombre d'habitants potentiellement concernés par les dépassements des valeurs limites journalières en PM₁₀ a diminué de 3,8 millions à environ 100 000, entre 2009 et 2017.

Malgré cela, les niveaux de plusieurs polluants dépassent toujours les valeurs limites dans Paris *intramuros*. Il y a eu, au total, 6 jours dans l'année où les concentrations en PM₁₀ ont dépassé les valeurs limites journalières sur l'ensemble des stations de fond parisiennes. Ce chiffre reste toutefois deux fois inférieur à celui de 2016 grâce aux conditions météorologiques. Sur les axes routiers, les concentrations dépassent toujours les valeurs annuelles (NO₂ et PM₁₀) et journalière (PM₁₀). Pour NO₂, les valeurs de concentrations les plus élevées mesurées à proximité des grands axes routiers peuvent toujours être jusqu'à 2 fois supérieures à la réglementation. La station de mesure située sur l'Autoroute A1 est donnée comme un exemple des sites où les dépassements des valeurs limites journalières pour les particules arrivent presque 1 jour sur 5 dans l'année.

Pour expliquer les tendances à la baisse de nos deux polluants d'intérêt par rapport aux années précédentes, il faut considérer la baisse du taux de véhicules diesel dans Paris *intramuros*, et l'évolution du parc routier vers un parc routier plus propre a permis de diminuer les taux de particules primaires retrouvés en agglomération.

Trafic

Quelques modifications significatives du réseau routier ont eu lieu depuis 2009. La Mairie de Paris a notamment fermé les voies sur berges en septembre 2016, soit l'équivalent de 3,5 km d'autoroutes qui traversaient, initialement, la ville. Airparif a mené deux campagnes de mesure⁴ (hivernale et estivale), comptabilisant 80 points de mesure, pour évaluer l'impact de la fermeture des voies sur berges sur la qualité de l'air. Ils ont pu mettre en avant une diminution des niveaux de NO₂ le long des quais malgré le fait que la circulation au niveau des quais haut soit restée la même. Par contre, ils ont constaté une diminution de la qualité de l'air dans la partie Est de la ville, le long du quai de Bercy, dû au report de la circulation et des phénomènes de congestion qui en découlent.

ZFE et ZCR

La zone à faibles émissions de Paris a été initiée en 2015 et renforcée en 2017. Par contre, c'est seulement à partir de 2019 qu'a été imposée une limitation aux Crit'air 4 et 5 de rouler dans la ZFE.

La ville de Paris a pris deux arrêtés de police en 2015 et en 2016 pour restreindre la circulation des poids lourds et des véhicules légers les plus polluants. En 2016, les véhicules pouvant rouler sans restriction dans la ville devaient respecter les normes Euro 3 ou plus récentes pour les poids lourds, Euro 2 ou plus récent pour les véhicules particuliers et les véhicules utilitaires légers, et Euro 1 ou plus récent pour les deux roues motorisées.

Réseau de mesure Airparif

Le nombre de stations de mesure disponibles pour la surveillance des niveaux de NO₂ et de PM₁₀ a augmenté entre 2009 et 2017. Au total, 5 nouvelles stations de mesure de trafic ont été installées, dont les sites « BP_EST », « HAUS », « OPERA », « SOULT » et « RD934 » comme listés dans le [Tableau 4](#). Ces 5 stations ont été intégrées dans les jeux de validation de NO₂ et de PM₁₀ de manière à ce que l'on ait un nombre de socle de données de validation plus large. Les étapes de validation en seront, de cette façon, plus robustes. Au niveau des stations de fond, une station supplémentaire a été ajoutée au réseau de mesure pour NO₂ et 3 pour les PM₁₀. Quatre sites ont été ajoutés au protocole d'apprentissage afin d'augmenter l'efficacité de la détermination d'un coefficient *k* pour les stations de fond du jeu de validation.

⁴ https://www.airparif.asso.fr/sites/default/files/pdf/Rapport_voiesurberges_rapportfinal_101117.pdf

2. Méthodologie

La méthodologie utilisée dans cette partie suit celle qui a été définie pour le travail sur l'année 2009. Nous n'avons pas cherché à appliquer directement les équations développées dans la section III à l'ensemble des sites disponibles pour 2017 car nous sommes partis du postulat que le calcul des coefficients α , β et donc k dépendent fortement des cadastres d'émissions considérés pour réaliser les simulations de CHIMERE. Comme nous l'avons vu dans la [Section I.](#), nous devons considérer une année de travail très différente de 2009. Nous avons donc basé notre calcul pour CHIMERE sur un inventaire d'émissions remis à jour : l'inventaire d'émission d'Airparif de 2015 (en lieu et place de celui de 2008). La principale complexité est que les méthodologies de construction et de formatage des inventaires (degré de déconvolution des secteurs d'activité, polluants à disposition...) avaient évolué entre les deux années et cela a affecté la structure et les propriétés de l'inventaire.

3. Retravail du cadastre d'émission de 2015

Nous avons pu obtenir d'Airparif un ensemble de données nous permettant de reconstruire un inventaire pour l'année 2017. Un long travail d'adaptation et de complétion de l'inventaire d'émission de 2015, dont le format différait de celui de 2008, a été nécessaire pour pouvoir simuler les niveaux de pollution de l'année 2017.

L'inventaire d'émission du cadastre de 2008 contenait les informations sur 8 polluants et gaz à effet de serre (NO_x, PM₁₀, PM_{2.5}, SO₂, COVNM, CO₂, CH₄ et N₂O) et sur l'ensemble des secteurs d'activités avec le niveau d'information sur les SNAPs le plus élevé (jusqu'à 4). En revanche, l'inventaire d'émission de 2015 contenait des informations plus agrégées, notamment en termes de secteurs d'activité. En l'occurrence, les secteurs d'activités ne suivaient pas les codes SNAP présentés précédemment mais étaient regroupés dans 5 catégories : (1) le transport routier, (2) les autres transports, (3) le secteur résidentiel, (4) le secteur tertiaire et (5) les autres types d'émissions. Nous avons, cependant, des informations sur les parts moyennes des différents sous-secteurs du transport routier, pour toutes les mailles du cadastre contenant des émissions de ce SNAP. Au-delà du caractère plus agrégé des données par SNAP, l'inventaire de 2015 ne contenait les informations que des NO_x, des PM₁₀ et des PM_{2.5}. Or, comme expliqué précédemment, les informations des autres polluants présents dans l'atmosphère urbaine, certains étant précurseurs d'une fraction importante des particules et donc indispensables pour pouvoir modéliser correctement les concentrations dans CHIMERE.

J'ai donc cherché à ajouter les informations manquantes dans cet inventaire d'émission. Deux façons d'agir s'offraient à moi : soit il m'était possible de trouver les totaux par polluant, ainsi que les répartitions sectorielles des émissions des polluants manquants dans la littérature, soit il me fallait devoir trouver ou calculer des taux d'évolutions des polluants entre 2009 et 2017. L'objectif étant d'appliquer ces coefficients d'abattement aux valeurs de l'inventaire de 2008 pour en déduire ceux de 2015. J'ai dirigé ma recherche bibliographique sur les articles ou rapports susceptibles de me renseigner sur les taux d'émissions de la région Île-de-France. Mais, de cette recherche bibliographique, je n'ai pu identifier aucun rapport donnant des informations suffisamment précises sur les émissions des polluants manquants sur la région et leur distribution en secteur. C'est pourquoi je me suis dirigée vers la recherche/détermination de coefficients d'abattement à appliquer au cadastre de 2008.

La principale source bibliographique adaptée était le bilan des émissions d'Airparif de l'année 2015. Ce bilan m'a permis de déduire des taux d'évolution pour le SO₂ et les COVNM pour chaque secteur d'activité, mais aucune information sur les autres polluants n'était mise à disposition, que ce soit dans le même rapport ou dans d'autres rapport. J'ai cherché à mettre en forme le cadastre d'émission en gardant de manière prioritaire les données présentes dans l'inventaire qui m'a été fourni pour 2015, et en déduisant celles qui me manquaient. En conséquence, je n'ai interpolé les émissions de 2008 que pour SO₂ et COVNM. Pour le reste, j'ai déduit des données (inventaires + rapport) à ma disposition, des tendances d'évolution pour tous mes polluants à partir des totaux de chaque secteur du cadastre de

2008 et des données présentées pour l'année 2015. Les résultats de mes calculs sont donnés dans le [Tableau 9](#) à titre informatif. Les totaux d'émissions ont été laissés à 0 pour les autres polluants (CH₄, CO₂ et N₂O) car aucune information sur leurs bilans ou leurs évolutions au cours du temps n'est fournie dans cette littérature.

Le secteur d'émissions « Autre type d'émission » du cadastre de 2015 représente la somme des émissions des SNAPs 01, 03, 04, 05, 06, 09, 10 et 11. Nous avons cherché à considérer au maximum les données fournies dans le cadastre de 2015 et, de ce fait, je n'ai pas cherché à répartir ces émissions dans leurs SNAPs respectifs. J'ai gardé les valeurs du secteur « Autre type d'émission » pour les NO_x, les PM₁₀ et les PM_{2,5}, et j'ai recréé ce secteur pour SO₂ et les COVNM en sommant les valeurs interpolées de chacun des SNAPs le composant.

Tableau 9 - Évolution des émissions de NO_x, PM₁₀, PM_{2,5}, SO₂ et COVNM entre 2008 et 2015 pour chaque secteur d'activité.

	NO _x	PM ₁₀	PM _{2,5}	COVNM	SO ₂
Secteur d'activité	Taux d'évolution 2008-2015	Taux d'évolution 2008-2015	Taux d'évolution 2008-2015	Taux d'évolution 2008-2015	Taux d'évolution 2008-2015
Industrie	-45%	-50%	-64%	-34%	-65%
Branche énergie	-58%	-71%	-71%	-10%	-57%
Déchets	4%	-84%	-91%		-19%
Résidentiel	-17%	-22%	-22%	-27%	-65%
Tertiaire	-21%		11%		-71%
Chantier	-32%	-16%	-21%	-39%	-96%
Transport routier	-23%	-27%	-33%	-64%	
Trafic ferroviaire et fluvial	-69%	0%	-10%		
Plateforme aéroportuaire	9%	15%	64%	-51%	27%
Agriculture	-50%	-4%	-33%	-94%	-96%
Emissions naturelles				25%	

La mise en forme des données de l'inventaire exige que les émissions soient redistribuées dans les secteurs de CHIMERE ([Tableau 1](#)). Le problème s'est posé d'attribuer le secteur (« Autres émissions ») à une catégorie de CHIMERE. C'est une décision qui n'est pas sans conséquences puisque chacun de ces secteurs est associé à des profils temporels d'activité (mensuels, hebdomadaires et journaliers) distincts. Avec l'expertise d'Airparif, j'ai pu définir que le profil temporel le plus adapté à ce secteur d'activité est celui qui correspond au SNAP 03 et donc le Secteur 5 des cadastres CHIMERE ([Tableau 2](#)).

Une fois le cadastre au format CHIMERE généré, j'ai pu réaliser les simulations en suivant les paramétrisations définies dans la [Section II.1](#).

4. Résultats du travail sur l'année 2017

La méthodologie appliquée ici est la même que celle du développement de l'approche pour l'année 2009. Seuls les coefficients et leur partage sont recalculés pour l'année 2017, le reste de l'équation *Cprox* est gardé tel quel. Afin de ne pas encombrer le manuscrit avec des répétitions d'explications méthodologiques, seuls les résultats et les discussions qui y sont associées sont présentés dans cette section du chapitre.

Phase d'apprentissage

Le calcul des coefficients des stations du jeu d'apprentissage suit la méthodologie présentée dans la [Section III.1](#). Les tableaux de résultats sont présentés en [Annexe 6](#) et les séries temporelles ainsi que les boxplots sont illustrés dans l'[Annexe 7](#). Les principales données sur les scores sont présentées dans les [Figure 26](#) et [Figure 27](#) ci-dessous.

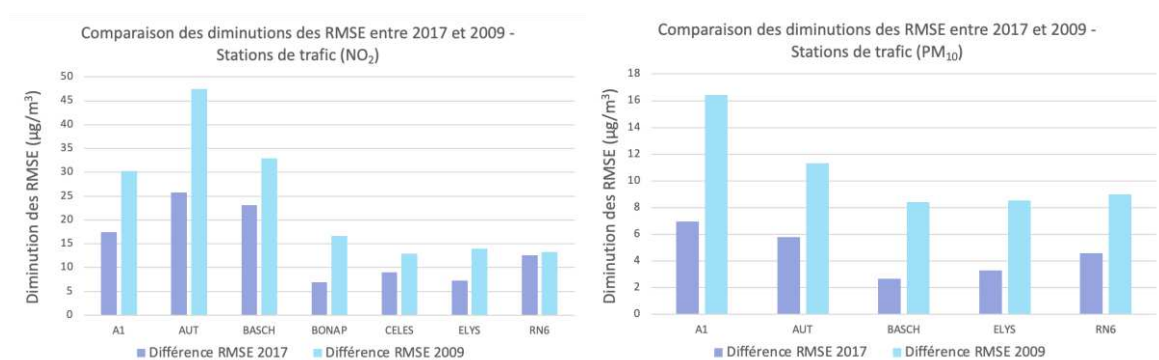


Figure 26 - Gains obtenus sur la RMSE ($RMSE_{CHIMERE} - RMSE_{Cprox}$) pour 2017 (en bleu foncé) et 2009 (en bleu clair) pour NO_2 (à gauche) et les PM_{10} (à droite).

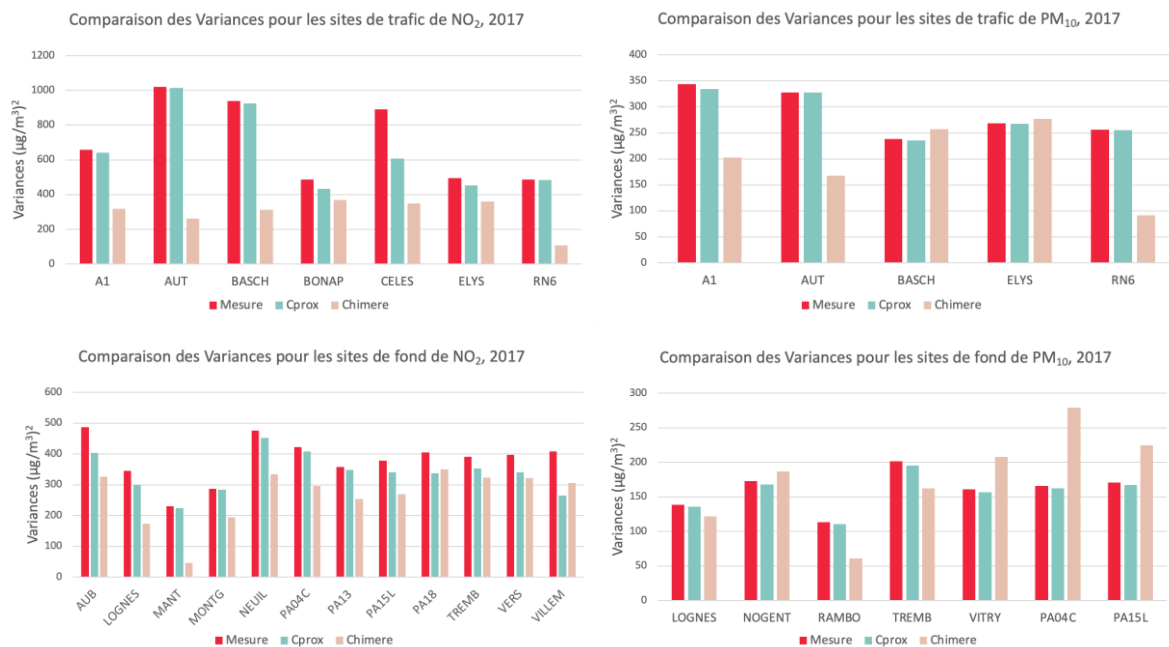


Figure 27 - Comparaison des variances pour les sites d'apprentissage de NO_2 (à gauche) et PM_{10} (à droite) pour l'année 2017. Les variances issues des mesures sont représentées en rouge, celles issues de *Cprox* sont en vert et celles issues de *CHIMERE* sont en beige.

Comme on le voit sur les figures ci-dessus, l'application de l'Équation 4 conduit à une amélioration des RMSE, qui diminuent en moyenne de 29% pour les sites de trafic de NO₂ et de 19% pour les stations de PM₁₀. Le gain obtenu pour 2017 sur les RMSE est plus faible que celui obtenu pour 2009, mais les RMSE de départ et d'arrivée sont en 2017 plus faibles également : le résultat est au final très semblable. Tous les sites de trafic et de fond, tant pour les PM₁₀ que pour le dioxyde d'azote, font état d'un ajustement significatif de leur variance par rapport à la mesure, que ce soit à la hausse ou à la baisse. Seuls deux sites sur tous ceux étudiés présentent encore un écart après correction, ce qui est satisfaisant. On note à nouveau la différence dans le besoin d'ajustement de la variance entre les sites de fond urbains (Paris 4e et Paris 15e) et les sites de fond périurbains à ruraux (Lognes, Rambouillet) qui confirment qu'une nécessaire évolution de l'approche serait de nuancer les coefficients de fond selon leur distance à la ville dense.

L'amélioration de la moyenne annuelle apportée par la correction *Cprox* est reportée sur le graphique suivant (Figure 28). Les résultats montrent que, pour le dioxyde d'azote, l'amélioration est sensible et la correction est favorable sur tous les sites, même si pour certains sites à très haute fréquentation routière (A1, périphérique Auteuil) les écarts avec la mesure ne sont pas complètement comblés. En proximité au trafic pour la particules, l'amélioration est significative mais il reste une sous-estimation de la valeur mesurée de l'ordre de 20 à 30% sur 4 des 5 sites. Enfin, la correction sur la valeur annuelle moyenne est moins satisfaisante sur les PM₁₀ en fond urbain, car l'effet de la correction est très faible.

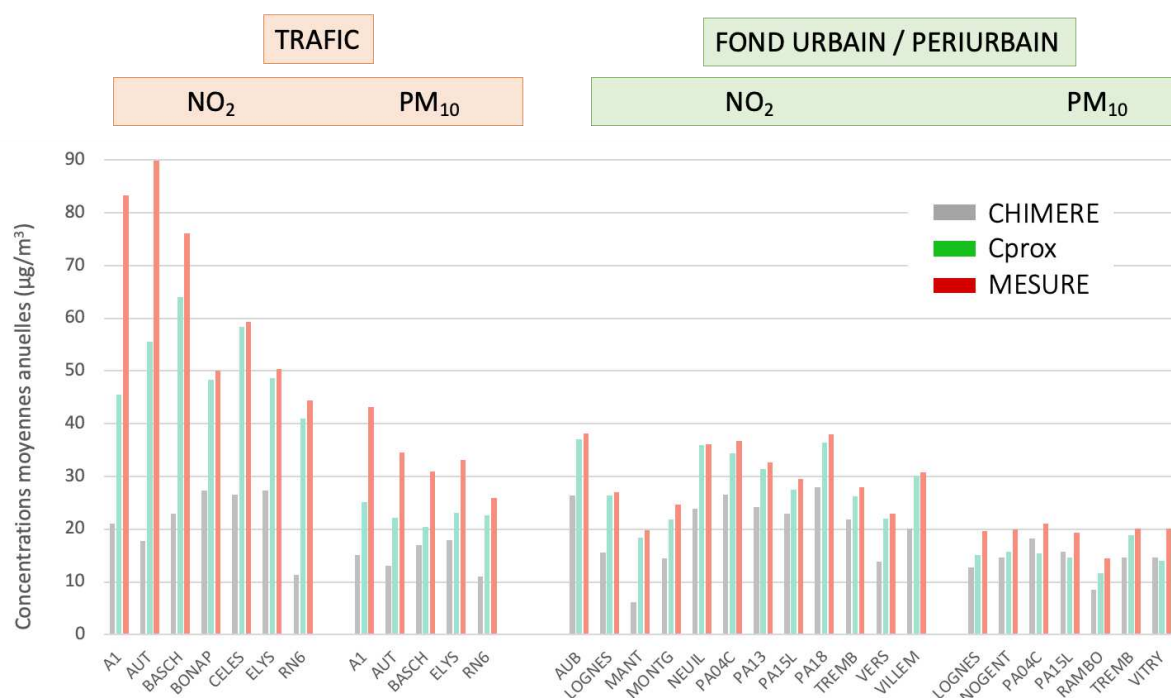


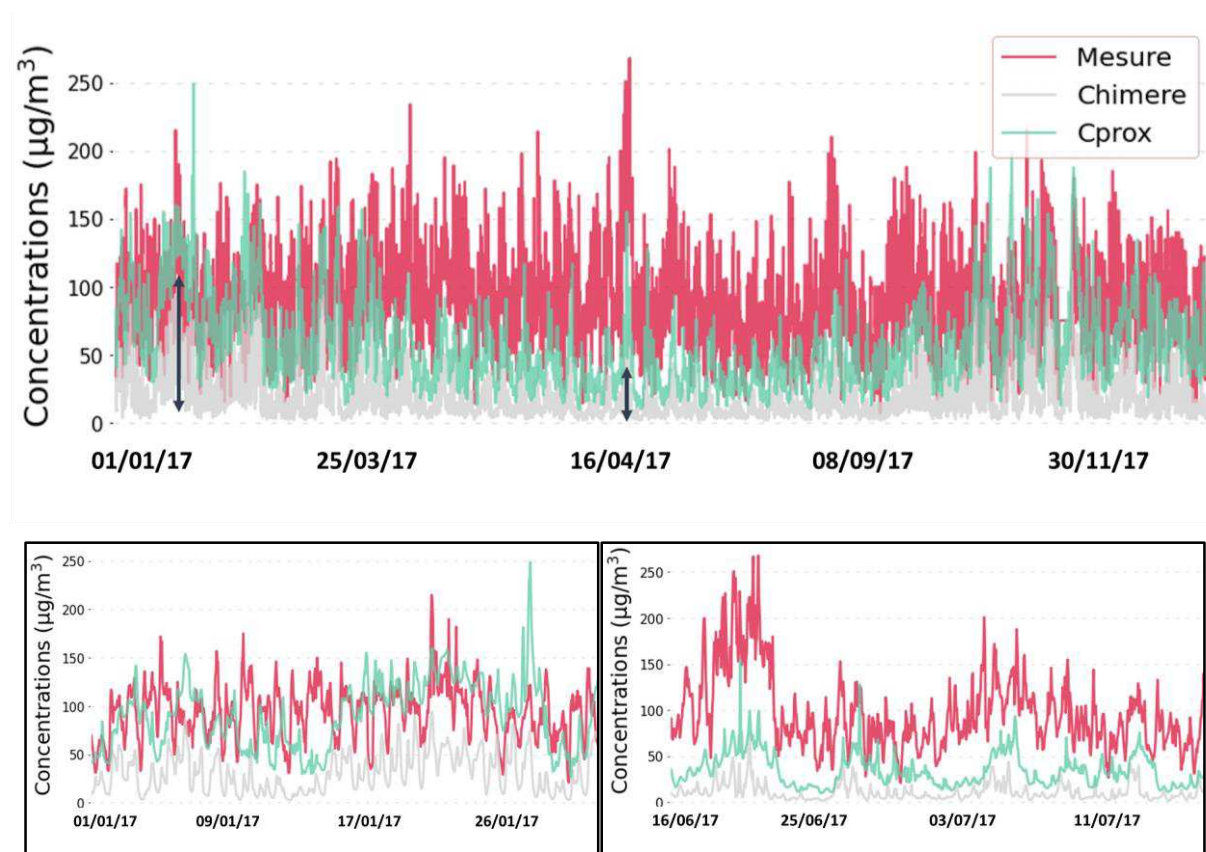
Figure 28 - Comparaison des moyennes annuelles calculées sur les sites d'apprentissage de trafic et de fond urbain, pour NO₂ et pour les PM₁₀ pour l'année 2017. Les moyennes issues des mesures sont représentées en rouge, celles issues de *Cprox* sont en vert et celles issues de CHIMERE sont en gris.

Les boxplots (montrés en Annexe 7) montrent enfin l'élargissement de la gamme sur laquelle sont distribuées les données, après l'application de *Cprox*, ce qui va dans le sens de la mesure.

Les séries temporelles issues de l'étape d'apprentissage sont également détaillées en Annexe 7. Un phénomène majeur apparu sur le cas d'étude 2017 doit être mentionné et discuté. On note en effet une tendance saisonnière sur la qualité de la simulation produite par CHIMERE, plus marquée en 2017 qu'en 2009, qui se traduit par une sous-estimation particulièrement forte des concentrations simulées pendant la période printemps-été, et ce sur la plupart des sites et pour les deux polluants. Ce biais

provenant de CHIMERE se répercute dans *Cprox* qui - comme on l'a mentionné - ne peut pas ajuster sa correction à la saison. Il est particulièrement visible sur le site de la porte d'Auteuil en [Figure 29](#). Notons que sur de nombreux sites la tendance à la diminution des niveaux en été est observée dans la mesure, mais le phénomène que l'on décrit est une augmentation du biais du modèle entre hiver et été, et pas uniquement une diminution des niveaux de concentration simulés.

Séries temporelles - NO₂ - Site de proximité au trafic routier Porte d'Auteuil



Séries temporelles - PM₁₀ - Site de proximité au trafic routier Porte d'Auteuil

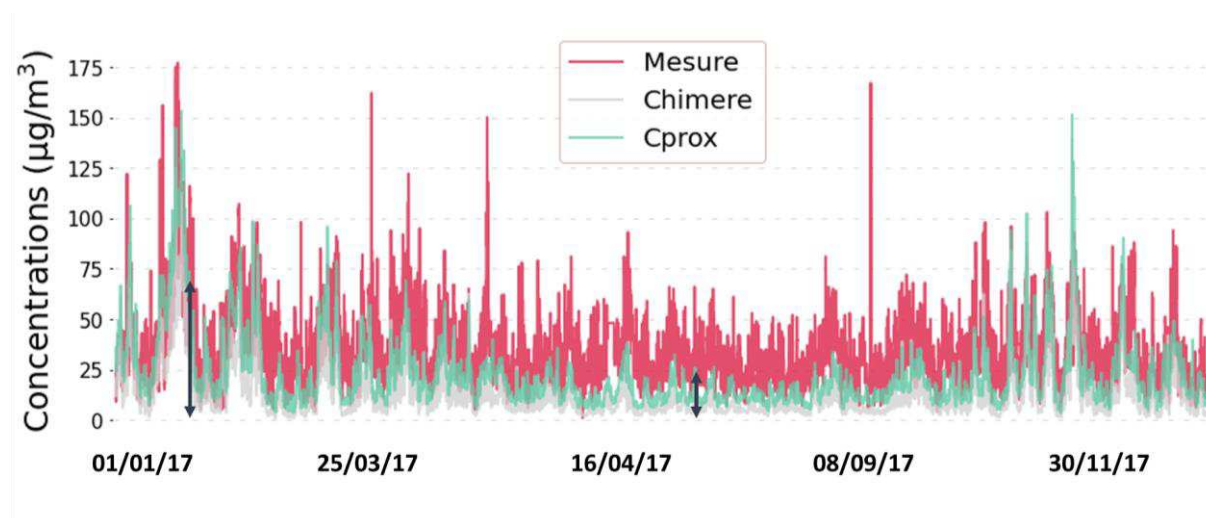


Figure 29 - Séries temporelles sur le site de trafic Porte d'Auteuil pour NO₂ (en haut) et pour les PM₁₀ (en bas) pour l'année 2017. Les deux graphiques du centre sont des zooms issus du graphique du haut, sur des périodes plus courtes. Les courbes de la mesure sont représentées en rouge, celles issues de *Cprox* sont en vert et celles issues de CHIMERE sont en gris.

J'ai représenté sur cette figure les données sur l'année (les deux flèches noires indiquent l'amplitude des données CHIMERE en hiver et en été), ainsi que deux zooms temporels, l'un en hiver et le second en été. Il est bien sûr possible que le faible niveau de détail de l'inventaire contribue à accentuer le biais saisonnier sur les émissions, puisque pour l'ensemble du secteur « Autre type d'émission » nous ne pouvons utiliser qu'un seul profil temporel correspondant à celui du secteur de la combustion dans les industries manufacturières mais cet effet est certainement de second ordre car le profil retenu a été choisi pour ressembler au plus grand nombre de sous-secteurs.

Après des échanges avec des utilisateurs de CHIMERE, ce problème est en fait fréquent dans les sorties des CTM et de CHIMERE en particulier, qui restituent parfois mal l'amplitude des variations saisonnières de la mesure. L'ampleur du phénomène est variable selon les années, ce qui signifie qu'il peut trouver son origine dans la restitution des paramètres météorologiques aux différentes saisons, en particulier la hauteur de la couche limite. Une des hypothèses retenues à ce stade par les utilisateurs de CHIMERE est en effet une tendance à utiliser une valeur trop basse de couche limite en hiver, ce qui entraîne des niveaux hivernaux artificiellement plus hauts en sortie de CHIMERE. Cette différence de qualité saisonnière ne peut pas être compensée dans *Cprox* sans implémentation d'un indicateur météorologique, ou la possibilité de mener des corrections distinctes sur des périodes temporelles plus courtes (saisonnières, mensuelles...) au lieu d'utiliser des couples de coefficients (α , β) moyens sur l'année. Bien sûr, l'amélioration des simulations du CTM est hors du cadre du travail défini initialement, puisqu'il s'agissait juste de produire des champs de concentration de polluants avec des gradients à fine échelle dans les espaces urbanisés. Mais les résultats montrent qu'il est sans doute pertinent de profiter de la descente en échelle pour corriger des biais saisonniers du CTM.

Ainsi, les tests sur l'année 2017 mettent en lumière la nécessité une des perspectives pour cette approche serait de définir des coefficients α , β et k pour chaque saison afin de contourner les problèmes de signaux initiaux du modèle. Il serait aussi intéressant d'analyser si une différenciation diurne des coefficients pourrait être un atout pour notre approche, par exemple pour restituer les valeurs nocturnes les plus basses mais ce n'est pas une problématique de premier ordre.

Automatisation de la correction

Comme précédemment, j'ai étudié la variabilité des coefficients α , β et k afin de vérifier si l'on pouvait utiliser des valeurs moyennes pour certains sous-groupes de mesure. Les détails des coefficients sont résumés en [Annexe 6](#). Ils montrent une faible variabilité de la valeur de k pour les sites de fond : ainsi, comme pour l'année 2009, j'ai considéré une valeur moyenne pour NO_2 ($k_{\text{fond}} = 1,54$) et pour les PM_{10} ($k_{\text{fond}} = 1,69$). J'ai tracé sur la [Figure 30](#) les parts des coefficients issus de l'apprentissage sur chaque site. On voit qu'elles permettent de proposer un partage moyen sur chacun des types de sites, qui diffère toutefois du partage retenu pour 2009 car les qualités et défauts des données de CHIMERE - et donc le besoin de correction - sont différents selon l'inventaire utilisé.



Figure 30 - Contribution des coefficients α et β à k , en pourcentage, pour les stations de trafic (à gauche) et de fond (à droite) des jeux d'apprentissage du NO_2 (en haut) et des PM_{10} (en bas) pour l'année 2017.

De la même façon que pour 2009, les rapports HE ont été corrélés aux coefficients k obtenus précédemment, et les corrélations sont illustrées dans la Figure 31. La même méthodologie a été utilisée pour les régressions de l'année 2009. Les valeurs des rapports HE présentées dans le Tableau 7 et ne sont utilisées que pour déterminer les coefficients k des sites de validation. Dans la suite du document, des figures illustreront des comparaisons entre les moyennes annuelles obtenues entre les sites d'apprentissage et de validation (Figure 32 et Figure 33). Du fait de la différence de méthodologie dans la détermination de leurs coefficients, les moyennes des sites d'apprentissage de trafic ont été calculées avec les coefficients k d'apprentissage, et celles des sites de validations ont été calculées avec les coefficients $k_{fité}$. Il n'est, bien sûr, pas pertinent de réaliser une comparaison poussée entre ces deux résultats mais les figures concernées permettent de montrer que l'approche de validation conduit à des résultats ayant des indicateurs de qualité similaires à ceux de l'étape d'apprentissage.

Lorsqu'on considère l'ensemble des stations d'apprentissage, le coefficient de corrélation de Pearson (R^2) obtenu pour les sites de NO_2 est de 0,56 et celui des PM_{10} est de 0,62. Le positionnement des sites A1 et RN6, de part et d'autre de la tendance donnée par AUT et BASCH, met en avant un des problèmes les plus délicats de l'approche : le manque de sites pour tracer des corrélations. Sur un ensemble plus large de points, on aurait pu discuter du poids à accorder à chacun de ces deux sites mais avec 7 points il est difficile de savoir lequel est le plus représentatif. Or les abscisses de ces points sont très sensibles à la valeur du TMJA simulé par OLYMPUS. J'ai choisi de retenir la droite ci-dessous qui passe par les 3 points à fort trafic les plus regroupés. Pour rendre l'approche plus robuste, il faudrait pouvoir compter sur plus de points.

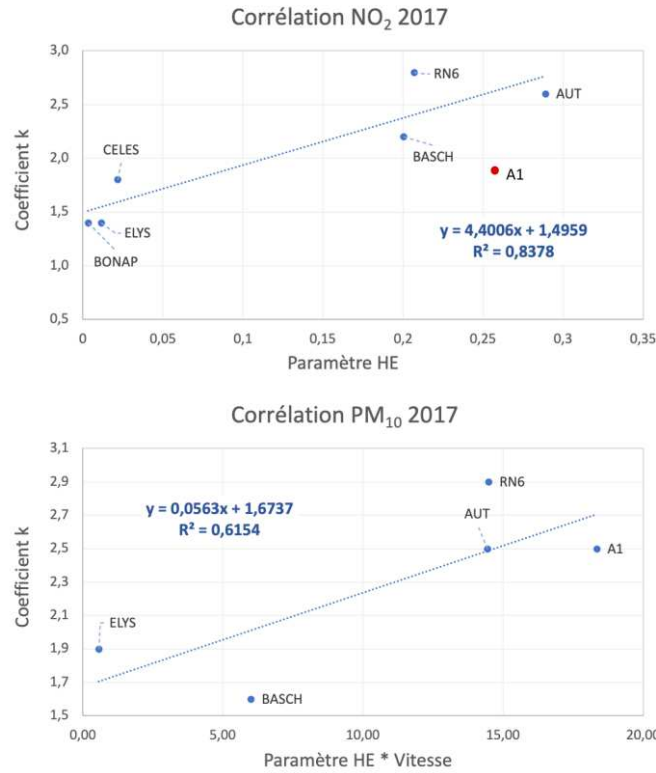


Figure 31 - Corrélations obtenues entre le paramètre HE et les coefficients k des stations de trafic de NO₂ (en haut) et de PM₁₀ (en bas) pour 2017.

Phase de validation

Les séries temporelles ainsi que les boxplots issus de l'application de *Cprox* sur les sites de validation sont présentés en Annexe 8. Les résultats sont illustrés par les courbes de deux sites de fond (péri-) urbain dans la Figure 33.

Nous avons de nouveaux sites de validation en proximité au trafic. Tous montrent une amélioration de l'estimation des moyennes annuelles, au moins aussi bonne que sur les sites d'apprentissage. Dans l'ensemble, la correction permet d'augmenter la variance pour lui donner plus de réalisme. Certains sont en revanche insatisfaisants, comme celui du site du boulevard périphérique Est (BP EST sur la Figure 32) dont la correction sur la moyenne annuelle est trop forte probablement en raison de l'estimation d'un trafic très fort sur le périphérique, alors que la mesure est en léger retrait du périphérique. Il est probable qu'il faille penser à plafonner la correction k en proximité au trafic, notamment lorsque les axes routiers sont très ouverts comme c'est le cas du périphérique francilien.

Le phénomène précédemment décrit de biais saisonnier est toujours présent. Pour NO₂, la courbe *Cprox* des sites OPERA et SOULT (en plein cœur de Paris) est très réaliste en début et fin d'année mais elle sous-estime les niveaux moyens en milieu d'année. La plupart des courbes *Cprox* corrigées présentent une diminution de la RMSE et une augmentation de leur variance, ce qui tend vers plus de réalisme. De par l'approche d'apprentissage choisie, qui s'appuie en partie sur l'indicateur « moyenne annuelle », les courbes *Cprox* voient pour la plupart leur moyenne annuelle améliorée. Mais la bonne restitution, que l'on observe sur RN2 par exemple, cache sur 2017 une succession de sous-estimations (estivales) et surestimations (hivernales) liées au biais de CHIMERE.

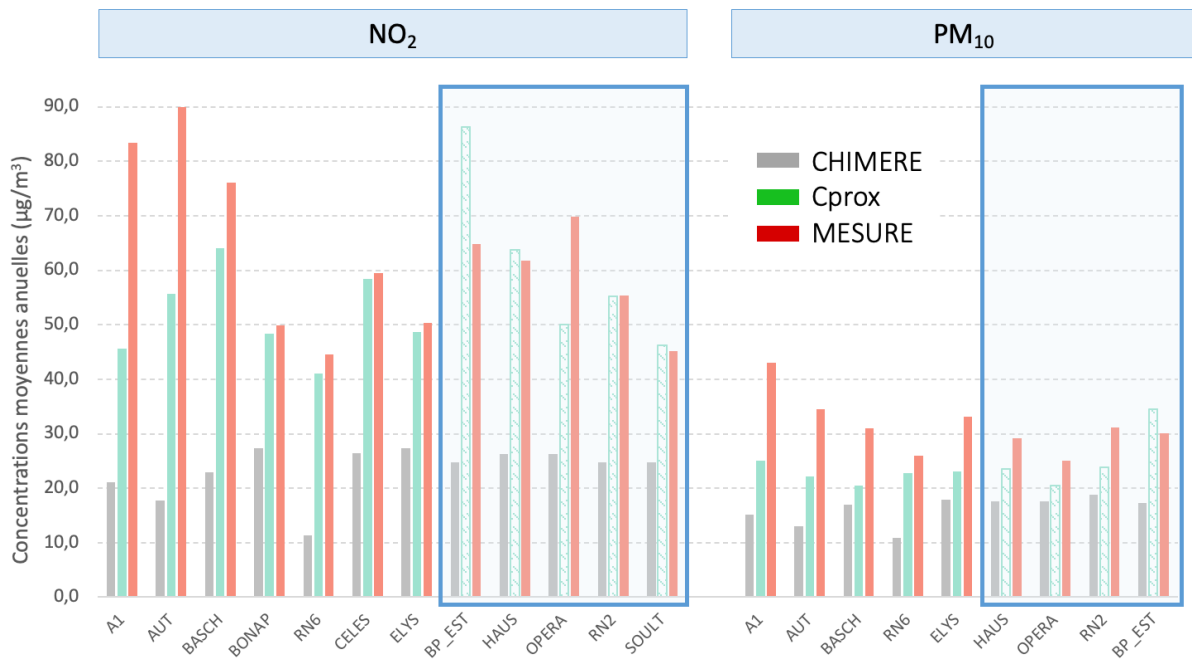


Figure 32 - Comparaison des moyennes annuelles calculées sur les sites d'apprentissage et de validation (encadrés en bleu) en proximité au trafic, pour NO_2 (à gauche) et pour les PM_{10} (à droite) pour l'année 2017. Les moyennes issues des mesures sont représentées en rouge, celles issues de *Cprox* fité sont en vert plein, celles issues de *Cprox* automatisé sont en vert pointillé et celles issues de CHIMERE sont en gris.

Les deux courbes présentées en exemple en Figure 33, sur une période non estivale, illustrent l'amélioration apportée en site de fond (péri-)urbain. Ces courbes sont associées à des boxplots qui confirment une meilleure correspondance des moyennes annuelles entre *Cprox* et la mesure qu'entre CHIMERE et la mesure. Il reste que, comme on le voit ci-dessous, les pics sont toujours mieux reproduits lorsque le site est en zone urbaine assez dense, proche de Paris (c'est le cas de Champigny) qu'à plus grande distance, dans des zones périurbaines (c'est le cas d'Evry).

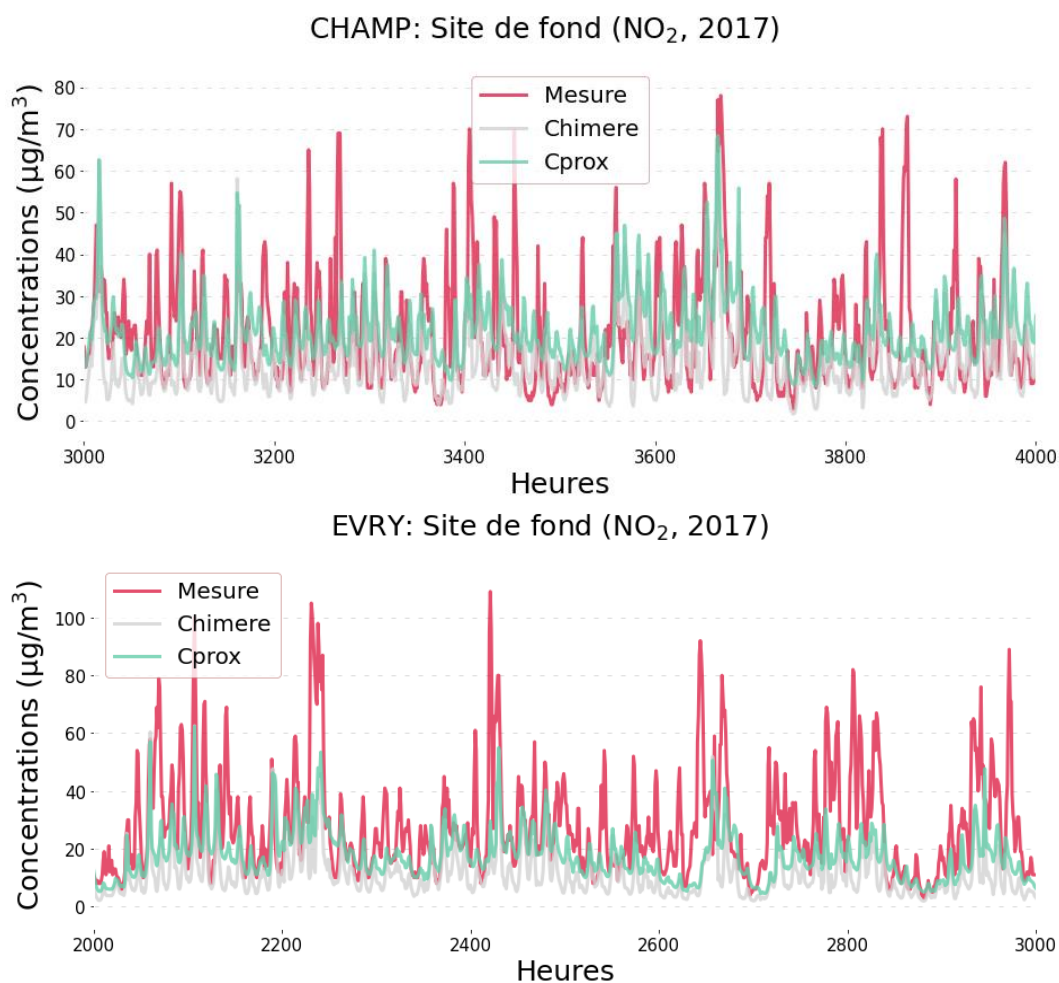


Figure 33 - Séries temporelles sur deux sites de validation de fond urbain, pour NO₂ pour l'année 2017. Les courbes de la mesure sont représentées en rouge, celles issues de *Cprox* sont en vert et celles issues de CHIMERE sont en gris.

Les mêmes conclusions peuvent être tirées pour particules (voir en [Annexe 8](#)).

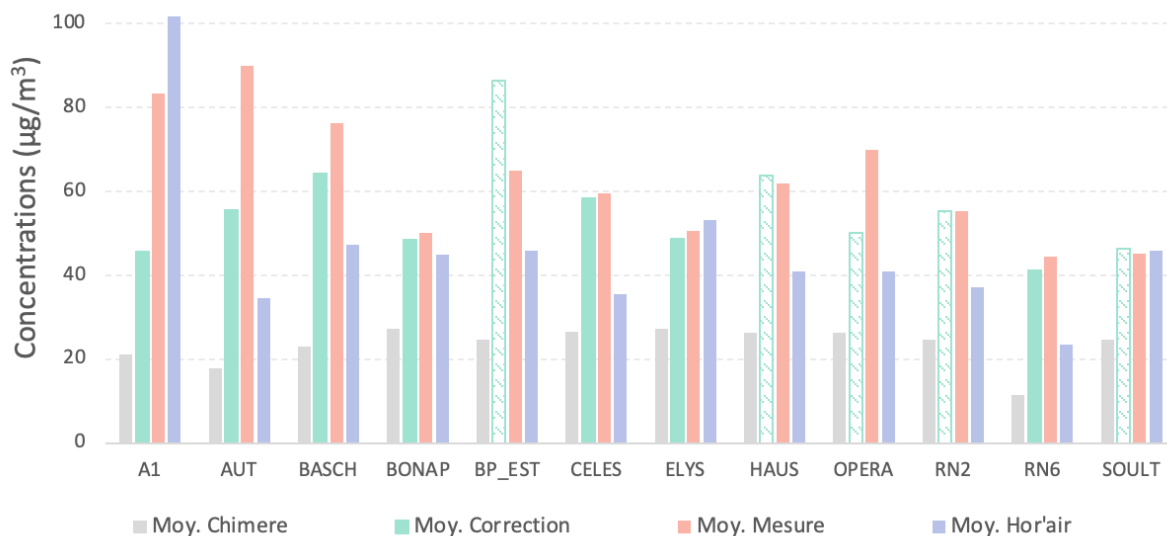
Comparaison avec une approche de référence

Pour compléter mon travail d'évaluation et de validation de *Cprox*, j'ai recherché des résultats d'autres démarches de modélisation/d'extrapolation à fine échelle sur la région parisienne. J'ai pu récupérer les sorties annuelles moyennes du modèle HOR'AIR d'Airparif pour l'année 2017, pour les deux polluants sur lesquels j'ai mené mon étude.

J'ai confronté les valeurs d'HOR'AIR avec les résultats de *Cprox*. Dans le cadre de cet exercice, j'ai considéré les valeurs HOR'AIR comme une référence afin de vérifier la qualité de mon approche statistique de descente en échelle. Les comparaisons avec les données de NO₂ et de PM₁₀ sont illustrées dans la [Figure 34](#). Sur ces figures, les valeurs moyennes annuelles de CHIMERE sont représentées en gris, celle de la mesure est en rouge et celle d'HOR'AIR est en bleu.

Deux types de résultats sont illustrés pour *Cprox*, les moyennes annuelles issues des calculs de coefficients *k* de l'étape d'apprentissage pour les sites d'apprentissage (en vert uni) et celles issues de l'étape d'automatisation de l'approche (en vert avec fond hachuré) pour les sites de validation.

Comparaison des moyennes annuelles NO₂ - Stations de trafic (2017)



Comparaison des moyennes annuelles PM₁₀ - Stations de trafic (2017)

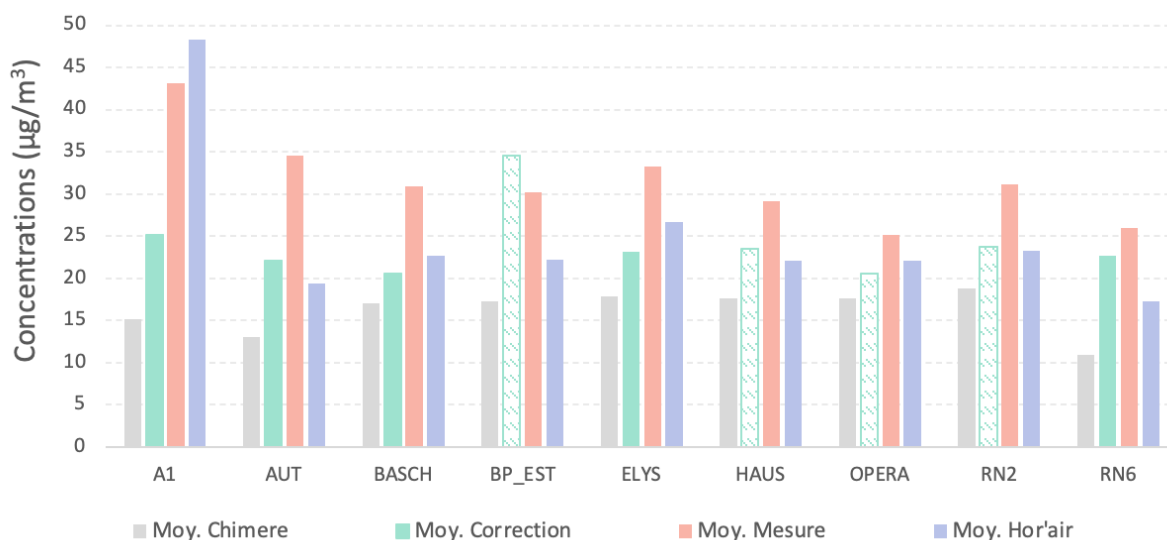


Figure 34 - Confrontation des moyennes annuelles obtenues avec les sorties du modèle HOR'AIR d'Airparif pour NO₂ (en haut) et pour les PM₁₀ (en bas). Les valeurs de CHIMERE sont représentées en gris, celles de la correction issue de l'apprentissage en vert, celles de la correction automatisée en vert pointillé, celles de la mesure en rouge et celles d'HOR'AIR en bleu.

Globalement, on constate sur ces figures le très grand potentiel de *Cprox* à prédire les niveaux annuels moyens d'exposition sur chaque site de proximité au trafic, *Cprox* proposant toujours une estimation des moyennes annuelles au moins aussi bonne que celle issue du modèle HOR'AIR, pour NO₂ et pour les PM₁₀. Les niveaux moyens calculés par l'approche automatisée de *Cprox* sont améliorés par rapport aux niveaux initiaux de CHIMERE des deux polluants, à l'instar de ce qui est observé pour les résultats de l'étape d'apprentissage de la correction, exception faite pour le site de BP_EST pour qui des surestimations de la moyenne annuelle résultent de la correction pour les deux polluants.

Pour autant, les sites de A1, AUT et BP_EST présentent de forts biais entre leurs valeurs respectives de *Cprox* et la mesure. Nous avons déjà supposé que l'hétérogénéité des émissions seule ne permettait pas de discriminer les types de stations entre elles ce qui explique les sur/sous-estimations observées. Il serait intéressant de caractériser les types de véhicules sur les sites à proximité de grosses voies routières (type « AUT », « A1 » et « BP_EST ») et de considérer l'ouverture et la densité végétale autour des stations. Une piste d'amélioration est à rechercher dans l'estimation de *k*, et probablement du TMJA sur l'autoroute A1 dont la capacité est peut-être mal exploitée dans le modèle OLYMPUS.

Ces résultats nous confortent dans l'idée que, à l'heure actuelle, *Cprox* permet de corriger significativement les sorties de CHIMERE, mais qu'il serait intéressant de chercher à affiner davantage l'étape d'automatisation de l'approche. En effet, malgré sa simplicité et malgré le grand nombre d'incertitudes liées à la formation du cadastre d'émissions, *Cprox* permet d'obtenir des résultats plus robustes tant sur la moyenne annuelle et la variance des concentrations que sur la réduction du biais du modèle - sur les deux cas étudiés dans ce chapitre de thèse (2009 et 2017). Elle permet, de plus, d'obtenir des résultats équivalents aux sorties d'Airparif utilisées pour la surveillance opérationnelle de la qualité de l'air, et qui nécessitent de lancer une chaîne complète de modélisation :

- Sa première partie est identique à notre étape CHIMERE, partant d'une modélisation du trafic (Système Heaven qui considère les données de trafic routier, les caractéristiques du parc roulant automobile et des facteurs d'émissions pour en déduire les émissions) et procédant à la modélisation de la composition atmosphérique de fond (plateforme Esméralda qui considère des données de forçage météo et de landuse et qui modélise la qualité de l'air grâce au modèle CHIMERE),
- Elle est couplée à une modélisation de la qualité de l'air à plus haute résolution, via l'utilisation du modèle de dispersion atmosphérique ADMS-Urban qui modélise l'impact du trafic routier. Ce modèle considère la qualité de l'air de fond simulée dans la première étape de la chaîne de modélisation ainsi que les données météo pour ses simulations,
- Enfin, les sorties de simulations sont corrigées par une interpolation optimale qui est une approche d'assimilation de données qui va comparer les sorties brutes simulées avec la mesure pour en déduire une correction à appliquer au modèle afin de diminuer les écarts existants entre les deux.

Les résultats donnés par *Cprox* par rapport à ceux du modèle HOR'AIR confirment que l'approche statistique que j'ai développé donne des résultats aussi bons qu'une approche plus longue, plus compréhensive de l'impact du trafic et plus déterministe. *Cprox* est donc une approche robuste et la comparaison avec les sorties d'HOR'AIR a permis de valider une méthodologie simple pour corriger les sorties de modélisation des modèles de qualité de l'air.

5. Cartographie des résultats

La même méthodologie que pour les cartographies de l'année 2009 a été utilisée pour réaliser celles de 2017 présentées ci-dessous.

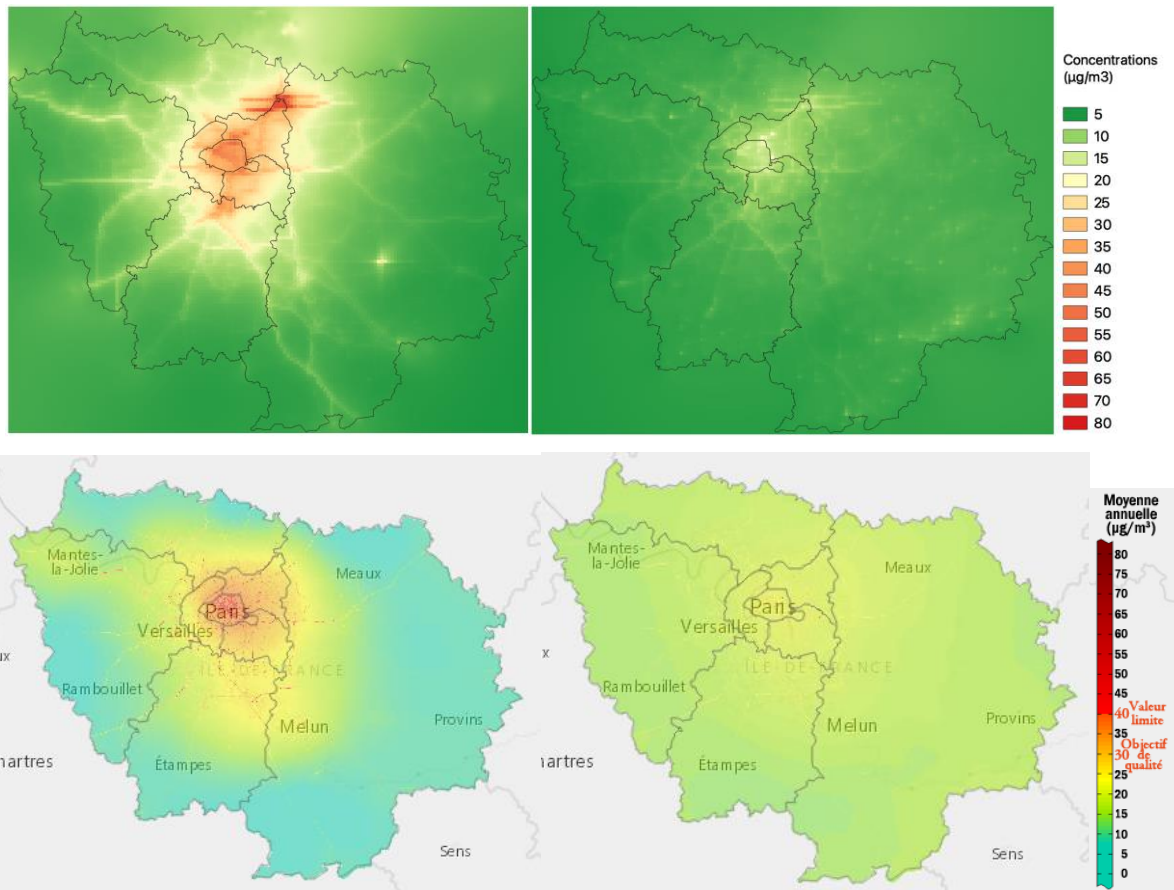


Figure 35 - Comparaison des cartographies annuelles des sorties de *Cprox* (en haut) et des bilans annuels d'Airparif (en bas) pour NO₂ (à gauche) et les PM₁₀ (à droite) pour l'année 2017.

Les niveaux de fond de NO₂ s'étendent moins autour de la petite couronne pour les sorties de *Cprox* que pour celles d'Airparif. En effet, les valeurs de concentration de la grande couronne sont en moyenne de $5 \mu\text{g}/\text{m}^3$ en périphérie du territoire et augmentent à $15-20 \mu\text{g}/\text{m}^3$ à forte proximité de la petite couronne. Les gradients de concentrations de cette zone sont plus prononcés pour le bilan d'Airparif. On peut constater très facilement sur les cartographies que l'aéroport Charles de Gaulle a, comme pour l'année 2009, un fort impact sur les concentrations de NO₂ en sortie de *Cprox* (environ $50 \mu\text{g}/\text{m}^3$) alors que rien n'est observé dans cette zone-là chez Airparif. On voit que les concentrations le long des routes principales du domaine sont assez bien reproduites notamment à proximité de Mantes-la-Jolie et Meaux.

Pour les particules, on peut constater qu'on sous-estime les concentrations de fond sur l'ensemble de la grande couronne $5 \mu\text{g}/\text{m}^3$ au lieu de $15-20 \mu\text{g}/\text{m}^3$ chez Airparif. Les concentrations sont aussi sous-estimées sur la partie en périphérie de la petite couronne. On peut remarquer que, comme pour NO₂, on remarque l'impact de l'aéroport sur les concentrations de particules alors que ça n'est pas observé sur la cartographie d'Airparif.

De manière générale, on obtient des tendances similaires avec les sorties d'Airparif.

V. Conclusion

Nous avons développé, dans cette partie du travail de thèse, une approche statistique ayant pour but de corriger les erreurs de modélisation effectuées par le modèle de qualité de l'air CHIMERE. Cette approche suit les lignes de construction des machines learning en réalisant une étape d'analyse et d'apprentissage de l'erreur du modèle par rapport à un jeu de données mesuré sur l'année 2009. Une première forme d'équation est proposée à la suite de cette étape d'apprentissage et elle nous a permis d'obtenir des résultats satisfaisants au vu des statistiques (écarts-types, RMSE, moyennes annuelles) qui en ressortent pour les sites de mesure de NO₂ et de PM₁₀. Il a, ensuite, été question d'automatiser cette équation pour qu'elle puisse être appliquée à n'importe quel site de la région Île-de-France sans pour autant nécessiter une comparaison avec la mesure. Nous avons étudié, pour cela, différents paramètres discriminant des sites de mesures du jeu d'apprentissage et il en est ressorti que les gradients d'émissions liés au trafic routier permettaient de différencier adéquatement les sites entre eux. Ces gradients d'intensité des émissions ont été exprimés en fonction des flux de trafic à proximité de la station de mesure par rapport à l'ensemble des flux de trafic de la maille CHIMERE considérée. La nouvelle forme d'équation a ensuite été appliquée à un jeu de données modélisé au niveau de station de mesure n'ayant pas été utilisée pour la phase d'apprentissage. Le site de trafic de validation présente de bons résultats statistiques pour les PM₁₀ mais est moins bon pour NO₂, l'hypothèse qui en découle concerne les biais sur les comptages routiers à proximité par le modèle (comptages modélisés sont deux fois plus élevés que les comptages observés). Pour les sites de fond, les variations de RMSE ne sont pas très importantes et les moyennes annuelles ont tendance à être améliorées pour correspondre à la mesure.

Ensuite, nous avons cherché à vérifier si cette approche pouvait être transposable à une nouvelle année simulée (2017). Les trois étapes (apprentissage, universalisation et validation) ont de nouveau été réalisées car nous avons renouvelé le cadastre d'émission pour ces nouvelles simulations. Une façon de valider nos résultats a été de les mettre en regard avec les sorties du modèle HOR'AIR d'Airparif et ça nous a permis de nous rendre compte que malgré les nombreux biais de l'approche (pas de différenciation des saisons dans les équations, pas de considération de la typologie de l'environnement local des sites de mesures), elle nous permet d'obtenir des résultats équivalents, voire parfois plus pertinents que l'utilisation d'une chaîne de modélisation entière.

J'ai rencontré certaines limitations lors du développement de cette approche. Il aurait été souhaitable d'avoir plus de données de mesure pour les étapes d'apprentissage pour une meilleure robustesse des résultats. Ensuite, actuellement, l'étape de discrimination des différents sites entre eux pour automatiser l'approche ne se base que sur des gradients de flux de trafic. Or, nous avons pu constater au cours de ce travail que ce paramètre seul ne permettait pas une bonne différenciation des sites entre eux. Il serait très intéressant de pouvoir considérer le degré d'aération de l'environnement proche des sites afin de mieux caractériser l'accumulation ou la dispersion de la pollution notamment pour mieux restituer la sensibilité de la qualité de l'air à l'intensité du vent. Enfin, une limitation importante est liée au fait qu'il a fallu beaucoup retravailler le cadastre d'émission utilisé pour les simulations CHIMERE de l'année 2017 pour pouvoir être utilisé par le modèle, ce qui a nécessité plusieurs hypothèses et induit une dégradation de la précision des données. Cela a, entre autres, pu être la raison pour laquelle une forte différence de niveaux moyens simulés a pu être observée dans la seconde partie du document.

Chapitre III

Estimation de l'exposition spatio-temporelle de la population : Prise en compte de la mobilité et des activités quotidiennes de la population

Table des matières

I.	Présentation du mode de fonctionnement d'OLYMPUS.....	103
II.	Représentation des trajets en transport en commun dans OLYMPUS	105
1.	Démarche suivie	105
2.	Données du réseau francilien de TC	106
3.	Création d'un module de construction des itinéraires	108
4.	Présentation et discussion des résultats obtenus.....	111
III.	Représentation de l'exposition dynamique de l'exposition des populations.....	115
1.	Construction d'une méthode de calcul de l'exposition.....	115
2.	Affectation d'une concentration à chaque localisation d'un individu.....	118
IV.	Résultats	124
1.	Caractéristiques de l'échantillon de population considéré pour cette étude	124
2.	Confrontation des déplacements modélisés avec les données de l'EGT 2010	124
3.	Visualisation des estimations d'exposition réalisées	125
V.	Autres développements.....	133
1.	Le réseau routier utilisé par OLYMPUS	133
2.	Construction et implémentation d'un nouveau réseau routier	134
VI.	Conclusions.....	138

L'intérêt de travailler sur l'exposition des individus plutôt que sur les concentrations de polluants, pour évaluer des situations environnementales, est double. Tout d'abord, la variable « exposition » permet d'amener la problématique de la qualité de l'air sur le terrain des enjeux de santé des populations, plutôt que de la limiter à la description physique d'un lieu. Mais surtout, en associant le diagnostic environnemental à un individu et non à un lieu, cette démarche permet d'intégrer la mobilité de chaque personne sur son territoire et donc la diversité des lieux fréquentés au cours d'une journée. Cette forme de diagnostic est parfois appelée « exposition dynamique » dans la littérature ([Dewulf et al., 2016](#); [Dias & Tchepel, 2018](#)). Ce qui est visé par cette démarche, à terme, c'est bien sûr de pouvoir aborder la dimension des inégalités environnementales, en questionnant l'exposition en regard des modes et lieux de vie, des activités et des choix de consommation d'énergie, et à la lumière des contraintes sociales, économiques et démographiques subies.

L'approche que j'ai développée et que je présente ci-dessous contribue à cette démarche. En termes d'outils, elle s'insère dans la plateforme de modélisation en cours de développement au LISA ([Figure 36](#)). Elle s'articule notamment autour du modèle d'émission et de construction de scénarios urbains **OLYMPUS**. OLYMPUS a été développé au LISA avec l'ambition de répondre à trois objectifs principaux: (1) permettre d'intégrer des données urbaines issues de différentes disciplines - économie, urbanisme, mobilité - dans la modélisation de la qualité de l'air, (2) construire des scénarios d'émissions plus ambitieux que les scénarios de réduction des émissions construits jusqu'à présent dans les modèles de qualité de l'air et (3) constituer une forme plus intégrée et plus opérationnelle d'aide à la décision environnementale, incluant une exposition basée sur l'activité individuelle ([Elessa Etuman & Coll, 2018](#)). Les scénarios qu'OLYMPUS ambitionne d'explorer incluent l'aménagement urbain, les politiques énergétiques et de transport, ou encore la modification des pratiques individuelles et collectives. Nous verrons ci-après quels jeux de données ou ensembles de modèles permettent de fournir ce contexte d'étude. Le rôle d'OLYMPUS est d'intégrer les spécificités de chaque situation urbaine étudiée, de simuler une activité anthropique en lien avec ces situations, et de générer les inventaires d'émissions de polluants qui en découlent. OLYMPUS est en effet un outil basé sur l'activité des individus, c'est-à-dire qu'il s'appuie sur la représentation des pratiques consommant de l'énergie, qu'elles soient domestiques, professionnelles ou de loisir, pour calculer les émissions correspondantes. Certaines de ces données, et notamment celles liées à la mobilité, seront ensuite réutilisées pour le calcul de l'exposition. Les émissions produites par OLYMPUS ont vocation à être intégrées dans un modèle de chimie-transport, en l'occurrence CHIMERE dans notre plateforme, pour produire un diagnostic sur la qualité de l'air dans la situation étudiée. Deux étapes sont ensuite nécessaires. Tout d'abord, les champs de concentration des polluants issus de CHIMERE possédant au mieux une résolution kilométrique, il faut les retravailler pour obtenir une qualité de l'air à fine échelle. Cela se fait à l'aide de l'algorithme de correction statistique que j'ai développé et présenté dans le chapitre précédent, et cela permet d'obtenir une cartographie distinguant les différents environnements urbains, toujours pour chaque situation simulée. Enfin, on procède à l'estimation de l'exposition dynamique des populations. Cette seconde étape consiste à associer les concentrations des polluants à fine échelle avec le positionnement des individus à tout instant, puis à cumuler ces valeurs sur une journée entière. Cette dernière étape m'a été confiée dans sa totalité. Au cours de ma thèse, ma contribution a consisté à récupérer, voire construire quand elles étaient manquantes, et mettre en forme, les données de mobilité quotidienne simulées par OLYMPUS. Puis à croiser espace fréquenté, temps de séjour dans cet espace et qualité de l'air locale, sur chaque pas de temps d'une journée et pour chacun des individus modélisés. Cela est fait à l'aide d'un code numérique que j'ai intégralement construit.

Les difficultés de ce travail étaient en grande partie liées à sa complexité technique. D'une part parce qu'il m'a demandé d'intervenir dans le code d'OLYMPUS pour l'adapter à mes besoins : j'ai notamment dû créer dans OLYMPUS une représentation des trajets en transports en commun (TC), incluant leur spatialisation et leur temporalité, puisqu'OLYMPUS ne calculait jusqu'alors que des matrices origine-destination pour ce mode de déplacement. Et d'autre part parce que calculer l'exposition dynamique demande de croiser des données de nature et de dimensions totalement différentes (temps de

parcours, brins routiers, concentrations horaires maillées, typologie des espaces, matrices origine-destination, itinéraires en TC...). Ce travail a enfin nécessité la conception d'un ensemble cohérent d'hypothèses pour proposer une estimation de l'exposition dynamique des populations qui soit pertinente, notamment au vu du manque de données sur certains champs (environnements intérieurs, trajets piétons...). Il est présenté ci-dessous en deux étapes : l'intervention sur le code d'OLYMPUS, et le développement d'un module d'exposition dynamique. Pour cette étape d'implémentation de l'exposition dynamique, ma contribution s'est limitée à deux polluants communément suivis en épidémiologie, fréquemment mesurés en ville et maîtrisés de manière satisfaisante par les modèles : le dioxyde d'azote et les particules fines (PM₁₀ essentiellement). Ce sont aussi les polluants qui ont été travaillés dans le cadre du raffinement des sorties de CHIMERE. Les résultats obtenus seront présentés et discutés, au vu de ce qu'ils apportent ou peuvent apporter pour la connaissance de l'exposition en zone urbaine, et au regard de leurs principales limites.

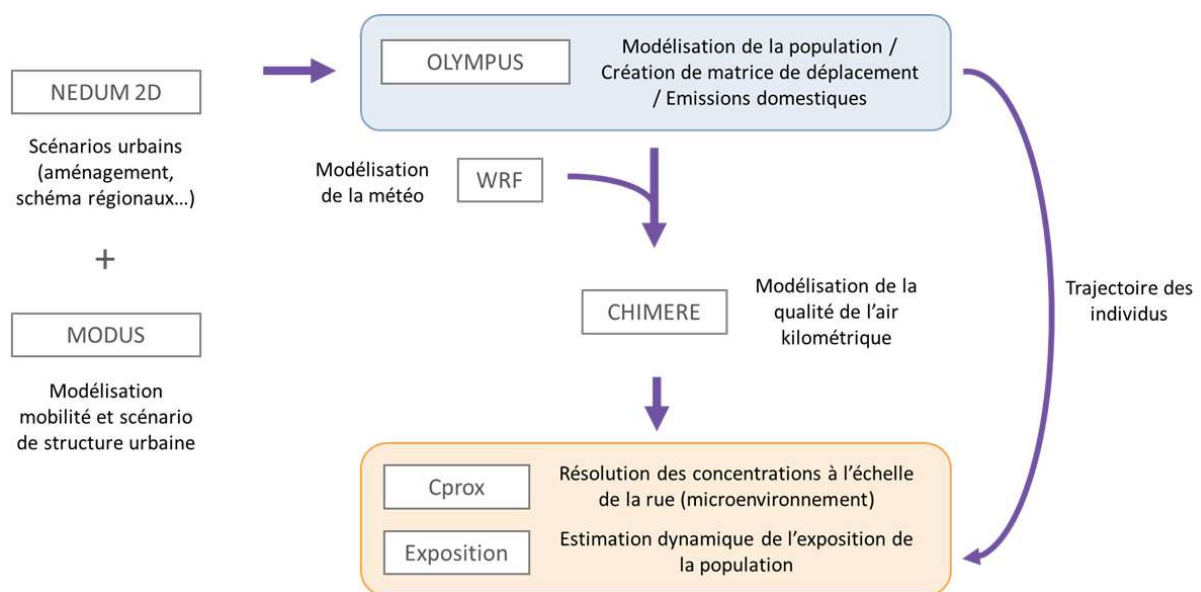


Figure 36 - Représentation schématique de la plateforme de modélisation du projet POLL-EXPO. Les noms des modèles sont indiqués dans les encadrés gris. En bleu figure le modèle auquel j'ai apporté des modifications et en orange figurent les parties de la plateforme que j'ai développées.

I. Présentation du mode de fonctionnement d'OLYMPUS

OLYMPUS utilise une approche de modélisation multi-agents prenant en compte l'usage des sols ainsi que la distribution spatiale des habitations et des services (voirie, transports, emplois, commerces...) pour créer une modélisation centrée sur les individus : leurs pratiques de mobilité, leur consommation d'énergie, les émissions de polluants qui en résultent.

Si la simulation se fait sur une situation existante, OLYMPUS peut s'appuyer sur des données d'enquête (recensement, Enquête Globale des Transports et différentes enquêtes ménages) pour reconstruire l'existant. En revanche, si l'on s'intéresse à une situation hypothétique, la configuration urbaine – ou le scénario urbain – qui sera exploité(e) par OLYMPUS peut être produit(e) par un modèle LUTI (Land-Use

Transport Interaction). Sur notre plateforme, c'est l'outil **NEDUM-2D (CIRED)⁵ / MODUS (LVMT - DRIEA)⁶** qui a été implémenté. Cet outil considère un contexte démographique, économique, environnemental et réglementaire (mise en œuvre de politiques publiques, d'aménagement du territoire, d'action sociale...) pour proposer une évolution de l'occupation des sols, de la densité du bâti et de la distribution des ménages à un horizon donné, sur le territoire considéré. Il s'appuie notamment sur un arbitrage entre le coût intégré de l'énergie, du foncier et de la mobilité. Au final, il propose à OLYMPUS une représentation de la structure, de l'organisation et de l'habitation du territoire urbaine issue d'un scénario d'expansion ou de densification, d'accroissement de l'offre de transports ou d'urbanisation ciblée, associé à une hypothèse d'évolution démographique. Dans le cas d'un scénario, OLYMPUS s'empare donc des données de sortie du LUTI pour construire la situation qui alimente CHIMERE.

Le travail de modélisation des émissions est découpé ainsi (Figure 37) :

- OLYMPUS construit tout d'abord une population « synthétique » à partir des sorties du LUTI - ou à partir des données d'enquête. La population synthétique inclut une représentation statistique des propriétés des ménages (taille, âge des individus, possession de véhicule...) et de leurs habitations (type d'habitat, surface, énergie pour le chauffage...), cohérente avec le scénario étudié. Cette étape détermine la spatialisation des habitations ainsi que celle des centres d'activité sur le domaine.
- L'accessibilité et l'attractivité des lieux de résidence/travail/activité sont des paramètres centraux pour simuler la mobilité des habitants de la zone urbaine. Ils sont attribués à chaque type de lieu sur la base d'indicateurs locaux de densité d'habitation ou d'emploi. A l'aide d'une approche gravitaire, OLYMPUS définit alors la géographie de la mobilité pour différents types d'activité.
- L'étape suivante consiste à calculer la demande en transport sur la base de la localisation des lieux de vie et des différentes activités des individus. Elle nécessite de simuler un agenda type pour chaque individu pour la réalisation de ses activités, et sur la base des pratiques constatées (ou imaginées dans le scénario). OLYMPUS associe à chaque individu un choix modal pour la réalisation de chaque activité. Celui-ci se base sur une approche probabiliste dite d'« utilité aléatoire », fonctionnant sur le principe de rationalité des décisions. La rationalité des décisions intègre notamment la possession ou non d'un véhicule, le besoin de mobilité pendulaire, la configuration des trajets planifiés, leur durée estimée, le coût des transports, la fréquentation des axes routiers et le déploiement des réseaux de transport en commun. La durée de chaque trajet, selon le choix modal retenu, est issue d'une itération avec le module d'affectation sur le réseau, décrit ci-après, et qui fournit notamment la congestion routière. OLYMPUS génère sur cette base les emplois du temps de la population, en fonction des activités effectuées au cours de la journée mais aussi en fonction de leur priorité pour l'individu.
- Comme je viens de le mentionner, OLYMPUS réalise une spatialisation de la demande en transport. C'est une modélisation des itinéraires qui consiste à les affecter sur leur réseau. Pour mener ces calculs, ce module se base sur la théorie des graphes, dans laquelle le réseau est représenté sous la forme de branches connectées entre elles par des nœuds, et à partir de laquelle l'algorithme de Dijkstra, identifiant le plus court chemin entre deux points (l'origine et la destination de chaque trajet), est mis en œuvre. Je reviendrai plus tard sur cet algorithme. Dans la version initiale du module, seuls les trajets réalisés en véhicule particulier (VP) étaient considérés. Les déplacements en transports collectifs n'étaient pas spatialisés et la structure des transports en commun n'était pas fournie au modèle. Le temps de trajet était issu de statistiques sur la relation entre distance parcourue et durée du trajet, en lien avec la densité urbaine.

⁵ <https://www.centre-cired.fr/axe-1-systemes-et-leurs-dynamiques-villes-energie-usage-des-sols/>

⁶ <https://www.drirea.ile-de-france.developpement->

[durable.gouv.fr/IMG/pdf/modelisation_des_deplacements_en_idf_avec_modus_3.1.pdf](https://www.drirea.ile-de-france.developpement-durable.gouv.fr/IMG/pdf/modelisation_des_deplacements_en_idf_avec_modus_3.1.pdf)

- Toujours en se basant sur le scénario territorial étudié, OLYMPUS réalise en parallèle une estimation de la consommation en énergie des bâtiments résidentiels, tertiaires et institutionnels. Il utilise une approche « bottom-up » basée sur les caractéristiques énergétiques de chaque bâtiment, son agenda d'occupation, son usage (chauffage et/ou activités domestiques incluant la production d'eau chaude), et s'appuyant sur des statistiques de consommation énergétique ainsi que sur les conditions climatiques. Les données régionales de consommation énergétique permettent également de contraindre la consommation totale sur le domaine.
- Enfin, le dernier module appelé lors des simulations procède au calcul des quantités de polluants émises par les déplacements motorisés individuels et du fait de l'utilisation d'énergie dans les bâtiments. Pour compenser la non-représentation explicite des trajets en transports en commun, OLYMPUS ajoute au flux général de véhicules particuliers précédemment calculé un flux spécifique représentant les transports en commun non ferrés (bus) et les activités commerciales (livraisons). Cet ajout est fait sur la base de statistiques d'occupation des voies par ces activités, selon les types de voies. Cette approche statistique est celle qui est utilisée dans la quasi-totalité des modèles de mobilité, mais elle n'est pas satisfaisante car le fret représente des flux de trafic importants et constitue un enjeu environnemental majeur. Un travail de représentation déterministe de la logistique urbaine est en cours d'implémentation dans OLYMPUS, mais il n'est pas intégré dans la version du modèle que j'ai utilisée.

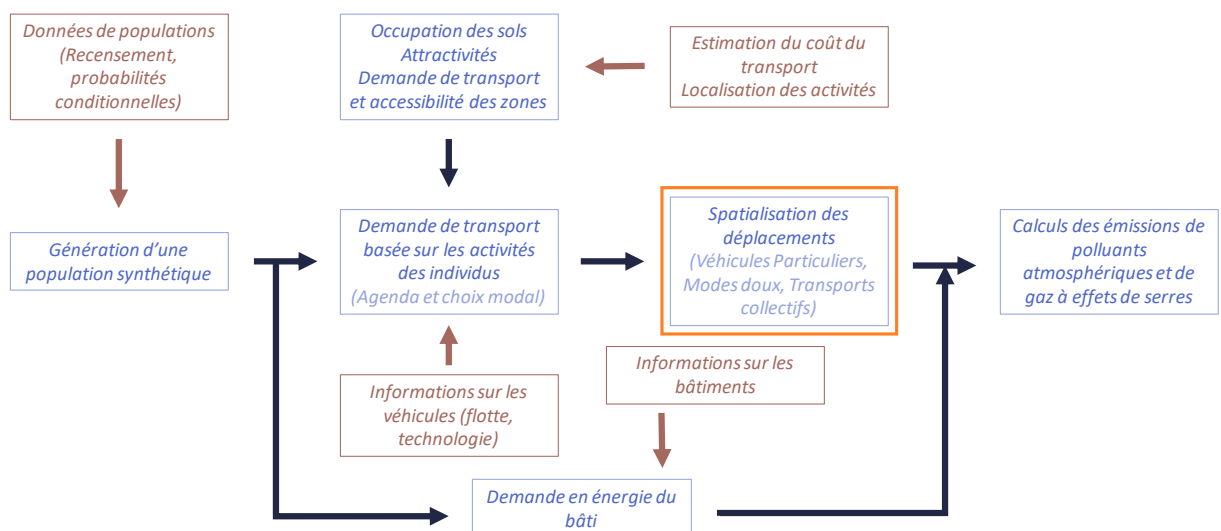


Figure 37 - Représentation schématique de l'organisation et des rôles des modules d'OLYMPUS. L'encadré orange indique la partie du code sur laquelle je suis intervenue. Les différents modules du modèle sont représentés en bleu et les fichiers d'entrées sont représentés en marron.

L'objectif principal de mes interventions dans le code d'OLYMPUS était de produire une représentation réaliste - et non statistique - des déplacements en TC, en considérant leur lien avec un réseau, qu'il soit routier ou souterrain. Le module d'intervention dans le code est entouré en orange dans la Figure 37.

II. Représentation des trajets en transport en commun dans OLYMPUS

1. Démarche suivie

Les calculs d'itinéraire en véhicule particulier sont réalisés dans OLYMPUS grâce au package Networkx et son module de calcul du plus court chemin « bidirectional_dijkstra » qui repose sur l'algorithme de Dijkstra (Dijkstra, 1959), développé dans les années 1950. Cet algorithme permet de trouver la distance

la plus courte entre deux nœuds d'un graphe pondéré, via la minimisation du poids des brins - ou arcs - reliant les nœuds. Il trouve une très grande utilité dans le domaine routier, pour lequel le poids des brins (voies parcourues) entre deux nœuds (croisements) peut être le temps, la distance ou encore des paramètres de coût ou de consommation de carburant. Depuis, de nombreux algorithmes ont été développés (Golden, 1976; Madkour et al., 2017) et le choix d'un algorithme plutôt qu'un autre dépend du type de graphe considéré. Dans OLYMPUS, la pondération des calculs du plus court chemin est réalisée en fonction des temps de déplacement entre l'origine et la destination du déplacement : l'itinéraire calculé entre les deux nœuds doit prendre le moins de temps possible. C'est cette pondération des déplacements qui a justifié le choix de cet algorithme pour les calculs du plus court chemin car il n'est adapté qu'aux réseaux ayant des poids positifs (D. B. Johnson, 1973). L'utilisation de la version bidirectionnelle de l'algorithme de Dijkstra permet un calcul des itinéraires plus rapide qu'avec la version unidirectionnelle.

J'ai souhaité étendre cette méthodologie au calcul des itinéraires pour les déplacements réalisés en TC. Il me fallait donc implémenter dans OLYMPUS un réseau de TC exploitable par l'algorithme de Dijkstra. Je présente ci-après la création et l'ajout, dans le modèle, d'un réseau de TC représentatif de l'offre francilienne actuelle, sous la forme de lignes de transport connectées entre elles, et l'implémentation d'un module de calcul des temps de trajet associés. J'ai ensuite connecté ce module à l'algorithme de recherche du trajet optimal en utilisant, dans cette première version, uniquement le temps total de déplacement comme critère d'optimisation. Cette configuration est évolutive : le nombre de correspondances entre lignes sur un trajet et l'inconfort que cela peut procurer, pourront être pris en compte dans le choix du trajet dans de futures versions d'OLYMPUS.

2. Données du réseau francilien de TC

Pour calculer les temps de trajet en transports en commun, la première version d'OLYMPUS utilisait des données statistiques tirées d'enquêtes – et notamment de l'Enquête Globale des Transports (EGT 2010) – pour attribuer des temps de déplacements moyens par kilomètre parcouru, en fonction de la densité urbaine des zones de départ et d'arrivée. Mon objectif étant de pouvoir désormais appuyer ces déplacements sur un réseau de transport, et de les optimiser avec l'algorithme de Dijkstra, j'ai recherché dans les bases de données existantes des représentations numériques du réseau de transport francilien, déjà transformées ou transformables en graphe.

À la date où j'ai travaillé sur ce point, la région Ile-de-France mettait à disposition des données récapitulatives de l'ensemble des informations relatives aux transports en communs. Ces données étaient disponibles au format GTFS (General Transit Feed Specification), format de fichier créé par Google dans le but d'unifier et de faciliter le traitement de données mondiales de transport en commun. C'est donc avec cette base de données que j'ai implémenté le calcul de mobilité en TC dans OLYMPUS. Aujourd'hui un deuxième format - Netex - a été mis à disposition. Comme GTFS, Netex permet de récapituler les informations relatives aux lignes de transport en commun mais dans un format qui se veut plus facilement traitable et plus intuitif. Mon travail étant achevé à la date de parution des fichiers Netex, je ne l'ai pas reconduit avec ce nouveau format. Toutefois, les développeurs du modèle pourront envisager de proposer une version Netex des scripts que j'ai développés si ce format finissait par s'imposer, et que les utilisateurs d'OLYMPUS ont besoin d'avoir des données de TC franciliennes à jour. Pour l'instant, qu'il s'agisse des fichiers GTFS ou des fichiers Netex, ils sont quotidiennement révisés par les plateformes d'Open Data des régions qu'ils couvrent et permettent de pouvoir travailler avec des données actualisées.

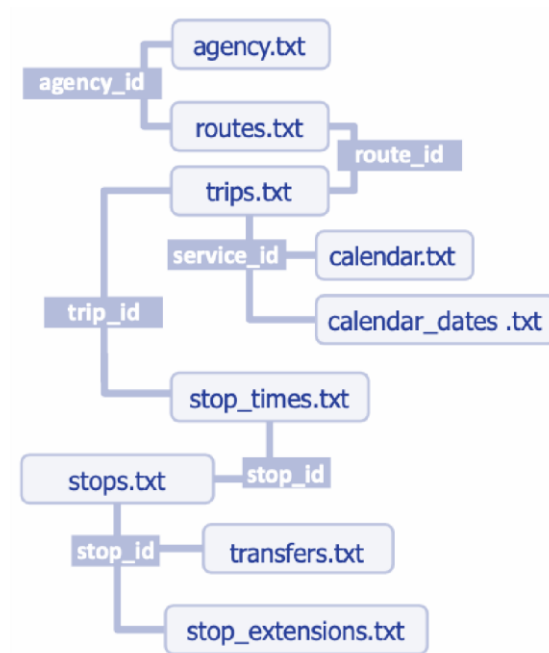


Figure 38 - Organisation des fichiers GTFS disponibles pour le réseau de transport en commun de l'Île-de-France.

Les fichiers GTFS que j'ai téléchargés et utilisés pour la construction du réseau de TC datent de Février 2020. Ils consistent en 9 fichiers textes contenant des tableaux pouvant être liés entre eux par des attributs communs (Figure 38). La création d'un réseau m'a demandé de travailler avec 6 de ces fichiers en particulier : agency.txt, routes.txt, trips.txt, stop_times.txt, stops.txt et transferts.txt. Leur rôle ainsi que le travail que j'ai mené avec ces fichiers sont décrits ci-dessous.

Dans la suite de cette section :

- Un « itinéraire » correspondra à l'ensemble du déplacement d'un individu sur une ou plusieurs lignes de transport en commun.
- Un « trajet » désignera une portion d'itinéraire constituée d'au moins 2 arrêts d'une même ligne.

Dans l'approche que j'ai développée, j'ai cherché à proposer des itinéraires pour chaque individu se déplaçant d'un point origine à un point destination, en utilisant une combinaison de trajets.

[Fichier routes.txt](#)

Le fichier *routes.txt* est un document permettant d'attribuer des identifiants à toutes les lignes régionales de transport. Il va surtout, pour nous, contribuer à l'identification du mode associé, grâce à son attribut *route_type*.

Sur l'Île-de-France, le paramètre *route_type* peut prendre 4 valeurs :

- 0 - Tramway ou métro léger
- 1 - Métro (il s'agit, plus précisément, de tout système ferroviaire souterrain)
- 2 - Train (permettant des trajets interurbains/longues distances)
- 3 - Bus (aucune différence n'est faite ici entre les bus courtes et longues distances)
- 7 - Funiculaire

[Fichier trips.txt](#)

Ce document permet de spécifier et d'identifier l'ensemble des trajets (« trips ») de chaque ligne de transport en commun étant desservie dans la journée. Il permet aussi de spécifier les directions de chaque trajet entre l'aller et le retour ce qui permettra, dans la suite du développement, de créer un réseau dirigé. Chaque brin de réseau créé sera donc à sens unique.

[Fichier stops.txt](#)

Ce document référence toutes les informations relatives aux localisations des arrêts de transport en commun : coordonnées latitude/longitude en WGS84, noms de stations et identifiants mais également un attribut nommé « *parent_station* ». Ce dernier donne l'information sur l'identifiant de la station précédente sur la ligne et permet ainsi d'identifier les deux terminus puisqu'ils n'ont pas de valeur pour l'attribut « *parent_station* ».

Cette donnée s'avère utile dans la suite de la méthodologie pour déterminer l'ordre de desserte d'un service (qui n'est pas explicitement indiqué dans les fichiers).

[Fichier stop_times.txt](#)

Ce fichier est le fichier le plus volumineux des fichiers GTFS et va condenser tous les horaires d'arrivées et de départs sur chacun des arrêts des différents services au cours d'une journée donnée. Il permet aussi de définir l'ordre des arrêts desservis lors d'un trajet particulier.

[Fichiers transferts.txt](#)

Le fichier *transferts.txt* renseigne les correspondances comme des zones multimodales où les agents peuvent se déplacer à pied entre deux arrêts de deux lignes différentes (Figure 39). Des temps de correspondance sont associés et exprimés en secondes.

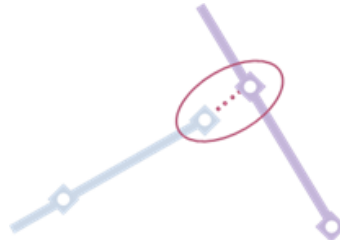


Figure 39 - Représentation schématique d'une correspondance entre deux arrêts de deux lignes de transport (entourés en rouge). La zone empruntée à pied par les agents est représentée par des pointillés.

3. Création d'un module de construction des itinéraires

J'ai développé un ensemble cohérent de scripts de calculs permettant de traiter les informations des fichiers GTFS pour les besoins d'OLYMPUS.

J'ai tout d'abord géolocalisé les arrêts à l'aide du fichier *stops.txt* et je leur ai ajouté l'ensemble des informations sur les trajets qui les empruntent, ce qui m'a permis de définir des nœuds du réseau.

J'ai associé les horaires d'arrivée et de départ dans l'ordre des stations, pour tous les trajets et selon le sens emprunté (aller ou retour) grâce aux fichiers *stop_times.txt* et *trips.txt*. L'ordre de desserte des stations a été défini grâce à l'information donnée par l'attribut « *parent_station* » du fichier *stops.txt*.

Pour cela, j'ai construit les brins routiers et ferrés reliant les arrêts d'une même ligne de transport. J'ai dû créer une base de données représentant chacun de ces arcs entre une station et sa station parent, à laquelle j'ai ajouté les informations relatives à l'identifiant de la ligne concernée, aux trajets possibles et aux temps de déplacement entre ces deux stations. Pour ce dernier paramètre, j'ai utilisé les horaires d'arrivée et de départ des bus ou rames reliant les stations, que j'ai simplement différenciés entre chaque station S1 et la station S0 la précédant, pour en déduire un temps de trajet. Cette étape a été menée sur l'ensemble du réseau de transport en commun francilien, et les valeurs de temps ont été associées à chaque arc.

J'ai ensuite créé les arcs représentant les déplacements des usagers lors des correspondances grâce au fichier *transferts.txt*, fichier contenant déjà les temps de déplacement entre deux stations de deux lignes différentes qu'il m'a simplement fallu ajouter à la base de données précédente (Figure 40). L'ajout des temps de trajet au niveau des brins permet donc de pondérer les calculs du plus court chemin directement dans le script d'estimation des plus courts chemins développé ici. A ce point, il est important de mentionner que ces arcs ne sont pas des routes complexes mais des lignes droites entre deux nœuds du réseau. Cela n'a aucune incidence pour le réseau ferré, mais cela veut dire qu'on considère un parcours routier à vol d'oiseau entre deux stations de bus, trajet qui n'est donc pas superposé au réseau routier d'OLYMPUS. Cela pose le problème de la localisation des individus entre deux arrêts, et donc du calcul de leur exposition. Ce problème n'est pas solutionnable car il demanderait au modèle d'affecter avec l'algorithme de Dijkstra un trajet routier entre chacun des arrêts, ce qui, même hors considérations de temps de calcul déjà réhibitores, n'est pas forcément une meilleure hypothèse. Car un bus n'emprunte pas le plus court chemin mais celui adapté à ses contraintes de circulation (couloir de bus, routes assez larges, pas de voies rapides...). Et les trajets réels ne sont pas renseignés dans les bases de données. Cela aura des conséquences sur le calcul de l'exposition le long des trajets en TC routiers, j'y reviendrai dans le paragraphe concerné.

Les correspondances constituent des éléments qui peuvent emporter la décision, pour le choix modal comme on l'a dit précédemment, mais aussi pour le choix des trajets composant l'itinéraire. Si on veut intégrer une contrainte sur le nombre de correspondances lors des calculs du plus court chemin, cela doit obligatoirement se faire en associant à chaque brin un paramètre nouveau représentant le cumul du nombre de brins dans le trajet. En effet, le type de calcul qui est mené ne peut prendre en compte comme facteur poids que des paramètres existant au niveau de chacun des brins. Or le cumul du nombre de brins n'est pas un paramètre intrinsèque au brin, il est lié au début et à la suite de l'itinéraire emprunté par l'individu. Pouvoir relier le nombre de correspondances à chaque brin est donc un premier verrou qu'il faudrait faire sauter. Une autre limitation à l'intégration de la contrainte « correspondances » vient du fait que le réseau ne me permet pas, à l'heure actuelle, de caractériser le temps d'attente des individus au niveau des quais ou des stations. C'est pourquoi ce paramètre n'a pu être intégré pour l'instant dans le choix de l'itinéraire. C'est pourtant un point limitant le choix d'un déplacement en TC car les niveaux de concentration sur les quais souterrains peuvent être très importants (Annexe 9). Il est donc important que ce paramètre soit intégré dans OLYMPUS dans le futur. Pour préparer cette évolution, j'ai donc tout de même gardé la mémoire de chaque correspondance afin que, dans une version plus avancée du modèle, toutes les conséquences puissent être intégrées.

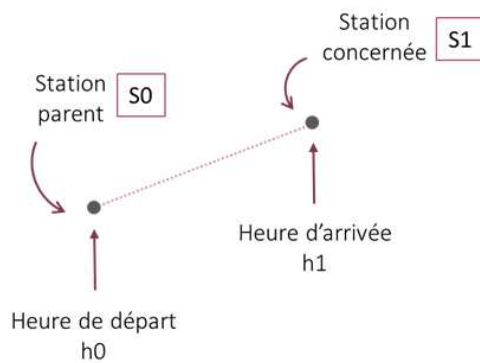


Figure 40 - Schéma des paramètres considérés pour les calculs des temps de déplacement entre deux stations d'une même ligne de transport.

Afin de pouvoir représenter la variation de l'offre de transport au cours de la journée - comme l'augmentation de la fréquence de desserte des transports en commun aux heures de pointe, leur diminution en heure creuse et la fermeture nocturne de certaines lignes - j'ai considéré 3 versions de mon réseau. La première a été construite en ne considérant que les lignes de transport en commun actives entre 6h et 7h du matin, et est utilisée pour représenter les déplacements aux heures de pointe. La deuxième utilise les lignes en service entre 15h et 16h pour représenter l'offre de déplacement en heures creuses. Et la troisième ne prend en considération que les lignes fonctionnant entre 2h et 3h du matin pour représenter les déplacements réalisables de nuit. De cette façon, j'ai pu limiter le nombre de réseaux-types à considérer pour les simulations, tout en gardant des situations représentatives des différentes plages horaires pendant lesquelles se déplacent les individus.

Pour pouvoir complètement configurer un itinéraire, un dernier élément doit être précisé : la première station de l'itinéraire. En effet, de la même façon que pour le réseau routier, rien ne peut être établi si on ne fournit pas au modèle l'information sur le nœud du réseau qui doit être considéré comme point de départ du trajet. Or, les points de départ des individus générés par OLYMPUS ne se trouvent pas nécessairement à la même localisation que les nœuds du réseau. J'ai donc ajouté un calcul permettant d'identifier le nœud le plus proche du point d'origine du déplacement de chaque individu. Comme pour les correspondances, la distance à la plus proche station du réseau peut être un facteur de choix important - que ce soit pour le mode de transport ou pour l'itinéraire. Le temps nécessaire pour rallier la plus proche station n'est aujourd'hui pas compté car cela impliquerait de coder la possibilité de rejoindre la station à pied, en voiture ou en vélo, et la multimodalité n'est pas encore considérée dans OLYMPUS. Cela fait partie des développements envisagés. J'ai tout de même codé un calcul qui permet à chaque individu de trouver le nœud du réseau le plus proche de son point de départ, même si le temps nécessaire pour parcourir cette distance et le mode utilisé par l'individu pour se rendre à l'arrêt considéré ne sont pas pris en compte. Cela permet de préparer l'implémentation de la multimodalité.

L'offre de transport ainsi créée est illustrée par son réseau en [Figure 41](#). Je l'ai intégrée au modèle OLYMPUS et elle permet désormais de caractériser les déplacements des individus grâce à un calcul du plus court chemin. Je me suis alors attachée à évaluer les résultats obtenus pour différents types de trajets. Ce travail est présenté et discuté dans la section qui suit.

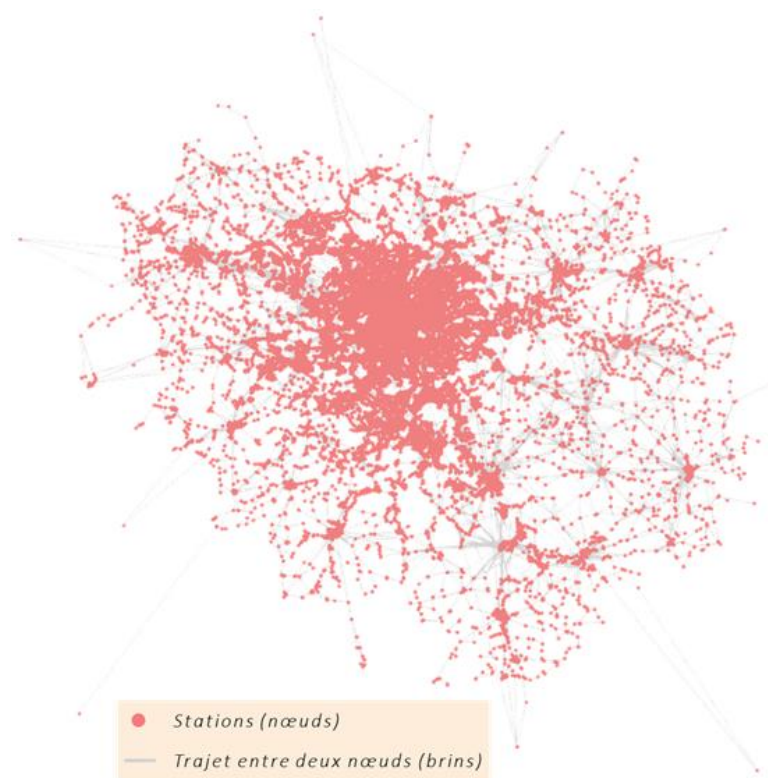


Figure 41 - Représentation du réseau de transport en commun ajouté au modèle OLYMPUS. Les stations de transports en commun sont illustrées sous forme de point rose et les brins reliant les stations d'une même ligne sont illustrés sous forme de lignes grises.

4. Présentation et discussion des résultats obtenus

A l'issue de sa création, le réseau de transport en commun d'OLYMPUS comporte 42 103 arrêts et 152 755 brins. Afin d'évaluer l'algorithme de choix d'itinéraire, j'ai réalisé 4 tests pour lesquels les origines et destinations ont été choisies aléatoirement sur les 42 103 stations du réseau. Les caractéristiques des itinéraires sont résumées dans le [Tableau 10](#), comparativement aux propositions de Google Maps considérées comme une référence dans le calcul du plus court chemin. Les résultats sont également présentés spatialement et comparés avec les propositions de parcours fournies par Google Maps dans les figures qui suivent.

Il convient de préciser que l'itinéraire retenu par Google Maps est sélectionné parmi un ensemble d'itinéraires qui ne sont pas tous proposés à l'utilisateur, et qui peuvent être proches en termes de temps de parcours. Leur durée totale dépend notamment de la manière dont le temps de correspondance est estimé. C'est pourquoi la comparaison précise des trajets et lignes empruntés ne constitue pas un critère pertinent d'évaluation de notre algorithme. Je chercherai donc plutôt des similitudes dans la spatialisation des trajets, qui indiqueraient que les deux algorithmes s'appuient sur les mêmes nœuds principaux du réseau pour permettre un transport rapide entre les points origine et destination. Mais je chercherai surtout la capacité de notre approche à proposer des itinéraires dont les durées sont cohérentes avec celles de la référence.

Tableau 10 - Confrontation des durées d'itinéraires calculées via mon réseau GTFS avec les informations données par Google Maps et le nombre de correspondances indiqué entre parenthèses pour chacun des deux.

	Source	Destination	Temps réseau GTFS (Nombre de correspondances)	Temps Google Maps (Nombre de correspondances)
Test 1	Braunstein (Yvelines)	Créteil Préfecture (Val-de-Marne)	1h41 (4)	1h56 (2)
Test 2	La Muette (Paris)	Bastille (Paris)	22 min (1)	31 min (0)
Test 3	Gillevoisin (Essone)	Rond Point (Seine-et-Marne)	2h04 (5)	2h47 (3)
Test 4	Gallieni-Métro (Seine-Saint-Denis)	Zone d'activité de l'Ormois (Val-de-Marne)	1h05 (3)	1h24 (2)

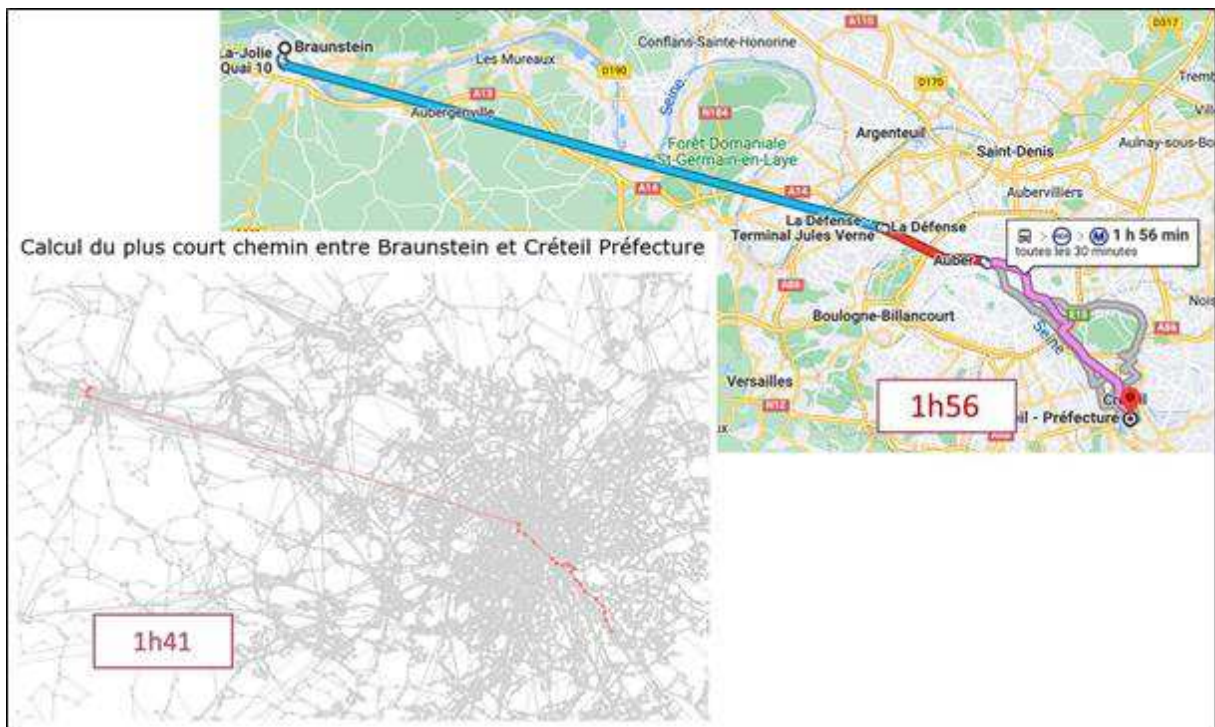
Sur les 4 tests menés, on peut constater que notre modèle propose la même hiérarchie des temps de transport que Google Maps, avec une même proportionnalité des durées de chaque itinéraire. Il montre toutefois une sous-estimation assez constante des temps totaux, de 34% par rapport au temps de référence. L'analyse plus fine des itinéraires (non détaillée ici) indique que cela est lié à la non prise en compte des délais d'accès aux stations ou des temps d'attente aux correspondances dans OLYMPUS. Ainsi, pour le test 2, Google Maps propose un trajet d'une vingtaine de minutes et sans correspondance, mais qui nécessite plus de 10 minutes de marche pour accéder à la station de départ. OLYMPUS a pour sa part sélectionné un itinéraire dont le temps total passé dans les transports est de 22 minutes. J'ai pu constater que cet itinéraire avait été proposé en 2^e choix par Google Maps : le temps de parcours hors correspondances y est estimé à exactement 22 minutes. De même, pour le test 3, Google Maps prévoit près de 40 minutes d'attente sur les trois correspondances de l'itinéraire, ce qui correspond à l'écart avec la prédiction d'OLYMPUS.

OLYMPUS a donc bien estimé le meilleur temps de parcours, mais pour un gain de quelques minutes il a fait le choix d'un autre itinéraire avec beaucoup de correspondances, ce qui le rend peu attractif dans la réalité. Cet ensemble de tests confirme donc la pertinence de notre algorithme pour l'estimation des temps de déplacement en TC sur la région francilienne. Il met également en lumière le besoin de prendre en compte les temps d'attente aux correspondances, afin de pouvoir cibler des itinéraires plus réalistes et de raffiner l'estimation du temps total nécessaire pour la réalisation du déplacement.

La spatialisation des déplacements présentée en [Figure 42](#) confirme que l'itinéraire d'OLYMPUS est toujours très similaire à celui de la référence. En particulier, l'algorithme que j'ai mis en place identifie toujours une même façon d'emprunter le réseau que Google Maps, en passant par exemple certaines fois par le centre (test 1) et d'autres fois par le Nord (test 4) de Paris lorsqu'il faut le traverser. Au final, ce travail me permet d'avoir accès à une représentation des déplacements des individus ne prenant pas leur véhicule particulier :

- Qui est représentative des temps réels de déplacement en TC en Île-de-France et des zones traversées, selon les points d'origine et de destination et le moment de la journée,
- Qui propose une décomposition de l'itinéraire en trajets sur des lignes de TC, ce qui permettra de calculer une exposition le long de ces trajets.

a. Test 1



b. Test 2

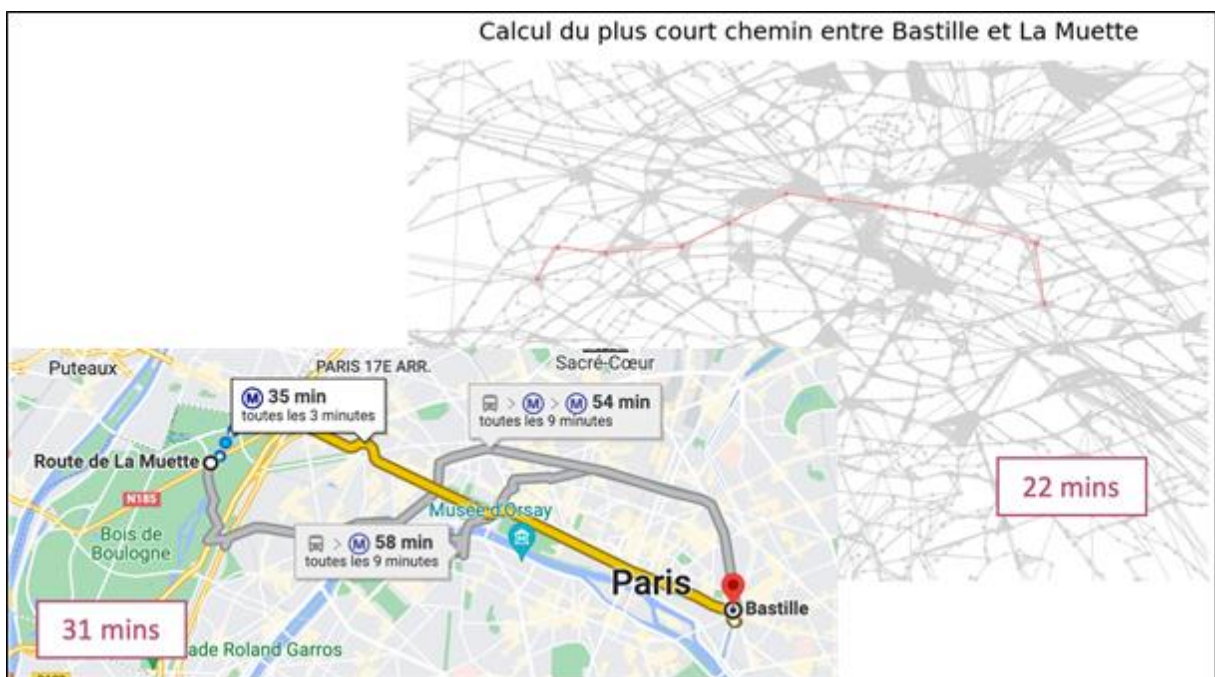


Figure 42 (a, b) - Confrontation visuelle des itinéraires trouvés pour le réseau GTFS (en gris) et Google Maps (en couleur).

c. Test 3



d. Test 4

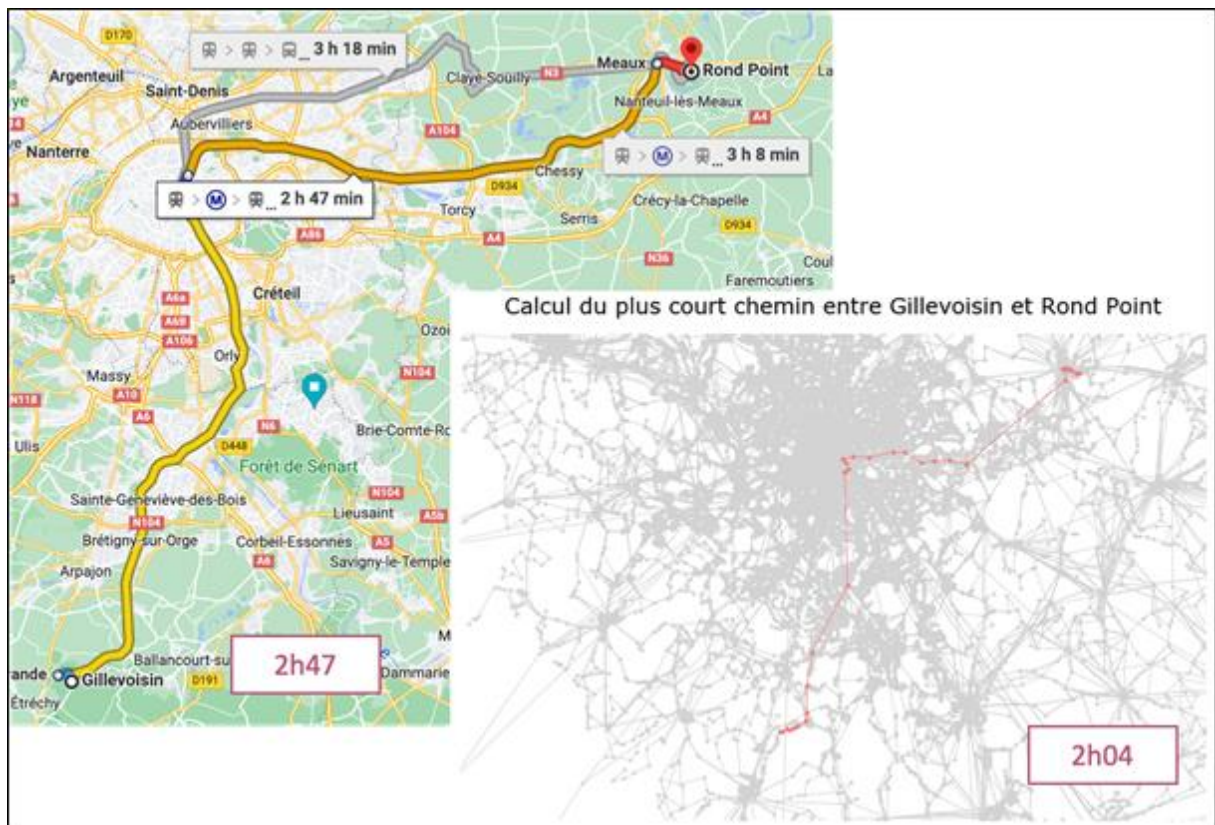


Figure 42 (c, d) - Confrontation visuelle des itinéraires trouvés pour le réseau GTFS (Réseau gris) et Google Maps (en couleur).

Cette approche permettra de construire des scénarios d'offre de transport beaucoup plus précis et pertinents, simulant par exemple le renforcement des lignes inter-banlieue (comme considéré dans le Grand Paris Express). Elle permet également d'envisager une réflexion sur le développement de la multi modalité dans les trajets des individus, qui est considérée comme un des leviers de la sobriété énergétique des déplacements et qui devient incontournable avec l'impossibilité d'accès à la ville dense pour de nombreux véhicules.

III. Représentation de l'exposition dynamique de l'exposition des populations

L'étape suivante a consisté à coupler les itinéraires simulés avec les niveaux de concentrations en NO₂ et PM₁₀ rencontrés durant le déplacement. Mon objectif était de pouvoir identifier à chaque instant les environnements-type traversés (proximité du trafic, fond urbain ou périurbain) et de leur associer des concentrations, dans l'idéal les concentrations à fine échelle produites dans le chapitre précédent.

J'ai organisé cette section de façon à introduire dans une première sous-section les objectifs et les méthodologies choisies, puis de les détailler dans les sous-sections qui suivent.

1. Construction d'une méthode de calcul de l'exposition

Méthodologie générale

J'ai commencé par construire un programme, codé en python, permettant d'exploiter les matrices de mobilité simulées par OLYMPUS ainsi que les concentrations en polluants issues de l'approche de correction statistique. Dans un premier temps, ce programme construit un tableau général qui récapitule, pour chaque agent et pour chacun de leurs itinéraires quotidiens, les informations sur leur choix modal, sur les activités effectuées, sur les communes d'origines et de destinations de leurs itinéraires, sur leurs IRIS et sur leurs coordonnées respectives (en WGS84) ainsi que sur les horaires des déplacements/activités. Pour des raisons de volume de données, la spatialisation des déplacements des individus n'est pas enregistrée dans les fichiers de sorties d'OLYMPUS. Par ailleurs, le temps de calcul nécessaire pour mener des calculs sur ces données devient rapidement important avec le nombre d'individus considéré. De ce fait, j'ai décidé de mener des tests sur un échantillon réduit de la population francilienne, plutôt que de considérer les presque 14 millions d'individus du territoire francilien. Dans cette étape test, sur un nombre restreint d'individus, j'ai donc récupéré les origines et destinations des itinéraires de chaque agent sélectionné ainsi que leurs choix modaux issus d'OLYMPUS, pour recalculer les plus courts chemins avec le même algorithme que décrit précédemment. Il faudra à terme envisager l'ajout d'une option à cocher lors du lancement d'OLYMPUS et permettant d'écrire dans un fichier dédié les données d'intérêt, plutôt que de devoir les recalculer.

Enfin j'ai calculé le temps passé sur chaque itinéraire, et pour la réalisation de chaque activité, ainsi que les niveaux de concentration correspondants. L'objectif était de multiplier les concentrations considérées par le temps de réalisation de l'activité, pour produire des $\mu\text{g}/\text{m}^3 \cdot \text{h}$ qui sont cumulés afin d'obtenir des expositions sur la journée (Figure 43).

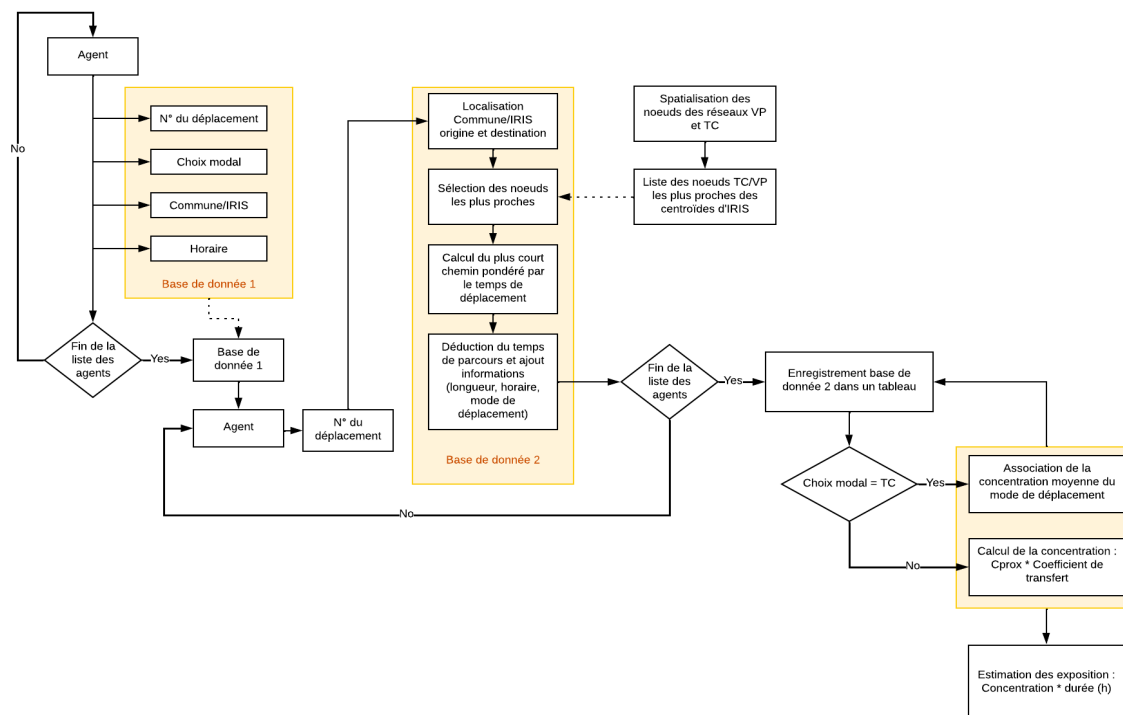


Figure 43 - Organigramme du script d'estimation des expositions.

Ce volet du travail recèle un verrou majeur qu'est l'estimation de l'exposition dans les transports en commun, notamment souterrains. Le CTM, tout comme l'approche de descente en échelle décrite dans le chapitre précédent, permettent en effet d'estimer des valeurs d'expositions extérieures, directement applicables pour les brins du réseau routier empruntés par les individus se déplaçant en modes doux. Mais pour tous les autres usages du réseau routier (véhicule particulier, transport en commun extérieurs), ces données doivent à minima être retravaillées. Et pour représenter les transports souterrains, ces données ne sont pas pertinentes et il faut trouver d'autres sources de données.

Pour ce qui est de l'exposition des conducteurs, la littérature propose des coefficients de transfert entre l'extérieur et l'intérieur des véhicules, qui peuvent être appliqués pour estimer les teneurs retrouvées dans les habitacles des véhicules particuliers ou dans les TC routiers comme les bus. Cela soulève en fait des questionnements supplémentaires, relatifs à la réelle qualité de l'air au milieu des voies, et que l'on peut soupçonner d'être - au moins dans certaines situations de congestion - moins bonne que sur le trottoir. En outre, pour les bus, comme je l'ai mentionné précédemment, les chemins empruntés entre deux arrêts ne sont pas fournis par les bases de données, et nous ne pouvons que proposer notre propre itinéraire, ce qui n'est pas forcément représentatif de la réalité. Nous ne pouvons donc pas relier le trajet à la qualité de l'air calculée par notre approche le long de chaque voie. Comme je le détaille ci-après, j'ai donc dû considérer d'autres options : soit prescrire des valeurs moyennes d'exposition, soit proposer à partir de ce qui est connu du trajet, un calcul des valeurs d'exposition. Enfin, pour les déplacements en TC souterrains, je n'avais aucun moyen d'estimer la qualité de l'air propre à chaque station ou ligne de transport, d'abord parce que c'est un questionnement émergent et qu'il n'existe pas encore assez de données pour proposer une valeur moyenne par station, ou par type de station, qui serait modulable (ou déjà modulée) en fonction de la fréquence de passage des rames ou en fonction du niveau de ventilation de la station. J'ai donc dû m'appuyer sur des valeurs moyennes de la littérature. En conséquence, j'ai produit et stocké dans OLYMPUS les informations géographiques relatives à ces

déplacements mais elles ne sont pas encore toutes utilisées pour le calcul de l'exposition en TC et dans les habitacles, qui s'appuie pour l'instant surtout sur des valeurs moyennes. L'évolution vers un couplage exposition- *Cprox* sur ses trajets reste possible, dans l'architecture que j'ai créée.

La mise en correspondance des déplacements et des niveaux d'exposition nécessite la manipulation de fichiers de très grande dimension, et est associée à des temps de calcul importants. Pour la construction, l'étude de faisabilité et l'évaluation de la méthodologie, j'ai donc travaillé sur un échantillon représentatif de la population.

Choix de la population étudiée

Parce que le travail rapide sur un échantillon de population peut être utile pour explorer à moindre coût numérique la variabilité de l'exposition dans un grand nombre de scénarios, je souhaitais que mes résultats de tests puissent être exploitables. J'ai donc veillé à ce que l'échantillon sélectionné soit extrapolable à l'ensemble de la population modélisée par OLYMPUS. J'ai pour cela décidé de retenir aléatoirement 1% de la population de chaque commune d'Ile-de-France dans mon échantillon de population. Python ne permet toutefois pas de faire de sélection aléatoire d'une population mais plutôt une sélection pseudo-aléatoire. En effet, le module python « `random()` » se base sur un algorithme de choix pseudo-aléatoire d'éléments dans une liste⁷, ce qui signifie que ce choix peut être reproduit puisqu'il est dirigé par l'algorithme. J'ai toutefois considéré cela comme satisfaisant pour la réalisation de mes tests.

Pour ce qui est de la représentativité de l'échantillonnage dans chaque commune, le chiffre de 1% permet de sélectionner un nombre suffisant d'individus pour que la variabilité des lieux de vie et des boucles d'activité quotidiennes puisse être représentée, même si la sélection aléatoire ne le garantit pas nécessairement. En respectant le poids relatif des habitants dans chaque zone de mon domaine, je peux extrapoler les résultats à l'ensemble de la population. Bien sûr, la représentativité de l'échantillon augmentant statistiquement avec sa taille, l'idéal serait de pouvoir considérer un échantillon de population plus large mais cela alourdit considérablement les temps de calculs.

Limitation des calculs d'itinéraires

Dans la version d'OLYMPUS que j'ai utilisée, seuls les centroïdes de commune sont considérés comme points d'origine et de destination des déplacements de la population simulée. Une telle simplification a plusieurs conséquences qui peuvent être préjudiciables pour les calculs de mobilité, et donc d'exposition. Ce point fait actuellement l'objet d'une évolution du code. Les travaux de Taos Benoussaid au LISA, qui s'intéressent à la modélisation des inégalités d'exposition avec notre plateforme, incluent en effet l'implémentation dans OLYMPUS d'une distribution des individus dans les quartiers ou logements d'un même IRIS. Ce travail n'étant pas terminé, je n'ai pas pu bénéficier de cette version mais j'ai mené quelques aménagements dans le code.

En effet, la combinaison d'un calcul par centroïde de commune et d'une représentation explicite du réseau de TC peut entraîner des biais dans le choix modal en ne prenant pas en compte la distribution de la proximité aux stations de transports en commun, parmi les individus. Cela n'était pas pénalisant dans la version initiale puisque les stations de transports en commun n'existaient pas et que l'utilité des TC était évaluée sur la base d'un temps de trajet statistique par type de zone. Avec l'amélioration que j'ai apportée, intégrer une distribution spatiale des points de départ dans chaque commune est devenu pertinent pour distribuer les individus sur les lignes de transport. Ne pas l'intégrer aurait même limité

⁷<https://realpython.com/python-random/#:~:text=Most%20random%20data%20generated%20with,random%20but%20still%20reproducible%20data.>

la plus-value de mon travail. Lors du développement de mon script d'estimation des expositions, j'ai donc distribué aléatoirement les points d'origine et de destination au niveau des centroïdes d'IRIS et non plus de ceux des communes en question. Cela m'a permis d'accéder à des trajets plus précis et plus complets.

Une autre conséquence du codage de la mobilité par centroïde de commune, plus problématique, est qu'OLYMPUS était dans l'incapacité de considérer des déplacements intra-communes puisque les points d'origine et de destination étaient identiques. La distribution des origines et destinations sur les centroïdes d'IRIS de la commune considérée permet désormais de maintenir beaucoup plus d'individus en mobilité dans la population étudiée, en forçant notamment la non-sélection d'un même IRIS de départ et d'arrivée au sein d'une commune. Cette méthode, efficace sur les régions densément peuplées du centre francilien, n'est en revanche que peu voire pas efficace pour les déplacements intra-communes situés les plus en périphérie de la grande couronne car, à cet endroit, plusieurs communes ne comportent qu'un seul IRIS. Je n'ai donc pas pu associer de trajet, dans mon échantillon de population, à la part d'individus se déplaçant au sein d'un même IRIS. Sur les tests que j'ai réalisés, la part de ces individus est élevée et peut excéder le tiers de l'échantillon. Puisque ce phénomène est essentiellement retrouvé sur la couronne externe de l'Île-de-France et qu'il est significatif, cela peut induire un biais spatial de représentativité des expositions calculées.

2. Affectation d'une concentration à chaque localisation d'un individu

Une fois la procédure de constitution de l'échantillon de population implémentée, je me suis attachée à la mise en relation des trajets et de la qualité de l'air locale. Cela imposait de prendre en compte la manière dont se déplaçait l'individu dans l'environnement concerné.

Pour les trajets urbains

Le travail de descente en échelle réalisé précédemment me permet de récupérer une information de qualité de l'air plus représentative que celle issue de CHIMERE, sur les tronçons de route du réseau considéré et pour chaque heure étudiée. Toutefois, au sein d'une rue, les individus se déplaçant en véhicule particulier, en bus, en deux roues ou à pied ne sont pas soumis de la même manière aux concentrations ambiantes puisqu'ils peuvent se trouver dans un habitacle plus ou moins ventilé. Il me fallait donc différencier les situations d'exposition selon le mode de déplacement.

J'ai d'abord considéré qu'il n'y avait pas de différence entre l'air respiré par les piétons et les cyclistes empruntant les voies routières ou leur proximité, puisqu'ils se trouvent dans le même espace, qui n'est pas clos. J'ai considéré que les valeurs de concentration auxquelles ils sont exposés sont égales aux valeurs de concentrations à fine échelle. La situation est plus complexe pour les conducteurs ou passagers des véhicules particuliers et des transports collectifs, qui sont dans des espaces clos susceptibles d'être mal ventilés mais dont l'air entrant est également possiblement filtré.

Exposition dans les véhicules particuliers

Comme mentionné dans l'introduction de cette section, l'idéal serait de pouvoir appliquer des coefficients de transfert (ou d'abattement) à la concentration des polluants que l'on trouve au milieu du flux de véhicules, ce qui pose la question de la concentration des polluants au cœur du trafic. A-t-on une continuité du gradient de qualité de l'air « urbain de fond → proximité du trafic », sur le transect entre le trottoir, donc à proximité du trafic, et le cœur du trafic? Que sait-on d'une potentielle dégradation de la qualité de l'air au milieu des véhicules en circulation? Cette question est très peu traitée dans la littérature, car si les mesures d'efficacité des filtres positionnés en entrée d'air de l'habitacle des véhicules deviennent plus fréquentes, elle s'appuie uniquement sur des mesures de concentration des polluants « capot versus habitacle » et ne s'accompagne pas d'une appréciation du gradient « trottoir -> centre de la voie ». De récentes études proposent toutefois des mesures

comparées « habitacle / extérieur du véhicule / station de mesure » en utilisant la station de mesure de proximité au trafic la plus proche, et un rapport d'un facteur 2 (légèrement inférieur pour les PM_{2,5} et légèrement supérieur pour NO₂) est observé entre la mesure dans le trafic et celle sur la station (Matthaios et al., 2020). Cette question reste toutefois assez inexplorée et nous ne pouvons pas pour l'instant considérer de facteur multiplicatif robuste à appliquer aux teneurs en proximité au trafic (équivalent de notre *Cprox* estimé précédemment). Il faut donc considérer que cela sera à traiter dans de futurs développements.

Pour appliquer des coefficients de transfert extérieur - intérieur du véhicule, je me suis appuyée sur le projet « CAPTIHV » (Mehel et al., 2019) qui a permis la mesure embarquée des concentrations de différents polluants à l'intérieur et à l'extérieur d'habitacles de voitures sur 40 circuits routiers et à différentes heures de la journée. Les données obtenues montrent qu'on peut représenter la qualité de l'air dans les habitacles en appliquant un facteur multiplicatif moyen unique : les valeurs des rapports Intérieurs/Extérieurs des concentrations obtenues issues de ce travail sont en moyenne de **0,82** pour NO₂ et de **1,07** pour les PM₁₀. J'ai utilisé ces coefficients de transfert pour en déduire les valeurs de concentrations dans les habitacles des véhicules particuliers. A nouveau, ces coefficients soulèvent la question de l'effet de la filtration de l'air en entrée du véhicule. L'équipe de Matthaios et al., 2020 a très récemment investigué les effets de la ventilation naturelle (vitre ouverte), de la ventilation mécanique (renouvellement d'air) et de l'usage d'un filtre à air sur l'exposition dans l'habitacle. Les résultats sont très variables : ils s'étalent entre 5% et 95% de réduction de NO₂ et des particules fines lorsque l'on utilise respectivement un filtre à pollen et un filtre à charbon (état neuf). A nouveau, ces effets demandent à être étudiés plus en détail afin de pouvoir intégrer ces éléments dans notre estimation, mais on peut déjà dire qu'il faudra les intégrer.

Du point de vue technique, chaque brin du réseau ayant un identifiant qui lui est propre, j'ai pu relier chaque brin sélectionné par le calcul du plus court chemin à sa valeur de qualité de l'air *Cprox*. La concentration simulée sur chaque brin, affectée du coefficient de transfert adéquat issu de CAPTIHV, a donc été ajoutée en tant que nouvel attribut du réseau.

Exposition dans les transports en communs

Pour les premiers tests de faisabilité des calculs, j'ai cherché à me baser sur des valeurs fixes moyennes issues de la littérature. J'ai réalisé une recherche bibliographique sur les niveaux moyens de concentrations mesurés dans les habitacles des transports en commun, en me concentrant principalement sur les études menées en France et en Europe afin de garder un maximum de réalisme sur les configurations des stations souterraines et la conception des trains - mais aussi sur les environnements routiers pour ce qui est des bus. J'ai étudié 24 articles/rapports au total dont une revue de la littérature de 63 articles (dont plus de la moitié portait sur des pays européens et 13 sur la France). Au total, 44 valeurs de concentrations de PM₁₀ ont été renseignées pour les différents modes de transport (voiture, vélo, marche à pied, bus, métro, RER, train, tramway et correspondance) et 16 pour NO₂. J'ai pu trouver un assez grand nombre d'articles étudiant les niveaux de concentrations en métro et en voiture notamment (16 et 11 respectivement, les deux polluants confondus), mais un nombre assez limité pour les tramways, les niveaux retrouvés dans les tunnels souterrains (correspondances) et pour les trains/RER (2, 3 et 3 respectivement). J'ai dû effectuer des choix pour retenir des valeurs.

Les rames des métros sont les espaces les moins abondamment décrits. Ainsi, 13 des valeurs moyennes proposées dans ces articles (et qui sont présentées en [Annexe 9](#)) ont été mesurées sur les quais ou dans les halls principaux des stations de métro et seulement 5 ont été effectuées dans les rames. Je n'ai pas considéré qu'il était pertinent d'appliquer les concentrations des quais aux rames de métro car je n'ai pas d'information sur les transferts de polluants entre les deux micro-environnements. Ainsi, les valeurs que j'ai reportées dans le [Tableau 11](#) sont celles obtenues dans les rames, même si elles sont moins robustes. Je n'ai enfin pas pu trouver d'études évaluant les niveaux de NO₂ retrouvés dans les tunnels souterrains pour les correspondances métro, notamment, ni pour les funiculaires. Pour le premier, j'ai

donc posé l'hypothèse que les niveaux de NO₂ dans les tunnels étaient équivalents à ceux retrouvés sur les plateformes de stations de métro. Pour le second, j'ai considéré que les niveaux de polluants retrouvés au sein des remontées mécaniques étaient équivalents à celles mesurées dans les tramways.

Pour les niveaux ambiants de NO₂ et de PM₁₀ dans les bus, les valeurs moyennes que j'ai retenues à l'issue de la bibliographie sont très élevées (160 µg/m³ pour NO₂ et 115 µg/m³ pour les PM₁₀) par rapport à ce qu'on simule la plupart du temps sur les routes. Dans le cas de NO₂, cela correspond à la moyenne de 3 résultats principaux d'observation qui sont reportés dans l'Annexe 9 et qui s'étalent sur une très large gamme de valeurs observées. La concentration la plus basse (26 µg/m³) est issue de la revue bibliographique menée lors du projet INTER'MODAL de l'INERIS en 2009. Sur plus d'une soixantaine d'articles passés en revue pour obtenir cette concentration de 26 µg/m³, seuls 2 sont relatifs à des concentrations de NO₂ dans les bus pour des pays Européens. Les deux autres articles que j'ai pris en compte (Delaunay et al., 2012; Molle et al., 2013) sont issus de campagnes de mesures sur environ 30 trajets respectivement de 40 minutes et d'une heure, dans la ville de Paris. Paris étant une ville très dense en termes de bâti urbain, des phénomènes de pollution locale par l'échappement du bus, et d'accumulation de la pollution dans la rue, peuvent avoir lieu et expliquer les fortes valeurs de concentrations qu'ils ont trouvées (203 et 120 µg/m³ respectivement). Étant donné que ces deux dernières études ont été menées dans le centre de l'Île-de-France, qui est mon domaine d'étude, j'ai retenu une valeur moyenne de 160 µg/m³ de NO₂ dans les bus. Néanmoins cette valeur doit être mise à jour car notamment, la flotte de bus a été rénovée et les effets potentiels de pollution par le bus lui-même pourraient ne plus être aussi importants. Ci-dessous se trouve une figure comparant les expositions totales au NO₂ obtenues en multipliant les durées de déplacement de la population en bus avec la concentration ambiante - à gauche de 160 µg/m³ et à droite de 26 µg/m³ (Figure 44). Les différences d'exposition sont, bien sûr, très importantes entre les deux. En particulier, si la valeur retenue peut être représentative de ce à quoi la majorité de la population peut être exposée dans Paris et peut-être sa proche banlieue, elle surestime forcément celle de la population prenant le bus en grande couronne. Et de même pour les particules. C'est donc un chiffre qu'il faudra raffiner.

Pour les mesures en tramway, seule une valeur médiane de PM₁₀ était citée dans l'article de Strasser et al., 2018. Étant donné que cet article est le seul que j'ai pu trouver pour les PM₁₀ dans cet environnement, j'ai retenu cette valeur médiane alors que je ne l'avais pas fait pour les autres environnements TC pour lesquels j'avais des valeurs moyennes.

Comparaison des expositions totales au NO₂ sur la durée des déplacements

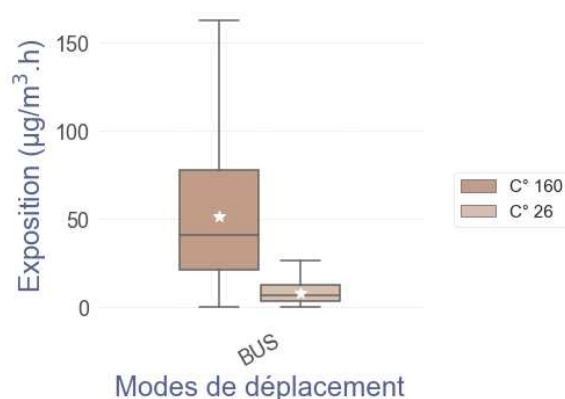


Figure 44 - Comparaison des expositions*temps, estimées pour les trajets en bus en considérant une concentration ambiante de 160 µg/m³ (à gauche) et de 26 µg/m³ (à droite) pour NO₂.

J'ai retenu des valeurs de NO₂ et de PM₁₀ représentant les concentrations moyennes auxquelles les usagers sont exposés lors de leurs déplacements, pour chaque mode de TC. Elles sont présentées dans le [Tableau 11](#). Quel regard poser sur les valeurs retenues ?

Tableau 11 – Valeurs moyennes de concentrations associées à chaque type de mode de transport collectif.

Mode	NO ₂ (µg/m ³)	PM ₁₀ (µg/m ³)
Métro	17	140
RER	35	118
Tramway/Funiculaire	30	9,3
Bus	160	115
Correspondance	45	196

Tout d'abord en termes de variabilité temporelle. Les écarts-types présentés pour les PM₁₀ dans le tableau de l'[Annexe 9](#) sont le plus souvent égaux à la moitié de la valeur moyenne, dénotant une variabilité modérée des valeurs sur une même campagne, exception faite de deux études pour lesquelles de très forts écarts-types sont calculés. Ces deux situations sont relatives à des mesures effectuées en extérieur (en bus et en voiture diesel) avec des moyennes et écarts-types de 144 et 170 µg/m³ pour la première et de 78,5 et 101 µg/m³ pour la seconde. On peut logiquement penser que c'est le mode de transport pour lequel on a le plus besoin de variabilité temporelle.

Il y a également de fortes variations sur l'ampleur des campagnes sous-tendant les articles, ce qui rend complexe le calcul d'une valeur moyenne. Le nombre de mesures effectuées est parfois plus d'un ordre de grandeur différent d'un article à l'autre : 6 articles rapportent une valeur issue de moins de 5 mesures pour certains modes de transport et 4 articles s'appuient sur une base de données de plus de 100 mesures. 4 articles ne donnent pas d'informations sur le nombre de mesures effectuées. Le seul moyen d'en tenir compte a été de calculer des moyennes pondérées. Cela ne résout pas tout, car les différentes campagnes sont en fait associées à des temporalités et lieux de mesures différents dont je ne peux tenir compte.

L'approche tabulée est la seule à pouvoir être mise en œuvre dans l'immédiat, car nous n'avons pas de projection du trajet des bus sur le réseau routier entre 2 arrêts. Mon principal problème reste toutefois le manque de capacité à proposer une variation spatiale, nécessaire pour distinguer les différentes lignes souterraines et extérieures, et avoir une variation temporelle de l'exposition, possiblement plus critique en extérieur.

Pour l'extérieur, plutôt que de s'appuyer uniquement sur de potentielles futures campagnes d'observation dans les bus, il faut trouver un moyen de pouvoir estimer l'exposition à partir de *Cprox*. Cela permettra d'être cohérent avec les calculs des expositions des piétons et conducteurs, mais aussi de prendre les variabilités spatiales et temporelles de la pollution en compte. Cela me permettra aussi de varier les valeurs selon le scénario considéré. Mais cela se heurte aujourd'hui au manque d'information sur le parcours des bus entre deux arrêts. L'évolution qui me semble la plus réaliste à envisager dans l'immédiat serait de considérer que l'air dans le bus est celui estimé au dernier arrêt, lors de l'ouverture des portes. On utiliserait alors *Cprox* tel quel, puisqu'à chaque arrêt le bus se trouve proche du trottoir. Cela demandera de complexifier un peu la boucle de calcul de l'exposition.

Exposition dans les bâtiments

Un des grands manques de l'approche traditionnelle de calcul de l'exposition est de ne pas prendre en compte la qualité de l'air dans les bâtiments, alors que le temps passé dans ces environnements intérieurs peut représenter jusqu'à 90% d'une journée. Ce choix est bien sûr un choix par défaut, puisque la littérature est très peu fournie sur le sujet, mais il peut se justifier par le fait que la pollution à l'extérieur d'un bâtiment affecte de manière significative la qualité de l'air à l'intérieur du logement. Dans le cadre de ma thèse, je me suis retrouvée confrontée à cette problématique, et ma démarche a consisté à implémenter une procédure pour estimer la qualité de l'air dans les bâtiments qui serait susceptible d'intégrer de nouvelles données et connaissances sur l'exposition aux polluants dans les bâtiments.

De la même façon que pour les transports collectifs, une des possibilités était d'utiliser les valeurs de concentrations extérieures modélisées et d'y appliquer un coefficient de transfert. J'ai recherché dans la bibliographie des informations relatives aux transferts des polluants entre les deux milieux. Une étude très détaillée de l'ANSES (ANSES, 2019), menée en collaboration avec le CSTB, a déterminé des coefficients de transfert des polluants de l'air entre l'extérieur et l'intérieur du bâti en France. L'ANSES a mis en place une convention de recherche avec le CSTB en complément de consultations d'organismes français et internationaux pour effectuer une revue de la littérature caractérisant l'état des connaissances actuelles sur les taux de transfert de polluants entre environnements extérieurs et intérieurs et à identifier les facteurs les influençant. Les résultats présentés dans leurs rapports sont basés sur une étude de 274 articles, dont les articles français ont été priorisés. Il en ressort que les taux de transfert sont influencés par quatre paramètres :

- Le premier concerne la caractérisation de l'environnement extérieur, à savoir : la distance du bâtiment en question par rapport aux sources d'émissions, la forme de la rue et les conditions météorologiques.
- Le deuxième concerne les caractéristiques du bâtiment en question : type de matériaux constituant le revêtement du bâtiment (car influence sur les phénomènes d'infiltration des polluants).
- Le troisième concerne les systèmes de ventilation/filtrations des bâtiments (à noter que ce point est décrit comme un des paramètres cruciaux des estimations des taux de transfert).
- Le quatrième concerne les habitudes des occupants des pièces du bâtiment.

Les auteurs mentionnent que le parc des bâtiments est relativement ancien, en France, à cause du faible taux de construction aperçu entre 1999 et 2013. Ils ont, de plus, pointé de nombreuses non-conformités des systèmes de ventilation des bâtiments les plus anciens comme les plus récents. Le renouvellement de l'air dans les pièces est dans les faits très souvent inférieur à la réglementation. Trois indicateurs de transferts peuvent généralement être considérés dans ce type d'études :

- Le facteur de pénétration (P) est un paramètre décrivant le passage d'un polluant extérieur à travers l'enveloppe et/ou le système de ventilation du bâtiment. Considérer ce paramètre nécessite de pouvoir caractériser la part de polluant qui est captée/arrêtée lors du passage extérieur-intérieur.
- Le rapport de concentration Intérieur/Extérieur (Rapport I/E en français et I/O en anglais). Ce paramètre se base sur des mesures de concentration réalisées en dehors et dans l'environnement intérieur considéré.
- Le facteur d'infiltration (F_{inf}) est un équivalent du rapport I/E mais il n'est caractérisable que s'il n'y a pas de source d'émission intérieure du polluant considéré.

La grande majorité des études cherchant à caractériser les transferts de polluant vont se baser sur le rapport I/E qui est plus simple à définir, même s'il est moins spécifique que les deux autres.

Au final, cette étude a pu conclure que, en moyenne dans toute la France, les rapports I/E de NO₂ et des PM₁₀/PM_{2.5} étaient de **1,32** et **1,61** respectivement pour toutes saisons confondues.

Plusieurs points assez limitants sont toutefois mis en avant dans leur étude. Il est expliqué, en premier lieu, que les études de taux de transfert de polluant entre environnements extérieurs et intérieurs suggèrent que les polluants proviennent exclusivement de l'extérieur car la caractérisation de la part des niveaux de polluants extérieur dont l'origine est intérieure s'avère difficile. S'ajoute à cela le fait que les pratiques des habitants eux-mêmes vont fortement influencer sur les niveaux de concentration retrouvés dans les pièces. En outre, pour appliquer ces coefficients, je rencontre deux verrous techniques :

- Il me faut localiser géographiquement le lieu d'activité de l'individu pour extraire la valeur locale de concentration en NO₂ ou en PM₁₀ extérieure. Or comme je l'ai mentionné, la version d'OLYMPUS sur laquelle j'ai travaillé ne fournit pas de précision sur la localisation des individus au sein des communes, même si j'ai statistiquement pu affiner la localisation jusqu'à l'échelle de l'IRIS.
- Cette approche est très consommatrice en ressource de calcul.

J'ai donc considéré une autre possibilité qui était d'utiliser, dans OLYMPUS, des concentrations intérieures moyennes provenant de l'observation et disponibles dans la littérature. Cette démarche a l'intérêt de prendre en compte indirectement la qualité de l'air extérieure - certes de façon moyenne mais elle est seulement une des composantes de ce qui est mesuré dans le logement - mais surtout de proposer une différenciation de l'exposition en fonction du type d'activité effectué.

OLYMPUS propose, aux agents qu'il simule, 9 activités : le logement, le travail, l'école, les courses, les loisirs, des activités dénommées « autre », des activités « secondaires » comme des visites à des amis/proches ainsi que l'accompagnement des enfants/adolescents à une de leurs activités. Mon but a donc été de rechercher des niveaux moyens en NO₂ et en PM₁₀ mesurés pour ces différentes activités. J'ai pu le faire pour certaines, mais j'ai rencontré des limitations : les activités « autres » et « secondaire », n'avaient pas de micro-environnements explicitement définis (ni même, à défaut, une différenciation entre les activités effectuées en intérieur ou en extérieur) pour me permettre d'orienter mes recherches. En l'occurrence, j'ai dû appliquer arbitrairement des valeurs basses pour ne pas leur donner trop d'importance. Un questionnement similaire s'est posé pour l'activité « accompagnement » à laquelle aucune information sur le type d'activité effectuée par le parent n'est proposée. Pour cette activité-là, j'ai considéré que les concentrations en polluants équivalaient à la moyenne des concentrations de toutes les autres activités.

Les détails des concentrations tirées de la littérature sont présentés en [Annexe 10](#). J'ai étudié, au total, 13 articles dont une revue de la littérature analysant 89 articles. Tout comme l'étude bibliographique sur les niveaux moyens dans les habitacles, j'ai principalement orienté mes recherches vers des études françaises ou européennes. Au total, 5 articles mesuraient les concentrations des polluants dans les logements, 4 articles dans les bureaux, 4 dans les écoles, 1 dans les magasins et 1 donnait une information sur les concentrations mesurées pendant les activités de loisir des participants (mesure effectuées tous loisirs confondus pendant 48h). Souvent, les informations sur les niveaux de particules concernaient les PM_{2.5} et peu concernaient directement les PM₁₀. Pour pallier à cela, j'ai considéré le rapport PM_{2.5}/PM₁₀ cité dans un document sur les études statistiques des PM_{2.5} produit par le LCSQA⁸ qui cite que ce rapport, en Ile-de-France, est compris entre 0,64 et 0,71. J'ai considéré la valeur moyenne de cette fourchette pour en déduire les niveaux de PM₁₀ pour les études ne discutant que des PM_{2.5}. Il ressort de mon travail bibliographique que cette littérature est surtout orientée vers les micro-environnements les plus visités par les individus (logement, lieu de travail/éducation). Presqu'aucun article, à ma connaissance, ne mesure les niveaux de concentrations retrouvés dans les lieux du type

⁸ https://www.lcsqa.org/system/files/media/documents/Etude_stat_PM2.5_mars07_v-finale_corr-27-09-07_1.pdf

magasins, grandes surfaces, cinéma/théâtre... Les résultats présentés et utilisés dans la suite du chapitre proviennent de l'implémentation de ces valeurs moyennes dans OLYMPUS. C'est une première proposition qui devra évoluer bien sûr, avec la littérature sur ces environnements clos.

Tableau 12 – Valeurs moyennes de concentrations associées à chaque type d'activité en $\mu\text{g}/\text{m}^3$ avec l'information sur les écarts-types entre parenthèses (quand celle-ci a été calculée).

Polluant	Logement	Travail	Accompagnement	Ecole	Courses	Loisirs	Autres	Secondaires	Visite à un ami
NO ₂	16,5 (9)	19,2 (7)	20	27,5	17,1 (8,6)	30	15	20	16,5 (9)
PM ₁₀	55 (60,4)	76,4 (74)	50	112 (69)	41,4 (14,9)	25	15	20	55 (60,4)

IV. Résultats

1. Caractéristiques de l'échantillon de population considéré pour cette étude

L'échantillon constitué aléatoirement via mon script de sélection des individus - et proportionnellement à la densité d'habitation - est composé de 140 273 individus. Comme mentionné précédemment, il n'a pas été possible d'affecter des itinéraires de déplacements pour les agents se déplaçant dans des communes n'ayant qu'un seul IRIS. J'ai donc eu à retirer ces agents de l'échantillon, la taille finale de mon échantillon était de 102 122 individus.

2. Confrontation des déplacements modélisés avec les données de l'EGT 2010

J'ai choisi d'effectuer ce travail d'évaluation des expositions sur l'année 2015 plutôt que sur l'année 2009 pour plusieurs raisons. D'une part, cela me permettait de paramétrer OLYMPUS et de préparer tous les fichiers d'entrée, pour une année proche de l'année sélectionnée pour mes scénarios urbains (voir le prochain chapitre). D'autre part, le réseau des transports en commun se base sur les fichiers GTFS, eux-mêmes générés grâce aux observations. Ainsi, les stations de transport en commun et les durées entre deux stations successives ont été calculées pour l'année 2020 qui est beaucoup trop éloignée de mon premier cas d'étude, sur 2009. Je comparerai dans cette partie les résultats issus des simulations OLYMPUS et du module d'estimation de l'exposition (2015), aux données publiées dans l'EGT 2010. Les résultats de l'EGT 2010 sont basés sur une enquête menée auprès de 18 000 ménages franciliens.

Les premières données sur la mobilité confirment que la simulation d'OLYMPUS est cohérente avec ce que l'on connaît de la réalité par les enquêtes de terrain. Les agents de mon échantillon de population effectuent en moyenne 3,46 déplacements par jours contre 3,87 pour l'EGT 2010 et le nombre de déplacements total journalier simulé par OLYMPUS (ramené au nombre total de la population recensée en 2010) est de 41,5 millions contre près de 41 millions pour l'EGT 2010.

Les individus de mon échantillon de population parcourent, en moyenne, 5,9 km par jour. Cette distance est de 4,4 km par jour dans l'EGT 2010 mais il faut prendre en considération le fait que les distances données dans l'EGT sont calculées à vol d'oiseau. Même si les valeurs ne peuvent pas être pertinemment confrontées, il reste tout de même intéressant de voir que la différence entre les deux représente est faible (moins d'un kilomètre) et cohérente avec l'approximation de l'EGT.

La distribution horaire des déplacements simulés permet de retrouver une structure avec des heures de pointe en matinée et en fin de journée (Figure 45). Une partie des individus de cet échantillon se déplace en soirée (au-delà de 20h) ce qui génère une traînée dans la distribution.

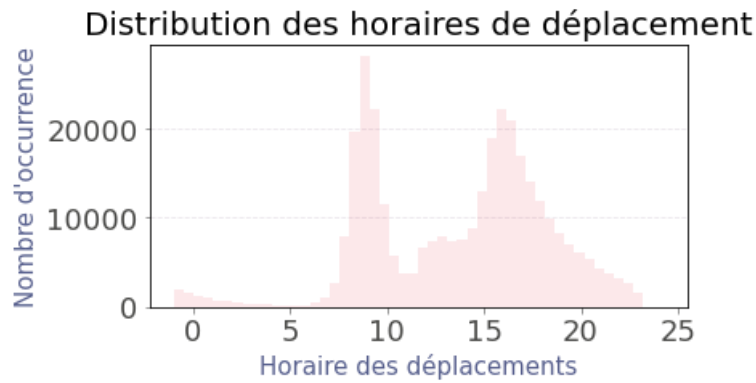


Figure 45 - Distribution des horaires de déplacement des individus de l'échantillon de population, tout mode de transport confondu.

Dans la Figure 46, les détails des distributions des durées de déplacement par modes de transport montrent que les choix de la population sont partagés entre les trois modes de déplacement pour les durées de déplacements les plus courtes, même si les modes doux l'emportent. À partir de 10 minutes de durée de déplacement et jusqu'à 25 minutes, OLYMPUS favorise l'utilisation des transports collectifs. On voit un pic d'utilisation des VP à 20 minutes qui est 3 fois plus important que celui des TC.

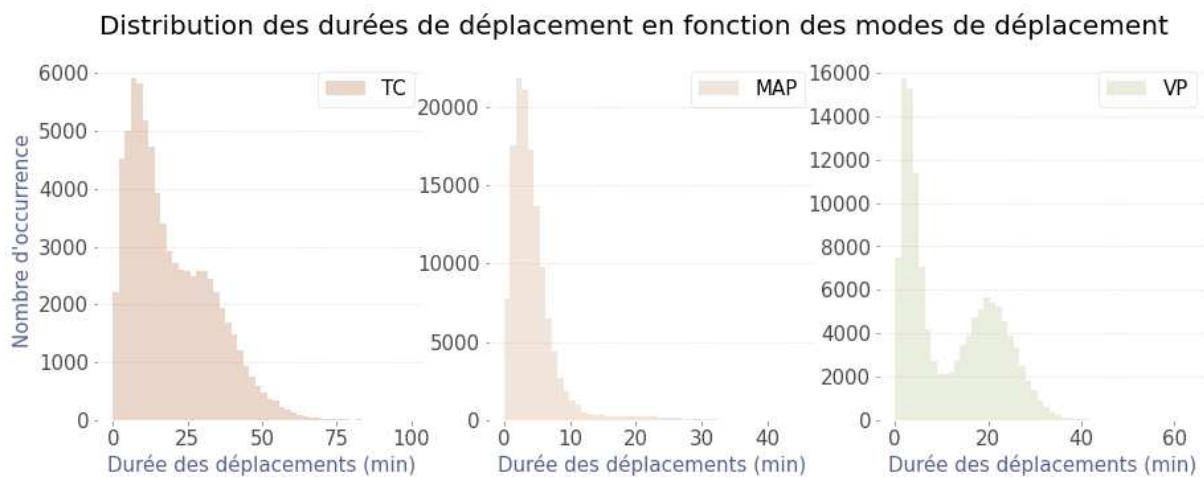


Figure 46 - Distribution des choix modaux en fonction des durées de déplacement pour les individus de l'échantillon de population.

3. Visualisation des estimations d'exposition réalisées

Dans cette section, je vais présenter les résultats des estimations d'expositions réalisées à partir de l'approche développée dans ce chapitre. Elles se présenteront sous deux formes :

- Les expositions totales sur la durée d'un trajet ou d'une activité, exprimées en $\mu\text{g}/\text{m}^3\cdot\text{h}$ et résultant du produit de la concentration des polluants par le temps de séjour ou de déplacement
- Les expositions journalières cumulées des individus sur 24 h, correspondant aux sommes des expositions totales sur les trajets et activités des agents au cours de leur journée, aussi exprimées en $\mu\text{g}/\text{m}^3\cdot\text{h}$.

Expositions durant un trajet ou une activité

Pour l'échantillon de population, les statistiques des expositions horaires sont présentées dans les graphiques de la [Figure 47](#), pour les différents modes de transport et pour les différentes activités, sous formes de boxplots. Les étoiles blanches représentent les moyennes.

Les expositions au dioxyde d'azote les plus fortes sont observées sur des déplacements en bus, notamment parce qu'ils couvrent une partie importante des longs trajets et parce que, comme expliqué précédemment, la valeur de concentration ambiante dans les bus est élevée. Les expositions sur les autres modes de déplacement représentent un cumul d'exposition similaire. Les gammes sur lesquelles s'étalent les expositions horaires pour les transports collectifs autres que le bus sont assez similaires entre elles, notamment du fait du choix de moyennes tabulées assez proches pour NO_2 . Pour les PM_{10} on constate le rôle déterminant des déplacements en métro et en train/RER dans l'exposition des individus. Les percentiles 50 et 95 des expositions horaires dans le métro sont très élevées pour les particules (50 et $125 \mu\text{g}/\text{m}^3\cdot\text{h}$ respectivement). De forts niveaux d'expositions sont aussi observés pour les déplacements en train/RER avec une moyenne légèrement supérieure à $50 \mu\text{g}/\text{m}^3\cdot\text{h}$ et une traînée de distribution qui atteint $100 \mu\text{g}/\text{m}^3$ pour le percentile 95. Cela est dû aux très fortes concentrations moyennes dans ces environnements.

Il n'y a pas, sur la journée que j'ai calculée, d'exposition très forte au NO_2 et aux particules pour les conducteurs dans leur habitacle. Ces valeurs d'exposition modérée peuvent être liées aux concentrations de polluants le long des routes plus faibles ce jour particulier. On doit aussi vérifier si les conducteurs de l'échantillon de population ne se trouvent pas en majorité sur des routes en moyenne et grande couronne, plutôt que dans Paris. Il faut enfin considérer le fait que les valeurs de *Cprox* ont été calculées à partir de la mesure réalisée sur les trottoirs. Or, comme je l'ai expliqué plus tôt dans mon manuscrit, les valeurs de concentrations au centre de la route sont potentiellement plus élevées que celles sur les trottoirs. Il est donc possible qu'on se trouve ici avec une valeur basse d'exposition, pour plusieurs raisons.

Enfin, le facteur temps joue bien sûr sur le cumul d'exposition. L'exposition cumulée lors des déplacements en marche à pied est très faible car ces déplacements sont très courts en temps ([Figure 47](#)). En sortie de simulation OLYMPUS, environ 50% des déplacements en VP et 95% des déplacements en mode doux ont des durées inférieures à 10 minutes ce qui correspond, respectivement, à 66 500 et 125 157 déplacements. Les valeurs d'expositions qui y sont associées sont, donc, faibles à très faibles.

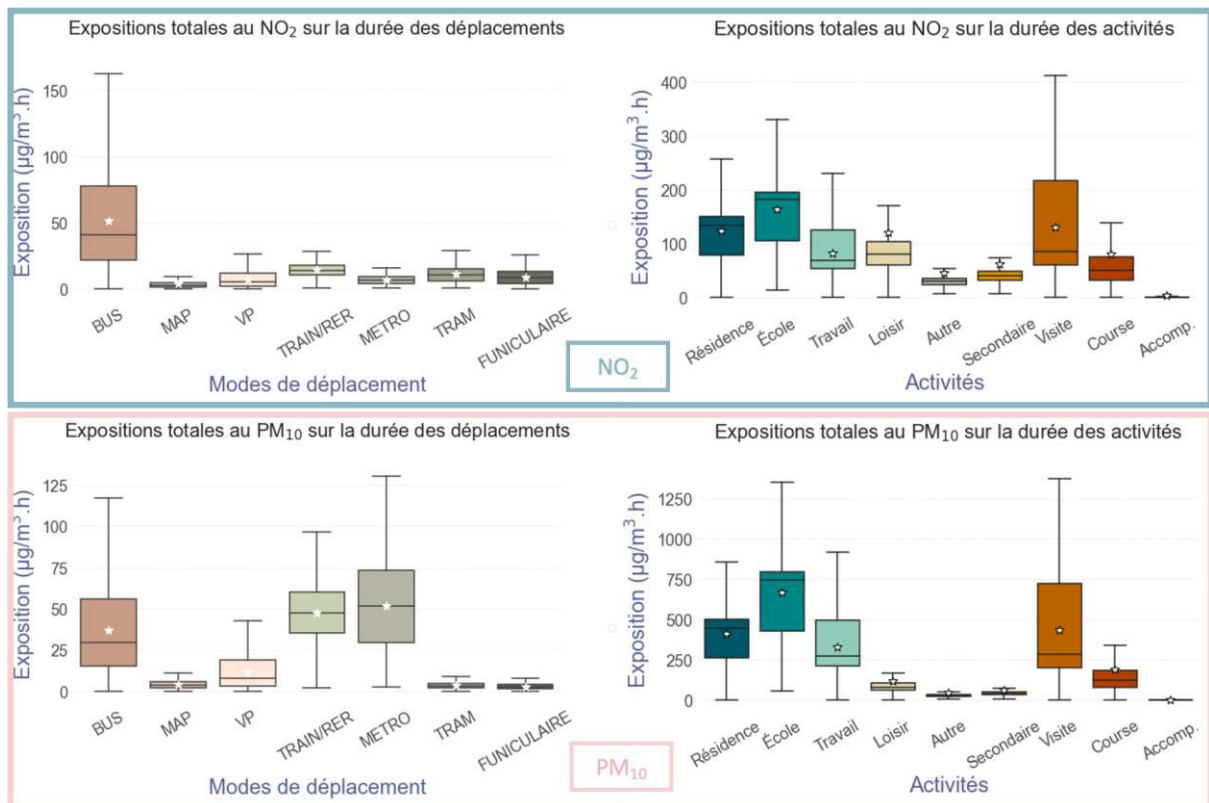


Figure 47 - Boxplots des expositions totales sur les durées de déplacements (à gauche) et d'activités (à droite) pour NO₂ (en haut) et les PM₁₀ (en bas).

Pour ce qui concerne les activités, les longs temps passés imposent des cumuls d'exposition significatifs. Ceux-ci dépassent 100 µg/m³.h pour NO₂ pour le temps passé dans le logement, l'école et les visites, et surtout ils s'étalent sur la gamme 500 – 800 µg/m³.h pour les journées d'école. En revanche, les activités « Secondaire », « Autre » et « Accompagnement » ont des expositions totales sur le temps d'activité très faibles car les concentrations qui y sont associées sont relativement faibles (Tableau 12). Ce ne sont toutefois pas des activités critiques pour nos statistiques d'exposition car dans notre échantillon, le nombre d'individus les effectuant est bas (respectivement 15 000, 18 000 et 22 500 contre un nombre d'individus 2 à 10 fois plus élevé pour les activités Résidence, École et Travail).

Les cumuls d'expositions dans les établissements scolaires (164,5 µg/m³.h pour NO₂ et 670 µg/m³.h pour les particules) sont donc problématiques dès lors qu'on considère le temps qui y est passé, qui plus lorsqu'on prend en compte le fait qu'ils concernent en majorité un public très jeune et sensible à la qualité de l'air inspiré. Les durées de cette activité sont, en majorité, comprises entre 7 et 10h et concernent sur notre échantillon environ 70 000 individus. Cela est lié au fait que les écoles sont associées aux moyennes les plus fortes de toutes les activités, et cela constitue un argument fort pour la multiplication des mesures de qualité de l'air dans les écoles. Notamment afin d'en observer la variabilité, la dépendance envers la qualité de l'air extérieur et les activités qui y sont pratiquées.

Au-delà de l'école, la figure met en avant l'exposition dans les logements - pour les deux polluants étudiés - que ce soit à domicile où lorsqu'ils visitent leurs proches. Les valeurs hautes de la distribution, pour l'activité « Visite », viennent du fait que les « Visites » sont plus nombreuses pour les durées les plus longues comme on le voit en Figure 48 ci-dessous.

Comparaison des distributions des durées entre les activités 'Résidence' et 'Visite'

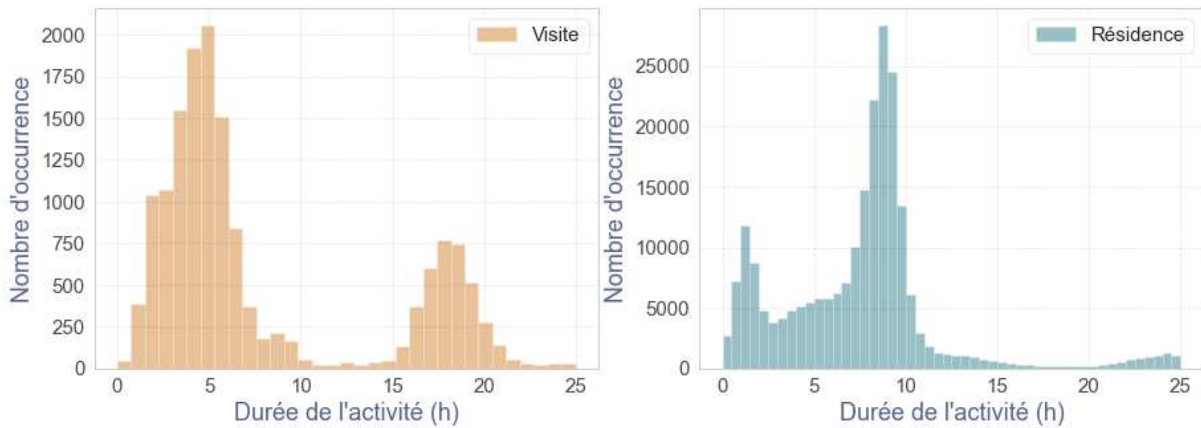


Figure 48 - Comparaison des distributions des durées de « Visite » (à gauche) et de « Résidence » (à droite).

Enfin, toujours dans la [Figure 47](#), on peut remarquer que les valeurs des percentiles 5 (que ce soit pour les modes de déplacement ou pour les activités) sont souvent très proches de 0. Après vérification, j'ai pu constater que, pour les déplacements, les durées les plus basses pour les Bus et Funiculaires étaient nulles. Cela vient du fait que les temps de déplacements entre certaines stations successives du réseau de transport en commun sont nulles car l'horaire de départ à la première station est identique à l'horaire d'arriver à la station suivante. Cette problématique est donc issue des fichiers GTFS en eux-mêmes mais je n'ai à ce jour pas d'explication sur l'origine de ces phénomènes. Le changement des fichiers d'entrées ou leur travail (en imposant un temps minimum entre deux stations) peut permettre de résoudre immédiatement ce problème.

Cumuls d'exposition sur une journée

J'ai cartographié les cumuls moyens d'expositions de la journée des individus sur leur commune de résidence dans les [Figure 49](#) et [Figure 50](#), pour NO_2 et les PM_{10} respectivement. Ces cartographies permettent de rendre compte des niveaux moyens d'expositions cumulés de la journée des habitants des différentes communes d'Île-de-France. La cartographie de NO_2 montre que les résidents de Paris et des communes situées en petite couronne sont plus exposés que ceux habitant en périphérie du domaine. Pour les particules, les expositions cumulées les plus fortes occupent un territoire plus vaste et sont plus homogènes sur ce territoire. Ceci est cohérent avec leur temps de séjour plus important que NO_2 dans l'atmosphère, et avec l'existence de fortes sources liées à la combustion du bois en périphérie parisienne, qui peuvent compenser l'exposition au trafic dans Paris. La plupart des communes de ce territoire sont associées à un cumul d'exposition moyen de 1550-1600 $\mu\text{g}/\text{m}^3\cdot\text{h}$, sur 24 heures.

J'ai étudié plus particulièrement les expositions cumulées des habitants de deux communes : (i) celle de Créteil, située en petite couronne, dans un environnement très urbanisé où l'accès aux modes de transport et aux activités est facilité et, (ii) celle d'Étampes, située en Essonne, dans la grande couronne. Les deux communes sont encadrées sur les cartes des [Figure 49](#) et [Figure 50](#). Seuls 58 individus d'Étampes sont représentés dans mon échantillon et j'ai donc sélectionné aléatoirement 58 individus de Créteil parmi ceux de mon échantillon, afin de comparer leurs niveaux d'exposition cumulés dans la journée. Pour illustrer ces cumuls, j'ai choisi de tracer des « heatmaps » : c'est une technique de visualisation des données qui repose sur l'utilisation de gradients de couleur pour pouvoir analyser rapidement les tendances sur un jeu de données complexe et/ou de grande dimension. Ces graphiques offriront un visuel simple pour comparer les expositions des individus sélectionnés. Les heatmaps

représentent en abscisse les différents individus, et en ordonnée, de haut en bas, le cumul d'exposition au cours de la journée.

J'ai mis en regard de ces figures de cumul d'exposition l'évolution des concentrations simulées par CHIMERE sur les deux communes. Les concentrations extérieures de NO₂ sur les communes varient entre 2 et 18 µg/m³ et celles des PM₁₀ sont au maximum de 28 µg/m³. Pour les deux polluants, les heatmaps mettent en avant non seulement la diversité des cumuls d'exposition des individus, mais aussi l'ampleur des cumuls d'exposition calculés par individu (800µg/m³.h pour NO₂ et jusqu'à 1400µg/m³.h pour les PM₁₀) par rapport à ce que l'on aurait obtenu avec les données CHIMERE (au maximum 300 et 500µg/m³.h respectivement sur la journée). Ces résultats soutiennent l'intérêt de procéder à une estimation dynamique et individuelle de l'exposition.

On constate également que les individus de Créteil ont des cumuls d'exposition au NO₂ légèrement plus élevés que ceux de d'Étampes : la différence est d'environ 100 µg/m³.h en moyenne entre les deux, pour un total qui atteint sur Créteil près de 800 µg/m³.h, soit 10 à 15%. Le nombre d'agents ayant de forts cumuls d'expositions aux PM₁₀ (> 1 400 µg/m³.h) est plus élevé pour la commune de Créteil.

On peut remarquer que certains agents ont des valeurs de cumuls d'exposition qui diminuent en fin de journée (phénomène très visible pour les heatmaps de la [Figure 50](#)). Ceci est dû au fait que des agents exercent des activités à cheval sur deux journées. Pour ces agents, leurs journées ne se déroulent pas de minuit à minuit mais de 17h à 17h par exemple. Dans les heatmaps de la [Figure 50](#), on peut lire la fin de leur journée à l'heure où les intensités des expositions sont les plus élevées.

Pour terminer, j'ai représenté les expositions cumulées sur la journée pour des individus ayant de long temps de déplacements. Ces cumuls d'expositions sont représentés dans la [Figure 51](#) pour 5 individus. Le tracé en rouge représente l'évolution des cumuls d'expositions aux PM₁₀ et en bleu, à NO₂. Les fonds hachurés représentent les activités effectuées et les fonds unis représentent les déplacements. Ce format de représentation permet de se rendre compte très facilement de l'impact des déplacements sur les expositions : si cet impact est très ponctuel dans le temps, on peut observer les ruptures de pentes imposées au cumul par l'activité « transports », qui mènent à un cumul sensiblement plus élevé en fin de journée. Les cumuls d'expositions aux PM₁₀ en µg/m³.h sont systématiquement plus élevés que le NO₂ à cause des niveaux moyens rencontrés dans les différentes activités et déplacements de la journée, notamment sur les réseaux ferrés.

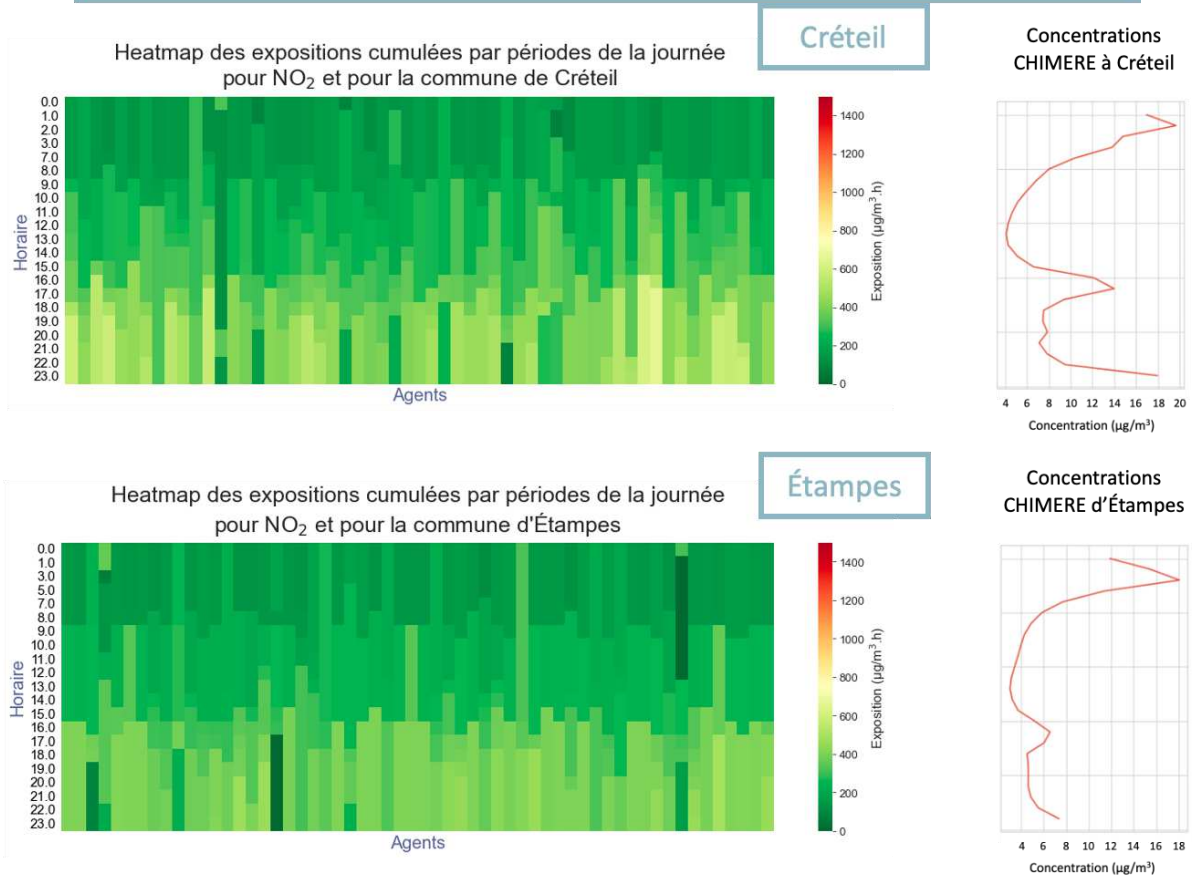
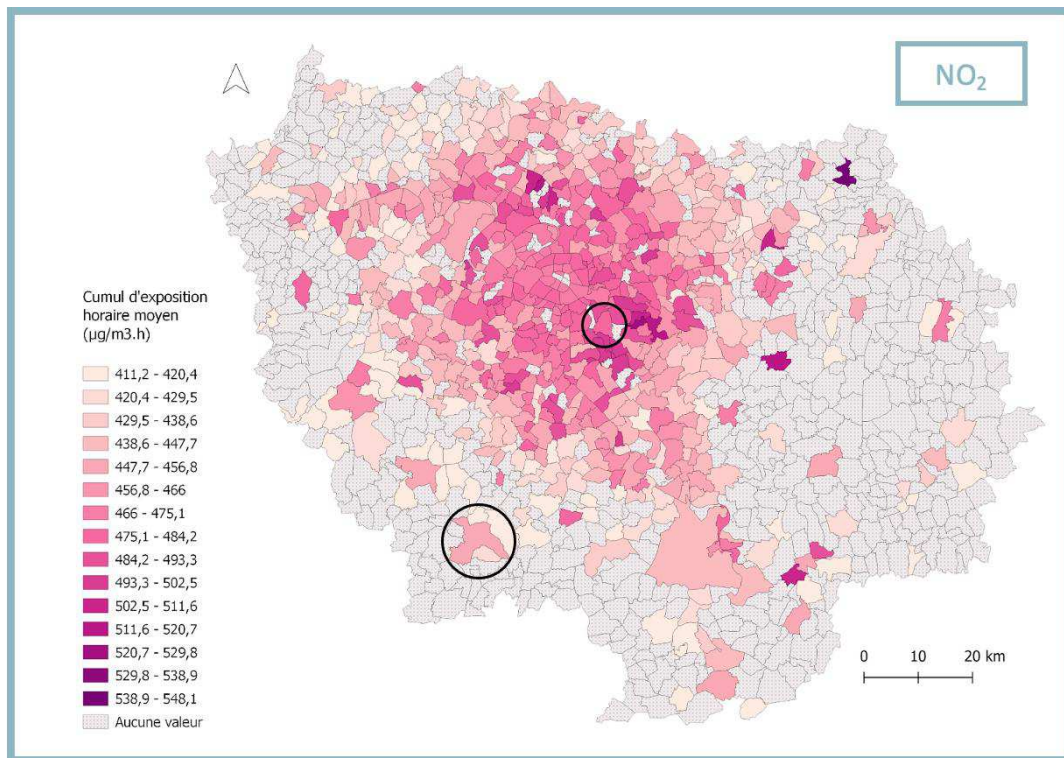


Figure 49 - En haut, cartographie des expositions dynamiques cumulées moyennes sur la journée des individus au niveau de leur lieu de résidence pour le NO_2 . En bas, comparaison des cumuls d'exposition des individus de Créteil et d'Étampes avec les concentrations CHIMERE simulées au niveau des deux communes pour NO_2 .

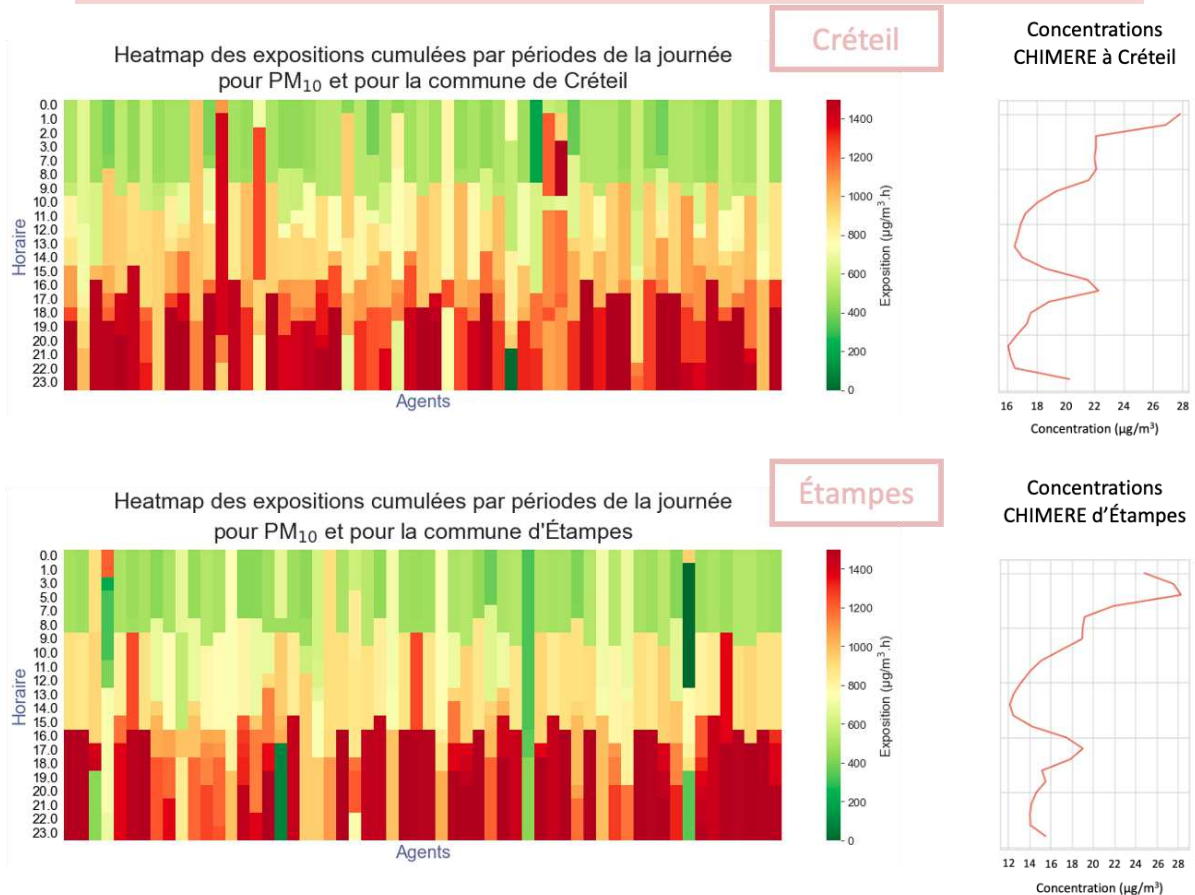
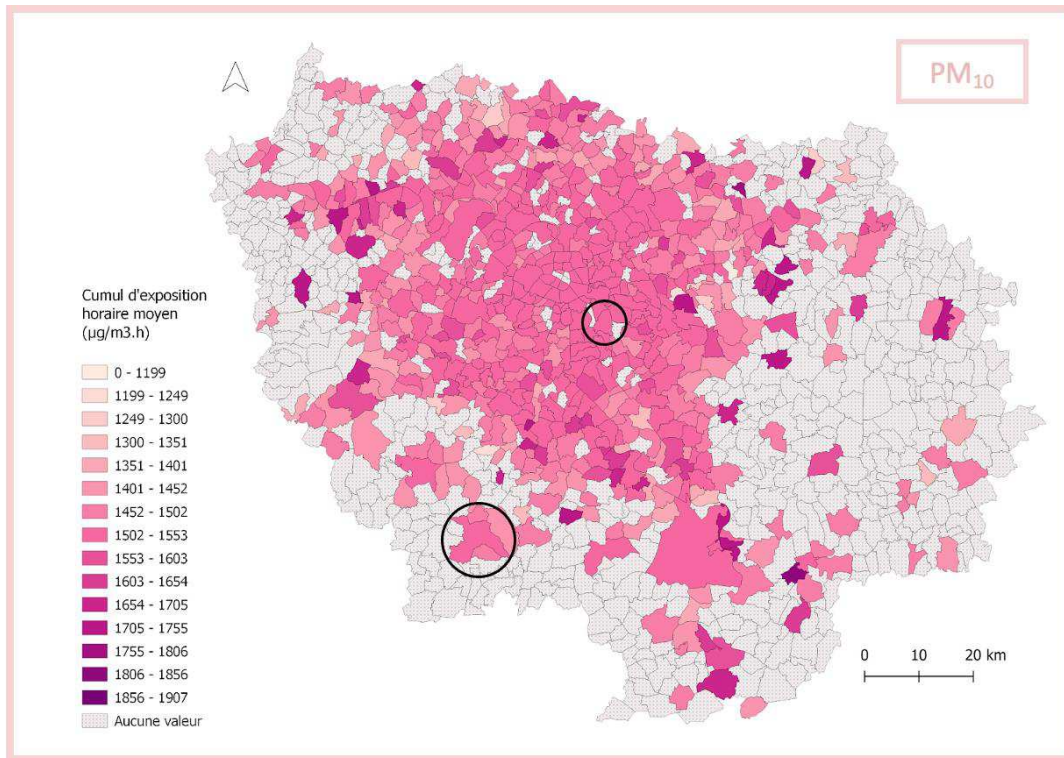


Figure 50 - En haut, cartographie des expositions dynamiques cumulées moyennes sur la journée des individus au niveau de leur lieu de résidence pour les PM₁₀. En bas, comparaison des cumuls d'exposition des individus de Créteil et d'Étampes avec les concentrations CHIMERE simulées au niveau des deux communes pour PM₁₀. Tout en bas, illustration du décalage de journée d'un individu.

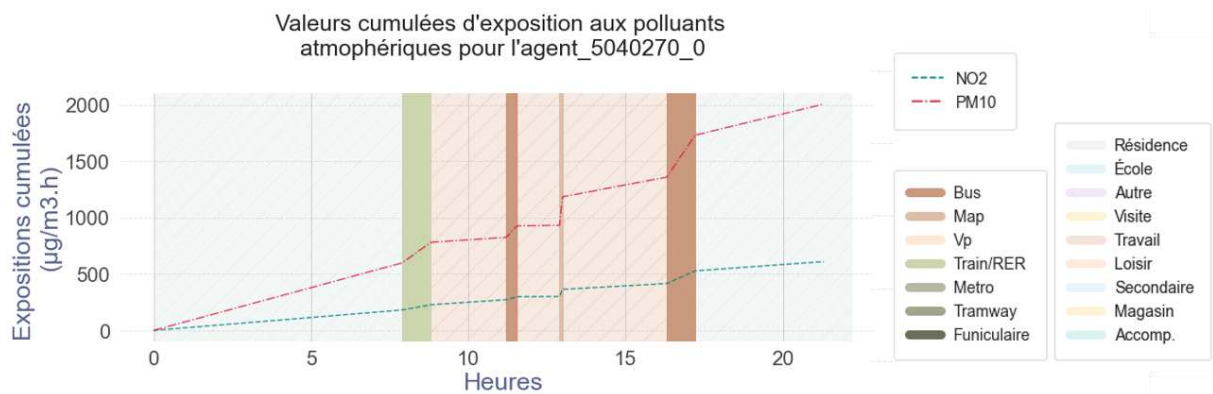
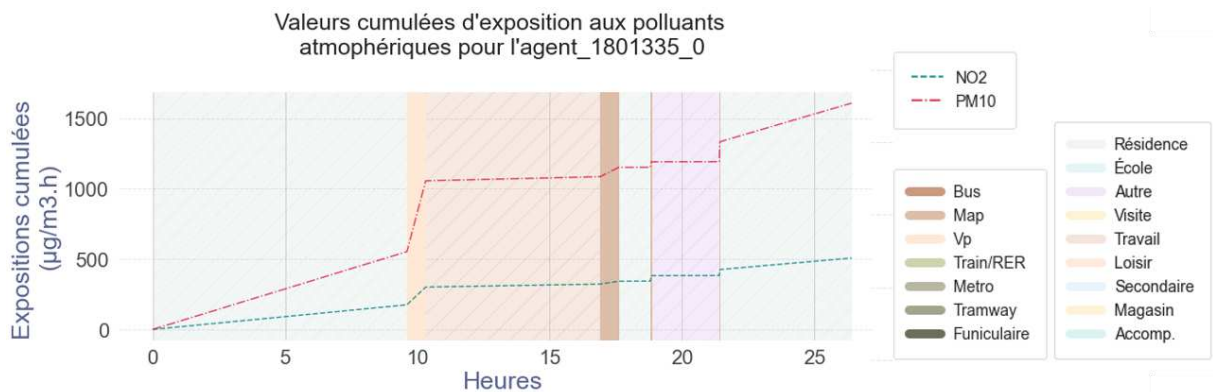
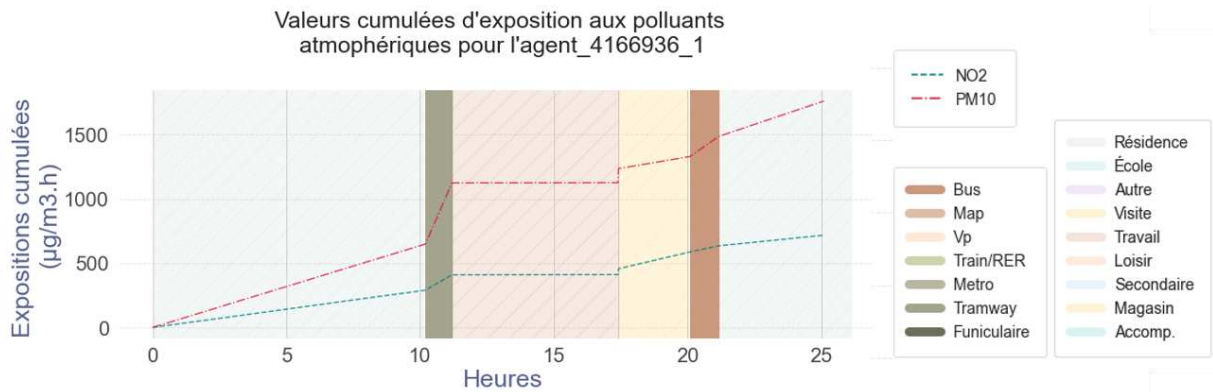
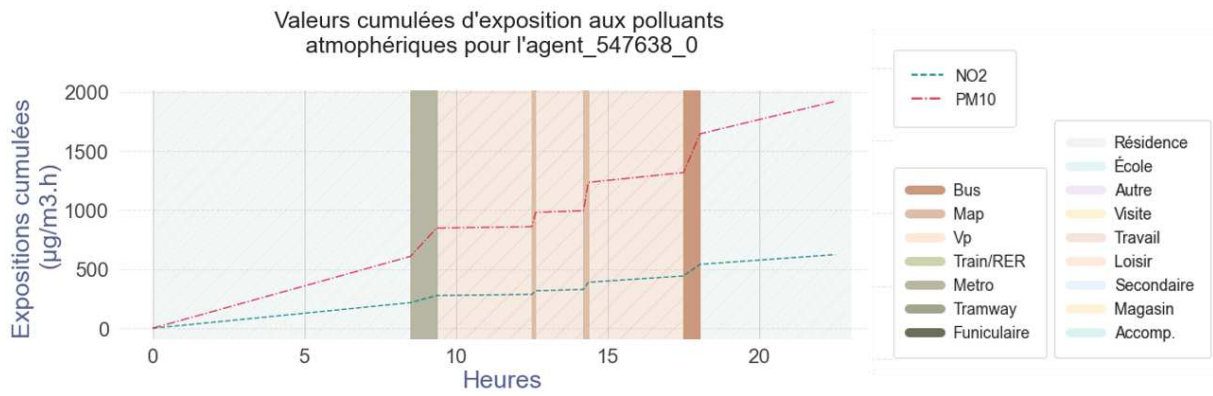


Figure 51 - Cumuls d'expositions dynamiques sur la journée pour 5 individus possédant de longs temps de transport. Le cumul d'exposition à NO₂ est présenté en vert pointillé, celui aux PM₁₀ en tirets rouges. Les activités et déplacements sont présentés en couleur, en plein pour la mobilité et en hachuré pour les activités sur un site donné.

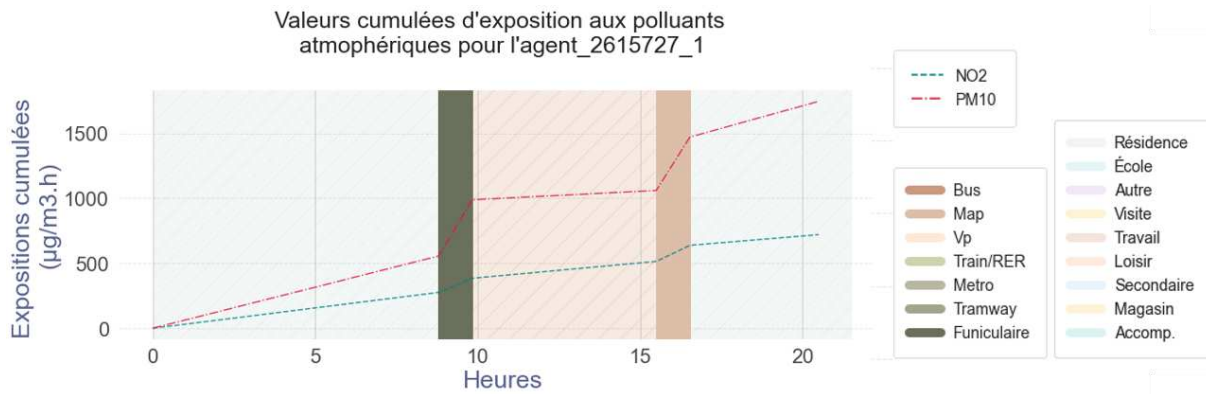


Figure 51 – (suite).

Les principaux points d'améliorations que l'on doit apporter à cette approche concernent en premier lieu la caractérisation des niveaux de concentrations des polluants dans les environnements intérieurs. Pour pouvoir prendre en compte le transfert intérieur/extérieur, en particulier pour NO₂, il faudra implémenter dans OLYMPUS une localisation géographiquement précise des logements et activités des individus, et utiliser les concentrations de l'approche *Cprox* affectées de coefficients de transfert. Cela permettra une meilleure caractérisation de la variabilité quotidienne des concentrations ambiantes subies par les individus au cours de leur temps passé dans les bâtiments. Ce point est en cours de développement dans l'équipe (thèse de Taos Benoussaid). Il pourra aussi affecter la mobilité des individus, en tant que point de départ pour les matrices OD - plutôt que les centroïdes de communes actuels.

Dans un second temps, il sera nécessaire de modifier le réseau routier afin de pouvoir passer à une approche non tabulée de l'affectation des concentrations dans les transports en commun non souterrains. En parallèle, une meilleure discrimination de la qualité de l'air dans différents environnements souterrains est souhaitable. Ces environnements sont très chargés en particules et il est important d'avoir des données locales et potentiellement associées à des lignes de transport spécifiques, pour mieux pouvoir discriminer l'exposition selon les lignes et les modes de déplacement. Notamment, les niveaux ambiants sont bien supérieurs sur les parties souterraines des réseaux que sur les parties aériennes. Enfin, un dernier point limitant de l'approche est la non-prise en compte des temps d'attente des individus lors de leurs déplacements en transports en communs, car cette information n'est pas indiquée dans les bases de données publiques. Or, comme le montrent les campagnes de mesure dans les tunnels de métro, les niveaux de concentrations ambiants dans ces micro-environnements peuvent être très élevés et peuvent donc peser davantage sur les cumuls d'expositions en fin de journée.

V. Autres développements

La construction de l'approche d'estimation des expositions m'a amenée à constater que la voirie prise en compte jusque-là dans OLYMPUS manquait de réalisme : non seulement en termes de nombre de brins routiers, mais aussi du point de vue de la configuration spatiale des brins et intersections. Je présente ici le travail mené pour améliorer ce point.

1. Le réseau routier utilisé par OLYMPUS

Lors du développement d'OLYMPUS, nous avons initialement utilisé un réseau routier utilisé par nos partenaires du LVMT, pour leur modèle de transport. Il est représenté par un ensemble de nœuds

distribués sur le domaine et reliant différents brins routiers qui sont les routes principales et secondaires de la région pour représenter un total de 43 819 tronçons de route sur le territoire de l'Île-de-France. Ce réseau routier qui se densifie en son centre est aussi utilisé comme support pour le calcul des déplacements effectués en mode doux (marche à pied et vélo). Créer un réseau supplémentaire spécifique aux modes doux aurait en effet excessivement alourdi les temps de calculs et les ressources nécessaires au stockage des informations liées aux déplacements, d'autant plus qu'il aurait nécessairement été beaucoup plus dense que celui créé pour les déplacements en VP. Cela induit, forcément, des biais au niveau des grands axes routiers qui ne peuvent pas être empruntés par des piétons mais permet, tout de même, de représenter efficacement la mobilité douce sans contraindre les durées des simulations.

Or ce réseau a démontré qu'il possédait des limitations pour des applications de mobilité des individus. On peut notamment voir sur la [Figure 52.A](#) (carte de gauche) que ce réseau est fortement simplifié hors de Paris et de la première couronne. En outre, même dans Paris il propose parfois de relier des intersections de manière totalement irréaliste. Ainsi, en bas de la [Figure 52](#) on constate que les tronçons du réseau LVMT (en rouge) traversent le bâti urbain (en marron). On le voit également dans le médaillon « B » en haut à gauche de la figure. Cette représentation simplifiée porte préjudice au calcul des trajets de tous les véhicules utilisant la voirie : non seulement en induisant un biais dans le calcul des distances et durées de déplacement parcourues par l'individu mais surtout, *in fine*, elle bloque l'estimation de l'exposition des individus.

2. Construction et implémentation d'un nouveau réseau routier

Pour résoudre ces problèmes j'ai tenu à implémenter, pour de futures simulations, un réseau routier plus complexe ([Figure 52C](#)) et plus réaliste afin de permettre à OLYMPUS d'évoluer vers une plus grande finesse de représentation des trajets.

J'ai choisi d'utiliser les données fournies par OpenStreetMap⁹ ([Boeing, 2017](#)) pour construire le nouveau réseau routier à ajouter à OLYMPUS via le module python OSMnx. Ce module modélise le réseau routier de n'importe quelle région grâce aux données d'OpenStreetMap¹⁰ et de Google Maps¹¹ (via son API). Ce choix se justifie, en premier lieu, par son degré de raffinement de la voirie que l'on peut constater sur la partie notée « C » de la [Figure 52](#). Ensuite, ce module permet de télécharger rapidement et facilement (en une ligne de code) n'importe quel type de voirie (routière, de service, piétonne ou cyclable) sur un territoire donné puis de les convertir en « graph ». Il permet aussi de simplifier les réseaux téléchargés en fusionnant les nœuds proches et assemblant leurs brins (ainsi que leur attributs) pour alléger les calculs (voir en [Figure 53](#)). Au final, ce réseau téléchargé et simplifié grâce au module OSMnx contient 313 313 brins de routes, soit 7 fois plus que le nombre de tronçon du réseau LVMT.

⁹ <https://github.com/gboeing/osmnx>

¹⁰ <https://www.openstreetmap.org/#map=11/10.5061/-66.9146>

¹¹ <https://developers.google.com/maps/documentation/elevation/overview>

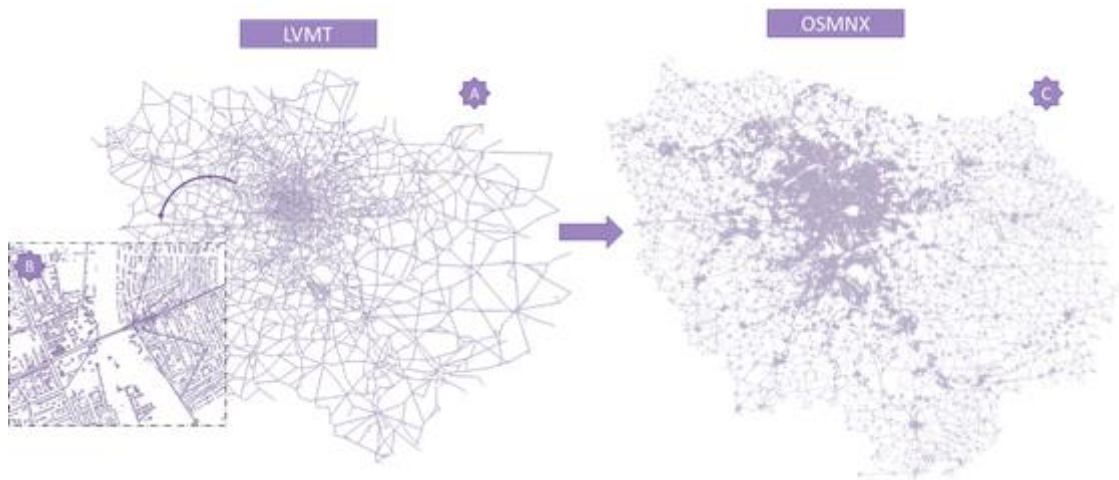


Figure 52 - Réseau routier généré par le LVMT sur le domaine Ile-de-France (A), zoom sur un exemple de traversée des zones d'habitation par les tronçons de route (B) et réseau routier généré grâce à OSMNX (C). En dessous, zoom sur une partie de la ville de Paris où le bâti est représenté en marron, les tronçons de route du réseau LVMT sont représentés en rouge et ceux du réseau OSMnx sont représentés en noir.

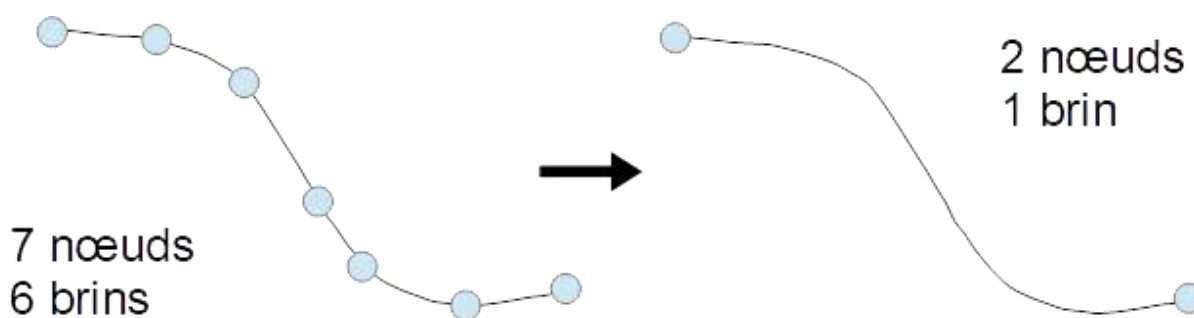


Figure 53 - Représentation schématique de la simplification du réseau OSMnx.

Les réseaux issus d'OSMnx contiennent les informations relatives à la longueur et la largeur (en m), au type de tronçons de route (autoroute, route principale/secondaire/tertiaire/non classée/résidentielle) et au nom du tronçon. Il contient aussi les informations sur les vitesses maximales autorisées (en km/h) par type de route et permet, ainsi, d'en déduire les temps de déplacement sur chacune d'entre elles. Enfin, des informations supplémentaires sont décrites avec, notamment, la présence ou non d'accès

particuliers aux routes, le sens de direction d'un tronçon, les identifiants des nœuds d'origine et de destination d'un brin ou encore la géométrie des brins de routes pour permettre de cartographier les résultats.

Je n'ai utilisé le module OSMnx que pour télécharger et simplifier le réseau routier de l'Île-de-France, puis j'ai utilisé Networkx¹², un autre module python permettant de créer, organiser et visualiser des graphs sous forme de réseaux orientés ou non. Dans notre cas, il est important de pouvoir considérer les tronçons de routes en sens-unique et, de ce fait, j'ai créé un réseau dirigé (appelé « DiGraph » pour « Directed Graph ») afin d'effectuer des calculs d'itinéraires plus réalistes. OSMnx organise alors les nœuds et les liens sous forme de Graph Networkx pour permettre à l'utilisateur de les manipuler, les visualiser et/ou les enregistrer sous différents formats de fichiers - ce qui est intéressant car du point de vue technique, OLYMPUS a besoin d'un réseau routier sous format Shapefile en entrée de simulation. J'ai choisi de créer un réseau regroupant toute la voirie francilienne en sélectionnant l'attribut « drive » du type de route demandé. Cette option permet de regrouper tous les tronçons de route pouvant être empruntés par des véhicules motorisés - qu'elles soient principales, secondaires, tertiaires, résidentielles ou autres.

Il s'avère toutefois que certaines données sont manquantes dans le réseau téléchargé par le module. Il manquait, notamment, l'information sur la vitesse maximale autorisée (VMA) sur certain brin de route. Or, un des paramètres les plus importants pour OLYMPUS est la vitesse moyenne de déplacement des véhicules sur les routes. Normalement, ces données de vitesses moyennes sur les routes lui permettent de prendre en compte la congestion lors des calculs d'itinéraires. Je n'ai pas pu obtenir ces données. À défaut, je me suis dirigée vers une attribution de VMA en fonction du type de route en question. Mes recherches m'ont conduite vers un projet appelé **OSRM** (Open Source Routing Machine) qui est une application de calcul d'itinéraire haute performance basée sur les réseaux routiers d'OpenStreetMap. Ils énumèrent, dans le code source de l'application, les valeurs de VMA retrouvées, en moyenne, par type de route (tout pays confondu) et avec un détail des VMA en zones rurales en fonction des pays. Les valeurs qu'ils proposent sont résumées dans le [Tableau 13](#). On peut se poser la question de la pertinence de ces valeurs appliquées aux routes franciliennes mais, comme expliqué précédemment, ce sont les seules données que j'ai pu me procurer. Toutefois, l'application de ces VMA n'était possible que lorsque le tronçon ne correspondait qu'à un seul type de route. Il arrive, en effet, qu'un tronçon de route corresponde à plusieurs types. Pour ces brins de routes, j'ai choisi d'appliquer une valeur de VMA correspondante à la moyenne des VMA des types de routes donnés ([Tableau 14](#)). Parmi ces brins, certains avaient pour type de route le type 'unclassified', pour lequel la VMA donnée par OSRM est de 25 km/h. Or, ce type de route était souvent associé à d'autres types bien identifiés (Exemple : 'unclassified' et 'motorway_link' ayant respectivement des VMA de 25 et 45 km/h). Pour ces brins, je n'ai considéré que le type de route identifié dans le calcul de la VMA.

Tableau 13 - Récapitulatif des valeurs limites utilisées pour compléter les données du réseau OSMNX en fonction du type de route des tronçons.

Type de route	VMA (km/h)
"motorway" (autoroute)	90
"motorway_link" (bretelle d'autoroute)	45
"trunk" (Frett)	85

¹² <https://networkx.org/documentation/stable/>

“trunk_link” (bretelle Frett)	40
“primary” (route principale)	65
“primary_link” (bretelle route principale)	30
“secondary” (route secondaire)	55
“secondary_link” (bretelle route secondaire)	25
“tertiary” (route tertiaire)	40
“tertiary_link” (bretelle route tertiaire)	20
“unclassified” (non classifiée)	25
“residential” (résidentielle)	25
“living_street” (?)	10
“service” (route de service)	15
“rural” (route rurale)	80

Tableau 14 - Moyenne des VMA pour les tronçons correspondant à plusieurs types de route.

Type de route 1	Type de route 2	Moyenne VMA (km/h)
‘primary’	‘trunk’	75
‘secondary’	‘tertiary’	47,5
‘secondary’	‘trunk_link’	47,5
‘secondary’	‘residential’	40
‘motorway_link’	‘primary_link’	37,5
‘tertiary’	‘primary_link’	35
‘tertiary’	‘residential’	32,5

Une représentation de la VMA sur le réseau OSMnx (Figure 54) avant (A, zoom sur la ville de Paris, seuls les tronçons ayant une information sur les VMA sont tracées) et après les avoir complété (B, sur l’ensemble du réseau) montre l’intérêt d’avoir mené cette correction.

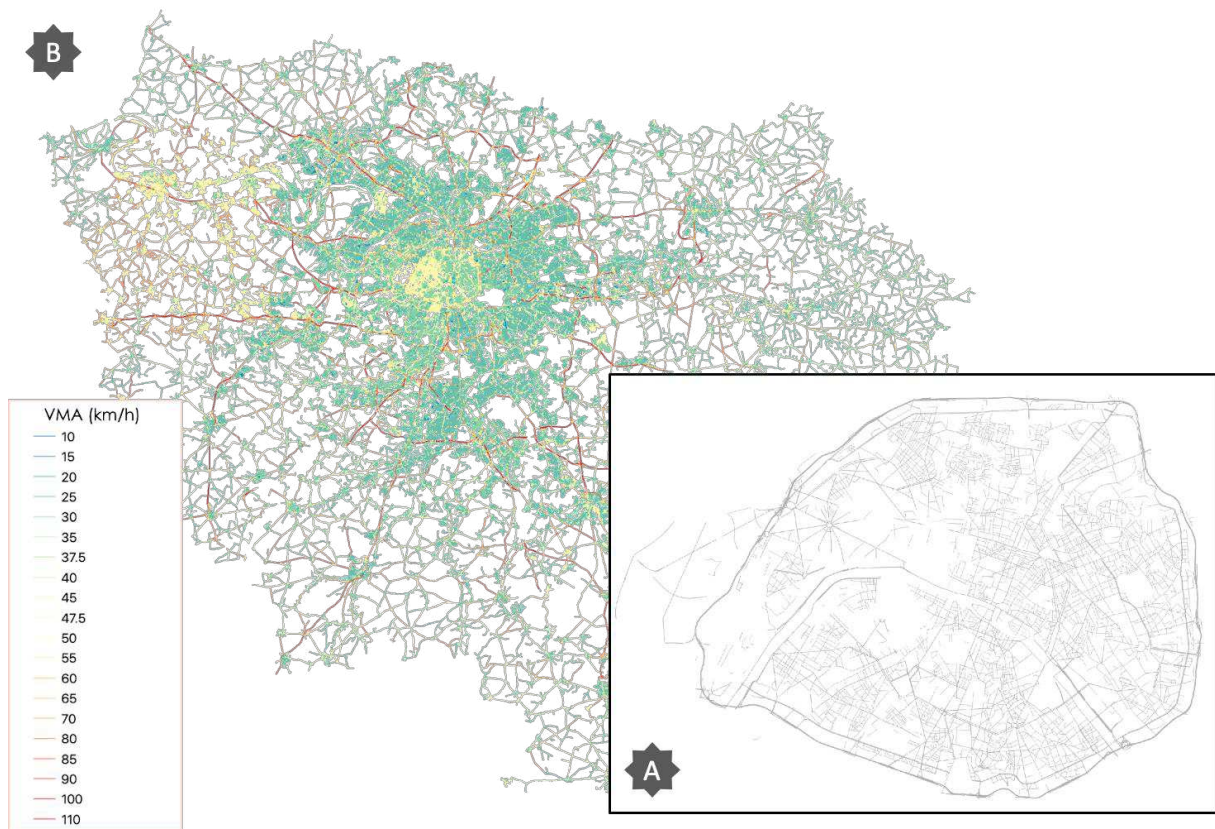


Figure 54 - Réseau OSMnx avec des VMAs manquantes au niveau de la ville de Paris (A) et complétées sur l'ensemble du réseau (B).

J'ai enfin intégré ce réseau au modèle OLYMPUS et ai déjà lancé des simulations « tests » avec cette configuration. Avant de pouvoir l'utiliser, il est nécessaire de mener un travail d'évaluation des résultats du modèle afin de vérifier que les statistiques de mobilité et de répartition modale simulées sont correctes. Les simulations prennent beaucoup plus de temps, et une réflexion sur un réseau intermédiaire (simplifié en grande couronne) doit être menée.

VI. Conclusions

J'ai développé, dans cette étape de mon travail de thèse, une méthodologie pour l'estimation des expositions des individus d'une région, considérant les déplacements effectués et les lieux visités par tous les individus au cours de leur journée. Si cette démarche souffre encore de nombreux verrous de connaissance (de la qualité de l'air intérieure et dans les transports), j'ai pu démontrer la faisabilité de la démarche et l'intérêt d'une approche de l'exposition liée à chaque individu. Nous sommes encore au début d'une démarche qui peut intéresser des épidémiologistes autant que des services d'aménagement urbain, et dans laquelle on voit déjà que les valeurs de concentrations dans les habitacles des véhicules particuliers, dans les transports en commun ferrés et routiers, et dans les lieux d'activité professionnelle ou domestique joueront un très grand rôle.

Une autre problématique qui est apparue lors de ce travail est l'inadéquation du réseau utilisé pour le calcul de mobilité d'OLYMPUS. J'ai initié le travail de remplacement du réseau routier dans le modèle OLYMPUS et les premiers tests de simulations montrent que la prise en compte du nouveau réseau par le modèle, ainsi que les calculs d'itinéraires, se font correctement. Un travail de calibration et

d'évaluation de cette configuration du modèle est en cours d'exécution. Il requiert de vérifier voire de recalculer les fonctions d'utilité qui permettent au modèle, lors des simulations, de déterminer les choix modaux des individus.

Chapitre IV

Mise en place et analyse d'un scénario d'aménagement urbain – Etude d'impact sur la qualité de l'air de l'implantation de Grand Paris Express dans le quartier de Créteil l'Échat

Table des matières

I.	Contexte	142
II.	Présentation de MUNICH et choix de son intégration dans la chaîne de modélisation	144
III.	Modification des données d'entrées d'Olympus	145
1.	Modification de la voirie	145
2.	Ajout des stations du GPE au réseau GTFS	147
IV.	Simulations OLYMPUS : Comparaison de la situation de référence et du scénario	148
V.	Simulations CHIMERE : Comparaison de la situation de référence et du scénario	149
1.	Comparaison des données de la référence et du scénario au niveau du quartier de gare..	149
2.	Comparaison des cartographies issues de l'approche statistique	150
VI.	Préparation des simulations MUNICH	152
3.	Fichiers d'entrés du modèle	152
4.	Verrous pour le lancement des simulations	153
VII.	Conclusion	153

I. Contexte

L'objectif de ce chapitre de thèse est de réaliser un exercice de diagnostic de la qualité de l'air sur un scénario impliquant des modifications de l'aménagement urbain, de la mobilité et des politiques publiques, réalisé dans le cadre de la collaboration avec le Département des Transports, de la Voirie et du Développement (DTVD) du Conseil Départemental du Val-de-Marne (CD94). Leur but premier est la production de science aux niveaux des quartiers de gare du Grand Paris Express en vérifiant si d'éventuels impacts sur la qualité de l'air sont observés à la suite de la mise en activité de ces dernières. À la suite de nombreuses discussions avec le CD94, le quartier de gare de Créteil l'Échat a été choisi comme zone d'étude du scénario (dont la zone géographique est répartie sur un rayon de 800m autour de la station de métro actuelle, et est illustrée dans la [Figure 55](#)). C'est ce quartier qui a été choisi parmi les différents quartiers en transformation, car il représente un des principaux pôles d'activité et d'emploi du département du Val-de-Marne (avec une densité de l'emploi comprise entre 15 000 et 28 000 par IRIS) et son taux d'urbanisation est très élevé (notamment grâce au mélange de ses pôles d'activités et de résidences). Aujourd'hui, 16 000 personnes travaillent dans ce quartier dont environ 8500 travaillant dans les services publics.

Les services du CD94 dans lesquels j'ai conduit ce projet ont pour but commun d'améliorer la qualité de l'air des quartiers de gare du Grand Paris Express (GPE) se trouvant sur leur territoire, en amenant les individus à reconsidérer leur mobilité, en la réalisant via des modes de transport moins polluants que la voiture. Pour cela, ils travaillent à remodeler l'architecture urbaine du quartier de gare de Créteil l'Échat, située le long de la future ligne 15 Sud, afin de contraindre l'utilisation de la voiture et favoriser l'utilisation des modes de déplacements doux. Le Plan de Déplacement du Val-de-Marne cite, en particulier, des objectifs de répartition de choix modaux pour ces quartiers à l'horizon 2030 : 20% d'utilisation des véhicules particuliers, 50% d'utilisation des modes de déplacements doux et 30% d'utilisation des transports collectifs. La favorisation des modes doux sera, en particulier, aidée par l'aménagement de parkings pour vélo qu'ils planifient d'installer à proximité de la gare pour faciliter les intermodalités. De nouveaux bâtiments seront construits à proximité de la gare pour accueillir des bureaux supplémentaires et, ainsi, augmenter l'activité de la zone.

L'objectif de ce chapitre de thèse était de répondre aux questionnements du CD94 via la mise en place d'un scénario d'aménagement urbain : l'objectif était de produire un diagnostic sur la future mobilité interne au quartier et sur l'amélioration de la qualité de l'air que l'on pouvait attendre, à l'horizon 2030. Cependant, dès le milieu de ma thèse, la crise sanitaire du COVID-19 a conduit à une modification des priorités du conseil départemental ce qui a considérablement freiné mes échanges avec eux. De fait, la mise à disposition prévue d'un scénario complet sur lequel je devais travailler n'a pas pu se faire, et je n'ai pu accéder qu'aux informations sur l'évolution des parts modales au niveau du département et qu'à des plans de restructurations du réseau routier. Je n'ai donc pu mener la construction d'un scénario que sur ces bases-là.

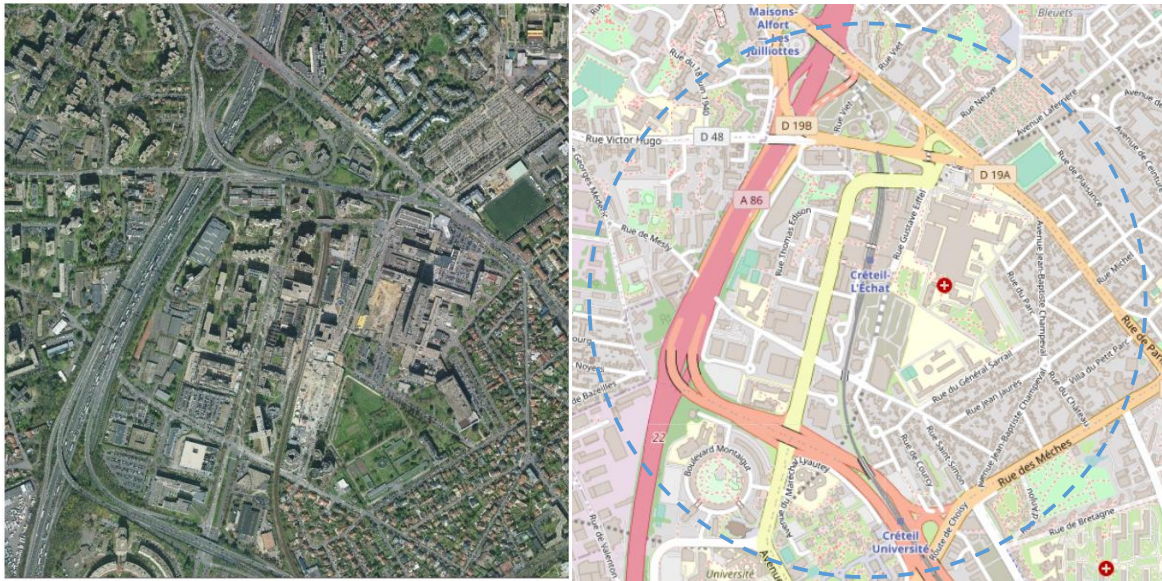


Figure 55 - À gauche, vue satellite du quartier de Crétell l'Échat. À droite, délimitation du quartier de gare de Crétell l'Échat.

Depuis le début de mon travail de thèse, deux options étaient considérées pour simuler la qualité de l'air dans le scénario du CD94. La première était d'utiliser l'approche de descente en échelle sur laquelle je travaillais, la seconde de mettre en œuvre une approche plus déterministe, de type modélisation urbaine, qui permettrait de prendre en compte explicitement la structure du quartier et de son bâti. Même si elle nous permet de retrouver des niveaux réalistes de pollution par rapport aux observations, l'approche de descente en échelle n'a pas pu être connectée à la structure du bâti : elle ne prend donc pas explicitement en compte l'influence du bâti et de l'orientation des rues sur l'accumulation ou la dispersion de la pollution locale. Or, nous estimons qu'il s'agit d'un paramètre important dans la précision de notre scénario. Cela sera encore plus déterminant lorsque nous aurons les données sur la modification du bâti dans le scénario. Nous avons donc opté pour l'utilisation d'un modèle urbain en sortie des simulations de CHIMERE. Malheureusement à nouveau, la crise sanitaire n'a pas permis d'avancer sur les collaborations que nous avons initiées en cours de thèse. En revanche, j'ai eu l'opportunité de pouvoir échanger et collaborer avec le CEREVA en cours d'année 2021 autour de l'utilisation de leur modèle urbain MUNICH, présenté ci-après. Le retard accumulé durant cette période ne m'a toutefois pas permis de pousser au maximum le développement de cette partie du scénario. Je ne présente ici que la configuration du scénario et non sa mise en œuvre.

Nous avons décidé de réaliser deux séries de simulations, sur l'année 2017, pour évaluer le potentiel impact sur la qualité de l'air de l'ouverture de la gare du GPE :

- Une première simulation dite de référence considère la voirie ainsi que le réseau de transport en commun actuels (identiques à ceux utilisés dans le Chapitre précédent d'estimation des expositions dynamiques de la population) pour la spatialisation de la mobilité des individus et le calcul des émissions dans le modèle OLYMPUS. Le réseau routier actuel, ainsi que le bâti actuel du quartier de gare, sont ensuite utilisés pour les simulations avec le modèle MUNICH.
- La seconde série de simulation considère un réseau routier modifié (suivant les indications du Plan de Déplacement du Val-de-Marne) et un réseau GTFS prenant en compte les stations du GPE. Ainsi, lors des simulations OLYMPUS, seules les valeurs d'émissions sur les routes sont impactées par la présence ou non de la gare et impacteront, par la suite, les simulations de concentrations de fond réalisées avec CHIMERE. Les autres données d'entrées nécessaires aux

modèles WRF et CHIMERE (météorologie, données de landuse) sont, quant à elles, identiques pour les deux situations.

II. Présentation de MUNICH et choix de son intégration dans la chaîne de modélisation

Le fait de mettre en œuvre un modèle urbain pour simuler la qualité de l'air au niveau du quartier de gare permet de produire un diagnostic plus fin que celui que l'on établit de manière statistique sur l'ensemble de la région. D'abord parce qu'il permet de moduler le diagnostic en fonction du bâti environnant - et nous avons démontré qu'il s'agissait là d'une limitation de notre approche - mais aussi parce qu'il inclut la réactivité des polluants à fine échelle et ne se limite pas à la correction de leurs concentrations. C'est donc plus adapté dans ce type d'exercice.

MUNICH¹³ (Model of Urban Network of Intersecting Canyons and Highways) est un modèle de qualité de l'air à l'échelle urbaine développé au CEREa et permet de calculer l'impact du trafic routier sur la qualité de l'air le long des routes et au niveau des intersections. Ce modèle a besoin d'informations concernant les hauteurs des bâtiments, des obstacles bordant les tronçons de routes considérés, du réseau routier du domaine ainsi que des émissions sur chaque brin de route (Figure 56) afin de prendre en compte les phénomènes d'accumulation/dispersion des émissions dans les calculs de concentration.

MUNICH a initialement été développé dans le but d'être connecté aux sorties du CTM Polair3D pour former le modèle Street-in-Grid (SinG) (Kim et al., 2018). Dans les versions actuelles, il peut prendre en compte la chimie des phases gazeuses et particulaires atmosphériques (par couplage avec le modèle ssh-aerosol¹⁴) ou uniquement celle des phases gazeuses (module photochemistry). C'est cette dernière option que j'ai choisie car elle est cohérente avec mes simulations CHIMERE. En effet, des travaux récents réalisés au LMD par Myrto Valari ont permis de connecter MUNICH aux sorties de CHIMERE, en intégrant le schéma chimique MELCHIOR2 au module photochemistry. Par ailleurs, au moment où j'ai commencé à prendre le modèle en main, le module ssh-aerosol n'était pas encore stabilisé dans MUNICH.

Concrètement, la chaîne de modélisation considérée dans le cadre de ce scénario se présente comme suit : (i) le modèle OLYMPUS est utilisé pour calculer les émissions le long des tronçons de route et générer les cadastres d'émissions de trafic, (ii) les champs météorologiques sont simulés par le modèle WRF, (iii) les émissions et les champs météo sont lus par CHIMERE qui simule les concentrations de fond urbain des espèces d'intérêt et (iv) ces valeurs de fond ainsi que les champs météo et les émissions le long des tronçons de routes sont lus par MUNICH pour calculer les concentrations des polluants à l'échelle des rues. L'usage du schéma chimique MELCHIOR2 et l'utilisation des concentrations de fond de CHIMERE étaient opérationnels dans la version du modèle qui m'a été fournie, mais il m'a fallu adapter certains des scripts de MUNICH afin de pouvoir lire et utiliser les données d'OLYMPUS.

¹³ <http://cerea.enpc.fr/munich/doc/munich-guide-v2.pdf>

¹⁴ <http://cerea.enpc.fr/ssh-aerosol>



Figure 56 - Cartographie des surfaces du quartier de gare (marron), de sa voirie (issue d'OSMnx en orange) et de l'emplacement de la gare (étoile rouge). Les mailles CHIMERE chevauchées par le quartier de gare sont représentées en hachuré dans le fond.

Comme je l'ai mentionné, la collaboration avec le CEREAs'est effectuée assez tardivement dans ma thèse et je n'ai pas pu réaliser les simulations de qualité de l'air à fine échelle avec MUNICH. La difficulté principale que j'ai rencontrée est liée au fait que le réseau routier considéré jusqu'ici dans les simulations d'OLYMPUS - et précédemment présenté comme le « réseau LVMT » - ne me permet pas de pouvoir réaliser des simulations cohérentes pour MUNICH étant donné que ses tronçons de routes sont trop simplifiés, et que certains traversent le bâti. Néanmoins, suite à plusieurs réunions avec l'équipe de développement du modèle, j'ai pu préparer la grande majorité des fichiers d'entrée nécessaires aux simulations basées sur le module photochemistry.

Ma première cible concernait l'évolution des niveaux de NO_2 dans le quartier de Créteil l'Échat. Les fichiers que j'ai eu à préparer sont : (i) les fichiers d'identification des tronçons de routes contenant la localisation géographique de leurs extrémités, (ii) les fichiers d'émissions le long des tronçons de route, (iii) le fichier contenant les niveaux moyens de concentrations simulées par CHIMERE au niveau du quartier de gare et (iv) sa météo.

III. Modification des données d'entrées d'Olympus

1. Modification de la voirie

J'ai sollicité le département du Val-de-Marne pour avoir des informations et/ou des documents sur les modifications prévues pour la voirie à proximité de la gare. Les données que j'ai pu récupérer sont présentées sous forme de plan des travaux à venir et sont résumés dans un document présentant en détail les modifications de l'aménagement du quartier (considérant les éléments relatifs à la mobilité, le stationnement et les pôles d'emploi.

Le but du réaménagement de la partie nord du quartier de gare, appelée le « triangle de Créteil l’Echât » est d’ajouter 3 nouveaux tronçons de routes au nord-ouest du triangle pour relier la rue des bleuets (au nord) avec l’avenue du Général de Gaulle (au sud) et d’en retirer deux sur sa partie est qui menaient vers la rue Gustave Eiffel. La construction de nouvelles surface bâtie est aussi prévue dans cette zone et est décrite dans la [Section VI.1](#).

Pour le moment et comme expliqué dans le chapitre précédent, les simulations de qualité de l’air ont été réalisées en considérant le réseau routier LVMT car le modèle n’a pas encore été calibré avec celui que j’ai créé via les données d’OpenStreetMap. Pour cela, je me suis basée sur les tracés présentés dans ces plans pour modifier manuellement le fichier SIG de la voirie actuelle grâce au logiciel ArcGIS. Mais le réseau LVMT n’est pas représentatif de la voirie réelle et, de ce fait, les modifications que j’y ai apportées sont restées approximatives. J’ai ajouté manuellement 3 nouveaux tronçons de routes (présentés en rouge dans la [Figure 57](#)). Ensuite, j’ai forcé la répartition modale du scénario pour correspondre aux projections du département du Val-de-Marne citées dans leur plan de Déplacement et résumées dans le [Tableau 15](#). Pour cela, j’ai trié dans le modèle tous les déplacements ayant une des communes du Val-de-Marne comme points d’origine et/ou de destination et ai forcé la répartition modale des déplacements en VP, TC et mode doux tout en laissant celle des autres départements telles qu’elles ont été calculées par les fonctions d’utilités d’OLYMPUS avec le réseau LVMT.

J’ai, quand même, déjà préparé la voirie prospective du réseau OSMnx pour que les fichiers soient déjà disponibles lorsque ce réseau sera intégré entièrement au modèle OLYMPUS. Il a été plus simple de réaliser les modifications ([Figure 57](#)) car les voies de ce réseau superposent celles de la réalité. Au total, j’ai eu à ajouter 3 nouveaux tronçons de route au réseau et à en supprimer deux. Cette façon de procéder, bien qu’assez approximative, m’a permis de générer un fichier dans le format qui pourra être lu par le modèle. J’ai attribué à chaque nouveau brin de nouveaux identifiants et ai eu à calculer leurs longueurs pour en déduire les temps de déplacements.

Tableau 15 – Parts modales décrites dans le Plan de Déplacement du Val-de-Marne (en %)

	TC	Vélo	Marche	2 RM	Voiture
Estimations 2016	23	1,7	41	2	32,3
Objectif 2030	30	9	41	2	18
Forçages OLP considérés	30	50		20	

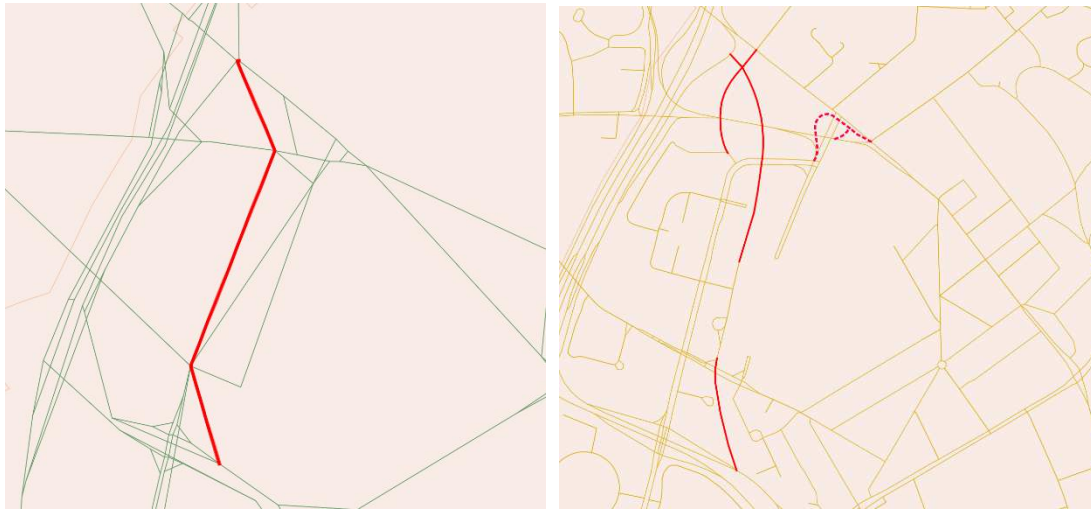


Figure 57 - À gauche, ajout des trois nouveaux tronçons de route au réseau LVMT (en rouge). À droite, ajout des trois nouveaux tronçons (en rouge plein) et suppression de deux tronçons du réseau OSMnx (en rouge pointillé).

Le CD94 souhaite également restructurer la circulation autour de la gare en modifiant les sens de circulations de certaines voies afin qu'elles deviennent des voies à sens unique. L'objectif est de simplifier et à terme limiter la circulation en véhicule particulier autour de la gare. Toutefois je n'ai pas eu accès à ces informations et n'ai donc pas considéré cette modification.

2. Ajout des stations du GPE au réseau GTFS

La Société du Grand Paris (SGP) a mis à disposition des données prospectives des stations du GPE sur leur portail de données. Cela inclut la localisation des gares¹⁵, les temps de correspondances¹⁶ avec les autres stations de transport en commun et les temps de parcours intergares¹⁷.

Les données fournies par la SGP ne sont toutefois pas totalement abouties. J'ai pu constater que de nombreux temps de parcours ne sont pas renseignés et que les listes de stations données pour les correspondances ne sont pas complètes. En particulier, dans le cas de la station de gare de Créteil l'Échat, le document ne cite qu'une seule correspondance possible avec la station de métro de la ligne 8. Or, actuellement, à l'extérieur même de la station de métro, il y a 5 lignes de bus dont les arrêts se trouvent à moins de 100 m de la gare (172, 281, O1, O2 et B) et dont on peut bien sûr penser qu'elles constitueront des correspondances stratégiques. Pour compléter les informations manquantes sur les correspondances autour de toutes les gares du GPE, j'ai donc cartographié les localisations des stations de transport en commun actuelles (GTFS) et du GPE. J'ai ensuite tracé une zone tampon de 500 m autour de chaque gare du GPE et ai extrait tous les arrêts de transport en commun qui s'y trouvent. J'ai calculé la distance entre les gares et les autres arrêts et, en considérant une vitesse de déplacement moyenne d'un individu de 6 km/h, j'ai pu en déduire des temps de correspondances. Par la suite, j'ai mis les résultats sous forme de tableau pour ajouter ces nouvelles informations au fichier GTFS transfer.txt initial.

¹⁵ <https://www.data.gouv.fr/fr/datasets/point-de-localisation-des-gares-du-grand-paris-express/>

¹⁶ <https://www.data.gouv.fr/fr/datasets/temps-de-correspondance-previsionnels-dans-les-gares/>

¹⁷ <https://www.data.gouv.fr/fr/datasets/temps-de-parcours-intergares-previsionnels/>

IV. Simulations OLYMPUS : Comparaison de la situation de référence et du scénario

La répartition modale obtenue en fin de simulation OLYMPUS pour la situation de référence et pour le scénario Val-de-Marne est présentée dans le [Tableau 16](#).

Tableau 16 – Répartition modale (en %) de la population simulée pour la situation de référence et pour le scénario Val-de-Marne sur l'ensemble de l'Île-de-France et sur le Val-de-Marne.

	Référence		Scénario	
	Île-de-France	Val-de-Marne	Île-de-France	Val-de-Marne
Véhicules particuliers	30	50	37,2	49,6
Transports collectifs	22,6	33,9	19,6	34,1
Modes doux	47,4	16,1	43,2	16,3

Le résultat que l'on obtient avec OLYMPUS apparaît surprenant. Lorsqu'on analyse les résultats pour l'ensemble de la région Île-de-France, l'ajout des stations du GPE au réseau de transport en commun ne modifie en effet que très peu les répartitions modales, et le sens des écarts obtenus est même inverse à celui escompté.

La part modale des transports collectif diminue ainsi de 3% et celle des modes doux de 4%, le tout au profit d'une augmentation de l'utilisation des véhicules particuliers de 7%. L'augmentation du nombre de stations de métro et la facilitation des déplacements entre les différentes banlieues de la petite couronne auraient, en théorie, dû favoriser une augmentation de l'utilisation des transports collectifs par la population se déplaçant sur ces territoires. Ce résultat est en fait dû à un effet rebond. L'augmentation du nombre de stations de métro amène un nombre important d'individus de se reporter vers les transports en commun. Mais alors que cela diminue la congestion routière, cela amène une autre partie de la population - pour laquelle l'utilisation des TC était surtout forcé par la congestion routière sur leur chemin habituel - de privilégier un déplacement en VP. Cet effet a été simulé et décrit par [Coulombel et al., 2019](#) dans le cadre d'un scénario de covoiturage ayant induit une augmentation du trafic routier, à cause de la place libérée sur les voies.

Sur le département du Val-de-Marne, cet effet d'augmentation du trafic n'est pas observé : on constate une légère diminution de l'utilisation des véhicules particuliers (à hauteur de 0,4%) au profit des transports collectifs et des modes doux ([Tableau 15](#)). Cela ne veut pas dire qu'il n'y a pas de changement modal, mais *a minima* qu'ils se compensent.

Pour de futures simulations, on peut imaginer deux façons de conduire notre scénario CD94 tel qu'il a été voulu et non tel qu'il pourrait être s'il amenait des effets rebond.

- La première façon est de simuler l'augmentation de l'utilité des transports pour l'ensemble de la population francilienne. En effet, le scénario que j'ai produit ne s'appuie que sur l'amplification du réseau de TC francilien via l'ajout de nouvelles lignes, et sur les nouvelles connexions créées autour du quartier de l'Échat. Il n'intègre pas les nouvelles connexions qui seront créées autour de toutes les nouvelles gares de banlieue du GPE, ni l'aménagement de ces nouveaux quartiers de gare qui seront également pensés pour rendre la mobilité douce, la

mobilité en TC et la multi modalité plus efficaces et plus agréables pour les individus, et rendre les VP moins utiles.

- La seconde façon est d'accompagner la simulation en prescrivant une augmentation de l'utilité des TC, afin de restituer l'effet de paramètres non codés dans OLYMPUS (trains plus confortables, correspondances bien agencées...) ou de restituer des changements de perception des individus (prise de conscience de l'importance de limiter l'usage des véhicules). On peut également imaginer toucher au coût du transport qui pourrait pencher en faveur des TC du fait d'une politique volontariste pour devenir plus compétitif face au coût des véhicules particuliers.

V. Simulations CHIMERE : Comparaison de la situation de référence et du scénario

1. Comparaison des données de la référence et du scénario au niveau du quartier de gare

J'ai réalisé les simulations pour ce scénario sur les 1 000 premiers pas de temps de l'année 2017 (ce qui équivaut à 42 jours de simulation). Le quartier de Créteil l'Échat s'étale sur deux mailles CHIMERE, j'ai donc extrait et moyenné leurs concentrations et je présente, dans les deux figures comparatives suivantes (Figure 58), les différences de concentrations entre la situation de référence et le scénario Val-de-Marne pour NO₂ et les PM₁₀. Cette figure représente les concentrations brutes de CHIMERE, moyennée entre les deux mailles, en gris. J'ai aussi appliquée la correction *Cprox* à ces valeurs pour pouvoir mieux évaluer de façon plus robuste les différences de concentrations entre les deux situations.

On voit, sur la Figure 58, des différences de concentrations qui restent négligeables entre les deux situations pour NO₂ (0,3 µg/m³ au plus) et des différences plus marquées pour les PM₁₀ (jusqu'à 5 µg/m³ de différence pour le signal corrigé). Qu'ils soient obtenus par CHIMERE ou par *Cprox*, les résultats sont d'amplitude et de signe très similaires. La différence étant réalisée dans le sens [Référence - Scénario] on peut conclure que les concentrations simulées sont presque toujours plus faibles pour les PM₁₀ dans le scénario du Val-de-Marne que dans la référence. Cet effet est trop partagé entre valeurs positives et valeurs négatives pour pouvoir en tirer des conclusions sur NO₂. Ces faibles différences ne sont pas étonnantes compte tenu de la forte inertie des choix modaux sur le département du Val-de-Marne (Tableau 16).

Tel qu'il est simulé par OLYMPUS, avec la possibilité d'un effet rebond sur toute l'Île-de-France, l'implémentation de la station l'Échat du GPE et la modification de l'aménagement du quartier n'amènent donc pas de changement significatif dans la qualité de l'air locale.

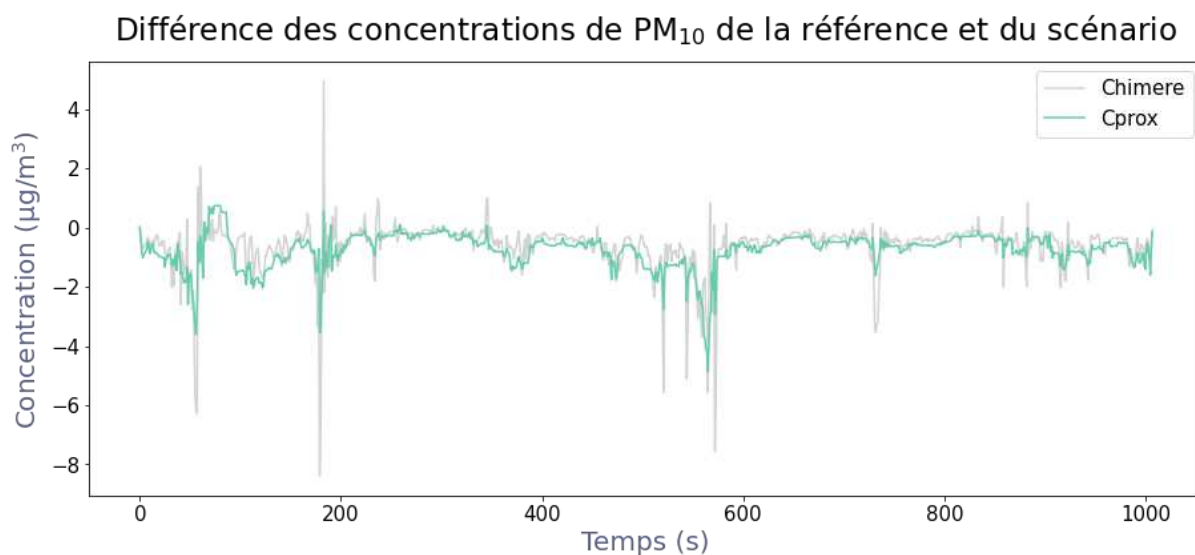
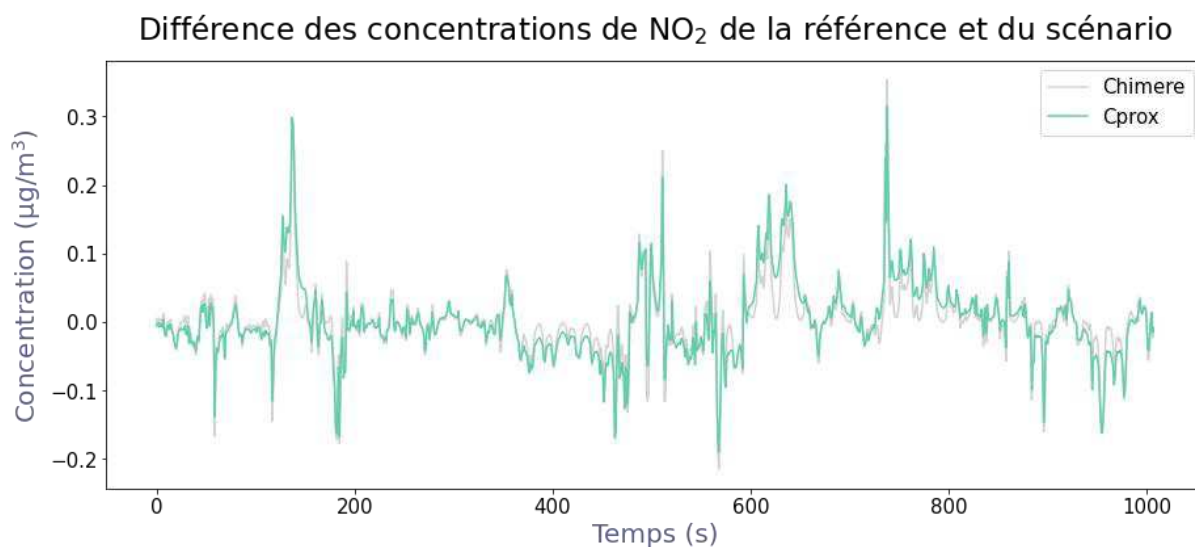


Figure 58 - Différences entre les niveaux de fond simulés pour la situation de référence et pour le scénario Val-de-Marne pour NO₂ (en haut) et les PM₁₀ (en bas), et calculées dans le sens [Référence - Scénario]. Elles sont présentées pour les sorties brutes de CHIMERE (en gris) et pour les sorties corrigées par l'approche statistique Cprox (en vert).

On voit une augmentation des concentrations de PM₁₀ pour le scénario et l'impact sur ces concentrations de fond semble être lié à l'augmentation de 7% des déplacements en véhicules particuliers de la population (au niveau de la région), seules sources d'émissions évoluant entre les deux situations. Même si cette augmentation des déplacements n'est observée qu'au niveau de la région (stagnation au niveau du département), le fait que les PM₁₀ soit un polluant régional montre l'importance d'avoir à faire évoluer l'intégralité des habitudes de mobilité de la population francilienne pour pouvoir, *in fine*, diminuer leurs niveaux de concentrations.

2. Comparaison des cartographies issues de l'approche statistique

Afin de pouvoir analyser spatialement les résultats du scénario, j'ai cartographié les différences de concentrations de NO₂ et de PM₁₀ sur le quartier. Les concentrations que j'ai tracées sont celles issues de Cprox, calculées le long des routes du quartier de Créteil l'Échat et non celles issues directement de CHIMERE (Figure 59).

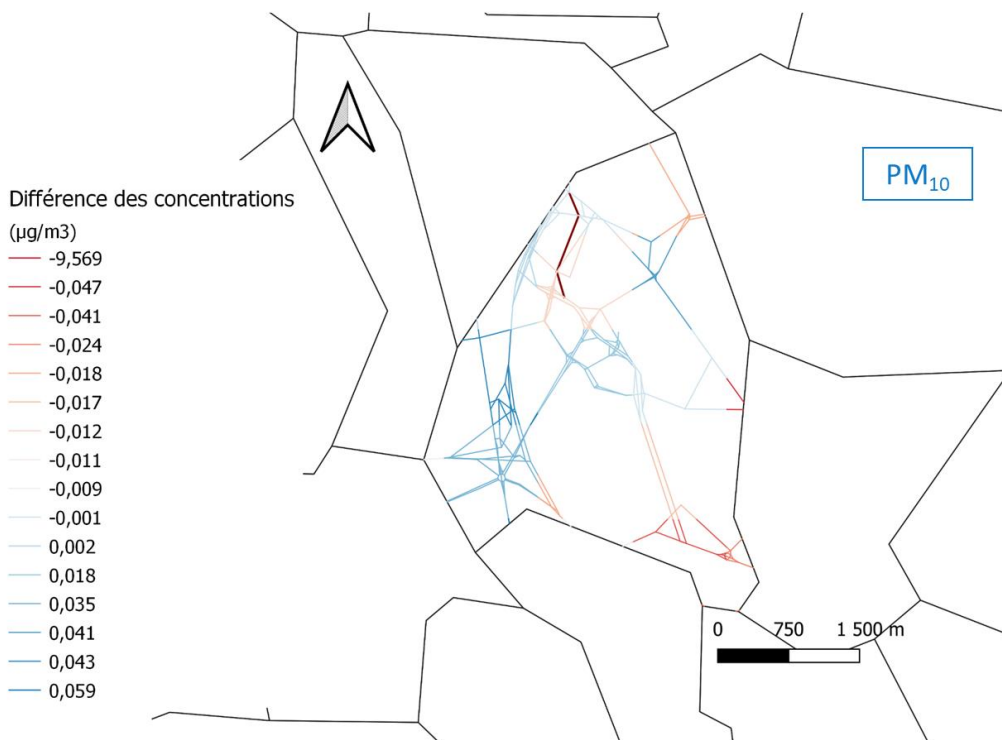
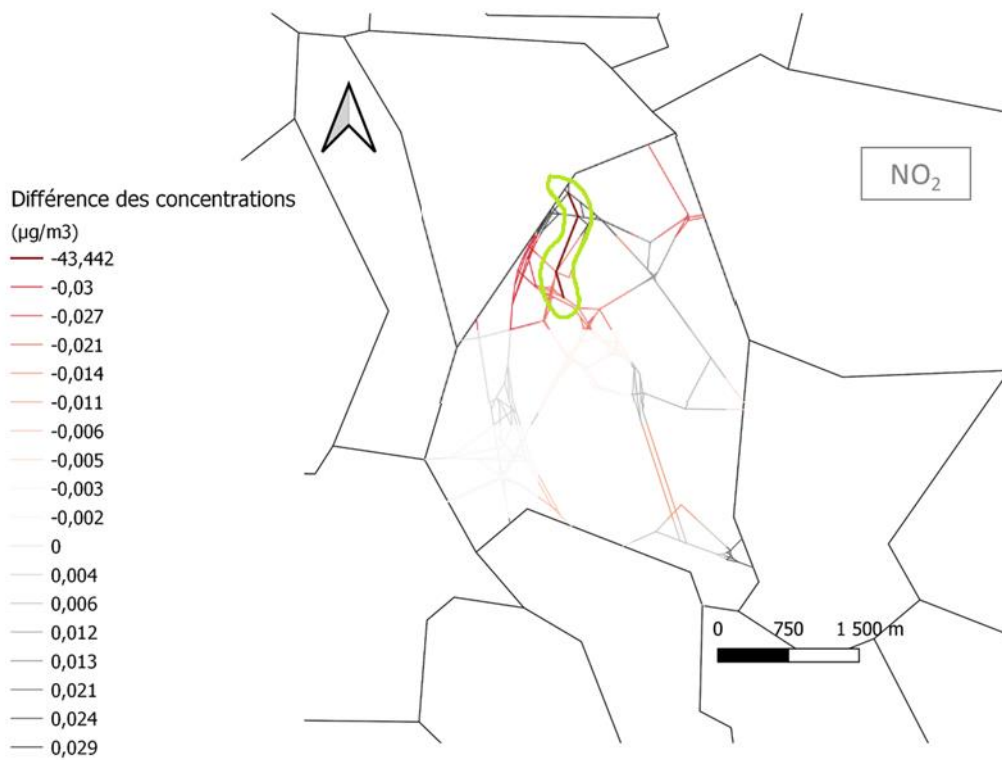


Figure 59 - Cartographies des différences de concentrations de NO₂ (en haut) et des PM₁₀ (en bas) entre la situation de référence et le scénario le long des routes du quartier de Créteil l'Échat.

En dehors des différences de concentrations observées au niveau des trois tronçons de routes ajoutés au réseau routier (en rouge gras et entouré en vert sur la [Figure 59](#) - au nord-ouest de la commune de Créteil), on peut remarquer que les concentrations diffèrent très peu entre les deux situations (différences $< 0,1 \mu\text{g}/\text{m}^3$ pour les deux polluants) dénotant le faible impact du scénario sur la qualité de l'air à forte proximité des routes. Le scénario a plus d'impact, en terme de qualité de l'air, sur les concentrations de fond urbain plutôt que celles à proximité du trafic routier avec une tendance à la dégradation de la qualité de l'air liée aux particules ([Figure 58](#)). Cette dégradation ne peut être liée qu'aux forçages de mobilité, au niveau de la région ou du département, car la seule donnée d'entrée ayant variée entre les simulations des deux situations sont les cadastres d'émissions issus d'OLYMPUS.

VI. Préparation des simulations MUNICH

1. Fichiers d'entrés du modèle

MUNICH nécessite plusieurs fichiers d'entrées, sous un format attendu, pour son étape de prétraitement des données :

- Les émissions calculées par le modèle OLYMPUS ne doivent pas être fournies à MUNICH sous forme de cadastres d'émissions mais plutôt comme des totaux d'émissions par polluant pour chaque tronçon de route. Il m'a fallu générer un fichier d'émission par pas de temps considéré et, ensuite, au cours de l'étape de prétraitement, réaliser une conversion de ces fichiers en format binaire.
- Comme pour les émissions, il faut générer un fichier résumant les niveaux moyens de NO_2 , NO et O_3 pour chaque pas de temps des simulations, ce que j'ai produit.
- Les réseaux routiers considérés dans le modèle MUNICH sont des versions zoomées sur un rayon de 800m autour du point de localisation de la gare comme il l'est illustré dans la [Figure 2](#).

Il n'y a pas de travail à effectuer pour les fichiers issus de WRF et doivent être utilisés tels quels.

Les simulations avec MUNICH nécessitent de fournir des données sur la surface bâtie du quartier, que j'ai récupéré dans la BDTOPPO. Comme pour le réseau routier, il faut adapter les surfaces bâties pour les simulations du scénario. Dans ce but, j'ai étudié l'évolution des surfaces bâties du quartier. Actuellement, il compte 1 539 logements.

Le document présentant les modifications de l'aménagement du quartier de Créteil l'Échat présente trois projets d'aménagements immobilier :

- Au niveau de l'ancien parking de la station de la ligne 8 (au nord du quartier) : il est prévu de créer 150 logements et 270 logements étudiants/de tourisme, 6 000 m^2 de bureaux (ce qui équivaut à 330 nouveaux salariés) ainsi qu'une crèche et un restaurant.
- Au nord du quartier (au niveau du « triangle de l'Échat ») : ils projettent de créer 1 300 logements avec près de 200 logements étudiants, 7 000 m^2 d'activités hôtelières, 24 000 m^2 de bureaux, 2 000 m^2 de commerces ainsi que d'ajouter un groupe scolaire et une crèche.
- Ensuite, à proximité de l'Hôpital H. Mondor (au sud du quartier) : des projets de constructions ont été établis avec la création de 550 logements, 12 000 m^2 qui seront attribués à des activités et 4 000 m^2 pour des équipements et des services.

L'agencement des nouveaux bâtiments n'est illustré dans les documents qui m'ont été fournis, que pour la partie nord du quartier. De ce fait, j'ai ajouté les nouveaux bâtiments, pour lesquels j'avais des informations, de la même façon que pour les modifications que j'ai apportées à la voirie (grâce à ArcGIS) et j'ai laissé telle quelle la partie sud du quartier par manque de données. J'ai estimé la hauteur des

nouveaux bâtiments en me référant à des figures présentées dans le Plan de Déplacement du Val-de-Marne qui illustrent, approximativement, leurs hauteurs par rapport aux bâtiments alentours déjà existant (pour lesquels les données se trouvent dans la BDTOPO). Il s'agit là d'un travail très approximatif mais j'ai dû m'en contenter puisque je n'ai pas pu avoir de plans futurs sur les surfaces bâties.

2. Verrous pour le lancement des simulations

Au cours de ce travail de scénarisation, j'ai rencontré plusieurs verrous techniques qui ne m'ont pas permis de mener jusqu'à leur terme réaliser les simulations de qualité de l'air à fine échelle avec MUNICH.

Au-delà des différences de réseau routier entre OLYMPUS et MUNICH, j'ai rencontré des problèmes de compatibilité informatique. En effet, j'ai cherché à mettre en œuvre mes simulations sur les serveurs du TGCC (Très Grand Centre de Calcul), serveurs sur lesquels j'ai l'habitude de réaliser les simulations de OLYMPUS, CHIMERE et WRF. Travailler ces serveurs permet de pouvoir utiliser une puissance de calcul importante ce qui réduit les temps totaux de simulations. Cependant, une partie de la version de MUNICH qui m'avait été fournie était codée en python2 et cela a généré plusieurs problèmes de compatibilité avec les bibliothèques du TGCC. Il m'a été conseillé d'utiliser python3 sur ces serveurs mais le travail de conversion du modèle était, à ce moment-là, toujours en cours de réalisation au CEREAS. J'ai donc pu avoir accès aux serveurs du CEREAS afin de lancer MUNICH dans sa version python2 mais le retard pris sur ce volet du travail ne m'a pas permis de pouvoir m'y atteler dans les délais.

VII. Conclusion

J'ai construit, dans ce chapitre de thèse, une procédure pour la simulation de l'impact de l'implémentation de gares du Grand Paris Express, sur la qualité de l'air dans le quartier considéré. Cette scénarisation inclut l'amplification de l'offre de TC et de sa connexion avec les réseaux existants, la prise en compte de l'évolution du bâti pour la dispersion des masses d'air et la modification du réseau routier local. Ce type de simulation doit permettre d'apposer un diagnostic de l'évolution de la qualité de l'air sur un projet d'aménagement du territoire.

Mon cas d'étude concernait l'ouverture de la station de la ligne 15 Sud du Grand Paris Express dans le quartier de Créteil l'Échat. Cet exercice cherche à répondre aux questions du Conseil Départemental du Val-de-Marne par la production d'une expertise sur les potentiels impacts, positifs comme négatifs, de l'installation de la gare et de la modification des modes de vie des habitants du quartier et des usagers des lignes de transports en commun. En effet, ces nouvelles lignes de métro visent à faciliter les déplacements inter-banlieues sans avoir à transiter par la ville de Paris et à diminuer la place de la voiture dans la mobilité quotidienne des franciliens. En conséquence, elles doivent permettre de désengorger les routes et de diminuer les niveaux de concentrations de polluants atmosphériques à proximité des voies, tout en augmentant l'offre de logement en Île-de-France. Le CD94 vise notamment à profiter de l'arrivée de cette gare pour favoriser les déplacements des Val-de-Marnais en transports collectifs et/ou en mode doux. Ils souhaitent par exemple ajouter des parkings à vélo à proximité de la gare pour inciter la population à recourir à la multi modalité plutôt qu'à utiliser leurs véhicules.

Initialement, afin de répondre à cet objectif, j'ai décidé de mettre en œuvre la chaîne de modélisation utilisée dans l'ensemble de mon travail de thèse (WRF-OLYMPUS-CHIMERE) et d'y ajouter une étape de simulation de la qualité de l'air à l'échelle de la rue grâce au modèle urbain MUNICH. Le but des simulations avec MUNICH est de pouvoir considérer les effets d'accumulation et de dispersion des polluants dû aux bâtiments environnants. Si ce point n'est pas fondamental dans une approche de calcul

statistique de l'exposition, il devient limitant lorsque l'on cherche à connaître l'effet du réaménagement d'un quartier sur l'exposition des habitants ou usagers de la voirie. Néanmoins, la pandémie du COVID-19 ainsi que des restructurations internes au sein du CD94 ont conduit à un fort retard dans le développement du scénario et ne m'ont pas permis de pousser à bout les simulations de la qualité de l'air avec MUNICH. J'ai, cependant, pu préparer la grande majorité des données nécessaires pour pouvoir avancer dessus dans un temps futur.

Dans un premier temps, j'ai pu utiliser les objectifs de mobilité cités dans le Plan de Déplacement du Val-de-Marne pour forcer la répartition modale des déplacements au sein du département en plus de la modification des réseaux de transport en commun et de la voirie du modèle OLYMPUS. Les modifications du réseau de transport collectif permettent de prendre en compte les nouvelles stations de métro du GPE et les correspondances qui y sont associées dans les déplacements quotidiens des franciliens. J'ai aussi modifié la voirie et les surfaces bâties autour de la station de gare en fonction des plans prospectifs de l'aménagement du quartier pour préparer les futures simulations avec MUNICH.

J'ai ensuite réalisé les simulations jusqu'à l'étape de calcul de la qualité de l'air avec CHIMERE et y ai appliqué la correction statistique pour caractériser les niveaux le long des routes, ce qui constituait un futur de comparaison avec MUNICH et qui me permettrait d'évaluer la qualité de mon approche statistique. Les résultats ont en fait montré très peu d'évolution des parts modales - que ce soit sur le département ou sur la région. Ces résultats sont probablement attribuables au faible gain en utilité des TC qui doit être repensé dans OLYMPUS pour restituer des phénomènes de gain en confort, ou encore l'accessibilité réduite à des parkings pour les VP, et/ou à une compensation des reports entre TC et VP sur l'ensemble des individus simulés.

On a pu constater une augmentation des niveaux de concentration de fond de particule pour le scénario par rapport à la référence. La pollution aux particules est une pollution régionale et cette augmentation peut être expliquée par l'augmentation des déplacements en véhicules particuliers entre les deux situations qui conduit donc à l'augmentation des émissions. Ce phénomène montre l'importance de devoir faire évoluer les habitudes de mobilité de l'ensemble de la population de l'Île-de-France pour diminuer, en particulier, les niveaux de fond des particules. C'est pourquoi les actions menées pour créer des Zones à Faibles Émissions dans la Métropole du Grand Paris et l'imposition de vignettes pour contrôler les véhicules se déplaçant vers et dans Paris et diminuer les émissions de PM₁₀ dans l'atmosphère. De plus, limiter les variations de vitesses permettra de limiter les émissions indirectes des PM₁₀ via leurs remises en suspension dans l'atmosphère.

Ce scénario fait partie d'un ensemble de scénarios fondamentaux visant à améliorer les conditions de mobilité des habitants vivant le plus en périphérie d'une région telle que la région parisienne, c'est-à-dire organisée autour d'un centre dense, riche en emploi et en zone de loisir, de culture et de commerce. En s'éloignant de ce centre, les communes sont de moins en moins densément desservies par les transports en communs, contraignant la population à dépendre des véhicules particuliers pour se déplacer, et faciliter l'abandon des véhicules particuliers pour une partie - au moins la partie pendulaire - des trajets, constitue un véritable enjeu pour la région. Les premiers résultats de cette étude ont montré qu'il était difficile de s'écarter d'un équilibre modal en grande partie contraint par la congestion routière. Pour tester des scénarios permettant une rupture dans l'usage du véhicule particulier, il serait intéressant de pouvoir également simuler des changements d'habitudes des populations, liés par exemple à un changement de perception des risques environnementaux et sanitaires, mais aussi d'y inclure une proposition de redistribution des pôles d'emploi, d'achat et de loisirs sur les nouveaux quartiers. Cela aurait pour effet d'augmenter l'attractivité locale et possiblement de soutenir une mobilité différente pour les habitants du département.

Conclusion générale

Avec plus de 12 millions d'habitants en 2022, l'Île-de-France est la région la plus peuplée de France. Centrée sur la ville de Paris dont la densité d'habitation dépasse 20 000 habitants par km², la Métropole du Grand Paris est également un pôle d'emploi 30 fois plus dense que la périphérie francilienne. Dans un tel contexte, il est logique que les problématiques relatives à la mobilité, à l'accès aux services, à l'environnement en zone urbaine dense et plus largement à la qualité de vie, soient devenues critiques. La surveillance de la qualité de l'air a mené depuis longtemps à l'identification d'un point noir environnemental majeur : la proximité au trafic routier, où les dépassements du seuil réglementaire annuel de NO₂ perdurent encore aujourd'hui. Mais même au-dessous des seuils, l'exposition chronique des habitants au dioxyde d'azote et aux particules fines (PM₁₀ notamment) reste une préoccupation majeure des habitants qui sollicitent toujours plus d'information à l'échelle de leur rue, de leur quartier pour pouvoir connaître et maîtriser leur exposition. Et c'est bien la question de l'exposition qui doit être soulevée car elle seule permet d'accéder au réel problème de pollution atmosphérique parisien. En effet, si l'on surveille aujourd'hui de près la qualité de l'air et que les approches de modélisation développées à Airparif permettent de reproduire les gradients de concentrations en extérieur, (i) il n'existe pas de suivi individuel de la qualité des environnements atmosphériques auxquels les individus sont successivement exposés durant leur journée et (ii) une partie importante de la journée des franciliens se passe dans des lieux pour lesquels la qualité de l'air est mal ou pas connue (logements, zones de commerce, habitacles, transports en commun, bureaux, parking...).

Il faut donc avancer sur la compréhension de ce qu'est l'exposition d'un individu, afin de pouvoir - dans une deuxième étape - travailler à son amélioration. Cela demande le développement de méthodologies, l'appui sur des outils permettant d'accéder à la fois à l'individu et à la globalité de la population, ainsi que la conduite d'une réflexion sur les besoins en données et en qualité des données, sur la résolution spatiale ou temporelle nécessaire à chaque étape, ou encore sur le sens des données produites. Construire une telle approche a été le cadre de mes travaux de thèse. À partir des outils sur lesquels mon équipe s'appuyait, j'ai identifié plusieurs points qu'il fallait pouvoir maîtriser : l'accès à des champs de concentration fine échelle robuste en termes de moyenne annuelle (la temporalité réglementaire), un état de l'art des connaissances sur les environnements intérieurs ou clos (habitacles par exemple), la modélisation des habitudes de mobilité quotidienne et l'agenda des activités des franciliens - à la fois au niveau de l'individu et intégré sur toute la population, et enfin une méthodologie d'intégration de l'ensemble des cas d'informations dans une plateforme de calculs pour produire des données d'exposition.

J'ai travaillé à implémenter le calcul de l'exposition dans une plateforme de modélisation utilisée au laboratoire. Cette plateforme permet de construire des situations d'occupation des sols, d'habitation de la région et d'organisation des transports et de l'emploi, ainsi que les émissions routières et domestiques qui s'y rattachent. Mon travail se découpe en deux grands chapitres de développement de la plateforme et de ses fonctionnalités, et en un cas d'étude localisé sur un quartier francilien.

I. Augmentation de la résolution spatiale des simulations de la qualité de l'air d'un CTM

Dans ce premier volet de mon travail, je me suis attachée à produire une représentation fine échelle des concentrations extérieures, capable de distinguer la proximité aux voies routières, avec la contrainte - voulue par l'équipe dans laquelle j'ai travaillé - de réfléchir à une démarche statistique qui pourrait s'appliquer à différents scénarios simulés par un CTM.

J'ai réalisé des simulations de la qualité de l'air grâce au modèle de chimie transport CHIMERE, fonctionnant sur l'Île-de-France à une résolution kilométrique et forcé à ses bornes par deux domaines de résolution décroissante fournissant les conditions aux limites sur toute l'Europe de l'Ouest. J'ai analysé les écarts des sorties de modèle avec la mesure, et démontré le besoin d'augmenter l'hétérogénéité des sorties du modèle pour représenter les gradients de concentrations effectivement observés dans les zones urbaines les plus denses. J'ai développé une approche statistique basée sur l'apprentissage de l'erreur du modèle, permettant de corriger le signal pour restituer l'effet de proximité au trafic.

Pour cela, j'ai initialement travaillé avec des données simulées et mesurées pour l'année 2009. J'ai séparé mes jeux de données mesurés en deux groupes : un premier groupe pour l'apprentissage, permettant l'établissement d'une équation de correction par site de mesure, et un second groupe permettant de vérifier la transposabilité de l'approche à d'autres sites urbains. L'équation de correction statistique a été développée en partant du constat que l'erreur affecte à la fois la concentration de fond urbain produite par CHIMERE, et l'intensité des pics de concentration lorsqu'on se rapproche des voies de circulation. La correction est - logiquement - dépendante du polluant considéré. Et le postulat que l'on pourrait sommer ces deux signaux, affectés de coefficients de correction, pour obtenir des valeurs de concentration fine échelle, offrant une variabilité d'échelle plus fine que celle de la maille de CHIMERE. Les proxys retenus pour représenter le bruit de fond urbain en NO₂ et celui en PM₁₀ sont respectivement la moyenne glissante sur 24h et le percentile glissant sur 24h des champs CHIMERE. Les coefficients de pondération ont été déterminés par un algorithme itératif contraint par la recherche d'un compromis entre la diminution des biais sur la moyenne annuelle, l'amélioration de la variance et la RMSE. Des aménagements ont été nécessaires, notamment l'ajout d'une variabilité de la correction en fonction du vent simulé, afin de tenir compte de l'accroissement de l'erreur du modèle avec la vitesse du vent.

La généralisation de cette équation permet de ne plus dépendre de la mesure pour calculer la concentration sur l'ensemble des voies du réseau francilien. Pour cela, j'ai cherché un paramètre discriminant les environnements des stations de mesures corrélé aux coefficients de correction. L'hétérogénéité des émissions dans la maille (plus précisément le rapport entre un proxy des émissions dans la rue et les émissions totales dans la maille) est le paramètre qui présentait la meilleure corrélation avec les valeurs de mes coefficients. Je l'ai utilisé pour deviner les coefficients sur des sites n'appartenant pas à l'apprentissage. La faiblesse de ce proxy est qu'il s'appuie sur les comptages des véhicules sur la voie, que j'obtiens via le modèle d'émission basée sur l'activité, OLYMPUS. Toutefois, les coefficients devinés pour de nouveaux sites ont montré une amélioration significative des critères de qualité (moyenne annuelle notamment, mais aussi RMSE) sur tous les sites de validation. J'ai pu, au fil du développement de l'approche, proposer une équation plus simple pour les stations de mesure de fond urbain, pour lesquelles la condition sur les vitesses de vent n'avait pas besoin d'être ajoutée. Tout d'abord parce que les champs de concentration CHIMERE sont assez proches des niveaux moyens de concentrations mesurés sur les sites de fond et que l'application d'un coefficient correcteur constant permettait d'améliorer suffisamment les résultats. Mais surtout parce que les environnements des sites de mesure de fond urbain sont situés dans des espaces bien ventilés et qu'ils ne présentaient pas d'erreur significativement croissante avec le vent.

Une seconde façon de valider l'approche a été de l'appliquer à une année plus récente. J'ai choisi de travailler sur l'année 2017. Cet exercice m'a demandé un travail majeur de mise en forme du cadastre d'émission d'Airparif. J'ai recalculé les coefficients obtenus par itération pour pondérer les représentations des concentrations locales et du bruit de fond urbain, puisque la modification du cadastre d'émission impliquait une nouvelle amplitude de l'erreur de CHIMERE. Les résultats que j'ai obtenus à la suite de l'étape de généralisation de l'approche sont très satisfaisants notamment en comparaison avec les sorties du modèle HOR'AIR d'Airparif, fonctionnant également sur l'ajustement des sorties de modèle par rapport à la mesure mais incluant au préalable la mise en œuvre d'un modèle urbain forcé par les sorties de CHIMERE.

Un point limitant du développement de cette approche est le faible nombre de stations de mesures utilisables pour l'apprentissage et la validation, notamment pour les particules fines à proximité des voies. C'est sans aucun doute le facteur influant le plus sur la construction de mon équation de correction finale, notamment dans la phase de mise en relation de ses coefficients avec les proxys des sites de mesure. Pour la suite et à des fins d'amélioration de l'approche, il serait intéressant de réaliser des simulations sur plusieurs années, encadrant l'année pour laquelle on a un cadastre d'émission, afin d'augmenter le nombre de mesure par station, voire considérer les nouvelles stations installées entre deux années successives. Ainsi, les résultats découlant de la phase d'apprentissage seront plus robustes. Bien sûr, la qualité des résultats est dépendante de la qualité de CHIMERE. Au-delà des erreurs spécifiques à certains sites complexes (en nature du trafic notamment), le travail sur 2017 a montré une saisonnalité forte de l'erreur de CHIMERE. Comme on l'a vu ici, la période printemps-été est beaucoup moins bien représentée par le modèle. Il semble essentiel d'introduire la question des saisons dans l'approche de correction. Enfin, l'absence de considération des surfaces bâties dans la construction des champs de concentration fine échelle (je n'ai pas identifié de paramètre qui pourrait être lié à la valeur des coefficients de correction) ne permet pas de considérer le facteur « accumulation locale » tel qu'on peut pourtant le voir dans certaines rues confinées entre de hauts immeubles. C'est un obstacle majeur pour la caractérisation de nouveaux types d'environnements urbains - autres que des environnements « trafic routier » et probablement une limitation intrinsèque à mon approche. Ce n'est pas pénalisant de mon point de vue pour un travail statistique sur l'ensemble d'une métropole, mais cela peut contraindre la précision de nos résultats si l'on travaille sur des problématiques associées à un quartier précis. C'est d'ailleurs pour cette raison que j'ai choisi de mettre en œuvre un modèle urbain dans mon cas d'étude.

Un des prochains travaux à mener sera de comparer les sorties de mon approche de correction statistique avec d'autres approches fine échelle : les sorties des simulations CHIMERE pour lesquelles le module de discrimination de l'usage des sols sous-maille est mis en œuvre, mais aussi des sorties de modèle urbain (MUNICH notamment). Cela permettrait ainsi de qualifier et de quantifier les atouts et biais potentiels de notre approche, au regard également de la complexité de la mise en œuvre de chaque approche.

II. Estimation des expositions dynamiques de la population

Dans un second temps, je me suis attachée à l'estimation d'une version « dynamique » de l'exposition des individus, prenant en compte les déplacements quotidiens et les lieux visités (statistiquement) par chaque individu. Ce travail a nécessité une part importante de révision de la manière dont la mobilité était codée dans le modèle d'activité et d'émissions OLYMPUS. A partir d'un contexte urbain et d'usage des sols donnés, OLYMPUS génère une population synthétique à qui il attribue diverses caractéristiques propres (âge, regroupement en ménage, possession d'un véhicule, activité salariée ou non...) et pour qui il produit un emploi du temps et qu'il spatialise et associe à un mode de déplacement afin de calculer les émissions de polluants qui en découlent. J'ai effectué un travail approfondi d'amélioration des réseaux de mobilité pour les véhicules particuliers et de création d'un réseau pour les transports en commun (TC), de façon à pouvoir disposer à l'échelle de chaque individu des distances, durées et étapes du déplacement.

Pour ce qui est du réseau routier, il s'agit d'un travail effectué en fin de thèse pour remédier à certains verrous de calcul de l'exposition liés à l'implicité excessive du réseau utilisé jusqu'alors dans OLYMPUS. J'ai implémenté un réseau routier plus dense et, surtout, plus proche de la réalité en considérant les données d'OpenStreetMap. Cette étape a nécessité de poser quelques hypothèses sur les vitesses maximales autorisées, manquantes sur une partie des brins du réseau routier.

Pour ce qui est des TC, l'objectif principal était d'avoir une représentation précise des trajets (et de leurs caractéristiques) pour les individus utilisant les transports en commun, mais également de pouvoir raffiner les éléments de choix modal à disposition de chaque individu en particulier pour de futurs scénarios d'offre de transport. J'ai implémenté dans OLYMPUS un réseau de transport en commun grâce aux données publiques GTFS, qui m'ont permis de proposer une représentation réaliste de l'intégralité des arrêts constituant le réseau des TC sur la région, de différencier le type de mode de transport associé à chacune des lignes et de tenir compte des informations relatives aux horaires et durées de déplacements des modes de transport sur une journée complète.

L'ajout d'une mobilité désormais bien décrite pour les personnes empruntant les TC a mené à l'obtention d'un descriptif complet et spatialisé de l'agenda de tous les individus modélisés, en l'occurrence un échantillon représentatif de la population francilienne - pour des questions de temps de calcul sur cette étape d'évaluation. J'ai procédé au calcul des temps passés dans la réalisation de chaque activité et déplacement.

Je me suis ensuite impliquée dans la proposition de valeurs pour les niveaux ambiants de pollution dans chacun des environnements intérieurs ou clos fréquentés par mes individus. Cette étape, qui s'est appuyée sur un travail détaillé de recherche bibliographique, met en lumière les nombreux points sur lesquels une meilleure connaissance de la qualité de l'air est souhaitée. En particulier, en l'absence de relations claires entre intérieur et extérieur pour certains environnements, et du fait de la variabilité temporelle (ou d'un endroit à un autre) pour ces environnements fermés, je n'ai pu que proposer des valeurs moyennes qui restent sujet à débat, et mettent en évidence le besoin de plus de données d'observation (rames de transport souterrain, quais de gare, logements, bureaux, écoles, cœur du trafic routier...). Exception faite des déplacements réalisés sur le réseau VP (mobilité en véhicule particulier ou en mode doux) pour lesquels j'ai considéré des valeurs d'exposition liées à *Cprox* (via un coefficient de transfert ou non), j'ai dû en effet utiliser des valeurs moyennes prescrites qui restent insatisfaisantes. Cette approche d'estimation des expositions peut être grandement améliorée en calculant les niveaux intérieurs à partir des concentrations extérieures. Des études comme celle menée par l'ANSES en collaboration avec CSTB (ANSES, 2019) ont établi des coefficients de transfert entre les logements et l'extérieur et une des voies à privilégier pour l'amélioration de mon approche est de distribuer les individus dans des bâtiments (et pas uniquement dans des communes) afin de pouvoir y associer une valeur de concentration extérieure ambiante adéquate, corrigée par l'approche de statistique de descente en échelle, et à laquelle on appliquera le coefficient de transfert. Un travail est en cours de réalisation dans l'équipe pour augmenter la précision de la mobilité des individus en leur attribuant des adresses de déplacement plutôt que de considérer les centroïdes de commune. Il est pour l'instant mené dans le but de mieux distinguer les individus, et d'apporter une différenciation des catégories sociales dans la population. Ce travail servira bien au-delà de cette question (majeure), puisqu'il amènera de la précision à la mobilité et offrira la possibilité de mieux traiter l'exposition à la pollution routière dans certains logements.

J'ai enfin écrit un programme permettant d'associer les temps d'activité/déplacements aux niveaux moyens de pollution dans les lieux concernés, afin d'en déduire les totaux d'expositions sur la durée d'une activité ou d'un trajet ainsi que les expositions cumulées au cours de la journée pour tous les agents de l'échantillon de population. Bien qu'affaiblie par l'absence de données robustes sur la plupart des environnements clos, la méthodologie que je propose pour le calcul de l'exposition présente tout de même l'intérêt de pouvoir s'adapter aux futures connaissances sur ces environnements. Elle permet surtout dès à présent d'évaluer la part que prennent ces environnements dans le cumul d'exposition des personnes concernées, et on peut apprécier l'impact qu'aura le réajustement de ces concentrations selon l'évolution des connaissances.

Une prospective moins prioritaire, mais nécessaire pour mener nos simulations, est de paralléliser les scripts que j'ai développés afin de pouvoir considérer facilement effectuer les calculs sur toute la population.

III. Construction et analyse d'un scénario d'aménagement urbain

Mes travaux de thèse ont également été pensés dans le cadre d'une collaboration entre le LISA et le conseil départemental du Val-de-Marne. L'objectif principal du CD94 était d'accéder à plus de savoirs sur les impacts de la mise en place des nouvelles lignes de métro du Grand-Paris Express (GPE) en Île-de-France et, plus spécifiquement, au niveau du quartier de gare de Créteil l'Échat, sur la qualité de l'air ambiante. Mais, les événements ayant survenu au cours de ma thèse (restructuration des services du département et la pandémie du COVID-19) ont induit des retards sur plusieurs volets de ce chapitre.

La construction du scénario était le premier verrou de ce chapitre et du fait de la re-priorisation des questions de mobilité/travail au CD94, je n'ai pas pu le construire durant les deux dernières années et je n'en propose ici qu'une ébauche.

Le second point critique de ce travail était d'identifier la démarche la plus adaptée à l'exercice. Comme nous travaillions sur l'organisation du bâti et des flux de trafic du quartier, et que seules quelques rues étaient concernées, l'usage d'un modèle statistique d'exposition était discutable. J'ai donc travaillé à la configuration d'une chaîne de simulations comprenant le modèle MUNICH développé au CERIA. L'essentiel de mon travail a consisté à préparer la simulation, à lever les verrous de format et de nature des données, et à implémenter la scénarisation en ma possession dans OLYMPUS. J'ai pu adapter les réseaux de déplacements au niveau de la région et au niveau du quartier de gare en question pour modifier les habitudes de mobilité des individus - même si les données que j'ai obtenues sur l'évolution de la voirie étaient assez limitées. J'ai imposé un forçage de la répartition modale du modèle pour les déplacements s'effectuant au sein du département du Val-de-Marne pour correspondre aux objectifs du CD94. J'ai pu récupérer les fichiers GTFS prospectifs mis en ligne par la Société du Grand Paris pour la représentation des lignes du GPE mais ils ont nécessité un travail de complétion car toutes les données (distances, temps de déplacement entre deux stations consécutives et les correspondances) n'étaient pas complètes. Ces modifications représentent pour l'instant les seules différences entre les simulations de la situation de référence et du scénario expliquant, très probablement, les faibles différences des niveaux de concentrations obtenus en sortie de chaîne de modélisation. La faisabilité a donc été démontrée, et le travail de scénarisation doit être approfondi. Le résultat a montré que l'évolution des parts modales pouvait être contraire à celle souhaitée par le scénario (en l'occurrence on observe une augmentation de la part de la voiture), ce que l'on pense devoir à un effet retour de la réduction de la congestion. Cela nous amène à deux conclusions :

- le scénario de modification routière et de réattribution du bâti ne doit pas être désolidarisé de son contexte et il sera pertinent de faire évoluer dans OLYMPUS des paramètres collatéraux comme l'évolution de l'accès au stationnement pour les vélos (croissant) ou les voitures (décroissant), le confort des nouvelles lignes de TC ou la facilitation de la multimodalité ou encore les incitations financières vers certains modes, afin de mieux restituer le choix qui s'offrira aux habitants et usagers de ces futurs quartiers de gare. De même, un travail de réévaluation de l'attractivité de la zone sera nécessaire. En effet, de nombreux emplois, logements et zones de loisirs vont voir le jour dans les années à venir, dont l'ouverture d'un centre commercial à proximité de la station. Cet ajout dans le scénario pourrait donc inciter davantage d'individus à se rendre à Créteil l'Échat et dimensionnera également la mobilité locale.
- Ma plus grande priorité concerne la mise en œuvre des simulations de qualité de l'air avec MUNICH. Cet outil nous amènera des résultats particulièrement intéressants sur le rôle du bâti local dans la dispersion des polluants émis sur les routes, et sur l'effet que l'on peut attendre de la modification de la structure bâtie du quartier. J'ai pu préparer la majorité des fichiers d'entrée pour les simulations. Il reste à calibrer le modèle OLYMPUS avec le réseau routier plus fin que j'ai implémenté durant ma thèse, pour pouvoir réaliser des simulations pertinentes avec MUNICH (les réaliser avec le réseau actuel induirait en effet des superpositions des tronçons de

routes avec les zones bâties). Au-delà de l'exercice de scénarisation, les résultats de cette simulation, nous fourniront également des points de comparaison avec l'approche *Cprox*, ce qui constituera un élément essentiel de sa validation.

Plus largement, ces travaux de thèse s'inscrivent dans une démarche de modélisation urbaine plus systémique, visant à explorer de nouvelles formes de scénarios urbains, pour mieux comprendre et hiérarchiser voire quantifier les relations qui existent entre les éléments structurels et organisationnels d'une métropole, les pratiques et usages des habitants, le contexte économique et réglementaire, et les questions environnementales et de risque sanitaire. Ils ont contribué au projet POLL-EXPO (PRIMEQUAL, ADEME) dans lequel ils ont participé au développement de l'outil OLYMPUS au sein d'une chaîne de modèles, autour de la question des expositions et des inégalités d'exposition des individus dans les métropoles.

Références

- Abu-Allaban, M., & Abu-Qudais, H. (2011). Impact Assessment of Ambient Air Quality by Cement Industry: A Case Study in Jordan. *Aerosol and Air Quality Research*, 11(7), 802–810. <https://doi.org/10.4209/AAQR.2011.07.0090>
- Achakulwisut, P., Brauer, M., Hystad, P., & Anenberg, S. C. (2019). Global, national, and urban burdens of paediatric asthma incidence attributable to ambient NO₂ pollution: estimates from global datasets. *The Lancet Planetary Health*, 3(4), e166–e178. [https://doi.org/10.1016/S2542-5196\(19\)30046-4](https://doi.org/10.1016/S2542-5196(19)30046-4)
- Airparif. (n.d.). *Effets sur la santé | Airparif*. Retrieved February 8, 2022, from <https://www.airparif.asso.fr/effets-sur-la-sante>
- Airparif. (2021). *Synthèse des connaissances sur les particules en Île-de-France*.
- Alimissis, A., Philippopoulos, K., Tzanis, C. G., & Deligiorgi, D. (2018). Spatial estimation of urban air pollution with the use of artificial neural network models. *Atmospheric Environment*, 191, 205–213. <https://doi.org/10.1016/J.ATMOENV.2018.07.058>
- Alkuwari, F. A., Guillas, S., & Wang, Y. (2013). Statistical downscaling of an air quality model using fitted empirical orthogonal functions. *Atmospheric Environment*, 81, 1–10. <https://doi.org/10.1016/j.atmosenv.2013.08.031>
- Andersen, Z. J., Wahlin, P., Raaschou-Nielsen, O., Ketzler, M., Scheike, T., & Loft, S. (2007). Size distribution and total number concentration of ultrafine and accumulation mode particles and hospital admissions in children and the elderly in Copenhagen, Denmark. *Occupational and Environmental Medicine*, 65(7), 458–466. <https://doi.org/10.1136/oem.2007.033290>
- André, M., Sartelet, K., Moukhtar, S., Andre, J. M., & Redaelli, M. (2020). Diesel, petrol or electric vehicles: What choices to improve urban air quality in the Ile-de-France region? A simulation platform and case study. *Atmospheric Environment*, 241, 117752. <https://doi.org/10.1016/J.ATMOENV.2020.117752>
- André, Michel, Lekhal, F., Allemand, A., Roux, C., Boutueil, V., & Hivert, L. (2019). *Connaissance et prospective des parcs automobiles*. www.ifsttar.fr
- André, Michel, Roche, A.-L., & Bourcier, L. (2014). *Statistiques de parcs et trafic pour le calcul des émissions de polluants des transports routiers en France*. www.ifsttar.fr
- Annesi-Maesano, I., Hulin, M., Lavaud, F., Raherison, C., Kopferschmitt, C., De Blay, F., Charpin, D. A., & Denis, C. (n.d.). *Poor air quality in classrooms related to asthma and rhinitis in primary schoolchildren of the French 6 Cities Study*. <https://doi.org/10.1136/thoraxjnl-2011-200391>
- ANSES. (2019). *AVIS et RAPPORT de l'Anses relatif à une caractérisation des transferts de pollution de l'air extérieur vers l'intérieur des bâtiments*.
- Apte, J. S., Messier, K. P., Gani, S., Brauer, M., Kirchstetter, T. W., Lunden, M. M., Marshall, J. D., Portier, C. J., Vermeulen, R. C. H., & Hamburg, S. P. (2017). High-Resolution Air Pollution Mapping with Google Street View Cars: Exploiting Big Data. *Environmental Science & Technology*, 51, 6999–7008. <https://doi.org/10.1021/acs.est.7b00891>
- Araki, S., Shima, M., & Yamamoto, K. (2018). Spatiotemporal land use random forest model for estimating metropolitan NO₂ exposure in Japan. *Science of The Total Environment*, 634, 1269–1277. <https://doi.org/10.1016/J.SCITOTENV.2018.03.324>
- AtmoSud. Mesures de la qualité de l'air intérieur par autodiagnostic Collège Puget - Toulon. Retrieved

- Bajwa, K. S., Arya, S. P., & Aneja, V. P. (2012). Modeling Studies of Ammonia Dispersion and Dry Deposition at Some Hog Farms in North Carolina. *https://doi.org/10.3155/1047-3289.58.9.1198*, 58(9), 1198–1207. <https://doi.org/10.3155/1047-3289.58.9.1198>
- Bakolis, I., Hammoud, R., Stewart, R., Beevers, S., Dajnak, D., MacCrimmon, S., Broadbent, M., Pritchard, M., Shiode, N., Fecht, D., Gulliver, J., Hotopf, M., Hatch, S. L., & Mudway, I. S. (2021). Mental health consequences of urban air pollution: prospective population-based longitudinal survey. *Social Psychiatry and Psychiatric Epidemiology*, 56(9), 1587–1599. <https://doi.org/10.1007/s00127-020-01966-x>
- Basagaña, X., Rivera, M., Aguilera, I., Agis, D., Bouso, L., Elosua, R., Foraster, M., de Nazelle, A., Nieuwenhuijsen, M., Vila, J., & Künzli, N. (2012). Effect of the number of measurement sites on land use regression models in estimating local air pollution. *Atmospheric Environment*, 54, 634–642. <https://doi.org/10.1016/J.ATMOSENV.2012.01.064>
- Batkin, E., & Godard, A. (2019). *Améliorer la qualité de l'air - Le programme d'expérimentations*.
- Batterman, S., Chambliss, S., & Isakov, V. (2014). Spatial resolution requirements for traffic-related air pollutant exposure evaluations. *Atmospheric Environment*, 94, 518–528. <https://doi.org/10.1016/J.ATMOSENV.2014.05.065>
- Batterman, S., Ganguly, R., & Harbin, P. (2015). High Resolution Spatial and Temporal Mapping of Traffic-Related Air Pollutants. *Int. J. Environ. Res. Public Health*, 12, 3646–3666. <https://doi.org/10.3390/ijerph120403646>
- Beelen, R., Hoek, G., Fischer, P., Brandt, P. A. van den, & Brunekreef, B. (2007). Estimated long-term outdoor air pollution concentrations in a cohort study. *Atmospheric Environment*, 41(7), 1343–1358. <https://doi.org/10.1016/J.ATMOSENV.2006.10.020>
- Beelen, R., Hoek, G., Pebesma, E., Vienneau, D., de Hoogh, K., & Briggs, D. J. (2009). Mapping of background air pollution at a fine spatial scale across the European Union. *Science of the Total Environment*. <https://doi.org/10.1016/j.scitotenv.2008.11.048>
- Beelen, R., Hoek, G., Vienneau, D., Eeftens, M., Dimakopoulou, K., Pedeli, X., Tsai, M. Y., Künzli, N., Schikowski, T., Marcon, A., Eriksen, K. T., Raaschou-Nielsen, O., Stephanou, E., Patelarou, E., Lanki, T., Yli-Tuomi, T., Declercq, C., Falq, G., Stempfelet, M., ... de Hoogh, K. (2013). Development of NO₂ and NO_x land use regression models for estimating air pollution exposure in 36 study areas in Europe - The ESCAPE project. *Atmospheric Environment*, 72, 10–23. <https://doi.org/10.1016/j.atmosenv.2013.02.037>
- Bell, M. L., & Davis, D. L. (2001). Reassessment of the lethal London fog of 1952: Novel indicators of acute and chronic consequences of acute exposure to air pollution. *Environmental Health Perspectives*, 109(SUPPL. 3), 389–394. <https://doi.org/10.2307/3434786>
- Bell, M. L., Davis, D. L., & Fletcher, T. (2004). A retrospective assessment of mortality from the London smog episode of 1952 - The role of Influenza and Pollution. *Environmental Health Perspectives*, 112(1), 3. <https://doi.org/10.1289/ehp.6539>
- Berghmans P, Bleux N, Panis LI, Mishra VK, Torfs R, Van Poppel M. Exposure assessment of acyclist to PM10 and ultrafine particles. *Sci Total Environ* 2009;407:1286–98
- Bernard, J., Bocher, E., Petit, G., & Palominos, S. (2018). Sky View Factor Calculation in Urban Context: Computational Performance and Accuracy Analysis of Two Open and Free GIS Tools. *Climate 2018*, Vol. 6, Page 60, 6(3), 60. <https://doi.org/10.3390/CLI6030060>

- Blondeau, P., Iordache, V., Poupard, O., Genin, D., & Allard, F. (2005). Relationship between outdoor and indoor air quality in eight French schools. *Indoor Air*, 15(1), 2–12. <https://doi.org/10.1111/J.1600-0668.2004.00263.X>
- Boeing, G. (2017). OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems*, 65, 126–139. <https://doi.org/10.1016/J.COMPENVURBSYS.2017.05.004>
- Bonafe', G., Stortini, M., Emilia-Romagna, A., Minguzzi, E., Deserti, M., & Emilia-Romagna, R. (2011). *Postprocessing of a CTM with observed data: Downscaling, unbiasing and estimation of the Subgrid scale pollution variability*. <http://www.arpa.emr.it/aria>
- Bonnafoous, A., & Brun, G. (1999). Les transports et l'environnement : vers un nouvel équilibre; rapport du groupe de travail. Documentation Française. Conseil National des Transports.
- Borge, R., Narros, A., Artífano, B., Yagüe, C., Gómez-Moreno, F. J., de la Paz, D., Román-Cascón, C., Díaz, E., Maqueda, G., Sastre, M., Quaassdorff, C., Dimitroulopoulou, C., & Vardoulakis, S. (2016). Assessment of microscale spatio-temporal variation of air pollution at an urban hotspot in Madrid (Spain) through an extensive field campaign. *Atmospheric Environment*, 140, 432–445. <https://doi.org/10.1016/j.atmosenv.2016.06.020>
- Borrego, C., Sá, E., Monteiro, A., Ferreira, J., & Miranda, A. I. (2009). Forecasting human exposure to atmospheric pollutants in Portugal - A modelling approach. *Atmospheric Environment*, 43(36), 5796–5806. <https://doi.org/10.1016/j.atmosenv.2009.07.049>
- Boulanger, G., Bayeux, T., Mandin, C., Kirchner, S., Vergriette, B., Pernelet-Joly, V., & Kopp, P. (2017). Socio-economic costs of indoor air pollution: A tentative estimation for some pollutants of health interest in France. *Environment International*, 104, 14–24. <https://doi.org/10.1016/j.envint.2017.03.025>
- Bowatte, G., Lodge, C. J., Knibbs, L. D., Lowe, A. J., Erbas, B., Dennekamp, M., Marks, G. B., Giles, G., Morrison, S., Thompson, B., Thomas, P. S., Hui, J., Perret, J. L., Abramson, M. J., Walters, H., Matheson, M. C., & Dharmage, S. C. (2017). Traffic-related air pollution exposure is associated with allergic sensitization, asthma, and poor lung function in middle age. *Journal of Allergy and Clinical Immunology*, 139(1), 122-129.e1. <https://doi.org/10.1016/J.JACI.2016.05.008>
- Boznar, M., Lesjak, M., & Mlakar, P. (1993). A neural network-based method for short-term predictions of ambient SO₂ concentrations in highly polluted industrial areas of complex terrain. *Atmospheric Environment. Part B. Urban Atmosphere*, 27(2), 221–230. [https://doi.org/10.1016/0957-1272\(93\)90007-S](https://doi.org/10.1016/0957-1272(93)90007-S)
- Braniš M. The contribution of ambient source to particulate pollution in spaces and trains of the Prague underground transport system. *Atmospheric Environment* 2006;40: 348e356
- Briggs, D. (2005). The role of GIS: Coping with space (and time) in air pollution exposure assessment. *Journal of Toxicology and Environmental Health - Part A*, 68(13–14), 1243–1261. <https://doi.org/10.1080/15287390590936094>
- Briggs DJ, de Hoogh K, Morris C, Gulliver J. Effects of travel mode on exposures to particulate air pollution. *Environ Int* 2008;34(1):12–22
- Briggs, D. J., Collins, S., Elliott, P., Fischer, P., Kingham, S., Lebret, E., Pryn, K., Van Reeuwijk, H., Smallbone, K., & Van Der Veen, A. (1997). Mapping urban air pollution using GIS: a regression-based approach. <https://doi.org/10.1080/136588197242158>, 11(7), 699–718. <https://doi.org/10.1080/136588197242158>
- Brokamp, C., Jandarov, R., Rao, M. B., LeMasters, G., & Ryan, P. (2017). Exposure assessment models for elemental components of particulate matter in an urban environment: A comparison of regression and random forest approaches. *Atmospheric Environment*, 151, 1–11.

<https://doi.org/10.1016/j.atmosenv.2016.11.066>

- Brunekreef, B., Beelen, R., Hoek, G., Schouten, L., Bausch-Goldbohm, S., Fischer, P., Armstrong, B., Hughes, E., Jerrett, M., & van den Brandt, P. (2009). Effects of long-term exposure to traffic-related air pollution on respiratory and cardiovascular mortality in the Netherlands: the NLCS-AIR study. *Research Report (Health Effects Institute)*, 139, 5–71; discussion 73. <https://europepmc.org/article/med/19554969>
- Buoli, M., Grassi, S., Caldiroli, A., Carnevali, G. S., Mucci, F., Iodice, S., Cantone, L., Pergoli, L., & Bollati, V. (2018). Is there a link between air pollution and mental disorders? *Environment International*, 118(May), 154–168. <https://doi.org/10.1016/j.envint.2018.05.044>
- Buonanno, G., Stabile, L., & Morawska, L. (2014). Personal exposure to ultrafine particles: The influence of time-activity patterns. *Science of The Total Environment*, 468–469, 903–907. <https://doi.org/10.1016/J.SCITOTENV.2013.09.016>
- Carter, W. P. L. (1990). A detailed mechanism for the gas-phase atmospheric reactions of organic compounds. *Atmospheric Environment. Part A. General Topics*, 24(3), 481–518. [https://doi.org/10.1016/0960-1686\(90\)90005-8](https://doi.org/10.1016/0960-1686(90)90005-8)
- 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*, 99, 293–302. <https://doi.org/10.1016/J.ENVINT.2016.12.007>
- Cepeda, M., Schoufour, J., Freak-Poli, R., Koolhaas, C. M., Dhana, K., Bramer, W. M., & Franco, O. H. (2017). Levels of ambient air pollution according to mode of transport: a systematic review. *The Lancet Public Health*, 2(1), e23–e34. [https://doi.org/10.1016/S2468-2667\(16\)30021-4](https://doi.org/10.1016/S2468-2667(16)30021-4)
- Cesaroni, G., Badaloni, C., Gariazzo, C., Stafoggia, M., Sozzi, R., Davoli, M., & Forastiere, F. (2013). Long-term exposure to urban air pollution and mortality in a cohort of more than a million adults in Rome. *Environmental Health Perspectives*, 121(3), 324–331. <https://doi.org/https://dx.doi.org/10.1289/ehp.1205862>
- Cesaroni, G., Forastiere, F., Stafoggia, M., Andersen, Z. J., Badaloni, C., Beelen, R., Caracciolo, B., De Faire, U., Erbel, R., Eriksen, K. T., Fratiglioni, L., Galassi, C., Hampel, R., Heier, M., Hennig, F., Hilding, A., Hoffmann, B., Houthuijs, D., Jöckel, K. H., ... Peters, A. (2014). Long term exposure to ambient air pollution and incidence of acute coronary events: Prospective cohort study and meta-analysis in 11 european cohorts from the escape project. *BMJ (Online)*, 348. <https://doi.org/10.1136/bmj.f7412>
- Challoner, A., & Gill, L. (2014). Indoor/outdoor air pollution relationships in ten commercial buildings: PM2.5 and NO2. *Building and Environment*, 80, 159–173. <https://doi.org/10.1016/j.buildenv.2014.05.032>
- Chang, J. C., & Hanna, S. R. (2004). Air quality model performance evaluation. *Meteorology and Atmospheric Physics*, 87(1–3), 167–196. <https://doi.org/10.1007/s00703-003-0070-7>
- Chen, G., Li, S., Knibbs, L. D., Hamm, N. A. S., Cao, W., Li, T., Guo, J., Ren, H., Abramson, M. J., & Guo, Y. (2018). A machine learning method to estimate PM2.5 concentrations across China with remote sensing, meteorological and land use information. *Science of The Total Environment*, 636, 52–60. <https://doi.org/10.1016/J.SCITOTENV.2018.04.251>
- Chen, J., de Hoogh, K., Gulliver, J., Hoffmann, B., Hertel, O., Ketzel, M., Bauwelinck, M., van Donkelaar, A., Hvidtfeldt, U. A., Katsouyanni, K., Janssen, N. A. H., Martin, R. V., Samoli, E., Schwartz, P. E., Stafoggia, M., Bellander, T., Strak, M., Wolf, K., Vienneau, D., ... Hoek, G. (2019). A comparison of linear regression, regularization, and machine learning algorithms to develop Europe-wide spatial models of fine particles and nitrogen dioxide. *Environment International*, 130, 104934.

<https://doi.org/10.1016/J.ENVINT.2019.104934>

- Choi, S., Park, J.-H., Bae, S.-Y., Kim, S.-Y., Byun, H., Kwak, H., Hwang, S., Park, J., Park, H., Lee, K.-H., Kim, W., & Park, D.-U. (2019). Characteristics of PM₁₀ Levels Monitored for More than a Decade in Subway Stations in South Korea. *Aerosol and Air Quality Research*, *19*, 2746–2756. <https://doi.org/10.4209/aaqr.2019.05.0263>
- CITEPA. (2021). *Gaz à effet de serre et polluants atmosphériques - Bilan des émissions en France*.
- Cohen, A. J., Brauer, M., Burnett, R., Anderson, H. R., Frostad, J., Estep, K., Balakrishnan, K., Brunekreef, B., Dandona, L., Dandona, R., Feigin, V., Freedman, G., Hubbell, B., Jobling, A., Kan, H., Knibbs, L., Liu, Y., Martin, R., Morawska, L., ... Forouzanfar, M. H. (2017). Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. *The Lancet*, *389*(10082), 1907–1918. [https://doi.org/10.1016/S0140-6736\(17\)30505-6](https://doi.org/10.1016/S0140-6736(17)30505-6)
- Colette, A., Granier, C., Hodnebrog, Jakobs, H., Maurizi, A., Nyiri, A., Bessagnet, B., D'Angiola, A., D'Isidoro, M., Gauss, M., Meleux, F., Memmesheimer, M., Mieville, A., Roüil, L., Russo, F., Solberg, S., Stordal, F., & Tampieri, F. (2011). Air quality trends in Europe over the past decade: A first multi-model assessment. *Atmospheric Chemistry and Physics*, *11*(22), 11657–11678. <https://doi.org/10.5194/acp-11-11657-2011>
- Coll, I., Lasry, F., Fayet, S., Armengaud, A., & Vautard, R. (2009). Simulation and evaluation of 2010 emission control scenarios in a Mediterranean area. *Atmospheric Environment*, *43*(27), 4194–4204. <https://doi.org/10.1016/j.atmosenv.2009.05.034>
- Colombi C, Angius S, Gianelle V, Lazzarini M. Particulate matter concentrations, physical characteristics and elemental composition in the Milan underground transport system. *Atmos Environ* 2013;70:166–78
- Copernicus Atmosphere Monitoring Service*. (n.d.). Retrieved May 10, 2022, from https://regional.atmosphere.copernicus.eu/index.php?category=ensemble&subensemble=hourly_ensemble&date=LAST&calculation-model=ENSEMBLE&species=o3&level=SFC&offset=000
- Coulombel, N., Boutueil, V., Liu, L., Viguié, V., & Yin, B. (2019). Substantial rebound effects in urban ridesharing: Simulating travel decisions in Paris, France. *Transportation Research Part D: Transport and Environment*, *71*, 110–126. <https://doi.org/10.1016/j.trd.2018.12.006>
- Cyrys, J., Pitz, M., Bischof, W., Wichmann, H. E., & Heinrich, J. (2004). Relationship between indoor and outdoor levels of fine particle mass, particle number concentrations and black smoke under different ventilation conditions. *Journal of Exposure Science & Environmental Epidemiology* *2004* *14*:4, *14*(4), 275–283. <https://doi.org/10.1038/sj.jea.7500317>
- Daly, A., & Zannetti, P. (2007). Air Pollution Modeling – An Overview. *Ambient Air Pollution*, *1*(2003), 15–28.
- de Hoogh, K., Korek, M., Vienneau, D., Keuken, M., Kukkonen, J., Nieuwenhuijsen, M. J., Badaloni, C., Beelen, R., Bolignano, A., Cesaroni, G., Pradas, M. C., Cyrys, J., Douros, J., Eeftens, M., Forastiere, F., Forsberg, B., Fuks, K., Gehring, U., Gryparis, A., ... Bellander, T. (2014). Comparing land use regression and dispersion modelling to assess residential exposure to ambient air pollution for epidemiological studies. *Environment International*, *73*, 382–392. <https://doi.org/10.1016/j.envint.2014.08.011>
- de Meij, A., Thunis, P., Bessagnet, B., & Cuvelier, C. (2009). The sensitivity of the CHIMERE model to emissions reduction scenarios on air quality in Northern Italy. *Atmospheric Environment*, *43*(11), 1897–1907. <https://doi.org/10.1016/j.atmosenv.2008.12.036>
- De Montjoye, Y. A., Hidalgo, C. A., Verleysen, M., & Blondel, V. D. (2013). Unique in the Crowd: The privacy bounds of human mobility. *Scientific Reports*, *3*, 1–5. <https://doi.org/10.1038/srep01376>

- de Nazelle, A., Bode, O., & Orjuela, J. P. (2017). Comparison of air pollution exposures in active vs. passive travel modes in European cities: A quantitative review. In *Environment International* (Vol. 99, pp. 151–160). Elsevier Ltd. <https://doi.org/10.1016/j.envint.2016.12.023>
- De Nazelle, A., Seto, E., Donaire-Gonzalez, D., Mendez, M., Matamala, J., Nieuwenhuijsen, M. J., & Jerrett, M. (2013). Improving estimates of air pollution exposure through ubiquitous sensing technologies. *Environmental Pollution*, 176, 92–99. <https://doi.org/10.1016/J.ENVPOL.2012.12.032>
- Delaunay, C., Goupil, G., Ravelomanantsoa, H., Person, A., Mazoue, S., & Morawski, F. (2010). Evaluation de l'exposition des citadins aux polluants atmosphériques au cours de leurs déplacements dans l'agglomération parisienne.
- Delaunay, C., Goupil, G., Ravelomanantsoa, H., Person, A., Mazoue, S., & Morawski, F. (2012). Évaluation de l'exposition des citadins aux polluants atmosphériques au cours de leurs déplacements dans l'agglomération parisienne City dwellers exposure to atmospheric pollutants when commuting in Paris urban area Cette étude a été menée dans le cadre d. *Pollution Atmospherique*.
- Derbez, M., & Dassonville, C. Campagne nationale de l'OQAI dans les écoles (2013-2016) : enjeux et résultats attendus (CSTB/OQAI). Retrieved April 20, 2022, from <https://www.fnogec.org/documents/oqai-campagne-nationale-2013-2016-enjeux-et>
- Derbez, M., Wyart, G., Douchin, F., Ramalho, O., Ribéron, J., Kirchner, S., & Mandin, C. (2016). Base de référence nationale sur la qualité de l'air intérieur et le confort des occupants de bâtiments performants en énergie – Description des premiers résultats de la qualité de l'air intérieur et du confort de bâtiments d'habitation performants en énergie.
- Derbez, M., Wyart, G., Douchin, F., Ramalho, O., Ribéron, J., Kirchner, S., & Mandin, C. (2016). Base de référence nationale sur la qualité de l'air intérieur et le confort des occupants de bâtiments performants en énergie – Deuxième état descriptif de la qualité de l'air intérieur et du confort de bâtiments d'habitation performants en énergie. CSTB-OQAI/, 100 pages.
- Derbez, M., Wyart, G., Le Ponner, E., Ramalho, O., Ribéron, J., & Mandin, C. (2018). Indoor air quality in energy-efficient dwellings: Levels and sources of pollutants. *Indoor Air*, 28(2), 318-338.
- Derognat, C., Beekmann, M., Baeumle, M., Martin, D., & Schmidt, H. (2003). Effect of biogenic volatile organic compound emissions on tropospheric chemistry during the Atmospheric Pollution Over the Paris Area (ESQUIF) campaign in the Ile-de-France region. *Journal of Geophysical Research: Atmospheres*, 108(D17), 8560. <https://doi.org/10.1029/2001JD001421>
- Dewulf, B., Neutens, T., Lefebvre, W., Seynaeve, G., Vanpoucke, C., Beckx, C., & Van de Weghe, N. (2016). Dynamic assessment of exposure to air pollution using mobile phone data. *International Journal of Health Geographics*, 15(1), 1–14. <https://doi.org/10.1186/s12942-016-0042-z>
- Diapouli E, Grivas G, Chaloulakou A, Spyrellis N. PM10 and ultrafine particles counts in- vehicle and On-road in the Athens area. *Water Air Soil Pollut Focus* 2008;8(1):89–97
- Dias, D., & Tchepel, O. (2018). Spatial and Temporal Dynamics in Air Pollution Exposure Assessment. *International Journal of Environmental Research and Public Health*, 15(558), 23. <https://doi.org/10.3390/ijerph15030558>
- Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, 1(1), 269–271. <https://doi.org/10.1007/BF01386390>
- Dons, E., Int Panis, L., Van Poppel, M., Theunis, J., Willems, H., Torfs, R., & Wets, G. (2011). Impact of time–activity patterns on personal exposure to black carbon. *Atmospheric Environment*, 45(21), 3594–3602. <https://doi.org/10.1016/J.ATMOENV.2011.03.064>
- Dreyfus, G., Martinez, J., Samuelides, M., Gordon, M., Badran, F., Thiria, S., & Hérault, L. (2002). Les

réseaux de neurones. *Mécanique Industrielle et Matériaux*, 51, 19.

- EEA. (2013). Air quality in Europe—2013 Report: EEA report no 9/2013. In *European Union* (Issue 9). <https://doi.org/10.2800/92843>
- Eeftens, M., Beelen, R., Hoogh, K. de, Bellander, T., Cesaroni, G., Cirach, M., Declercq, C., Dédélé, A., Dons, E., Nazelle, A. de, Dimakopoulou, K., Eriksen, K., Falq, G., Fischer, P., Galassi, C., Gražulevičienė, R., Heinrich, J., Hoffmann, B., Jerrett, M., ... Hoek, G. (2012). Development of Land Use Regression Models for PM_{2.5}, PM_{2.5} Absorbance, PM₁₀ and PM_{coarse} in 20 European Study Areas; Results of the ESCAPE Project. *Environmental Science and Technology*, 46(20), 11195–11205. <https://doi.org/10.1021/ES301948K>
- El-Harbawi, M. (2013). Air quality modelling, simulation, and computational methods: A review. *Environmental Reviews*, 21(3), 149–179. <https://doi.org/10.1139/er-2012-0056>
- Elessa Etuman, A., & Coll, I. (2018). OLYMPUS v1.0: Development of an integrated air pollutant and GHG urban emissions model-methodology and calibration over greater Paris. *Geoscientific Model Development*, 11(12), 5085–5111. <https://doi.org/10.5194/gmd-11-5085-2018>
- European Commission. (2010). *Survey on perception of quality of life in 75 European cities EN*.
- Eze, I. C., Schaffner, E., Fischer, E., Schikowski, T., Adam, M., Imboden, M., Tsai, M., Carballo, D., von Eckardstein, A., Künzli, N., Schindler, C., & Probst-Hensch, N. (2014). Long-term air pollution exposure and diabetes in a population-based Swiss cohort. *Environment International*, 70, 95–105. <https://doi.org/10.1016/j.envint.2014.05.014>
- Fable, S., Fraboulet, I., Godefroy, F., Jantolek, G., & Queron, J. (2009). INTER'MODAL - Vers une meilleure maîtrise de l'exposition individuelle par inhalation des populations à la pollution atmosphérique lors de leurs déplacements urbains.
- Faustini, A., Rapp, R., & Forastiere, F. (2014). Nitrogen dioxide and mortality: Review and meta-analysis of long-term studies. *European Respiratory Journal*, 44(3), 744–753. <https://doi.org/10.1183/09031936.00114713>
- Firket, J. (1936). Fog along the Meuse Valley. In *January* (Vol. 32).
- Freire, C., Ramos, R., Puertas, R., Lopez-Espinosa, M.-J., Julvez, J., Aguilera, I., Cruz, F., Fernandez, M.-F., Sunyer, J., & Olea, N. (2010). Association of traffic-related air pollution with cognitive development in children. *Journal of Epidemiology & Community Health*, 64(3), 223–228. <https://doi.org/10.1136/JECH.2008.084574>
- Gaffin, J. M., Petty, C. R., Hauptman, M., Kang, C. M., Wolfson, J. M., Abu Awad, Y., Di, Q., Lai, P. S., Sheehan, W. J., Baxi, S., Coull, B. A., Schwartz, J. D., Gold, D. R., Koutrakis, P., & Phipatanakul, W. (2016). Modeling indoor particulate exposures in inner-city school classrooms. *Journal of Exposure Science & Environmental Epidemiology* 2017 27:5, 27(5), 451–457. <https://doi.org/10.1038/jes.2016.52>
- Gariazzo, C., Carlino, G., Silibello, C., Renzi, M., Finardi, S., Pepe, N., Radice, P., Forastiere, F., Michelozzi, P., Viegi, G., & Stafoggia, M. (2020). A multi-city air pollution population exposure study: Combined use of chemical-transport and random-Forest models with dynamic population data. *Science of The Total Environment*, 724, 138102. <https://doi.org/10.1016/J.SCITOTENV.2020.138102>
- Geiss O, Barrero-Moreno J, Tirendi S, Kotzias D. Exposure to particulate matter in vehicle cabins of private cars. *Aerosol Air Qual Res* 2010;10:581–8
- Gerharz, L. E., Klemm, O., Broich, A. V., & Pebesma, E. (2014). Spatio-temporal modelling of individual exposure to air pollution and its uncertainty. *Atmospheric Environment*, 64(2), 56–65. <https://doi.org/10.1016/j.atmosenv.2012.09.069>
- Glasgow, M. L., Rudra, C. B., Yoo, E. H., Demirbas, M., Merriman, J., Nayak, P., Crabtree-Ide, C., Szpiro,

- A. A., Rudra, A., Wactawski-Wende, J., & Mu, L. (2014). Using smartphones to collect time–activity data for long-term personal-level air pollution exposure assessment. *Journal of Exposure Science & Environmental Epidemiology* 2016 26:4, 26(4), 356–364. <https://doi.org/10.1038/jes.2014.78>
- Golden, B. (1976). Shortest-Path Algorithms: A Comparison. *Operations Research*, 24(6), 1164–1168. <https://doi.org/10.1287/OPRE.24.6.1164>
- Goldsmith, J. R. (1969). Air Pollution Epidemiology. *Archives of Environmental Health*, 18(4), 516–522. <https://doi.org/10.1080/00039896.1969.10665446>
- Goldsmith, J. R., & Lester Breslow, M. (1959). Epidemiological Aspects of Air Pollution. *Journal of the Air Pollution Control Association*, 9(3), 129–132. <https://doi.org/10.1080/00022470.1959.10467886>
- Grange, S. K., Carslaw, D. C., Lewis, A. C., Boleti, E., & Hueglin, C. (2018). Random forest meteorological normalisation models for Swiss PM10 trend analysis. *Atmospheric Chemistry and Physics*, 18(9), 6223–6239. <https://doi.org/10.5194/ACP-18-6223-2018>
- Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P. I., & Geron, C. (2006). Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature). *Atmospheric Chemistry and Physics*, 6(11), 3181–3210. <https://doi.org/10.5194/ACP-6-3181-2006>
- Guillas, S., Bao, J., Choi, Y., & Wang, Y. (2008). Statistical correction and downscaling of chemical transport model ozone forecasts over Atlanta. *Atmospheric Environment*, 42(6), 1338–1348. <https://doi.org/10.1016/j.atmosenv.2007.10.027>
- Gulliver J, Briggs DJ. Personal exposure to particulate air pollution in transport microenvironments. *Atmospheric Environment* 2004;38:1–8
- Gulliver, J., & Briggs, D. J. (2005). Time-space modeling of journey-time exposure to traffic-related air pollution using GIS. *Environmental Research*, 97(1), 10–25. <https://doi.org/10.1016/j.envres.2004.05.002>
- Gulliver J, Briggs DJ. Journey-time exposure to particulate air pollution. *Atmospheric Environment* 2007;41:7195–207
- Gustafsson, M., Blomqvist, G., Gudmundsson, A., Dahl, A., Jonsson, P., & Swietlicki, E. (2009). Factors influencing PM10 emissions from road pavement wear. *Atmospheric Environment*, 43(31), 4699–4702. <https://doi.org/10.1016/J.ATMOSENV.2008.04.028>
- Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B. M., & Gräler, B. (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, 2018(8). <https://doi.org/10.7717/PEERJ.5518>
- Héroux, M.-E., Clark, N., Ryswyk, K. Van, Mallick, R., Gilbert, N. L., Harrison, I., Rispler, K., Wang, D., Anastassopoulos, A., Guay, M., Macneill, M., & Wheeler, A. J. (2010). Predictors of Indoor Air Concentrations in Smoking and Non-Smoking Residences. *Int. J. Environ. Res. Public Health*, 7, 7. <https://doi.org/10.3390/ijerph7083080>
- Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., & Briggs, D. (2008). A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric Environment*, 42(33), 7561–7578. <https://doi.org/10.1016/j.atmosenv.2008.05.057>
- Hoek, G., Fischer, P., Van Den Brandt, P., Goldbohm, S., & Brunekreef, B. (2001). Estimation of long-term average exposure to outdoor air pollution for a cohort study on mortality. *Journal Of Exposure Analysis and Environmental Epidemiology*, 11(6), 459–469. <https://doi.org/10.1038/sj.jea.7500189>
- Honoré, C., Rouïl, L., Vautard, R., Beekmann, M., Bessagnet, B., Dufour, A., Elichegaray, C., Flaud, J. M., Malherbe, L., Meleux, F., Menut, L., Martin, D., Peuch, A., Peuch, V. H., & Poisson, N. (2008).

- Predictability of European air quality: Assessment of 3 years of operational forecasts and analyses by the PREV'AIR system. *Journal of Geophysical Research Atmospheres*, 113(4). <https://doi.org/10.1029/2007JD008761>
- Hood, C., MacKenzie, I., Stocker, J., Johnson, K., Carruthers, D., Vieno, M., & Doherty, R. (2018). Air quality simulations for London using a coupled regional-to-local modelling system. *Atmospheric Chemistry and Physics*, 18(15), 11221–11245. <https://doi.org/10.5194/acp-18-11221-2018>
- Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359–366. [https://doi.org/10.1016/0893-6080\(89\)90020-8](https://doi.org/10.1016/0893-6080(89)90020-8)
- Host, S., & Chatignoux, E. (2009). *Pollution atmosphérique de proximité liée au trafic : expositions et effets sanitaires*.
- Host, S., Chatignoux, E., Leal, C., & Grémy, I. (2012). Exposition à la pollution atmosphérique de proximité liée au trafic : Quelles méthodes pour quels risques sanitaires ? *Revue d'Epidémiologie et de Santé Publique*, 60(4), 321–330. <https://doi.org/10.1016/j.respe.2012.02.007>
- Host, S., Chatignoux, E., & Saunal, A. (2012). Impacts sanitaires de la pollution atmosphérique urbaine et des expositions à proximité du trafic routier dans l'agglomération parisienne. *Pollution Atmosphérique, SPEC. ISS.*, 49–53.
- Hourdin, F., & Armengaud, A. (1998). The Use of Finite-Volume Methods for Atmospheric Advection of Trace Species. Part I: Test of Various Formulations in a General Circulation Model. *Monthly Weather Review*, 127.
- Hrust, L., Klaić, Z. B., Križan, J., Antonić, O., & Hercog, P. (2009). Neural network forecasting of air pollutants hourly concentrations using optimised temporal averages of meteorological variables and pollutant concentrations. *Atmospheric Environment*, 43(35), 5588–5596. <https://doi.org/10.1016/j.atmosenv.2009.07.048>
- Huang, C., Hu, J., Xue, T., Xu, H., & Wang, M. (2021). High-Resolution Spatiotemporal Modeling for Ambient PM_{2.5} Exposure Assessment in China from 2013 to 2019. *Environmental Science and Technology*, 55(3), 2152–2162. <https://doi.org/10.1021/acs.est.0c05815>
- IAU. (2016). *Les îlots morphologiques urbains (IMU)*. <http://www.iau-idf.fr>
- IAU, & ORS. (2014). *Territoires, incubateurs de santé ?*
- Ifop. (2018). *Etude de perception des Franciliens à l'égard de la qualité de l'air*.
- Int Panis L, de Geus B, Vandenbulcke G, Willems H, Degraeuwe B, Bleux N, et al. Exposure to particulate matter in traffic: a comparison of cyclists and car passengers. *Atmos Environ* 2010;44(19):2263–70
- Jerrett, M., Arain, A., Kanaroglou, P., Beckerman, B., Potoglou, D., Sahsuvaroglu, T., Morrison, J., & Giovis, C. (2005). A review and evaluation of intraurban air pollution exposure models. *Journal of Exposure Analysis and Environmental Epidemiology*, 15, 185–204. <https://doi.org/10.1038/sj.jea.7500388>
- Johansson, C., & Johansson, A. (2003). Particulate matter in the underground of Stockholm. In *Atmospheric Environment* (Vol. 37).
- Johnson, D. B. (1973). *A Note on Dijkstra's Shortest Path Algorithm*.
- Johnson, M., Brook, J. R., Brook, R. D., Oiamo, T. H., Luginaah, I., Peters, P. A., & Spence, J. D. (2020). Traffic-Related Air Pollution and Carotid Plaque Burden in a Canadian City With Low-Level Ambient Pollution. *Journal of the American Heart Association*, 9(7). <https://doi.org/10.1161/JAHA.119.013400>
- Joly, D., Gilbert, D., Diaz-de-Quijano, M., Hilal, M., Joly, M., & Bernard, N. (2021). Enhancing air quality

- forecasts by geomatic downscaling: an application to daily PM10 concentrations in France. *Theoretical and Applied Climatology*, 143(1–2), 327–339. <https://doi.org/10.1007/s00704-020-03418-7>
- Kang, C., Ma, X., Tong, D., & Liu, Y. (2012). Intra-urban human mobility patterns: An urban morphology perspective. *Physica A: Statistical Mechanics and Its Applications*, 391(4), 1702–1717. <https://doi.org/10.1016/j.physa.2011.11.005>
- Katsoulis, M., Dimakopoulou, K., Pedeli, X., Trichopoulos, D., Gryparis, A., Trichopoulou, A., & Katsouyanni, K. (2014). Long-term exposure to traffic-related air pollution and cardiovascular health in a Greek cohort study. *Science of the Total Environment*, 490, 934–940. <https://doi.org/10.1016/j.scitotenv.2014.05.058>
- Khaniabadi, Y. O., Goudarzi, G., Daryanoosh, S. M., Borgini, A., Tittarelli, A., & De Marco, A. (2017). Exposure to PM10, NO2, and O3 and impacts on human health. *Environmental Science and Pollution Research*, 24(3), 2781–2789. <https://doi.org/10.1007/s11356-016-8038-6>
- Khreis, H., Kelly, C., Tate, J., Parslow, R., Lucas, K., & Nieuwenhuijsen, M. (2017). Exposure to traffic-related air pollution and risk of development of childhood asthma: A systematic review and meta-analysis. In *Environment International* (Vol. 100, pp. 1–31). Elsevier Ltd. <https://doi.org/10.1016/j.envint.2016.11.012>
- Khreis, H., & Nieuwenhuijsen, M. J. (2017). Traffic-related air pollution and childhood asthma: Recent advances and remaining gaps in the exposure assessment methods. *International Journal of Environmental Research and Public Health*, 14(3). <https://doi.org/10.3390/ijerph14030312>
- Kim, Y., Wu, Y., Seigneur, C., & Roustan, Y. (2018). Multi-scale modeling of urban air pollution: Development and application of a Street-in-Grid model (v1.0) by coupling MUNICH (v1.0) and Polair3D (v1.8.1). *Geoscientific Model Development*, 11(2), 611–629. <https://doi.org/10.5194/gmd-11-611-2018>
- Klepeis, N. E., Nelson, W. C., Ott, W. R., Robinson, J. P., Tsang, A. M., Switzer, P., Behar, J. V., Hern, S. C., & Engelmann, W. H. (2001). The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants. *Journal of Exposure Science & Environmental Epidemiology* 2001 11:3, 11(3), 231–252. <https://doi.org/10.1038/sj.jea.7500165>
- Kohonen, T. (1988). An introduction to neural computing. *Neural Networks*, 1(1), 3–16. [https://doi.org/10.1016/0893-6080\(88\)90020-2](https://doi.org/10.1016/0893-6080(88)90020-2)
- Kotzias, D., Geiss, O., Tirendi, S., & Barrero-Moreno, J. (2009). Exposure to multiple air contaminants in public buildings, schools and kindergartens-the European indoor air monitoring and exposure assessment (airmex) study ICARUS (EU H2020 project): Integrated Climate forcing and Air pollution Reduction in Urban Systems View project Mediterranean Scientific Association of Environmental Protection (MESAEP) View project. *Fresenius Environmental Bulletin*. <https://www.researchgate.net/publication/231814278>
- Künzli, N., Kaiser, R., Medina, S., Studnicka, M., Chanel, O., Filliger, P., Herry, M., Horak Jr, F., Puybonnieux-Textier, V., Quénel, P., Schneider, J., Seethaler, R., Vergnaud, J.-C., & Sommer, H. (2000). Public-health impact of outdoor and traffic-related air pollution: a European assessment. *Activitas Nervosa Superior Rediviva*, 356, 795–801.
- Langer, S., Ramalho, O., Derbez, M., Ribéron, J., Kirchner, S., & Mandin, C. (2016). Indoor environmental quality in French dwellings and building characteristics. *Atmospheric Environment*, 128, 82–91.
- Larkin, A., Geddes, J. A., Martin, R. V., Xiao, Q., Liu, Y., Marshall, J. D., Brauer, M., & Hystad, P. (2017). Global Land Use Regression Model for Nitrogen Dioxide Air Pollution. *Environmental Science and Technology*, 51(12), 6957–6964. <https://doi.org/10.1021/acs.est.7b01148>
- Lattuati, M. (1997). *Impact des émissions européennes sur le bilan de l’ozone troposphérique a l’interface*

de l'europe et de l'atlantique nord : apport de la modelisation lagrangienne et des mesures en altitude. <http://www.theses.fr/1997PA066416>

- Lee, S.-S., & Keener, T. C. (2008). Dispersion Modeling of Mercury Emissions from Coal-Fired Power Plants at Coshocton and Manchester, Ohio. *The Ohio Journal of Science*, 108(4), 65–69. <http://www.netl>.
- Leelőssy, Á., Molnár, F., Izsák, F., Havasi, Á., Lagzi, I., & Mészáros, R. (2014). Dispersion modeling of air pollutants in the atmosphere: a review. *Central European Journal of Geosciences*, 6(3), 257–278. <https://doi.org/10.2478/S13533-012-0188-6/HTML>
- Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D., & Pozzer, A. (2015). The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature*, 525(7569), 367–371. <https://doi.org/10.1038/nature15371>
- Letinois, L., & Malherbe, L. (2010). Etude des relations entre les concentrations dans l'air intérieur et extérieur.
- Liang, L., Gong, P., Cong, N., Li, Z., Zhao, Y., & Chen, Y. (2019). Assessment of personal exposure to particulate air pollution: The first result of City Health Outlook (CHO) project. *BMC Public Health*, 19(1), 1–12. <https://doi.org/10.1186/S12889-019-7022-8/FIGURES/4>
- Lipsett, M., Hurley, S., & Ostro, B. (1997). Air pollution and emergency room visits for asthma in Santa Clara County, California. *Environmental Health Perspectives*, 105(2), 216–222. <https://doi.org/10.1289/EHP.97105216>
- Logan, W. P. D. (1953). MORTALITY IN THE LONDON FOG INCIDENT, 1952. *The Lancet*, 261(6755), 336–338. [https://doi.org/10.1016/S0140-6736\(53\)91012-5](https://doi.org/10.1016/S0140-6736(53)91012-5)
- Madkour, A., Aref, W. G., Rehman, F. U., Rahman, M. A., & Basalamah, S. (2017). *A Survey of Shortest-Path Algorithms*. <http://arxiv.org/abs/1705.02044>
- 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*, 10(6), 2397–2423. <https://doi.org/10.5194/gmd-10-2397-2017>
- Mandin, C., Trantallidi, M., Cattaneo, A., Canha, N., Mihucz, V. G., Szigeti, T., ... & Bartzis, J. (2017). Assessment of indoor air quality in office buildings across Europe—The OFFICAIR study. *Science of the Total Environment*, 579, 169–178.
- Manisalidis, I., Stavropoulou, E., Stavropoulos, A., & Bezirtzoglou, E. (2020). Environmental and Health Impacts of Air Pollution: A Review. *Frontiers in Public Health*, 8. <https://doi.org/10.3389/fpubh.2020.00014>
- Mao, P., Li, J., Xiong, L., Wang, R., Wang, X., Tan, Y., & Li, H. (2019). *Characterization of Urban Subway Microenvironment Exposure—A Case of Nanjing in China*. <https://doi.org/10.3390/ijerph16040625>
- Martin, A. E. (1964). *Mortality and Morbidity Statistics and Air Pollution*.
- Mastroianni, P., Monechi, B., Servedio, V. D. P., Liberto, C., Valenti, G., & Loreto, V. (2016). Individual mobility patterns in urban environment. *COMPLEXIS 2016 - Proceedings of the 1st International Conference on Complex Information Systems*, 81–88. <https://doi.org/10.5220/0005907000810088>
- Matthaios, V. N., Kramer, L. J., Crilley, L. R., Sommariva, R., Pope, F. D., & Bloss, W. J. (2020). Quantification of within-vehicle exposure to NO_x and particles: Variation with outside air quality, route choice and ventilation options. *Atmospheric Environment*, 240. <https://doi.org/10.1016/j.atmosenv.2020.117810>
- McCreddin, A., Alam, M. S., & McNabola, A. (2014). Modelling personal exposure to particulate air

- pollution: An assessment of time-integrated activity modelling, Monte Carlo simulation & artificial neural network approaches. *International Journal of Hygiene and Environmental Health*, 218(1), 107–116. <https://doi.org/10.1016/j.ijheh.2014.08.004>
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics* 1943 5:4, 5(4), 115–133. <https://doi.org/10.1007/BF02478259>
- Mehel, A., Murzyn, F., & Cuvelier, P. (2019). *Caractérisation et analyse des polluants issus du transport automobile s'infiltrant dans les habitacles des véhicules* (Issue June).
- Meier, R., Schindler, C., Eeftens, M., Aguilera, I., Ducret-Stich, R. E., Ineichen, A., Davey, M., Phuleria, H. C., Probst-Hensch, N., Tsai, M. Y., & Künzli, N. (2015). Modeling indoor air pollution of outdoor origin in homes of SAPALDIA subjects in Switzerland. *Environment International*, 82, 85–91. <https://doi.org/10.1016/J.ENVINT.2015.05.013>
- Menut, L., Bessagnet, B., Khvorostyanov, D., Beekmann, M., Blond, N., Colette, A., Coll, I., Curci, G., Foret, G., Hodzic, A., Mailler, S., Meleux, F., Monge, J.-L., Pison, I., Siour, G., Turquety, S., Valari, M., Vautard, R., & Vivanco, M. G. (2013). CHIMERE 2013: a model for regional atmospheric composition modelling. *Geoscientific Model Development*, 6(4), 981–1028. <https://doi.org/10.5194/gmd-6-981-2013>
- Menut, Laurent, Bessagnet, B., Briant, R., Cholakian, A., Couvidat, F., Mailler, S., Pennel, R., Siour, G., Tuccella, P., Turquety, S., & Valari, M. (2021). *The CHIMERE v2020r1 online chemistry-transport model*.
- Menut, Laurent, Goussebaile, A., Bessagnet, B., Khvorostyanov, D., & Ung, A. (2012). Impact of realistic hourly emissions profiles on air pollutants concentrations modelled with CHIMERE. *Atmospheric Environment*, 49(x), 233–244. <https://doi.org/10.1016/j.atmosenv.2011.11.057>
- Molle, R., Mazoué, S., Géhin, É., & Ionescu, A. (2013). Indoor-outdoor relationships of airborne particles and nitrogen dioxide inside Parisian buses. *Atmospheric Environment*, 69, 240–248. <https://doi.org/10.1016/j.atmosenv.2012.11.050>
- Monks, P., Granier, C., Fuzzi, S., Stohl, A., Williams, M., Akimoto, H., Amann, M., Baklanov, A., Baltensperger, U., Bey, I., Blake, N., Blake, R., Carslaw, K., Cooper, O., Dentener, F., Fowler, D., Fragkou, E., Frost, G., Generoso, S., ... von Glasow, R. (2009). Atmospheric composition change - global and regional air quality. *Atmospheric Environment*, 5268–5350. <https://doi.org/10.1016/j.atmosenv.2009.08.021>
- Morawska, L., Thai, P. K., Liu, X., Asumadu-Sakyi, A., Ayoko, G., Bartonova, A., Bedini, A., Chai, F., Christensen, B., Dunbabin, M., Gao, J., Hagler, G. S. W., Jayaratne, R., Kumar, P., Lau, A. K. H., Louie, P. K. K., Mazaheri, M., Ning, Z., Motta, N., ... Williams, R. (2018). Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: How far have they gone? *Environment International*, 116, 286–299. <https://doi.org/10.1016/J.ENVINT.2018.04.018>
- Naidoo, R. N. (2017). NO₂ increases the risk for childhood asthma: a global concern. In *The Lancet Planetary Health* (Vol. 3, Issue 4, pp. e155–e156). Elsevier B.V. [https://doi.org/10.1016/S2542-5196\(19\)30059-2](https://doi.org/10.1016/S2542-5196(19)30059-2)
- Nedelec, V., Mosqueron, L., Desqueyroux, H., & Jeannee, N. (2010). Effets des normes EURO IV et V sur la réduction des impacts sanitaires du trafic routier urbain en France. *Environnement, Risques et Santé*, 9(6), 503–515. <https://doi.org/10.1684/ers.2010.0394>
- Nezis, I., Biskos, G., Eleftheriadis, K., & Kalantzi, O. I. (2019). Particulate matter and health effects in offices-A review. *Building and Environment*, 156, 62-73.
- Ni Riain, C. M., Mark, D., Davies, M., Harrison, R. M., & Byrne, M. A. (2003). Averaging periods for indoor-outdoor ratios of pollution in naturally ventilated non-domestic buildings near a busy road.

Atmospheric Environment, 37, 4121–4132. [https://doi.org/10.1016/S1352-2310\(03\)00509-0](https://doi.org/10.1016/S1352-2310(03)00509-0)

- OECD. (2020). *Non-exhaust Particulate Emissions from Road Transport : An Ignored Environmental Policy Challenge*. <https://doi.org/https://doi.org/10.1787/4a4dc6ca-en>
- Olmo, N. R. S., Saldiva, P. H. do N., Braga, A. L. F., Lin, C. A., Santos, U. de P., & Pereirai, L. A. A. (2011). A review of low-level air pollution and adverse effects on human health: Implications for epidemiological studies and public policy. *Clinics*, 66(4), 681–690. <https://doi.org/10.1590/S1807-59322011000400025>
- Özkaynak, H., Baxter, L. K., Dionisio, K. L., & Burke, J. (2013). Air pollution exposure prediction approaches used in air pollution epidemiology studies. *Journal of Exposure Science and Environmental Epidemiology*, 23(6), 566–572. <https://doi.org/10.1038/jes.2013.15>
- Park, Y. M. (2020). Assessing personal exposure to traffic-related air pollution using individual travel-activity diary data and an on-road source air dispersion model. *Health & Place*, 63, 10. <https://doi.org/10.1016/J.HEALTHPLACE.2020.102351>
- Park, Y. M., & Kwan, M. P. (2017). Individual exposure estimates may be erroneous when spatiotemporal variability of air pollution and human mobility are ignored. *Health and Place*, 43(February 2016), 85–94. <https://doi.org/10.1016/j.healthplace.2016.10.002>
- Pascal, M., Yvon, J.-M., Medina, S., Blanchard, M., Corso, M., Host, S., & Lagarrigue, R. (2019). *Pollution atmosphérique. Guide pour la réalisation d'une évaluation quantitative des impacts sur la santé (EQIS). EQIS avec une exposition modélisée*. Santé Publique France. <https://www.santepubliquefrance.fr/determinants-de-sante/pollution-et-sante/air/documents/guide/pollution-atmospherique.-guide-pour-la-realisation-d-une-evaluation-quantitative-des-impacts-sur-la-sante-eqis-.eqis-avec-une-exposition-modelisee>
- Passi, A., Nagendra, S. M. S., & Maiya, M. P. (2021). Characteristics of indoor air quality in underground metro stations: A critical review. *Building and Environment*, 198, 107907. <https://doi.org/10.1016/J.BUILDENV.2021.107907>
- Perez, L., Declercq, C., Inñiguez, C., Aguilera, I., Badaloni, C., Ballester, F., Bouland, C., Chanel, O., Cirarda, F. B., Forastiere, F., Forsberg, B., Haluza, D., Hedlund, B., Cambra, K., Lacasanã, M., Moshhammer, H., Otorespec, P., Rodríguez-Barranco, M., Medina, S., & Kuñzli, N. (2013). Chronic burden of near-roadway traffic pollution in 10 European cities (APHEKOM network). *European Respiratory Journal*, 42(3), 594–605. <https://doi.org/10.1183/09031936.00031112>
- Petetin, H., Sciare, J., Bressi, M., Gros, V. rie, Rosso, A., Sanchez, O., Sarda-Est ve, R., Petit, J. E., & Beekmann, M. (2016). Assessing the ammonium nitrate formation regime in the Paris megacity and its representation in the CHIMERE model. *Atmospheric Chemistry and Physics*, 16(16), 10419–10440. <https://doi.org/10.5194/acp-16-10419-2016>
- Pilla, F., & Broderick, B. (2015). A GIS model for personal exposure to PM10 for Dublin commuters. *Sustainable Cities and Society*, 15, 1–10. <https://doi.org/10.1016/j.scs.2014.10.005>
- Poom, A., Willberg, E., & Toivonen, T. (2021). Environmental exposure during travel: A research review and suggestions forward. *Health & Place*, 70, 102584. <https://doi.org/10.1016/j.healthplace.2021.102584>
- Pope III, C. A., Burnett, R. T., Thun, M. J., Calle, E. E., Krewski, D., & Thurston, G. D. (2002). Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *The Journal of the American Medical Association*, 287(9), 1132–1141. <https://doi.org/10.1001/jama.287.9.1132>
- Powers, J. G., KlemP, J. B., sKamaroCK, william C., davis, C., dudhia, J., Gill, david, Coen, J., GoChis, david J., ahmadov, ravan, PeCKham, steven, Grell, G., miChalaKes, J., trahan, samuel, BenJamin, stanley G., alexander, C., dimeGo, G. J., wanG, wei, sSchwartz, C., romine, G., ... duda, miChael G. (2017).

The weather research and forecasting model - Overview, System Efforts, and Future Directions.
<https://doi.org/10.1175/BAMS-D-15-00308.1>

Prevaire. (n.d.). *PREVAIR*. Retrieved May 10, 2022, from <http://www2.prevaire.org/>

Pun, B. K., Griffin, R. J., Seigneur, C., & Seinfeld, J. H. (2002). Secondary organic aerosol 2. Thermodynamic model for gas/particle partitioning of molecular constituents. *Journal of Geophysical Research: Atmospheres*, 107(D17), AAC 4-1. <https://doi.org/10.1029/2001JD000542>

Querol X, Moreno T, Karanasiou A, Reche C, Alastuey A, Viana M, et al. Variability of levels and composition of PM₁₀ and PM_{2.5} in the Barcelonametro system. *Atmos Chem Phys* 2012;12:5055–76

Raaschou-Nielsen, O., Andersen, Z. J., Beelen, R., Samoli, E., Stafoggia, M., Weinmayr, G., Hoffmann, B., Fischer, P., Nieuwenhuijsen, M. J., Brunekreef, B., Xun, W. W., Katsouyanni, K., Dimakopoulou, K., Sommar, J., Forsberg, B., Modig, L., Oudin, A., Oftedal, B., Schwarze, P. E., ... Hoek, G. (2013). Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE). *The Lancet Oncology*, 14(9), 813–822. [https://doi.org/10.1016/S1470-2045\(13\)70279-1](https://doi.org/10.1016/S1470-2045(13)70279-1)

Rajagopalan, S., & Brook, R. D. (2012). Air pollution and type 2 diabetes: Mechanistic insights. *Diabetes*, 61(12), 3037–3045. <https://doi.org/10.2337/db12-0190>

Raut J-C, Chazette P, Fortain A. Link between aerosol optical, microphysical and chemical measurements in an underground railway station in Paris. *Atmos Environ* 2009;43: 860–8

Reich, B. J., Chang, H. H., & Foley, K. M. (2014). A spectral method for spatial downscaling. *Biometrics*, 70(4), 932–942. <https://doi.org/10.1111/biom.12196>

Ribalaygua, J., Torres, L., Pórtolos, J., Monjo, R., Gaitán, E., & Pino, M. R. (2013). Description and validation of a two-step analogue/regression downscaling method. *Theoretical and Applied Climatology*, 114(1–2), 253–269. <https://doi.org/10.1007/S00704-013-0836-X/TABLES/2>

Ripanucci G, Grana M, Vicentini L, Magrini A, Bergamaschi A. Dust in the underground Railway Tunnels of an Italian Town. *J Occup Environ Hyg* 2006;3:16–25

Rivas, I., Kumar, P., Hagen-Zanker, A., Andrade, M. de F., Slovic, A. D., Pritchard, J. P., & Geurs, K. T. (2017). Determinants of black carbon, particle mass and number concentrations in London transport microenvironments. *Atmospheric Environment*, 161, 247–262. <https://doi.org/10.1016/J.ATMOENV.2017.05.004>

Roberts, H., & Helbich, M. (2021). Multiple environmental exposures along daily mobility paths and depressive symptoms: A smartphone-based tracking study. *Environment International*, 156, 106635. <https://doi.org/10.1016/J.ENVINT.2021.106635>

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*, 1(X), 100010. <https://doi.org/10.1016/j.aeaoa.2019.100010>

Rolstad Denby, B., Gauss, M., Wind, P., Mu, Q., Grøtting Wærsted, E., Fagerli, H., Valdebenito, A., & Klein, H. (2020). Description of the uEMEP_v5 downscaling approach for the EMEP MSC-W chemistry transport model. *Geoscientific Model Development Discussions*, 1–38. <https://doi.org/10.5194/gmd-2020-119>

Rotko, T., Oglesby, L., Künzli, N., Carrer, P., Nieuwenhuijsen, M. J., & Jantunen, M. (2002). Determinants of perceived air pollution annoyance and association between annoyance scores and air pollution (PM_{2.5}, NO₂) concentrations in the European EXPOLIS study. *Atmospheric Environment*, 36(29), 4593-4602.

Rybarczyk, Y., & Zalakeviciute, R. (2018). Machine Learning Approaches for Outdoor Air Quality

- Modelling: A Systematic Review. *Applied Sciences* 2018, Vol. 8, Page 2570, 8(12), 2570. <https://doi.org/10.3390/APP8122570>
- Salonen, H., Salthammer, T., & Morawska, L. (2019). Human exposure to NO₂ in school and office indoor environments. *Environment International*, 130, 104887. <https://doi.org/10.1016/J.ENVINT.2019.05.081>
- Schaap, M., Cuvelier, C., Hendriks, C., Bessagnet, B., Baldasano, J. M., Colette, A., Thunis, P., Karam, D., Fagerli, H., Graff, A., Kranenburg, R., Nyiri, A., Pay, M. T., Rouil, L., Schulz, M., Simpson, D., Stern, R., Terrenoire, E., & Wind, P. (2015). Performance of European chemistry transport models as function of horizontal resolution. *Atmospheric Environment*, 112, 90–105. <https://doi.org/10.1016/J.ATMOSENV.2015.04.003>
- Schmitz, O., Beelen, R., Strak, M., Hoek, G., Soenario, I., Brunekreef, B., Vaartjes, I., Dijst, M. J., Grobbee, D. E., & Karssenberg, D. (2019). Data descriptor: High resolution annual average air pollution concentration maps for the netherlands. *Scientific Data*, 6, 1–12. <https://doi.org/10.1038/sdata.2019.35>
- Schoof, J. T., & Pryor, S. C. (2001). Downscaling temperature and precipitation: a comparison of regression-based methods and artificial neural networks. *International Journal of Climatology*, 21(7), 773–790. <https://doi.org/10.1002/JOC.655>
- Schulz, H., Harder, V., Ibaldo-Mulli, A., Khandoga, A., Koenig, W., Krombach, F., Radykewicz, R., Stampfl, A., Thorand, B., & Peters, A. (2005). Cardiovascular effects of fine and ultrafine particles. In *Journal of Aerosol Medicine: Deposition, Clearance, and Effects in the Lung* (Vol. 18, Issue 1, pp. 1–22). <https://doi.org/10.1089/jam.2005.18.1>
- Seaton A, Cherrie J, Dennekamp M, Donaldson K, Hurley JF, Tran CL. The London under- ground: dust and hazards to health. *Occup Environ Med* 2005;62:355–62
- Serveau, L., Taïeb, N., Chang, J.-P., Vincent, J., Allemand, N., Boutang, J., Andre, J.-M., Bort, R., Bouchard, D., Deflorenne, E., Druart, A., Dulhoste, S., Durand, A., Gavel, A., Gueguen, C., Jacquier, G., Jeannot, C., Lemaire, A., Mathias, E., ... Ringuet, J. (2015). *Inventaire des émissions de polluants atmosphériques et de gaz à effet de serre en France-Séries sectorielles et analyses étendues*. <http://www.citepa.org/fr/activites/inventaires-des-emissions/secten>
- Setton, E., Marshall, J. D., Brauer, M., Lundquist, K. R., Hystad, P., Keller, P., & Cloutier-Fisher, D. (2010). The impact of daily mobility on exposure to traffic-related air pollution and health effect estimates. *Journal of Exposure Science & Environmental Epidemiology* 2011 21:1, 21(1), 42–48. <https://doi.org/10.1038/jes.2010.14>
- Silibello, C., Carlino, G., Stafoggia, M., Gariazzo, C., Finardi, S., Pepe, N., Radice, P., Forastiere, F., & Viegi, G. (2021). Spatial-temporal prediction of ambient nitrogen dioxide and ozone levels over Italy using a Random Forest model for population exposure assessment. *Air Quality, Atmosphere and Health*, 14(6), 817–829. <https://doi.org/10.1007/S11869-021-00981-4>
- Silveira, C., Ferreira, J., & Miranda, A. I. (2019). The challenges of air quality modelling when crossing multiple spatial scales. *Air Quality, Atmosphere and Health*, 12(9), 1003–1017. <https://doi.org/10.1007/s11869-019-00733-5>
- Simoni, M., Annesi-Maesano, I., Sigsgaard, T., Norback, D., Wieslander, G., Nystad, W., ... & Viegi, G. (2010). School air quality related to dry cough, rhinitis and nasal patency in children. *European Respiratory Journal*, 35(4), 742–749.
- Siour, G., Colette, A., Menut, L., Bessagnet, B., Coll, I., & Meleux, F. (2013). Bridging the scales in a eulerian air quality model to assess megacity export of pollution. *Environmental Modelling & Software*, 46, 271–282. <https://doi.org/10.1016/J.ENVSOFT.2013.04.001>
- Smolak, K., Rohm, W., Knop, K., & Siła-Nowicka, K. (2020). Population mobility modelling for mobility

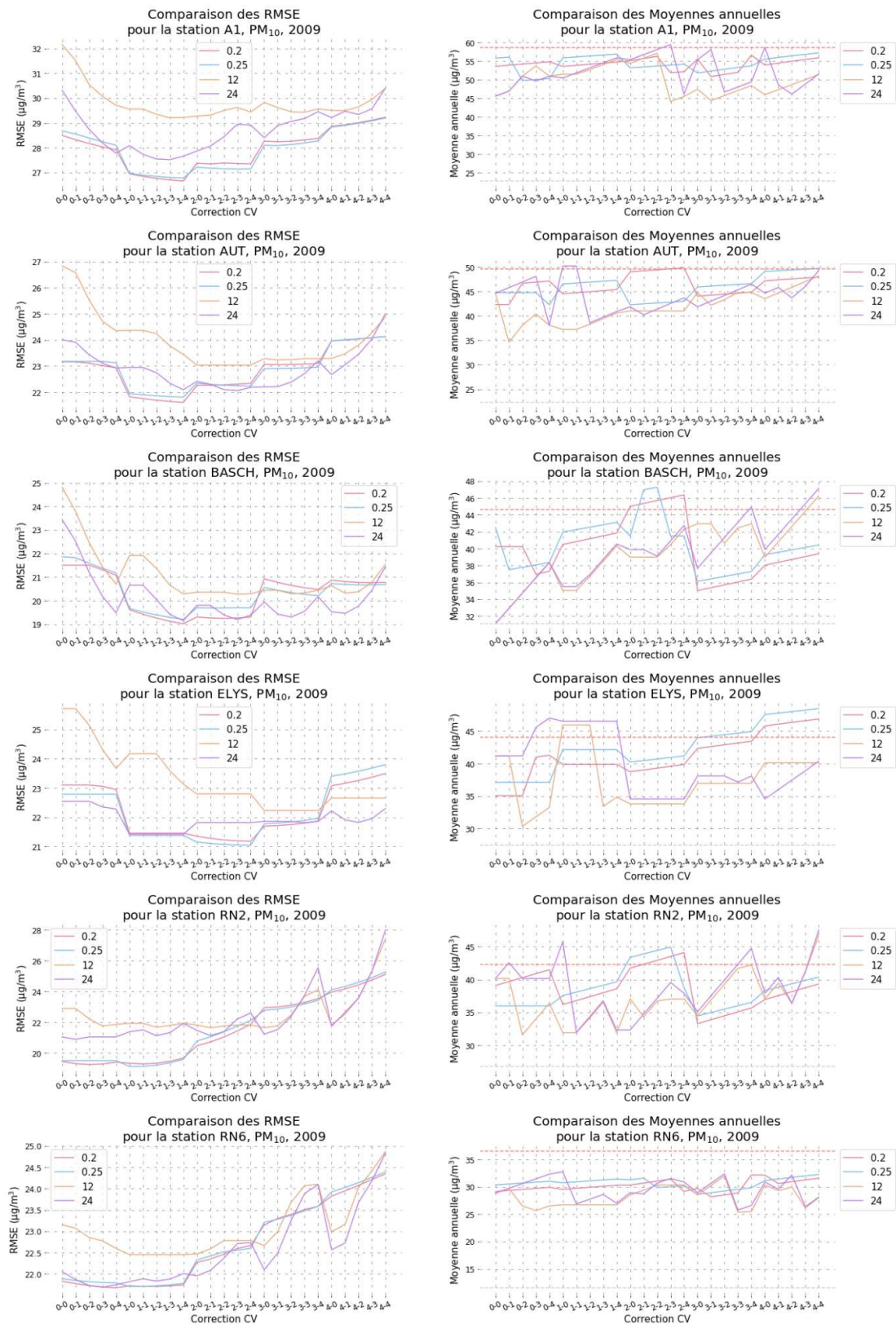
- data simulation. *Computers, Environment and Urban Systems*, 84, 101526. <https://doi.org/10.1016/J.COMPENVURBSYS.2020.101526>
- Snyder, L. P. (1994). "The Death-Dealing Smog Over Donora, Pennsylvania": Industrial Air Pollution, Public Health Policy, and the Politics of Expertise, 1948–1949. *Environmental History Review*, 18(1), 117–139. <https://doi.org/10.2307/3984747>
- Stewart, I. D., & Oke, T. R. (2012). Local climate zones for urban temperature studies. *Bulletin of the American Meteorological Society*, 93(12), 1879–1900. <https://doi.org/10.1175/BAMS-D-11-00019.1>
- Strak M, Boogaard H, Meliefste K, Oldenwening M, Zuurbier M, Brunekreef B, et al. Respiratory health effects of ultrafine and fine particle exposure in cyclists. *Occup Environ Med* 2010;67(2):118–24
- Strasser, G., Hiebaum, S., & Neuberger, M. (2018). Commuter exposure to fine and ultrafine particulate matter in Vienna. *Wiener Klinische Wochenschrift*, 130(1–2), 62–69. <https://doi.org/10.1007/S00508-017-1274-Z>
- Su, X., Gough, W., & Shen, Q. (2016). Correlation of PM2.5 and meteorological variables in Ontario cities: statistical downscaling method coupled with artificial neural network. *Air Pollution XXIV*, 1(Air), 215–226. <https://doi.org/10.2495/air160201>
- Terrenoire, E., Bessagnet, B., Rouil, L., Tognet, F., Pirovano, G., Létinois, L., Beauchamp, M., Colette, A., Thunis, P., Amann, M., & Menut, L. (2015). High-resolution air quality simulation over Europe with the chemistry transport model CHIMERE. *Geoscientific Model Development*, 8(1), 21–42. <https://doi.org/10.5194/gmd-8-21-2015>
- Thunis, P., Rouil, L., Cuvelier, C., Stern, R., Kerschbaumer, A., Bessagnet, B., Schaap, M., Builtjes, P., Tarrason, L., Douros, J., Moussiopoulos, N., Pirovano, G., & Bedogni, M. (2007). Analysis of model responses to emission-reduction scenarios within the CityDelta project. *Atmospheric Environment*, 41(1), 208–220. <https://doi.org/10.1016/j.atmosenv.2006.09.001>
- Tonne, C., Milà, C., Fecht, D., Alvarez, M., Gulliver, J., Smith, J., Beevers, S., Ross Anderson, H., & Kelly, F. (2018). Socioeconomic and ethnic inequalities in exposure to air and noise pollution in London. *Environment International*, 115, 170–179. <https://doi.org/10.1016/J.ENVINT.2018.03.023>
- Troen, I. B., & Mahrt, L. (1985). *A simple model of the atmospheric boundary layer - Sensitivity to surface evaporation*.
- Urban air pollution – what are the main sources across the world? | EU Science Hub*. (n.d.). Retrieved February 7, 2022, from <https://ec.europa.eu/jrc/en/news/what-are-main-sources-urban-air-pollution>
- Valari, M., & Menut, L. (2008). Does an increase in air quality models' resolution bring surface ozone concentrations closer to reality? *Journal of Atmospheric and Oceanic Technology*, 25(11), 1955–1968. <https://doi.org/10.1175/2008JTECHA1123.1>
- Valari, M., & Menut, L. (2010). Transferring the heterogeneity of surface emissions to variability in pollutant concentrations over urban areas through a chemistry-transport model. *Atmospheric Environment*, 44(27), 3229–3238. <https://doi.org/10.1016/J.ATMOENV.2010.06.001>
- 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. *Journal of the Air and Waste Management Association*, 61(2), 164–179. <https://doi.org/10.3155/1047-3289.61.2.164>
- Vardoulakis, S., Fisher, B. E. A., Pericleous, K., & Gonzalez-Flesca, N. (2003). Modelling air quality in street canyons: A review. *Atmospheric Environment*, 37(2), 155–182. [https://doi.org/10.1016/S1352-2310\(02\)00857-9](https://doi.org/10.1016/S1352-2310(02)00857-9)

- Vardoulakis, S., Giagloglou, E., Steinle, S., Davis, A., Smeuwenhoek, A., Galea, K. S., Dixon, K., & Crawford, J. O. (2020). Indoor Exposure to Selected Air Pollutants in the Home Environment: A Systematic Review. *Int. J. Environ. Res. Public Health*, *17*, 8972. <https://doi.org/10.3390/ijerph17238972>
- Vedal, S., Brauer, M., White, R., & Petkau, J. (2003). Air Pollution and Daily Mortality in a City with Low Levels of Pollution. *Environmental Health Perspectives*, *111*(1). <https://doi.org/10.1289/ehp.5276>
- Vienneau, D., de Hoogh, K., Beelen, R., Fischer, P., Hoek, G., & Briggs, D. (2010). Comparison of land-use regression models between Great Britain and the Netherlands. *Atmospheric Environment*, *44*(5), 688–696. <https://doi.org/10.1016/J.ATMOSENV.2009.11.016>
- Vilà-Guerau de Arellano, J., Dosio, A., Vinuesa, J.-F., Holtzlag, A. A. M., & Galmarini, S. (2004). The dispersion of chemically reactive species in the atmospheric boundary layer. *Meteorology and Atmospheric Physics* *2004 87:1*, *87*(1), 23–38. <https://doi.org/10.1007/S00703-003-0059-2>
- Wang, D., & Sun, B. (2015). Detecting Activity Types and Trip Purposes from Passive GPS Data: A Data Mining Approach. *Space-Time Integration in Geography and GIScience: Research Frontiers in the US and China*, 211–234. https://doi.org/10.1007/978-94-017-9205-9_13
- Wang, R., Henderson, S. B., Sbihi, H., Allen, R. W., & Brauer, M. (2013). Temporal stability of land use regression models for traffic-related air pollution. *Atmospheric Environment*, *64*, 312–319. <https://doi.org/10.1016/j.atmosenv.2012.09.056>
- WHO. (2013). *Review of evidence on health aspects of air pollution-REVIHAAP Project Technical Report*. <http://www.euro.who.int/pubrequest>
- WHO. (2016). *Ambient air pollution: A global assessment of exposure and burden disease*.
- Wilson, W. E., & Suh, H. H. (1997). Fine particles and coarse particles: Concentration relationships relevant to epidemiologic studies. *Journal of the Air and Waste Management Association*, *47*(12), 1238–1249. <https://doi.org/10.1080/10473289.1997.10464074>
- Xu, B., & Hao, J. (2017). Air quality inside subway metro indoor environment worldwide: A review. *Environment International*, *107*, 33–46. <https://doi.org/10.1016/J.ENVINT.2017.06.016>
- Xu, M., Jin, J., Wang, G., Segers, A., Deng, T., & Lin, H. X. (2021). Machine learning based bias correction for numerical chemical transport models. *Atmospheric Environment*, *248*, 118022. <https://doi.org/10.1016/j.atmosenv.2020.118022>
- Yang, F., Kaul, D., Wong, K. C., Westerdahl, D., Sun, L., Ho, K. fai, Tian, L., Brimblecombe, P., & Ning, Z. (2015). Heterogeneity of passenger exposure to air pollutants in public transport microenvironments. *Atmospheric Environment*, *109*, 42–51. <https://doi.org/10.1016/j.atmosenv.2015.03.009>
- Zhang, Q. J., Beekmann, M., Drewnick, F., Freutel, F., Schneider, J., Crippa, M., Prevot, A. S. H., Baltensperger, U., Poulain, L., Wiedensohler, A., Sciare, J., Gros, V., Borbon, A., Colomb, A., Michoud, V., Doussin, J. F., Denier Van Der Gon, H. A. C., Haeffelin, M., Dupont, J. C., ... Perrussel, O. (2013). Formation of organic aerosol in the Paris region during the MEGAPOLI summer campaign: Evaluation of the volatility-basis-set approach within the CHIMERE model. *Atmospheric Chemistry and Physics*, *13*(11), 5767–5790. <https://doi.org/10.5194/acp-13-5767-2013>
- Zhang, X., Craft, E., Zhang, K., & Zhang, (K. (2017). *Characterizing spatial variability of air pollution from vehicle traffic around the Houston Ship Channel area*. <https://doi.org/10.1016/j.atmosenv.2017.04.032>
- Zuurbier, M., Hoek, G., Oldenwening, M., Lenters, V., Meliefste, K., van den Hazel, P., & Brunekreef, B. (2010). Commuters' exposure to particulate matter air pollution is affected by mode of transport, fuel type, and route. *Environmental Health Perspectives*, *118*(6), 783–789.

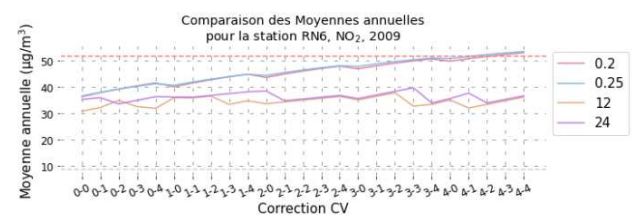
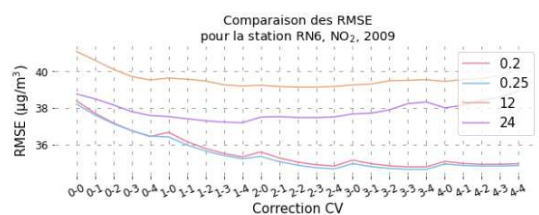
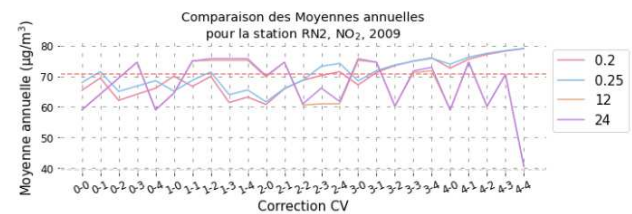
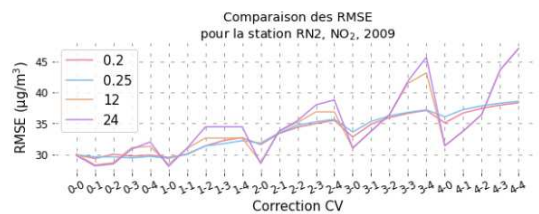
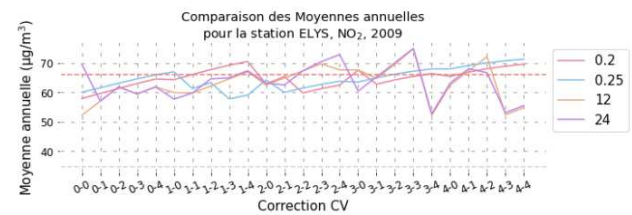
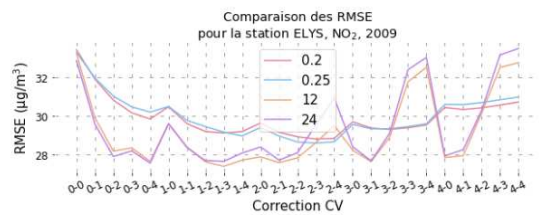
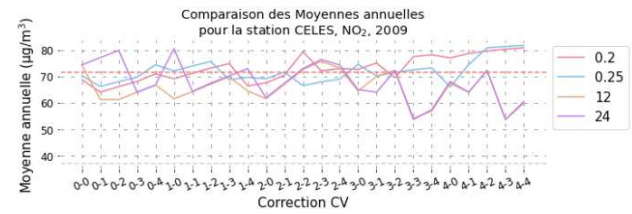
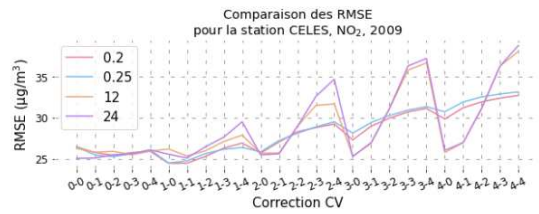
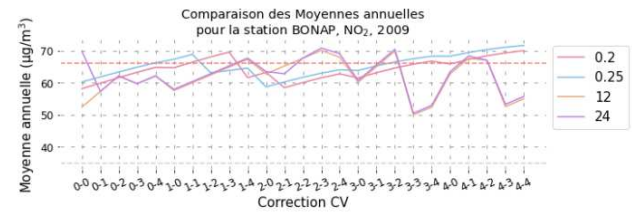
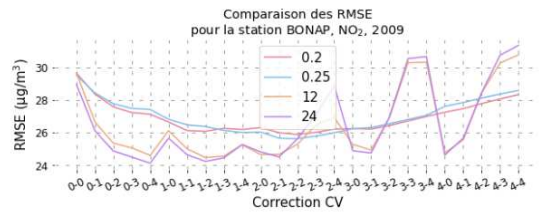
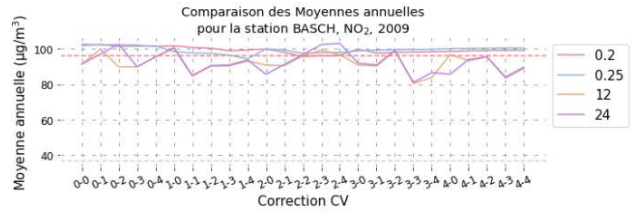
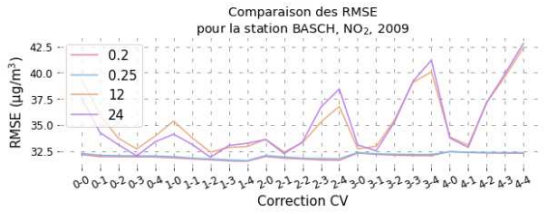
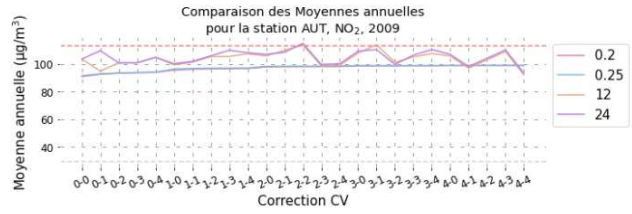
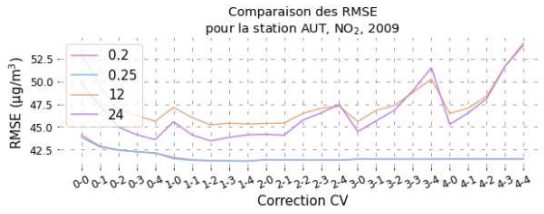
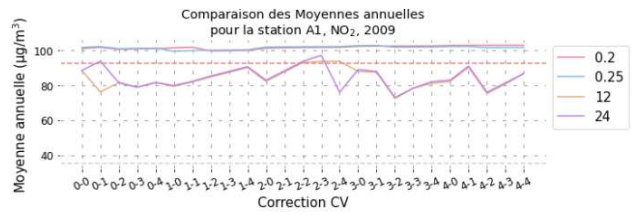
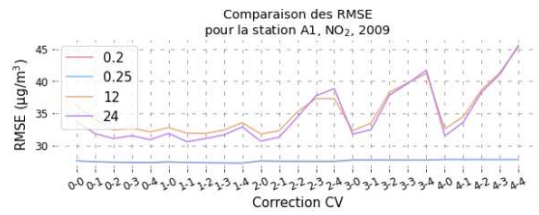
<https://doi.org/10.1289/ehp.0901622>

ANNEXE 1

Comparaison des RMSE et moyennes annuelles obtenus pour PM10



Comparaison des RMSE et moyennes annuelles obtenus pour NO2



ANNEXE 2

Données comparées à la mesure pour l'année 2009.

Tableaux de données de NO₂ – Les moyennes sont écrites en bleu dans les tableaux.

Stations de trafic										
Station	Coef. α	Coef. β	Coef. k	RMSE CHIMERE ($\mu\text{g}/\text{m}^3$)	RMSE Cprox ($\mu\text{g}/\text{m}^3$)	Moyenne CHIMERE ($\mu\text{g}/\text{m}^3$)	Moyenne Cprox ($\mu\text{g}/\text{m}^3$)	Moyenne Mesure ($\mu\text{g}/\text{m}^3$)	% α	% β
A1	1,5	0,5	2	63,1	32,8	35,5	90,8	93,2	75	25
AUT	2,4	0,7	3,1	93,8	46,4	29,6	113,9	114,1	77	23
BASCH	1	0,9	1,9	68,1	35,1	36,6	96	96,9	53	47
BONAP	0,8	0,7	1,5	46,3	29,7	34,9	70,4	71,6	53	47
CELES	0,6	0,8	1,4	43,5	30,6	37,2	73,4	74,2	43	57
ELYS	0,6	0,8	1,4	44,2	30,2	34,8	67,5	67,9	43	57
RN6	2,7	0,8	3,5	50,5	37,2	8,8	39,4	51,9	75	25
									57	43

Stations de fond										
Station	Coef. α	Coef. β	Coef. k	RMSE CHIMERE ($\mu\text{g}/\text{m}^3$)	RMSE Cprox ($\mu\text{g}/\text{m}^3$)	Moyenne CHIMERE ($\mu\text{g}/\text{m}^3$)	Moyenne Cprox ($\mu\text{g}/\text{m}^3$)	Moyenne Mesure ($\mu\text{g}/\text{m}^3$)	% α	% β
AUB	0,4	0,6	1	20,4	19,2	40,4	40,4	42,8	40	60
CACH	0,4	0,6	1	21,2	20,3	32,5	32,4	34,3	40	60
CERGY	1	0,6	1,6	15,3	12,9	11,4	18,2	19,4	63	38
GARCH	0,4	0,6	1	15,7	14,6	23,7	23,7	23,8	40	60
LOGNES	0,4	0,5	0,9	19,6	18,1	36	32,4	35,5	44	56
MANT	1,7	0,9	2,6	19,9	15,6	7,8	20,3	21,7	65	35
MONTG	0,7	0,6	1,3	17	15,6	19,3	25,1	26,8	54	46
NEUIL	0,4	0,7	1,1	22,6	21,2	35,1	38,6	42	36	64
NOGENT	0,6	0,7	1,3	21,5	19,1	27,7	36	37,5	46	54
PA13	0,8	0,7	1,5	19,8	17,4	17,7	26,6	29,2	53	47
PA18	0,6	0,6	1,2	22,2	20,4	29,1	34,9	35,7	50	50
TREMB	0,6	0,7	1,3	22,3	19,3	32,2	41,9	43,2	46	54
VERS	0,6	0,6	1,2	21,7	19,8	35,1	42,1	42,5	50	50
VILLEM	0,4	0,6	1	19,8	18,6	34,7	34,7	37,9	40	60
			1,29						48	52

Tableaux de données des PM₁₀ – Les moyennes sont écrites en bleu dans les tableaux.

Stations de trafic										
Station	Coef. α	Coef. β	Coef. k	RMSE CHIMERE ($\mu\text{g}/\text{m}^3$)	RMSE Cprox ($\mu\text{g}/\text{m}^3$)	Moyenne CHIMERE ($\mu\text{g}/\text{m}^3$)	Moyenne Cprox ($\mu\text{g}/\text{m}^3$)	Moyenne Mesure ($\mu\text{g}/\text{m}^3$)	% α	% β
A1	2,9	0,5	3,4	43	26,6	22,8	55,3	58,6	85	15
AUT	2	0,5	2,5	34,7	23,4	22,3	38,9	49,6	80	20
BASCH	1,7	0,2	1,9	28	19,6	31,1	44	44,8	89	11
ELYS	1,9	0,3	2,2	29,6	21,1	27,5	43,4	44,7	86	14
RN6	2,7	0,6	3,3	31	22	11,5	28,9	36,8	82	18
									85	15

Stations de fond											
Station	Coef. α	Coef. β	Coef. k	RMSE CHIMERE ($\mu\text{g}/\text{m}^3$)	RMSE Cprox ($\mu\text{g}/\text{m}^3$)	Moyenne CHIMERE ($\mu\text{g}/\text{m}^3$)	Moyenne Cprox ($\mu\text{g}/\text{m}^3$)	Moyenne Mesure ($\mu\text{g}/\text{m}^3$)	% α	% β	
GON	1,9	0,2	2,1	16,8	13,2	18	24,8	27,5	90	10	
LOGNES	2	0,2	2,2	17,6	13	16,4	23,9	27	91	9	
MELUN	2,1	0,3	2,4	18,1	14	11,6	18,4	25	88	13	
NOGENT	1,7	0,2	1,9	18,6	14,3	22,3	27,1	27,4	89	11	
TREMB	1,4	0,2	1,6	19	14,9	23	23,1	27,3	88	13	
VITRY	1,3	0,2	1,5	20,5	13,6	28,3	26,8	28,5	87	13	
			1,95							89	11

Tableaux de données de NO₂ – Comparaison des variances des sites de trafic.

Station	Étape d'apprentissage			Étape de généralisation
	Variance mesure (µg/m ³) ²	Variance Cprox (µg/m ³) ²	Variance CHIMERE (µg/m ³) ²	Variance Cprox (µg/m ³) ²
A1	990,3	926	435,5	698,8
AUT	2094,1	1470,5	366,8	518,8
BASCH	1451,1	937,2	408,2	1754,6
BONAP	859,9	515,5	396,4	413
CELES	684,1	642,3	401,2	460,2
ELYS	998,6	546,7	381,9	479
RN6	693,6	682,1	87,7	343

Tableaux de données de NO₂ – Comparaison des variances des sites de fond.

Station	Étape d'apprentissage			Étape de généralisation
	Variance mesure (µg/m ³) ²	Variance Cprox (µg/m ³) ²	Variance CHIMERE (µg/m ³) ²	Variance Cprox (µg/m ³) ²
AUB	545,9	254,7	403,8	352,7
CACH	566,6	324,8	453,4	478,3
CERGY	291,8	202,9	135,5	137,1
GARCH	365,8	223,1	316	326,6
LOGNES	411,6	330,7	225,9	200,1
MANT	279	243,4	61,7	63,4
MONTG	317,9	278,7	262,1	268,2
NEUIL	644,1	344,1	405,7	394,3
NOGENT	503	307,8	309,2	270
TREMB	411,9	248,6	276,5	244,3
VERS	399,2	380,5	252,6	275,2
VILLEM	523,9	268,8	333	288,1
PA07	531,9	406,2	376,9	364,4
PA13	464,8	241,9	375,4	338,6

Tableaux de données de PM₁₀ – Comparaison des variances des sites de trafic.

Station	Étape d'apprentissage			Étape de généralisation
	Variance mesure ($\mu\text{g}/\text{m}^3$) ²	Variance Cprox ($\mu\text{g}/\text{m}^3$) ²	Variance CHIMERE ($\mu\text{g}/\text{m}^3$) ²	Variance Cprox ($\mu\text{g}/\text{m}^3$) ²
A1	701,2	700,2	375,0	291,2
AUT	575,5	574,7	379,4	492,4
BASCH	438,3	428,5	613,9	813,1
ELYS	490,9	488,2	495,9	439,5
RN6	426,4	423,7	113,9	297,0

Tableaux de données de PM₁₀ – Comparaison des variances des sites de fond.

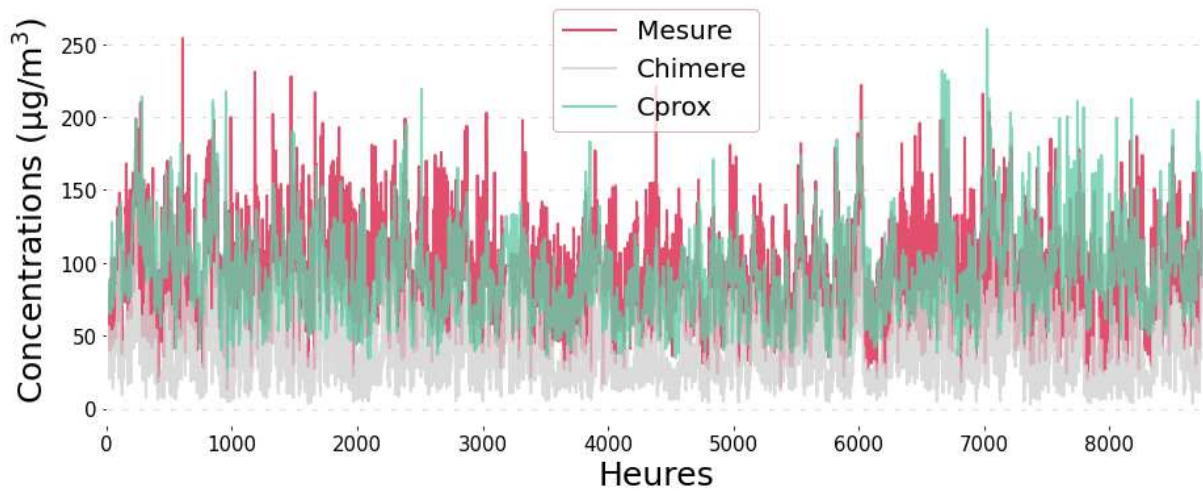
Station	Étape d'apprentissage			Étape de généralisation
	Variance mesure ($\mu\text{g}/\text{m}^3$) ²	Variance Cprox ($\mu\text{g}/\text{m}^3$) ²	Variance CHIMERE ($\mu\text{g}/\text{m}^3$) ²	Variance Cprox ($\mu\text{g}/\text{m}^3$) ²
GON	233,9	233,7	198,8	209,0
LOGNES	231,4	228,8	200,4	187,0
MELUN	225,8	217,6	116,6	155,8
NOGENT	282,8	274,4	364,0	297,5
TREMB	358,1	356,1	192,7	188,6
VITRY	266,2	254,8	417,6	389,5

ANNEXE 3

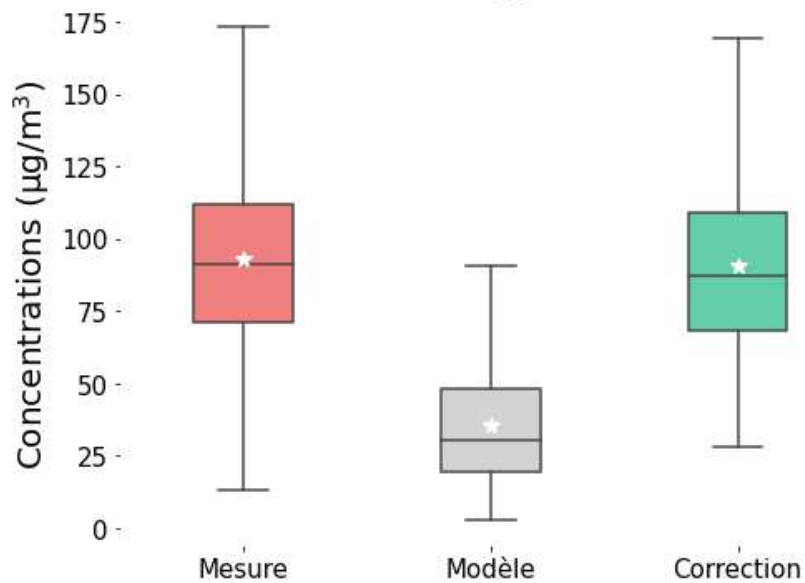
Résultat de la construction de C_{prox} via comparaison des données à la mesure.

Séries temporelles NO_2 – Sites de trafic.

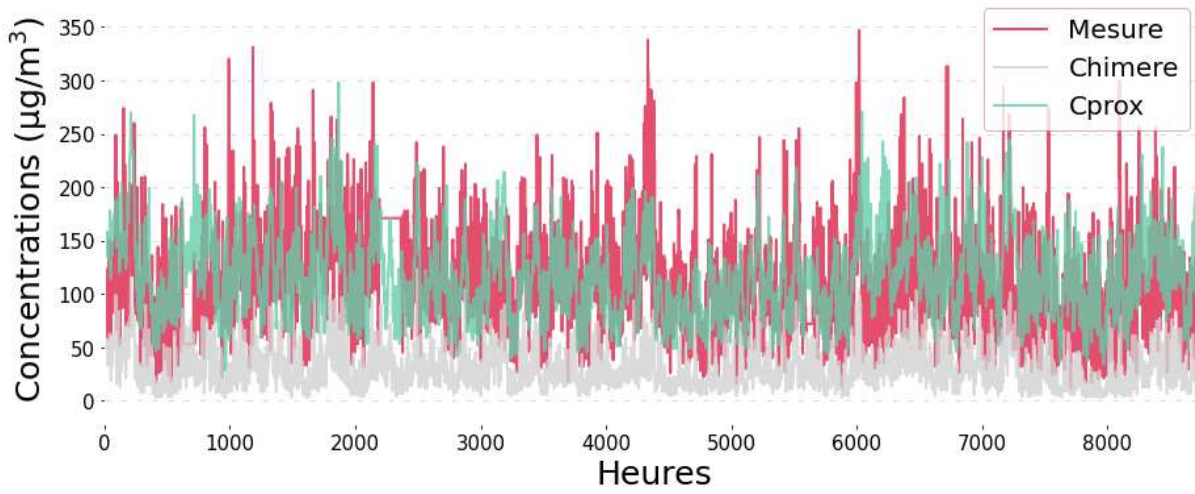
A1: Site de trafic (NO_2 , 2009)



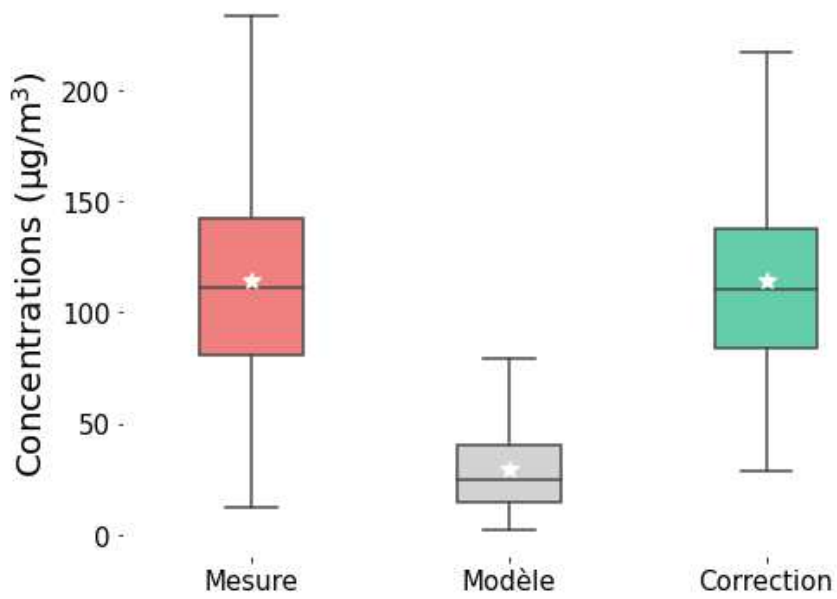
Comparaison des boxplots
A1 : NO_2 , 2009



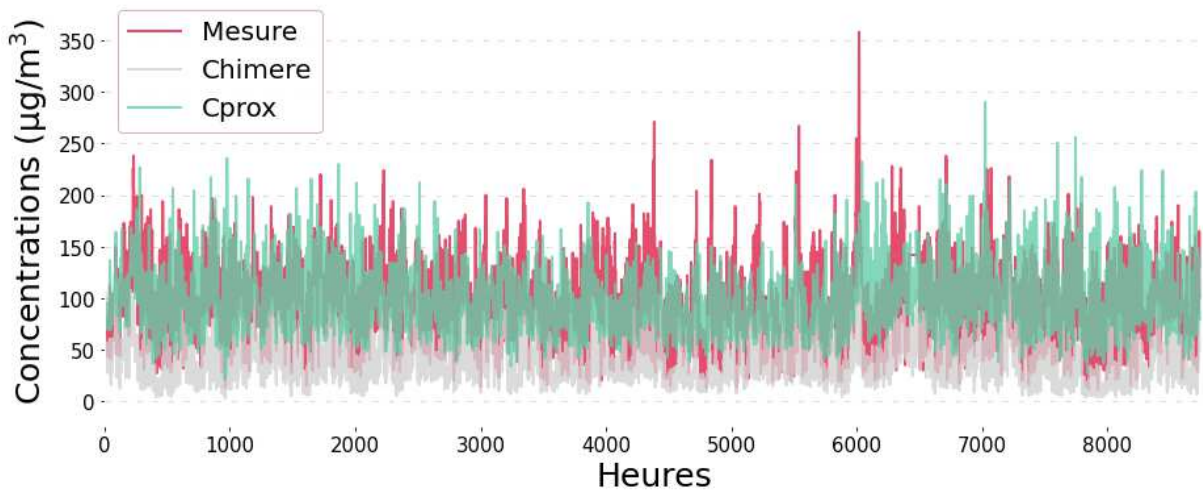
AUT: Site de trafic (NO₂, 2009)



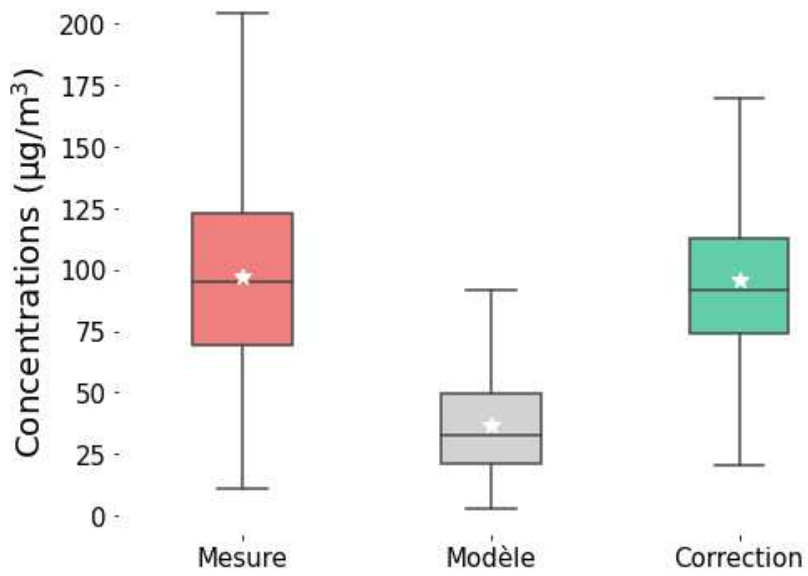
Comparaison des boxplots
AUT : NO₂, 2009



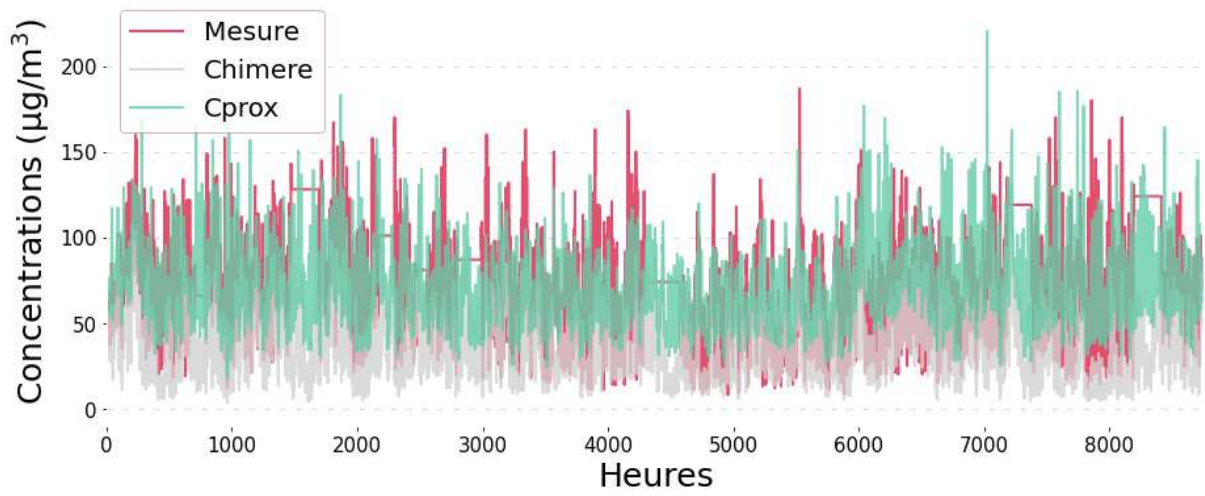
BASCH: Site de trafic (NO₂, 2009)



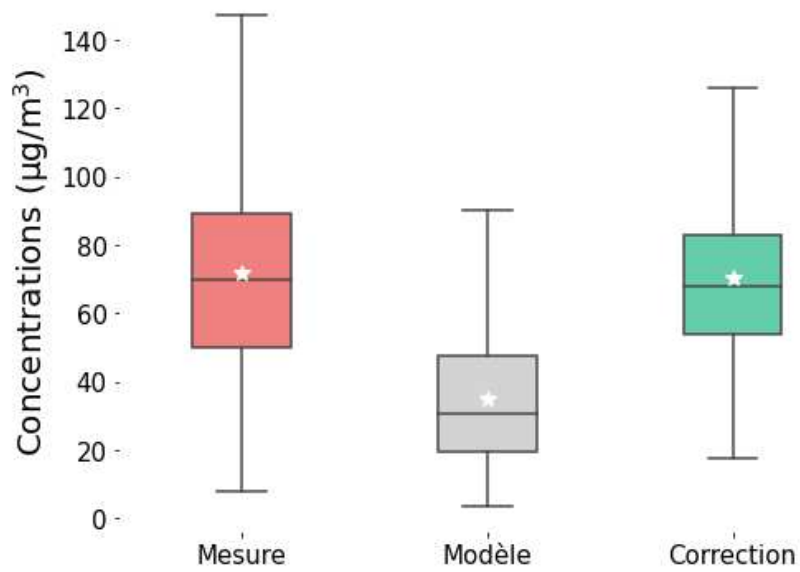
Comparaison des boxplots BASCH : NO₂, 2009



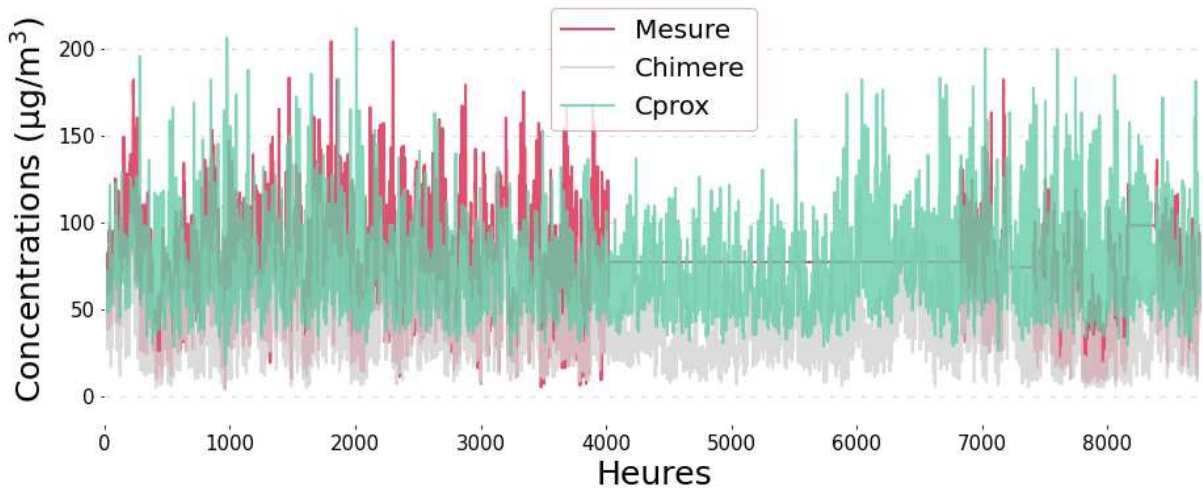
BONAP: Site de trafic (NO₂, 2009)



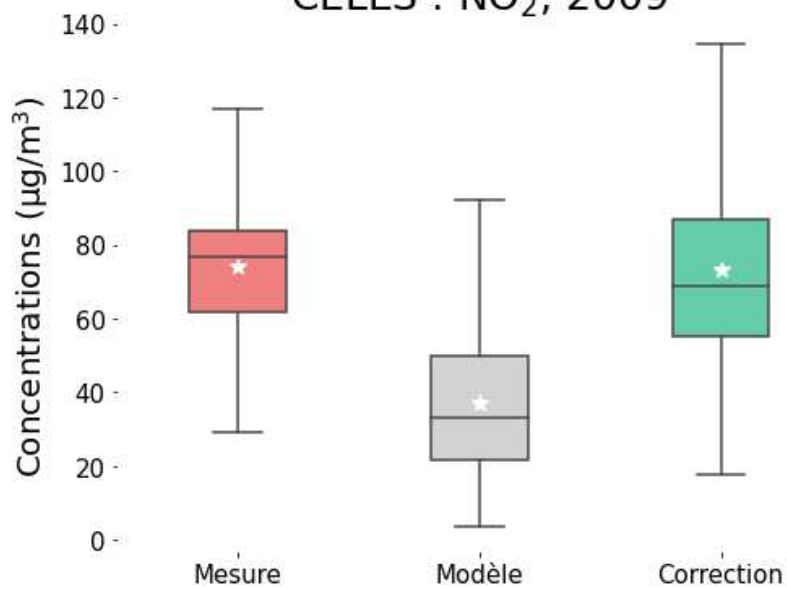
Comparaison des boxplots BONAP : NO₂, 2009



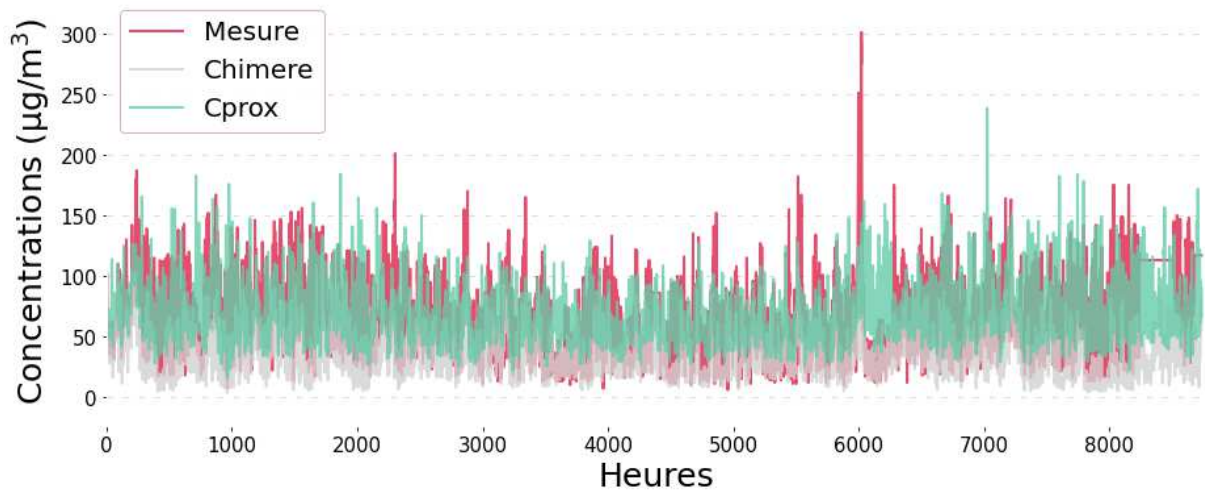
CELES: Site de trafic (NO₂, 2009)



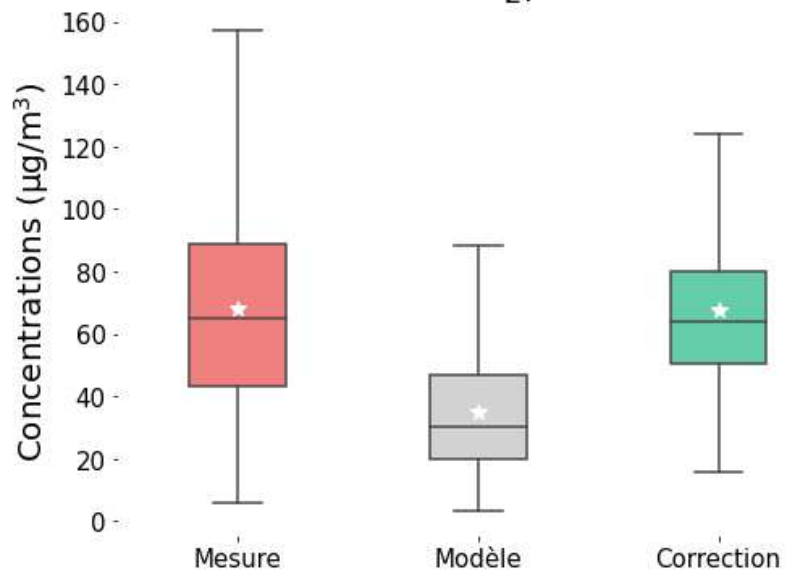
Comparaison des boxplots CELES : NO₂, 2009



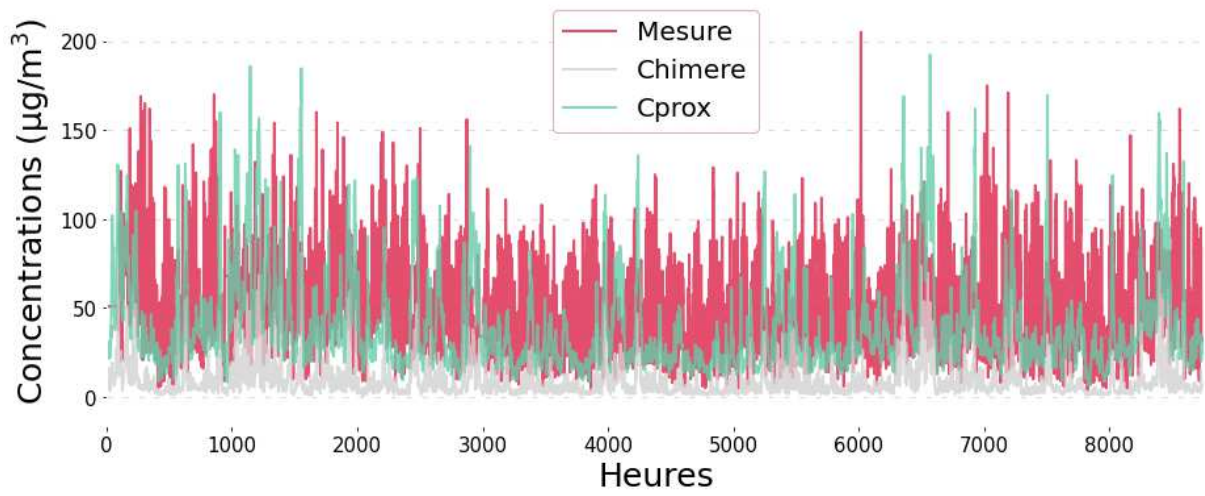
ELYS: Site de trafic (NO₂, 2009)



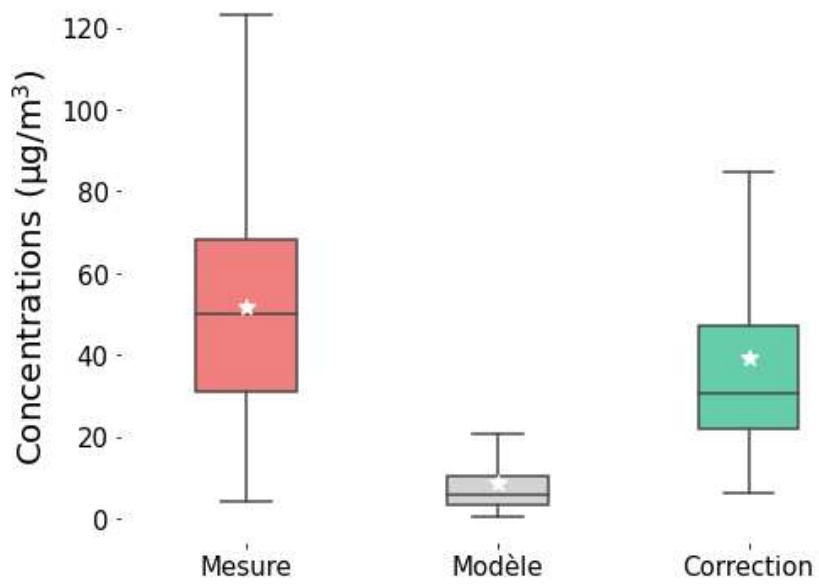
Comparaison des boxplots ELYS : NO₂, 2009



RN6: Site de trafic (NO₂, 2009)

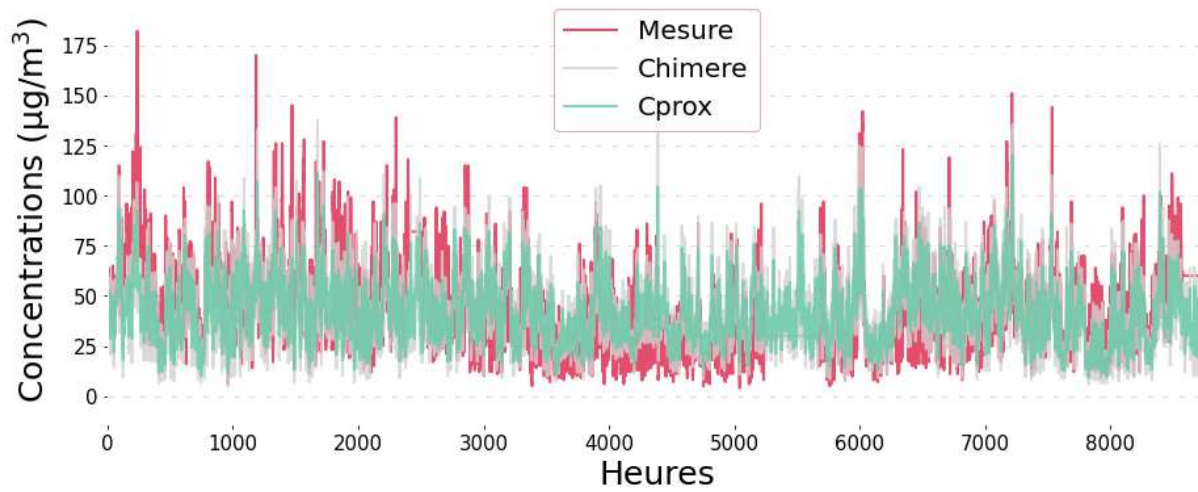


Comparaison des boxplots RN6 : NO₂, 2009

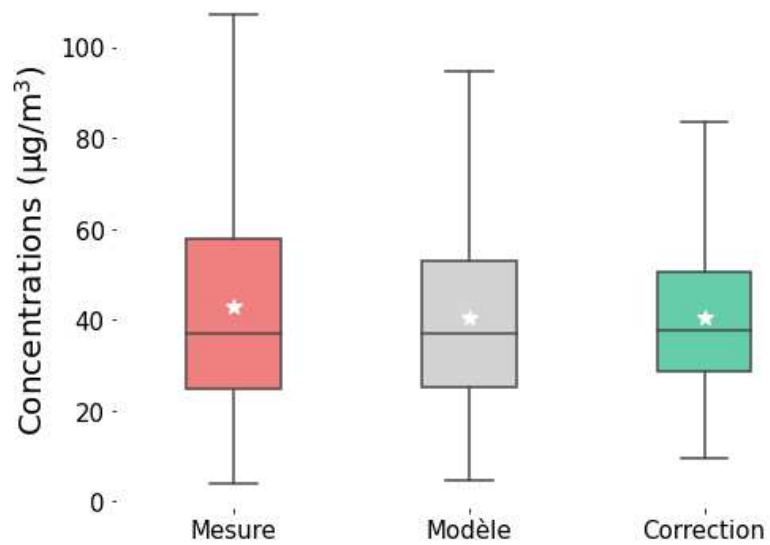


Séries temporelles NO₂ – Sites de fond.

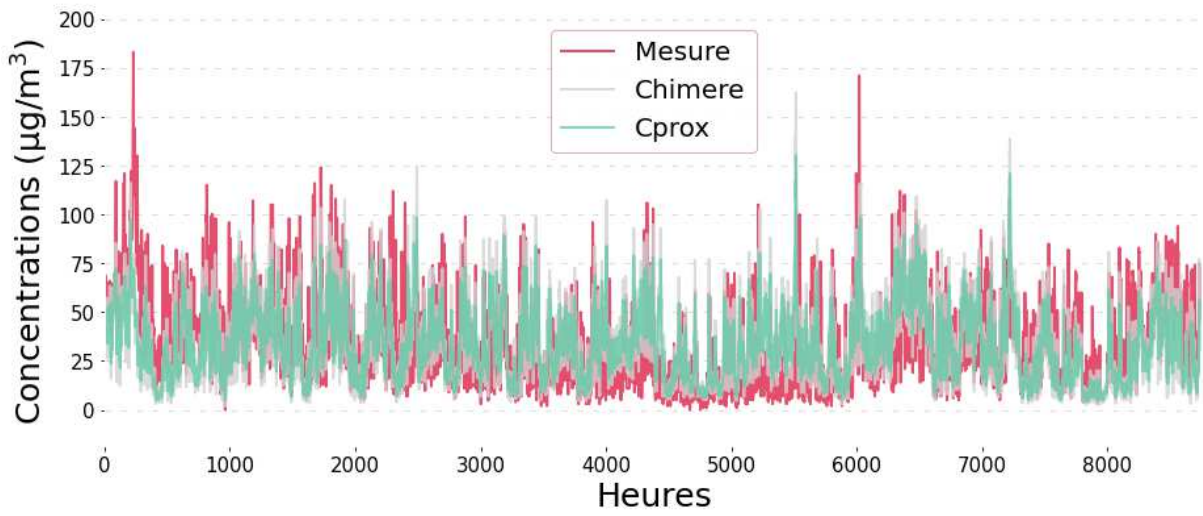
AUB: Site de fond (NO₂, 2009)



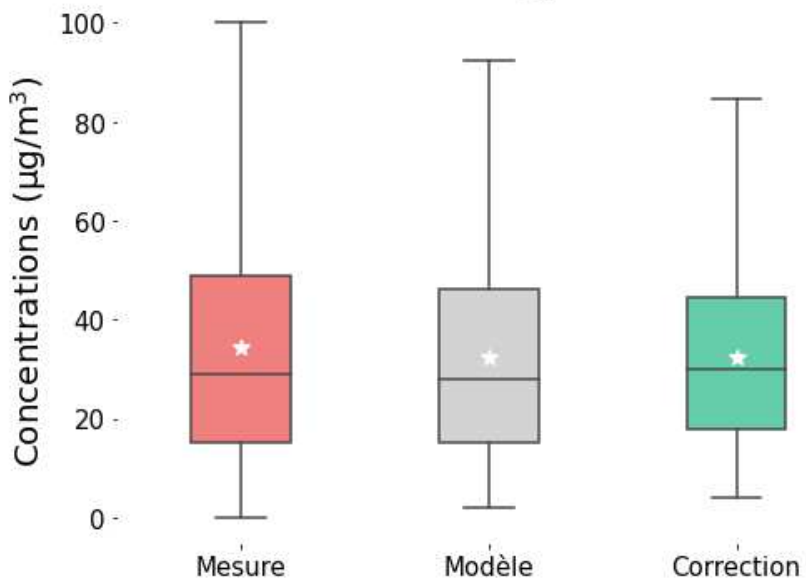
Comparaison des boxplots
AUB : NO₂, 2009



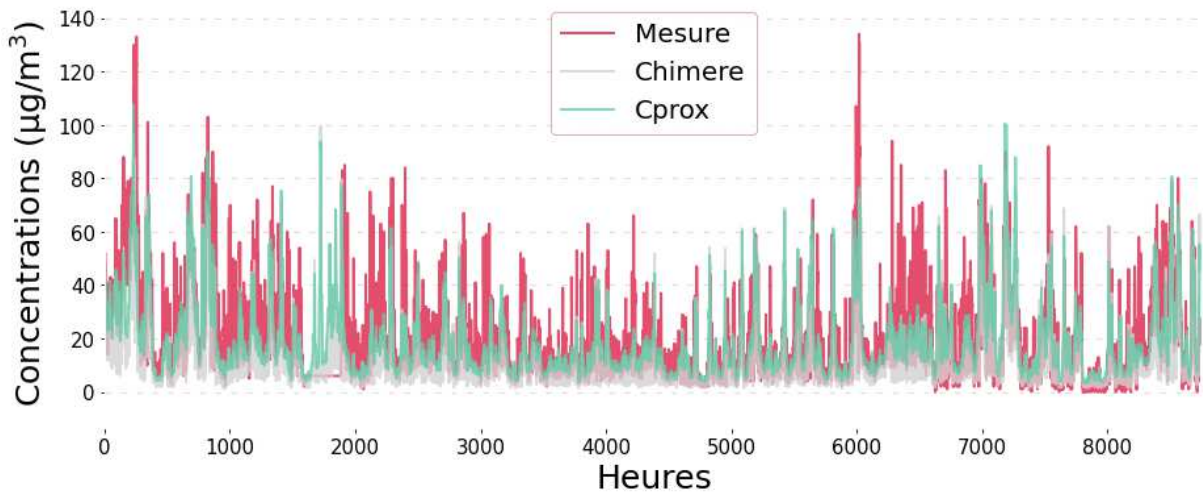
CACH: Site de fond (NO₂, 2009)



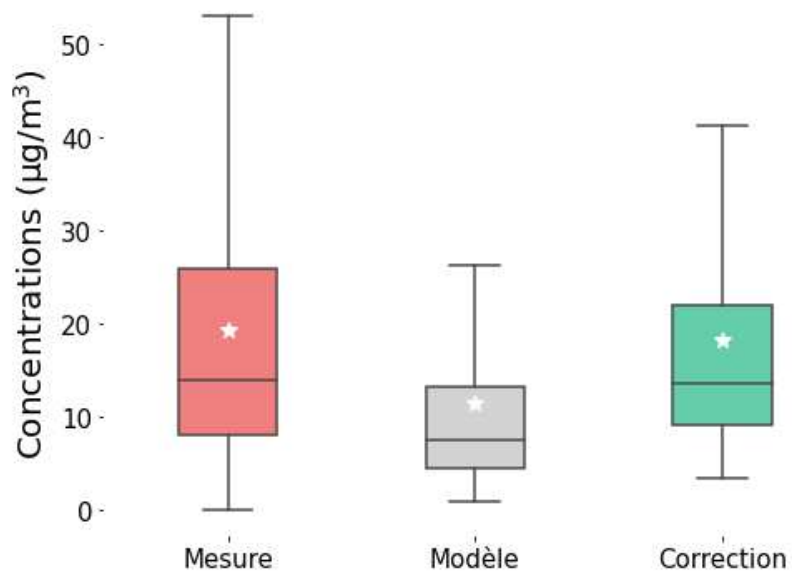
Comparaison des boxplots
CACH : NO₂, 2009



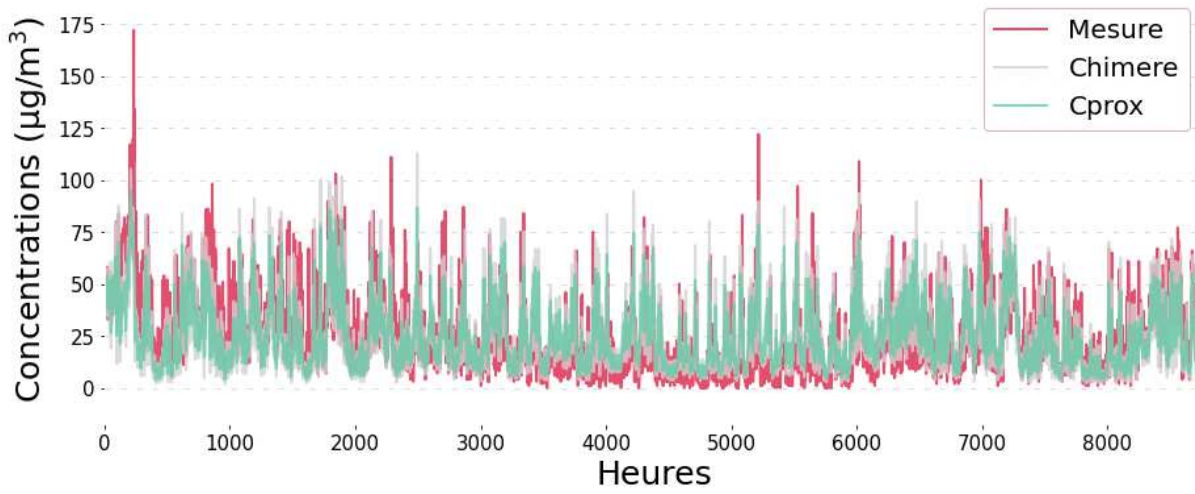
CERGY: Site de fond (NO₂, 2009)



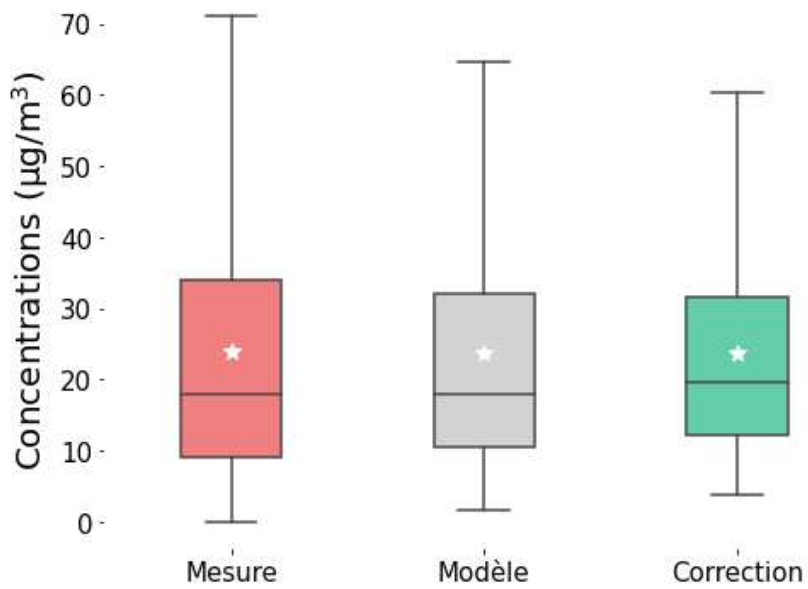
Comparaison des boxplots CERGY : NO₂, 2009



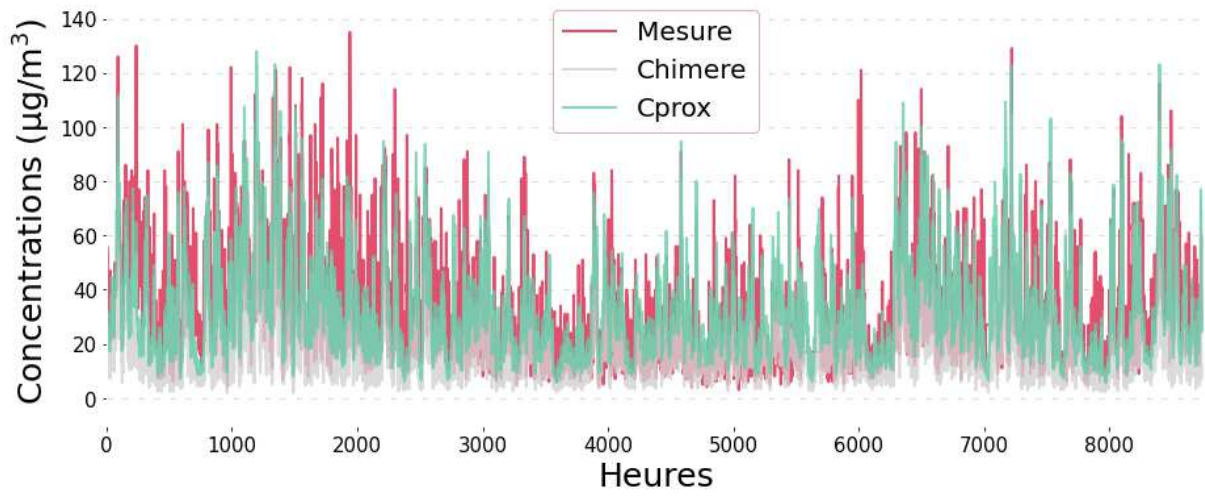
GARCH: Site de fond (NO₂, 2009)



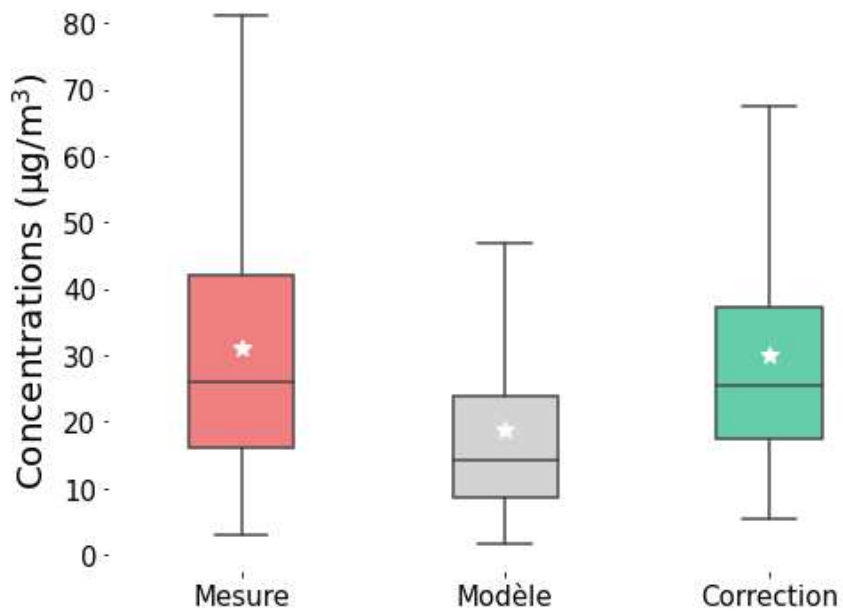
Comparaison des boxplots GARCH : NO₂, 2009



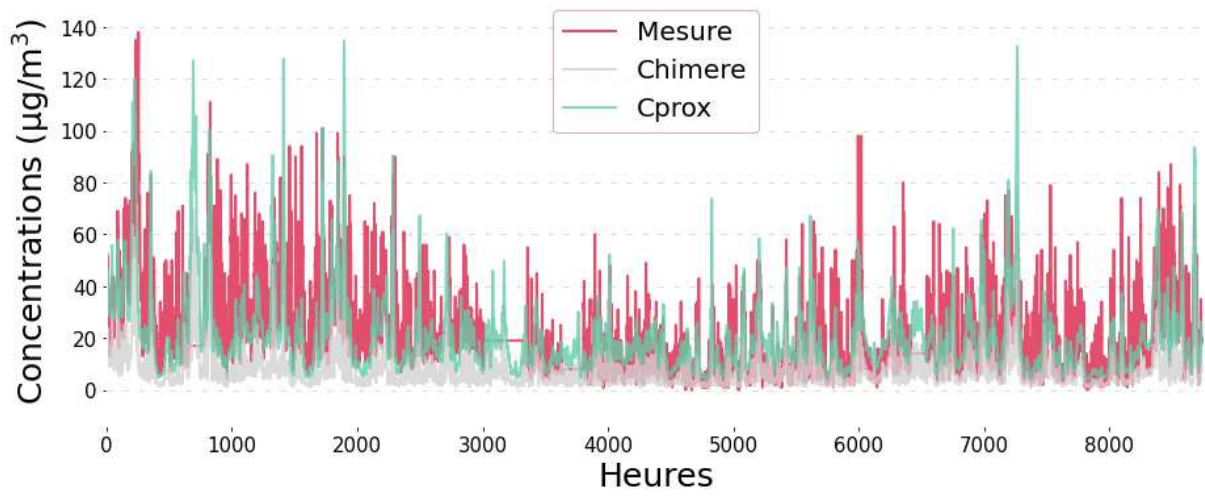
LOGNES: Site de fond (NO₂, 2009)



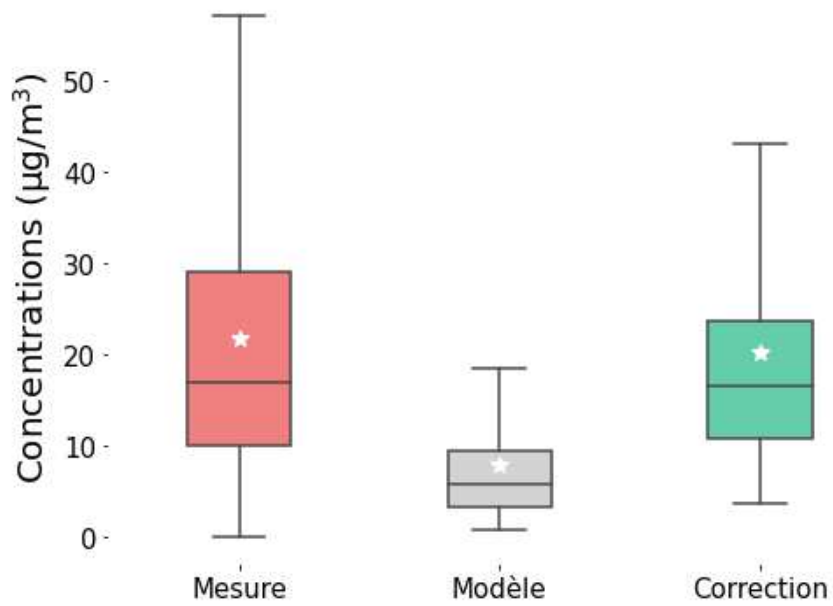
Comparaison des boxplots LOGNES : NO₂, 2009



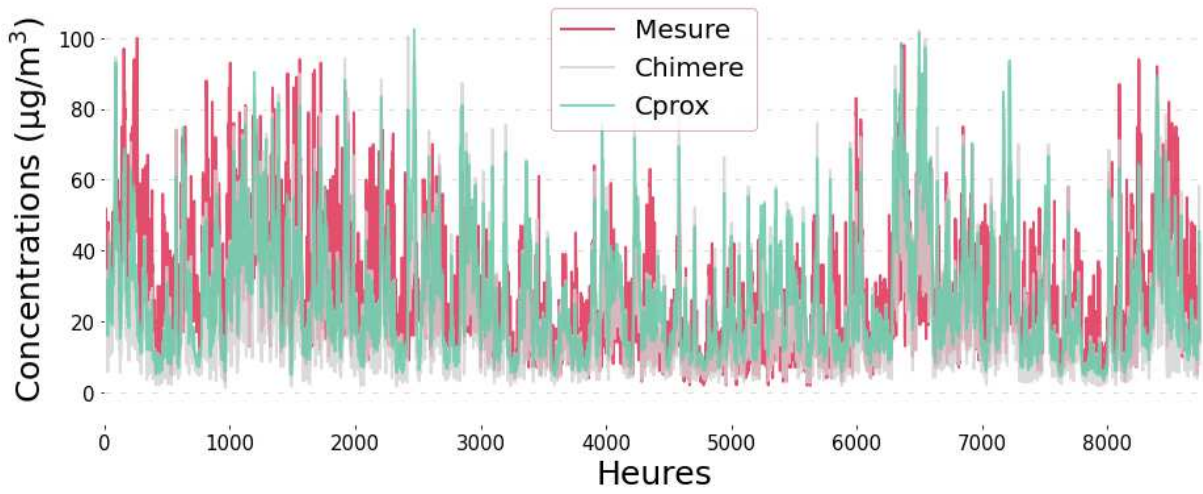
MANT: Site de fond (NO₂, 2009)



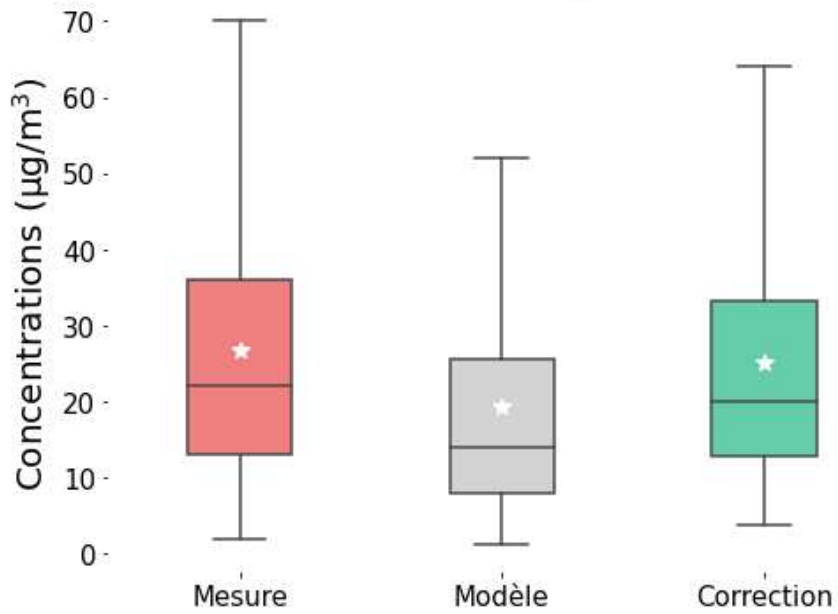
Comparaison des boxplots MANT : NO₂, 2009



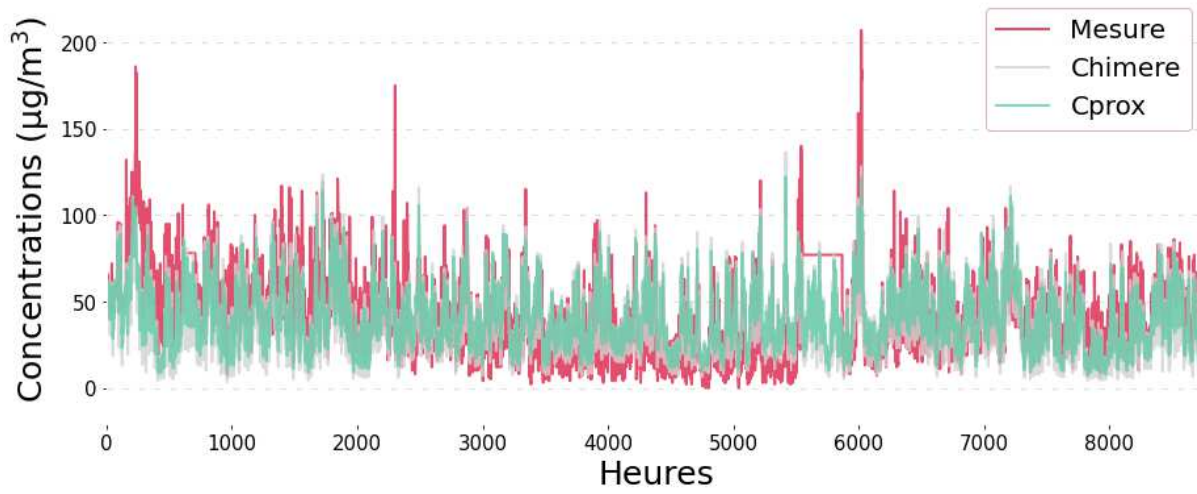
MONTG: Site de fond (NO₂, 2009)



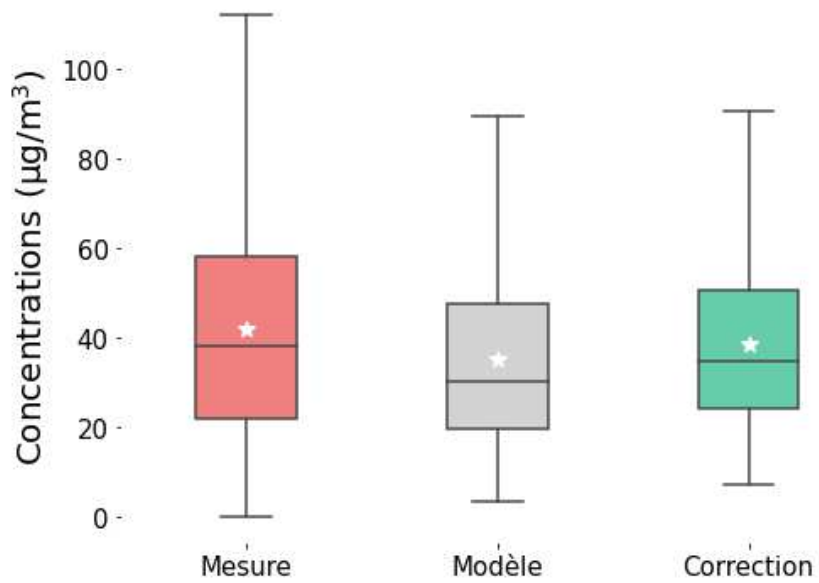
Comparaison des boxplots MONTG : NO₂, 2009



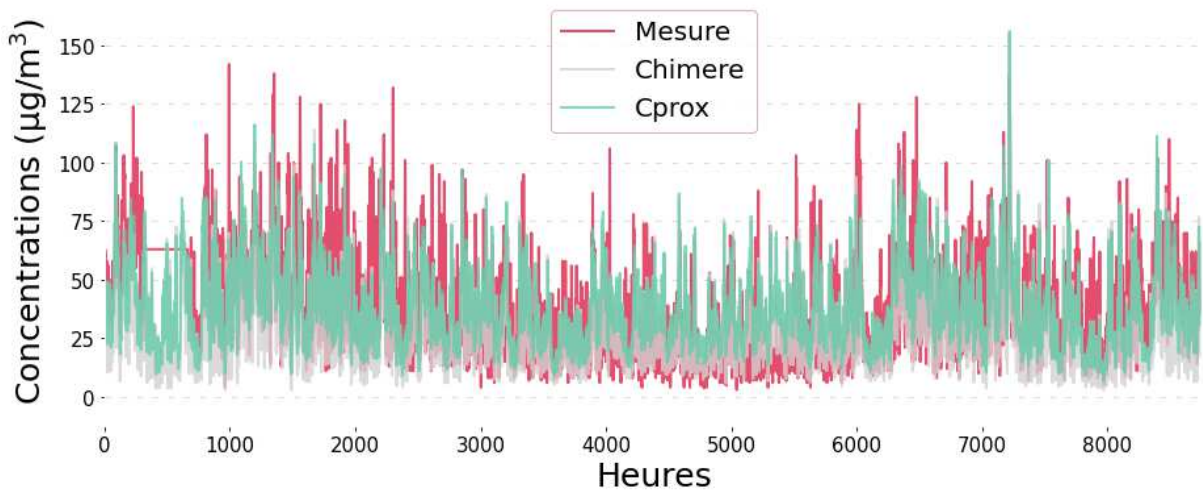
NEUIL: Site de fond (NO₂, 2009)



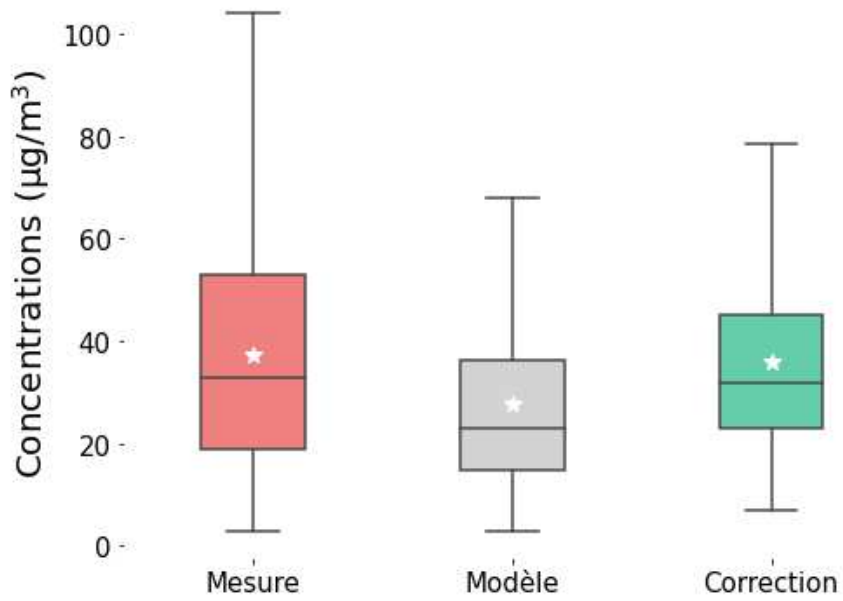
Comparaison des boxplots NEUIL : NO₂, 2009



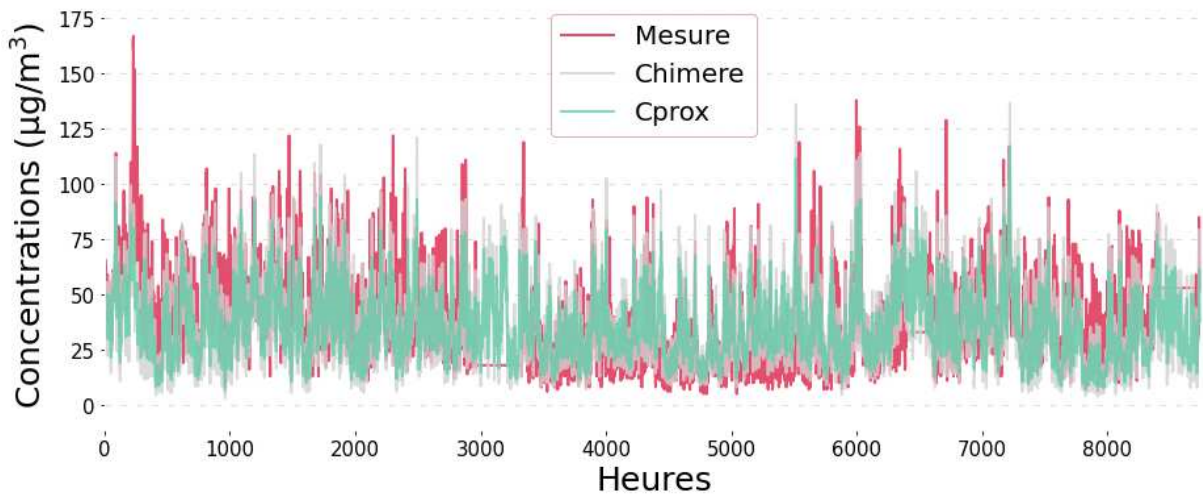
NOGENT: Site de fond (NO₂, 2009)



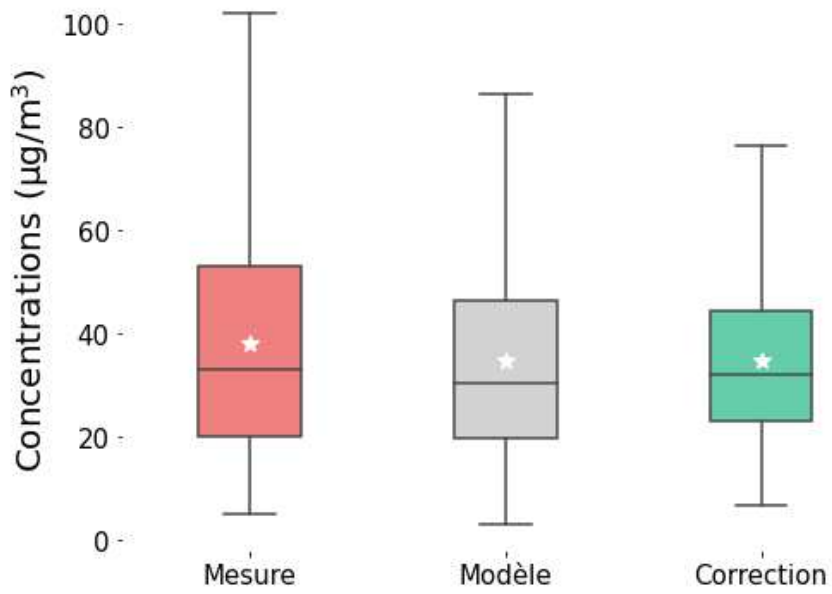
Comparaison des boxplots NOGENT : NO₂, 2009



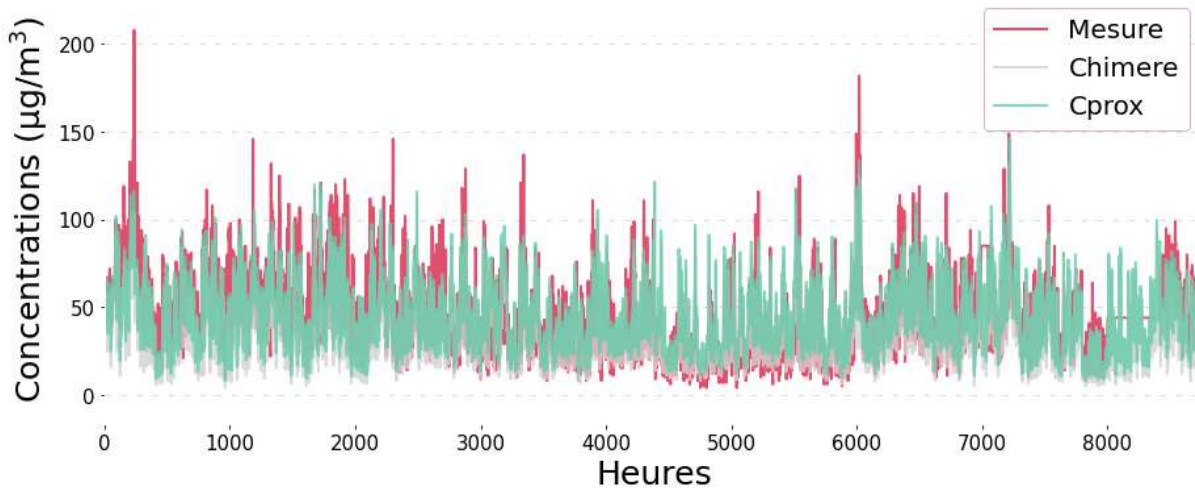
PA13: Site de fond (NO₂, 2009)



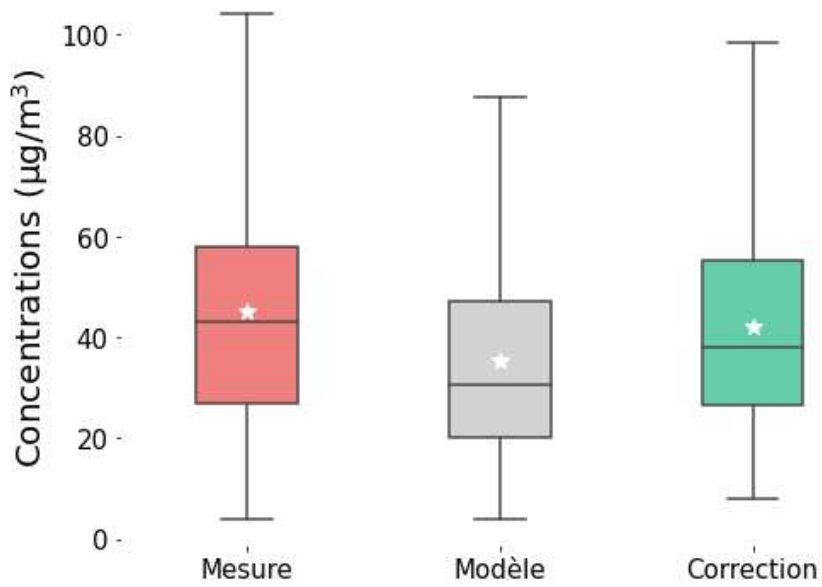
Comparaison des boxplots
PA13 : NO₂, 2009



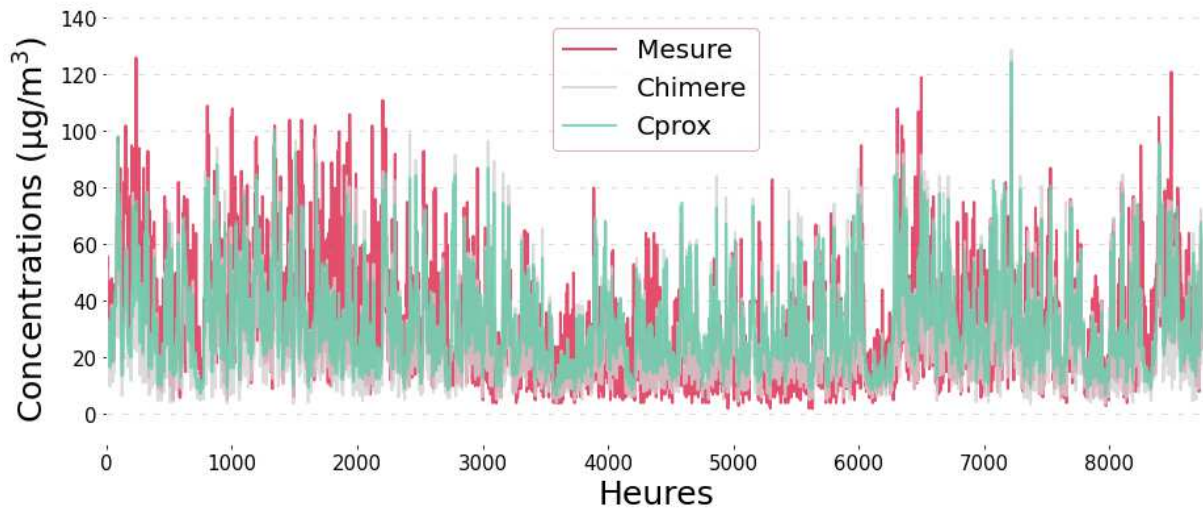
PA18: Site de fond (NO₂, 2009)



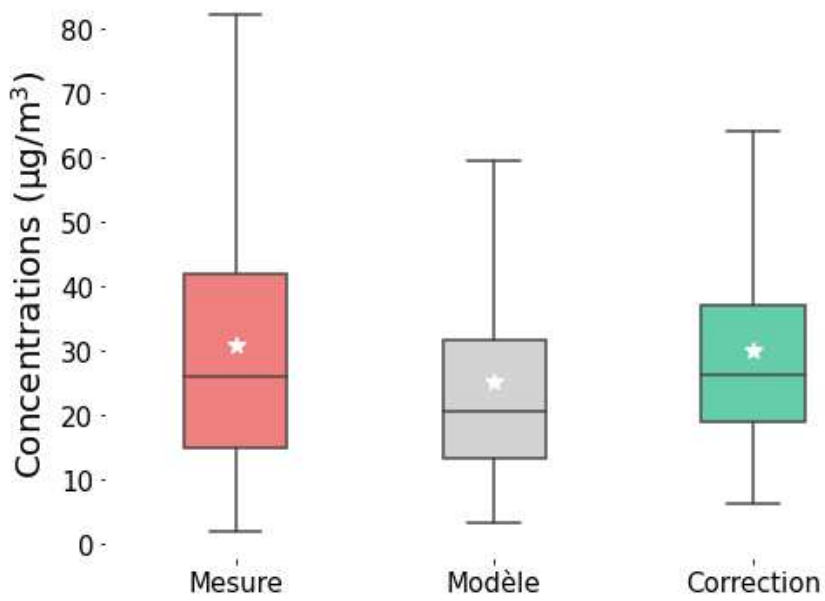
Comparaison des boxplots PA18 : NO₂, 2009



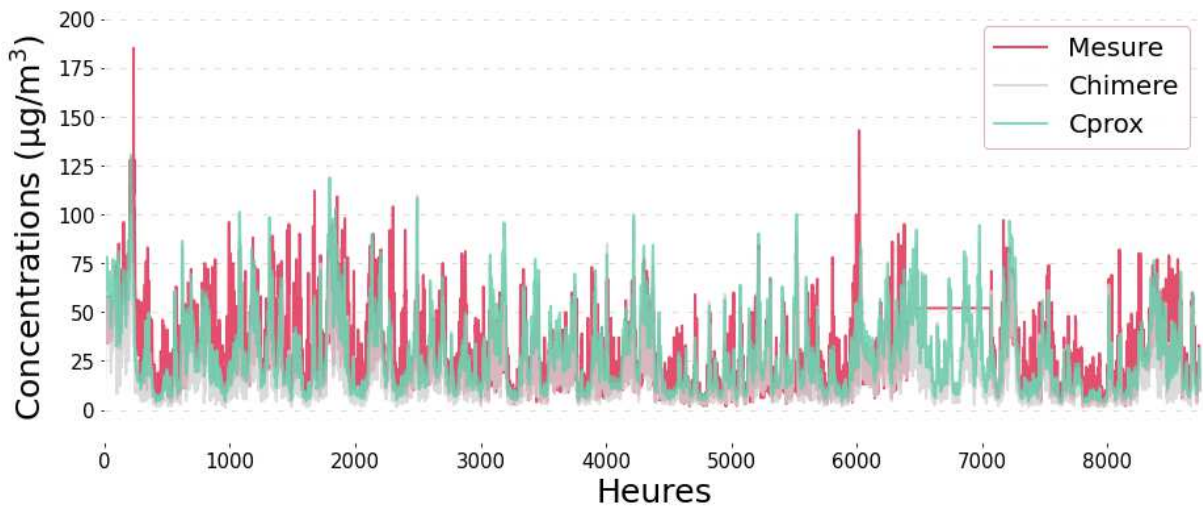
TREMB: Site de fond (NO₂, 2009)



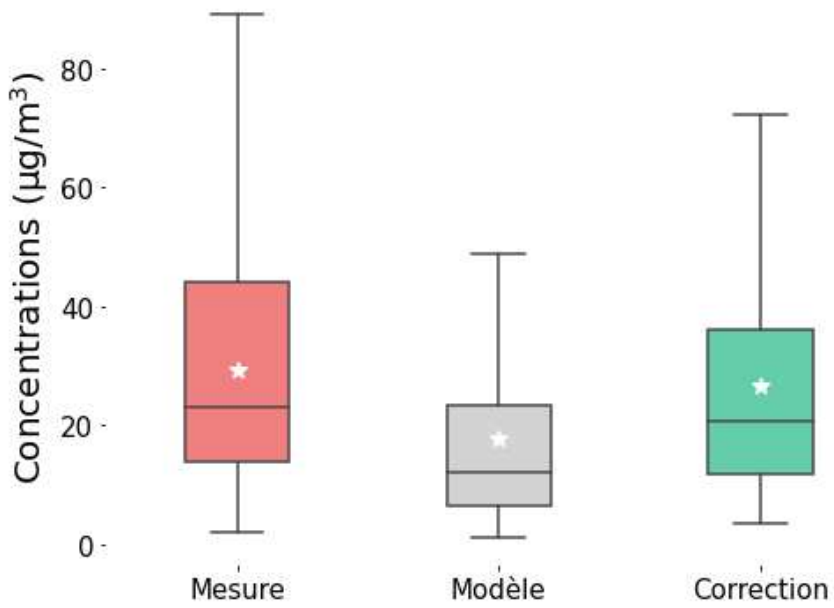
Comparaison des boxplots
TREMB : NO₂, 2009



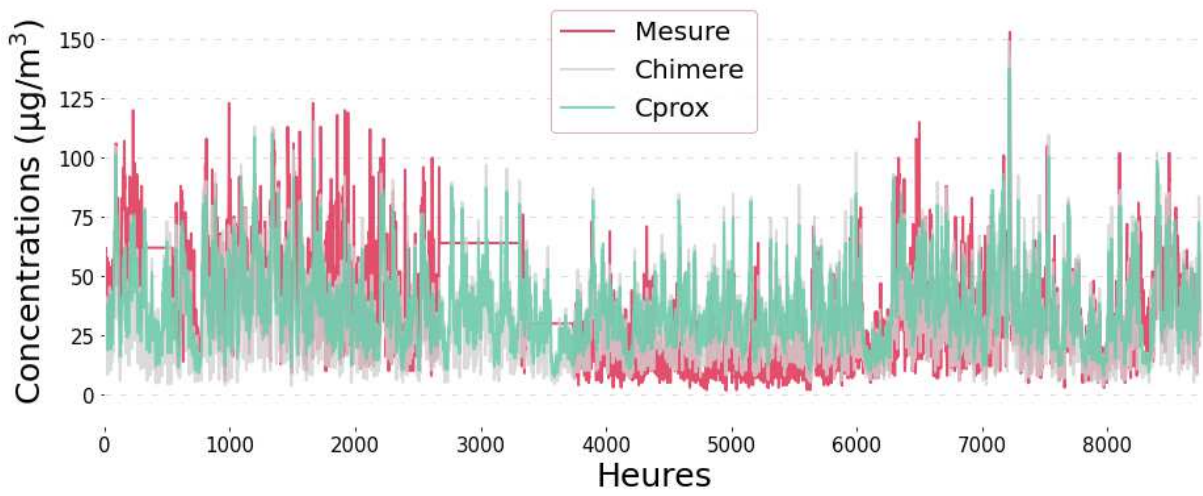
VERS: Site de fond (NO₂, 2009)



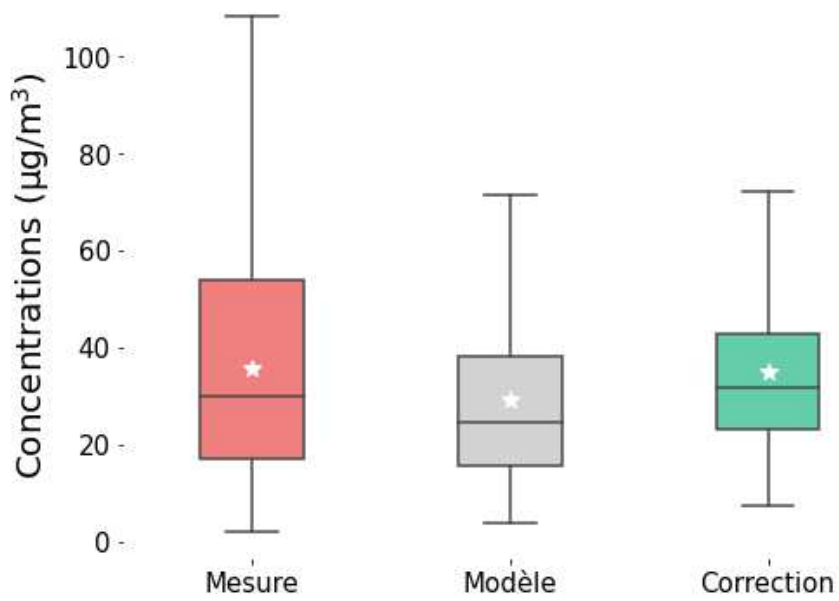
Comparaison des boxplots VERS : NO₂, 2009



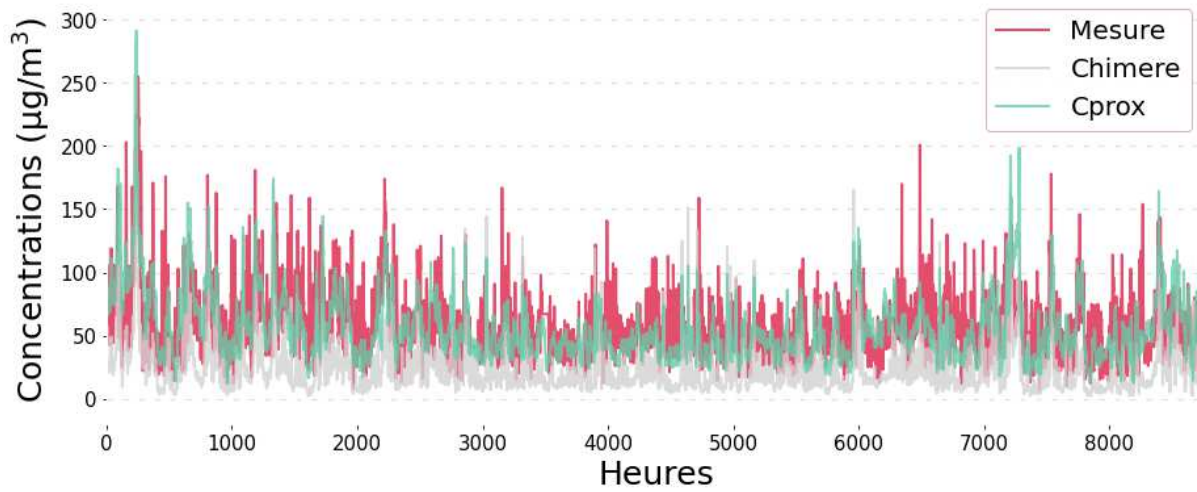
VILLEM: Site de fond (NO₂, 2009)



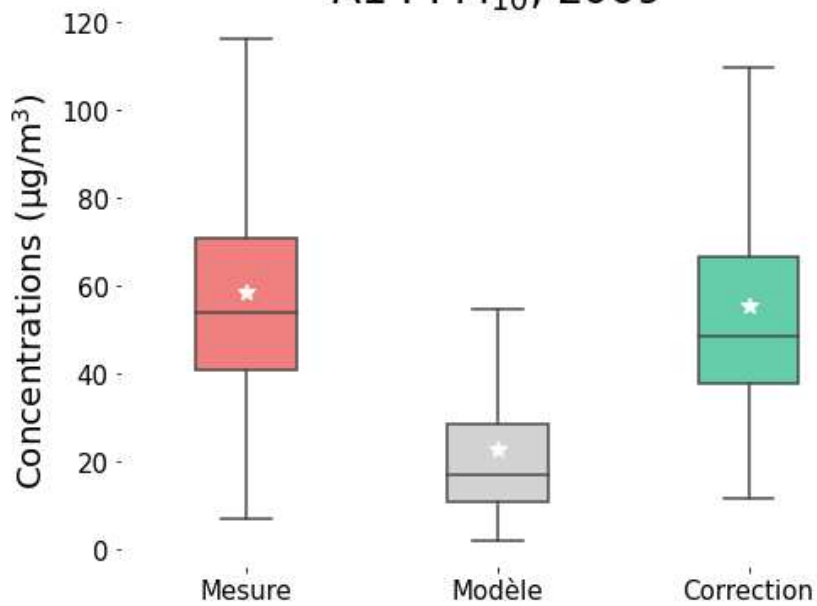
Comparaison des boxplots VILLEM : NO₂, 2009



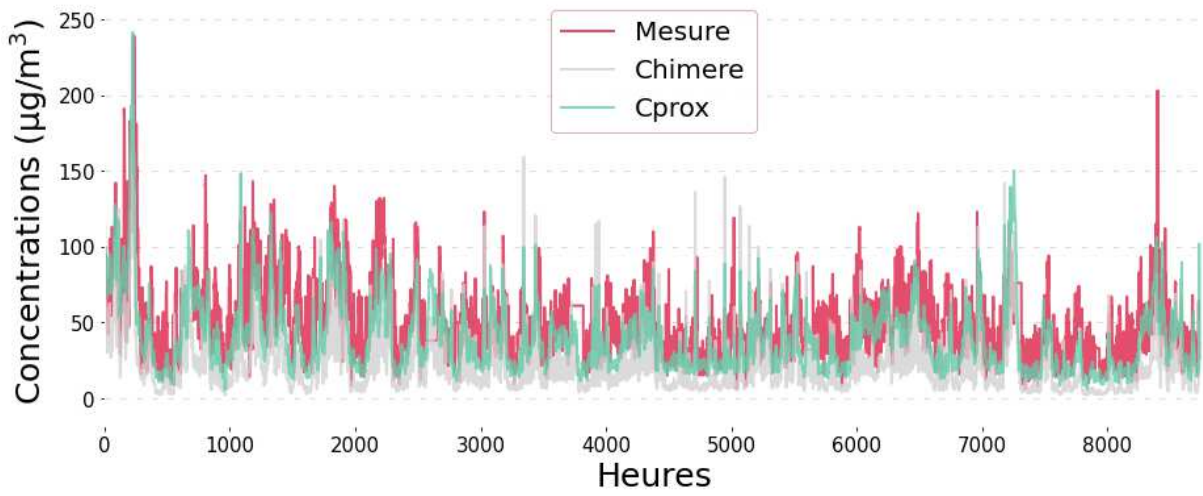
A1: Site de trafic (NO₂, 2009)



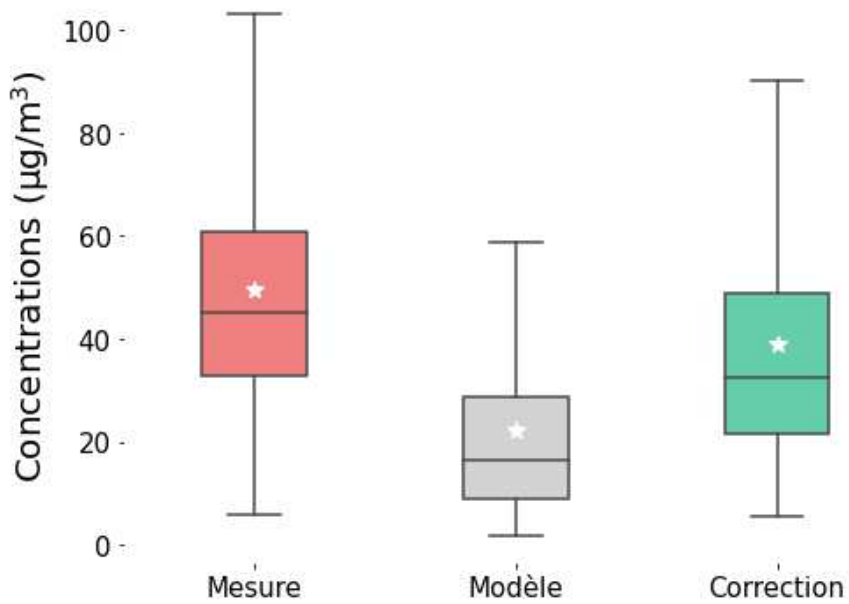
Comparaison des boxplots
A1 : PM₁₀, 2009



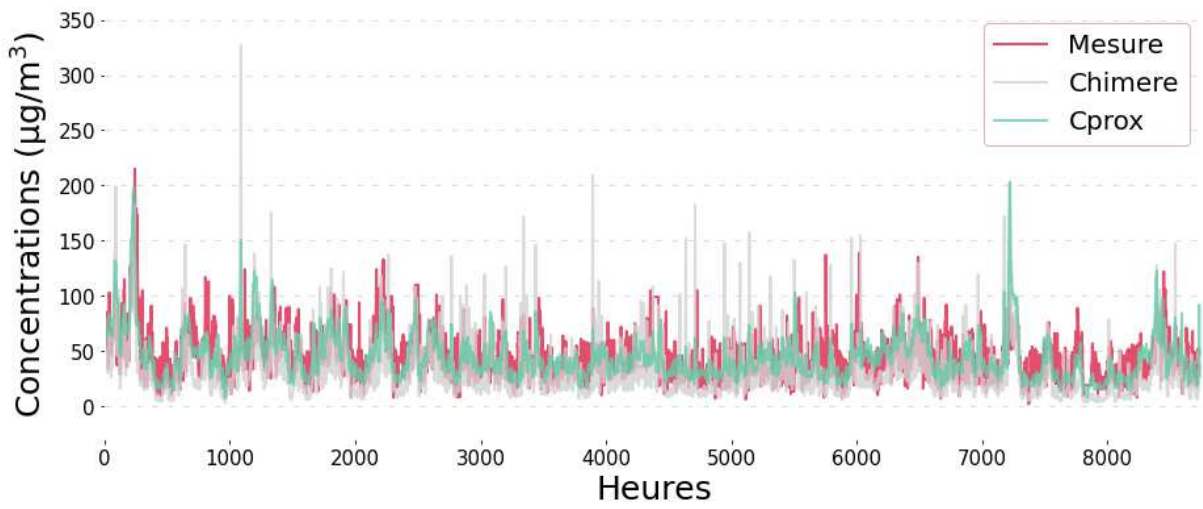
AUT: Site de trafic (NO₂, 2009)



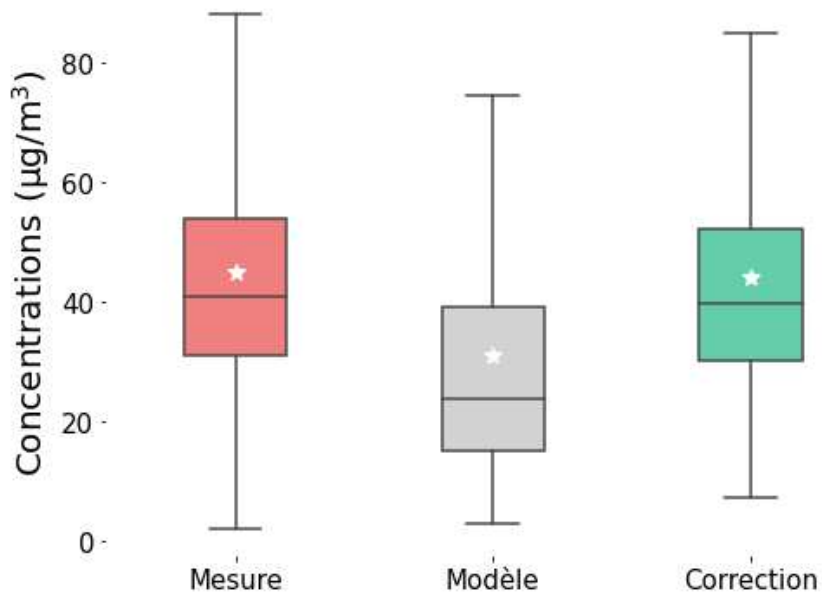
Comparaison des boxplots
AUT : PM₁₀, 2009



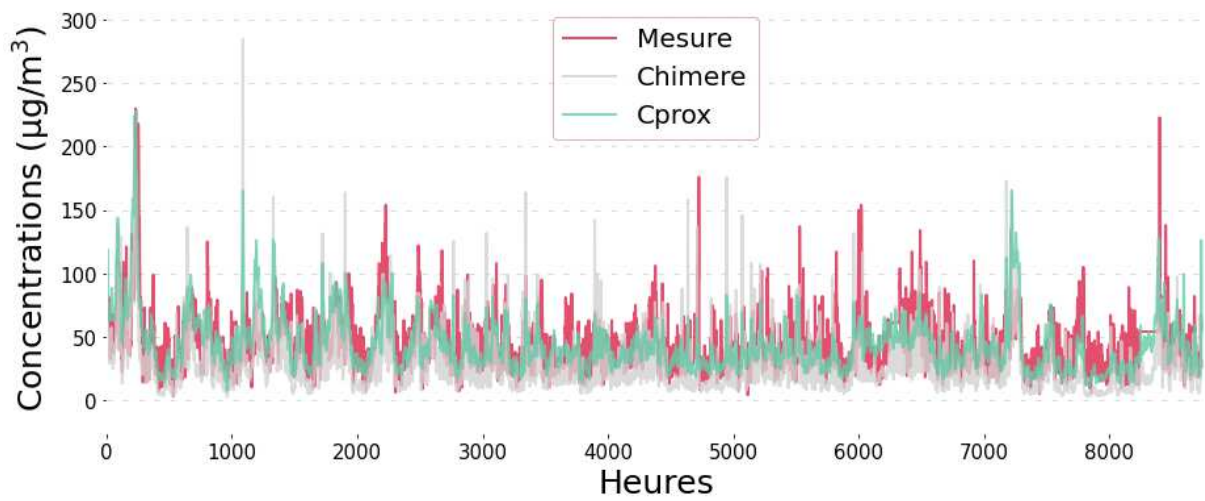
BASCH: Site de trafic (NO₂, 2009)



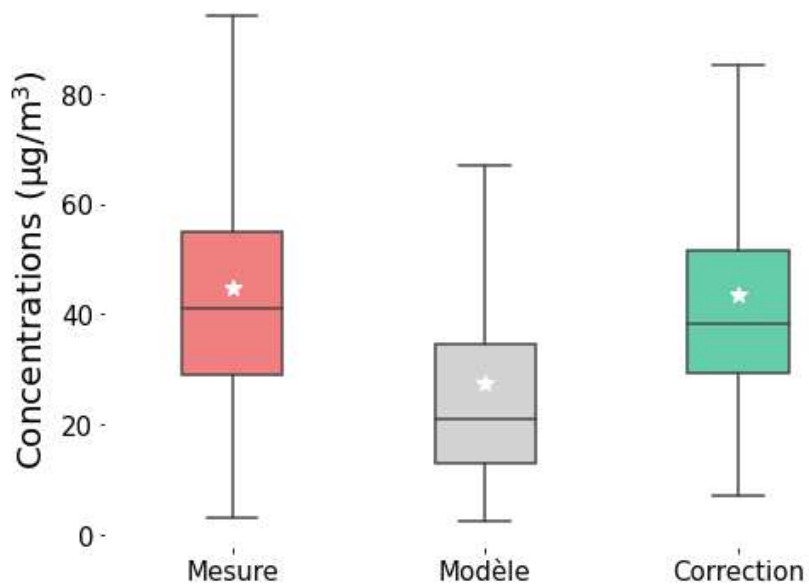
Comparaison des boxplots
BASCH : PM₁₀, 2009



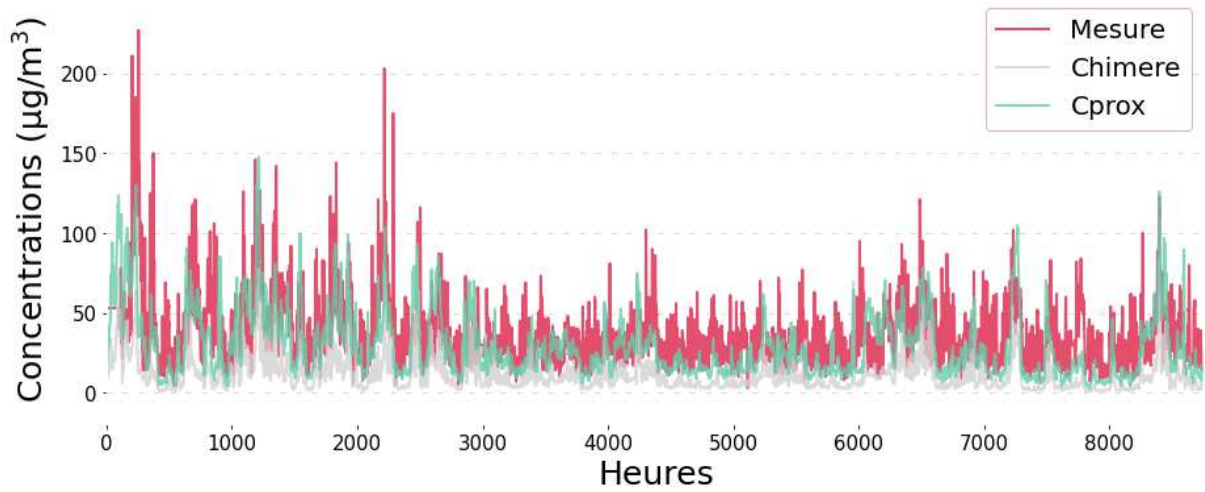
ELYS: Site de trafic (NO₂, 2009)



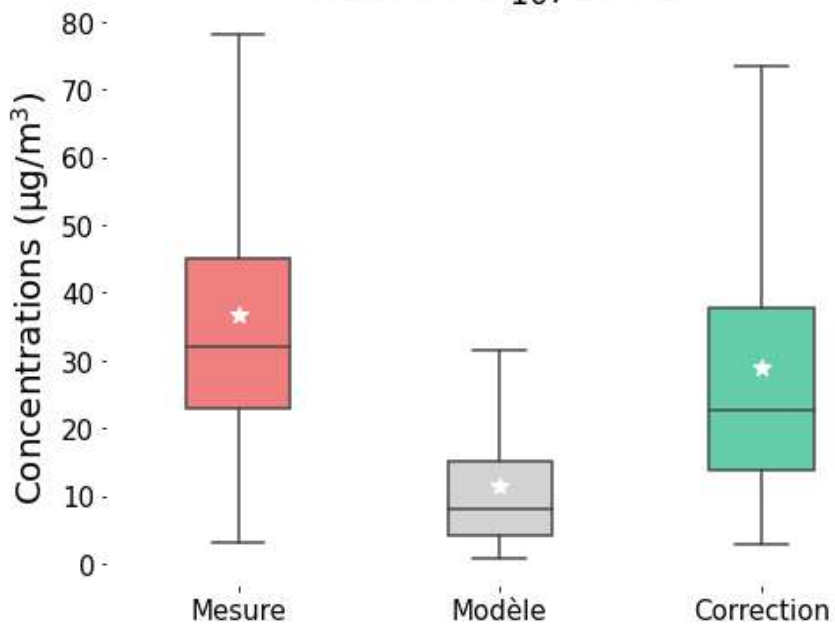
Comparaison des boxplots ELYS : PM₁₀, 2009



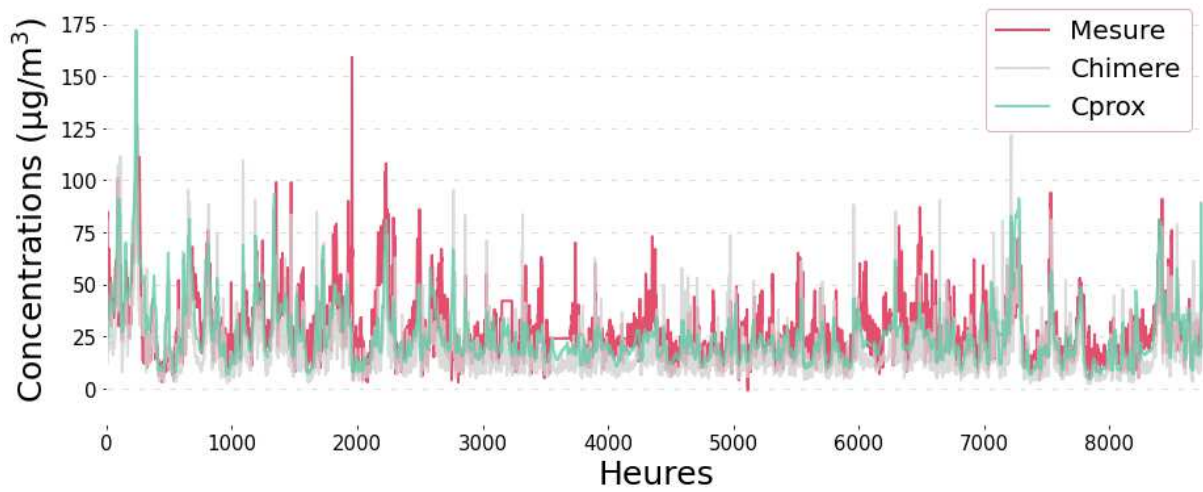
RN6: Site de trafic (NO₂, 2009)



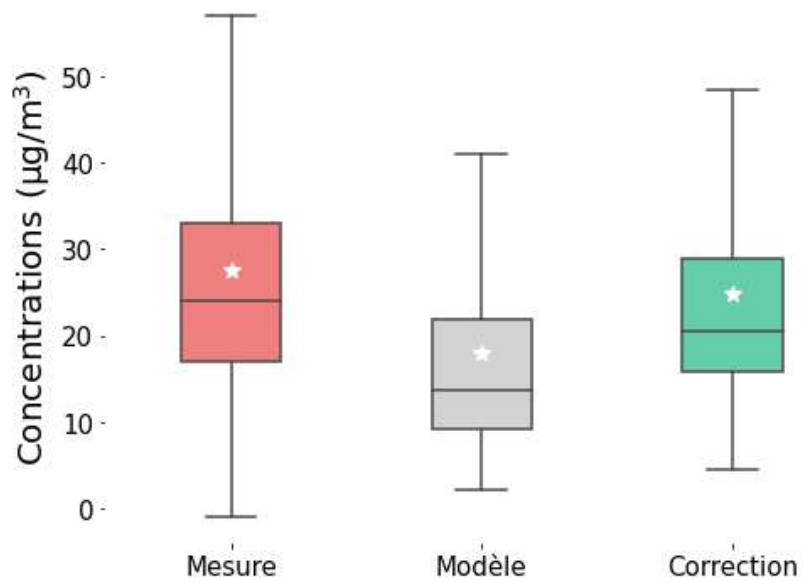
Comparaison des boxplots
RN6 : PM₁₀, 2009



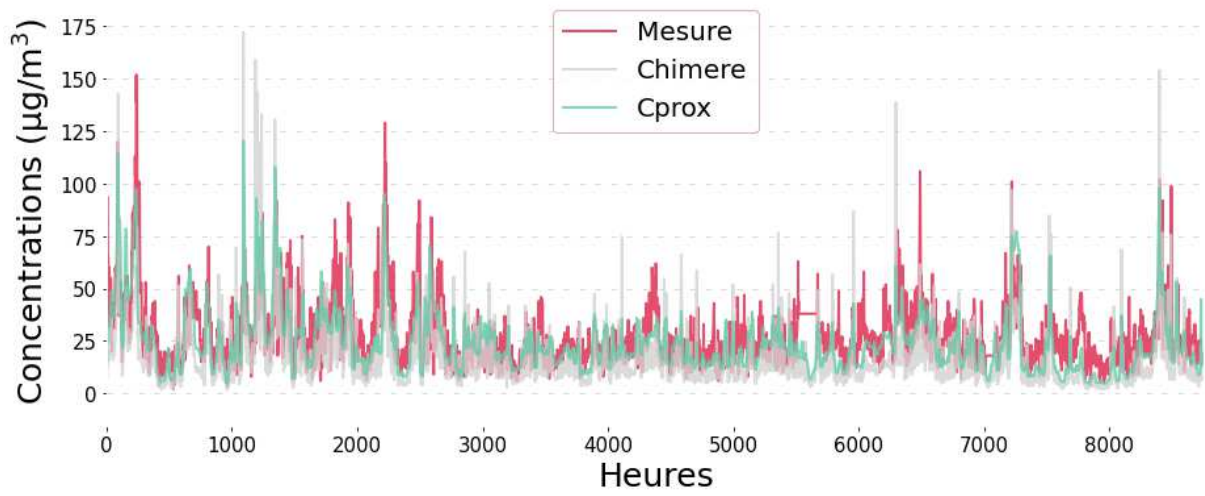
GON: Site de fond (NO_2 , 2009)



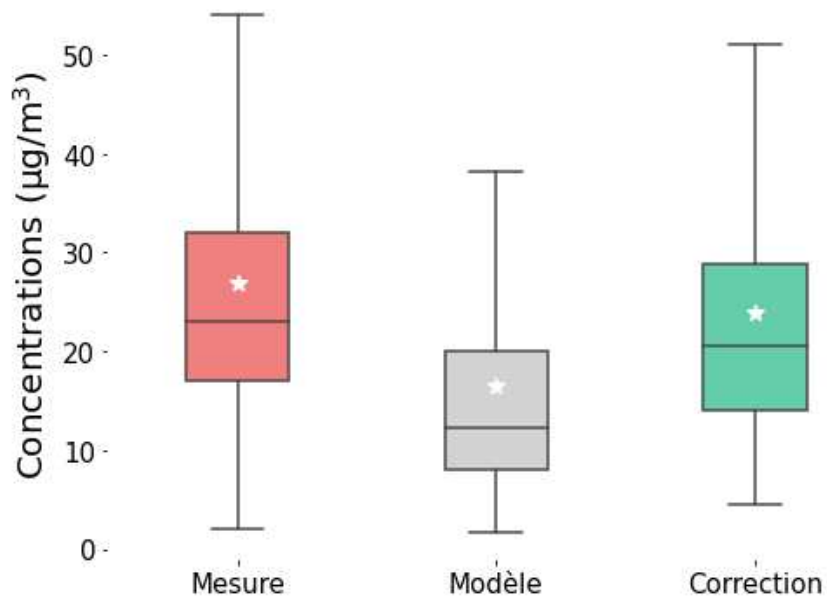
Comparaison des boxplots
GON : PM_{10} , 2009



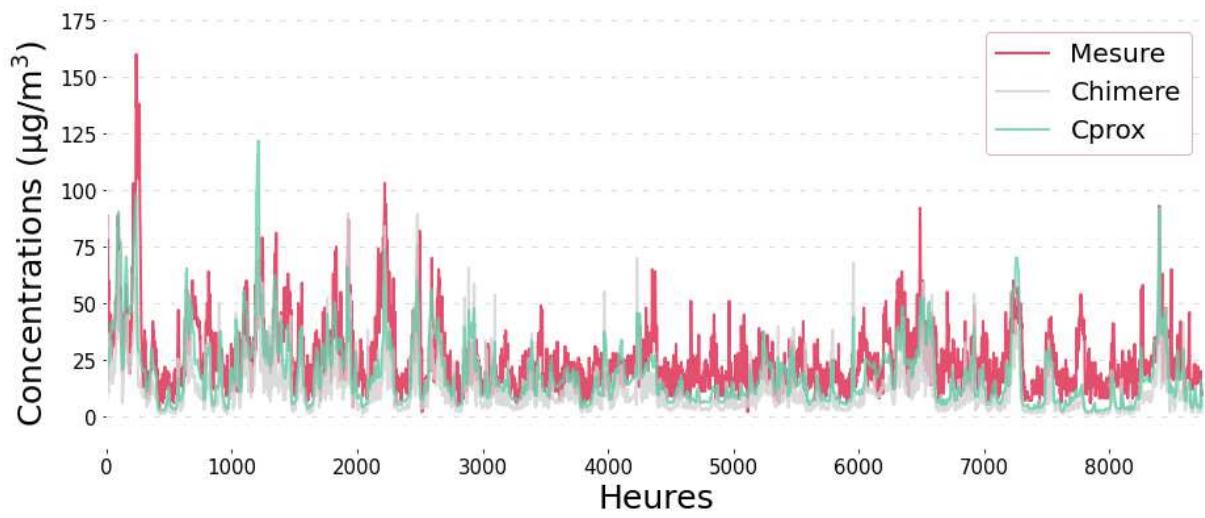
LOGNES: Site de fond (NO₂, 2009)



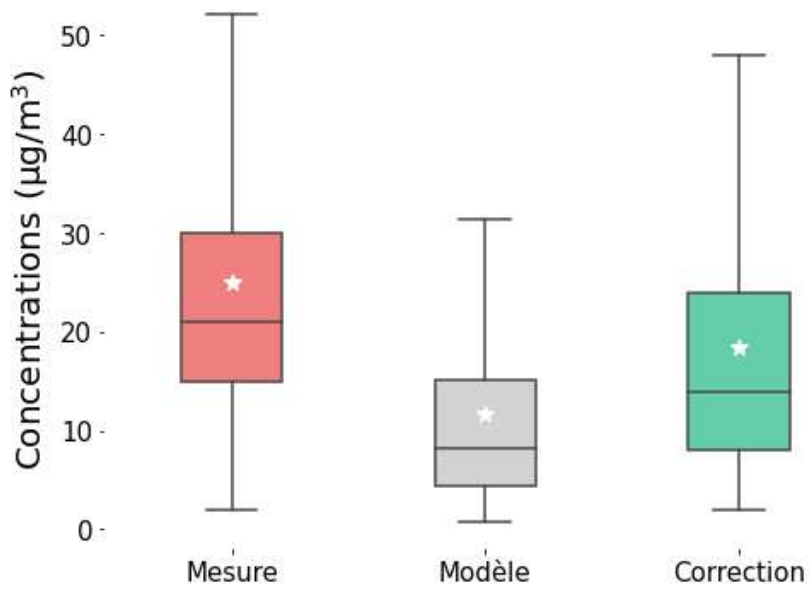
Comparaison des boxplots LOGNES : PM₁₀, 2009



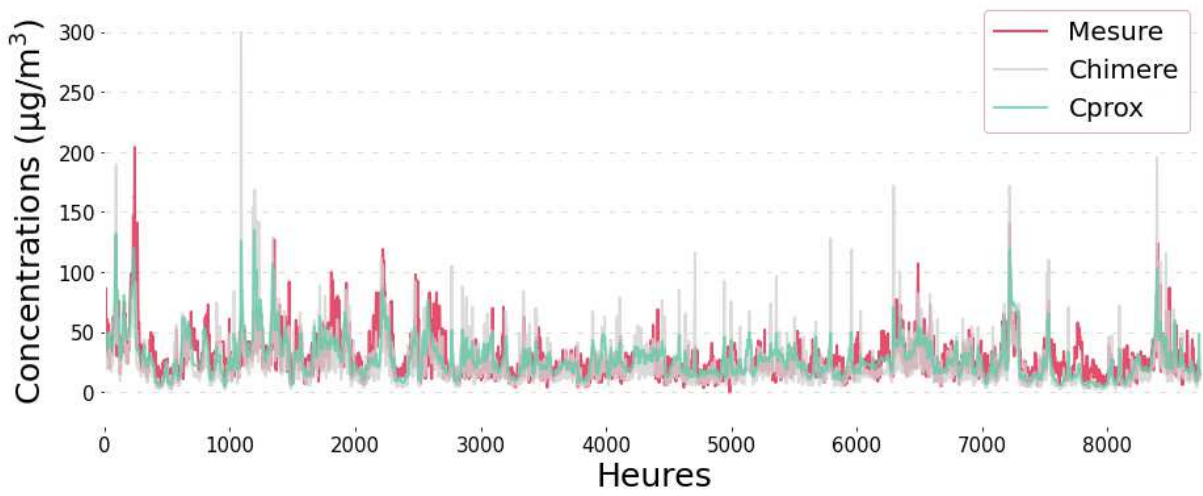
MELUN: Site de fond (NO₂, 2009)



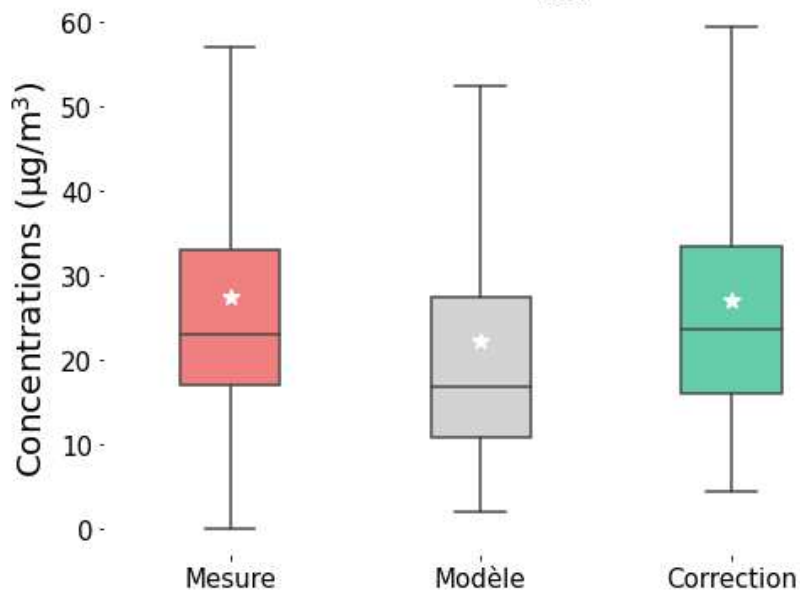
Comparaison des boxplots
MELUN : PM₁₀, 2009



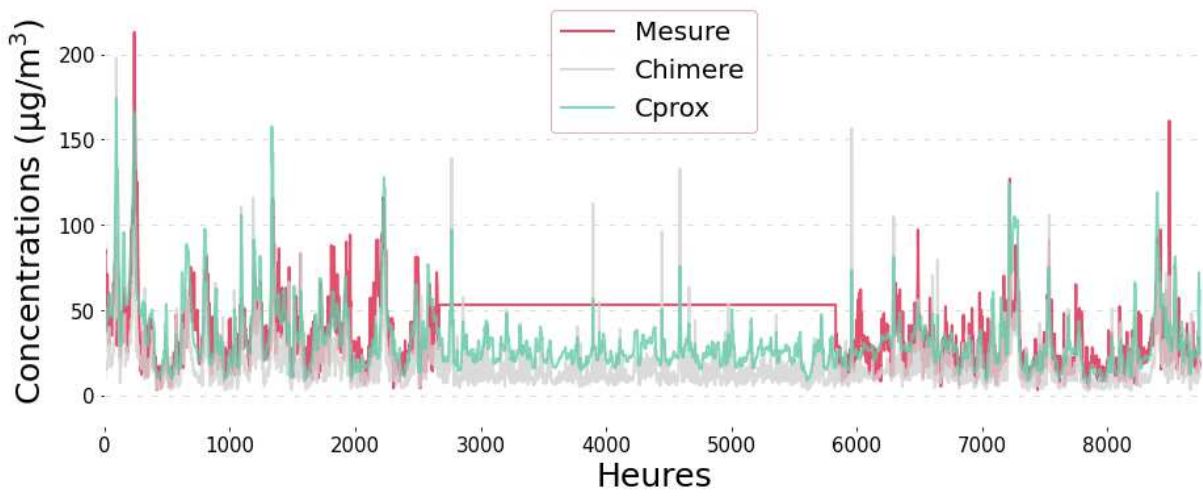
NOGENT: Site de fond (NO₂, 2009)



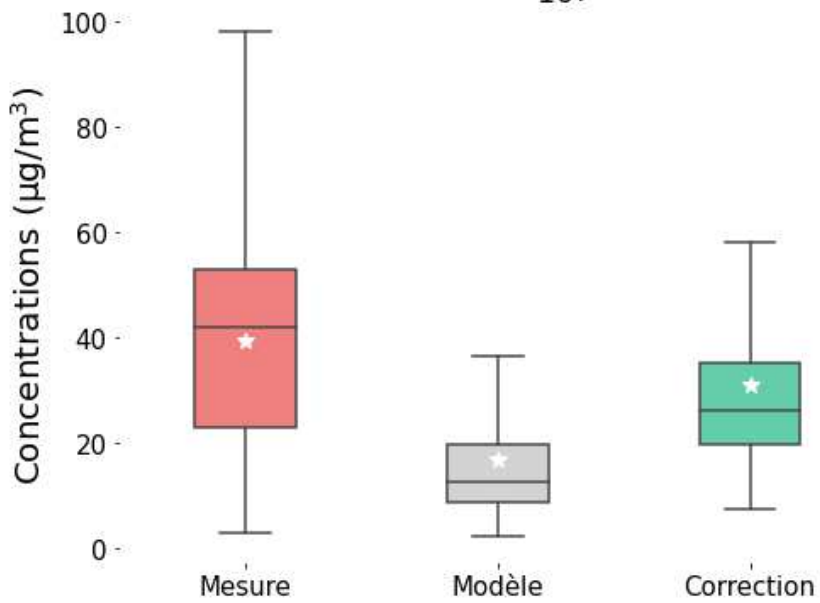
Comparaison des boxplots NOGENT : PM₁₀, 2009



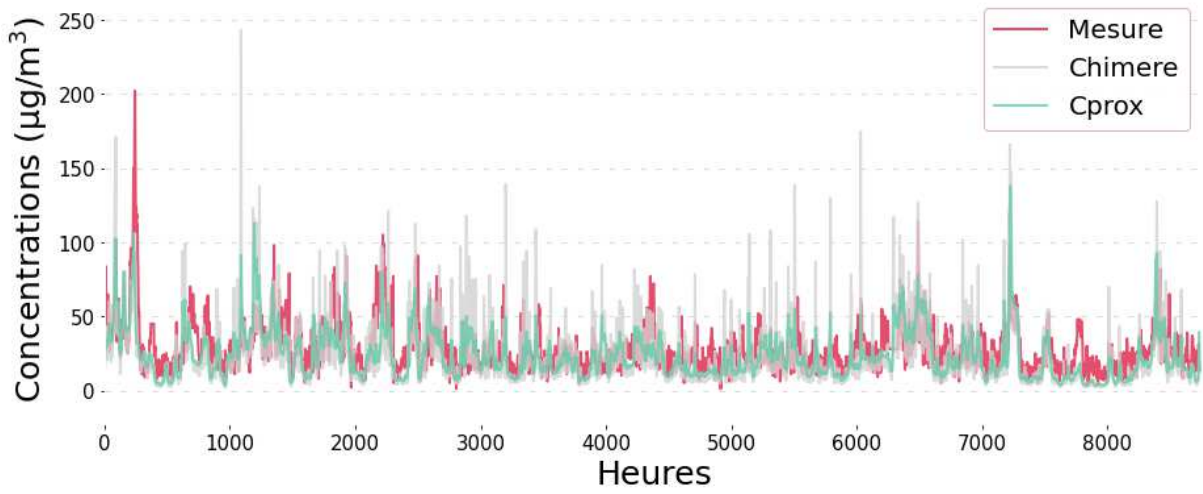
TREMB: Site de fond (NO₂, 2009)



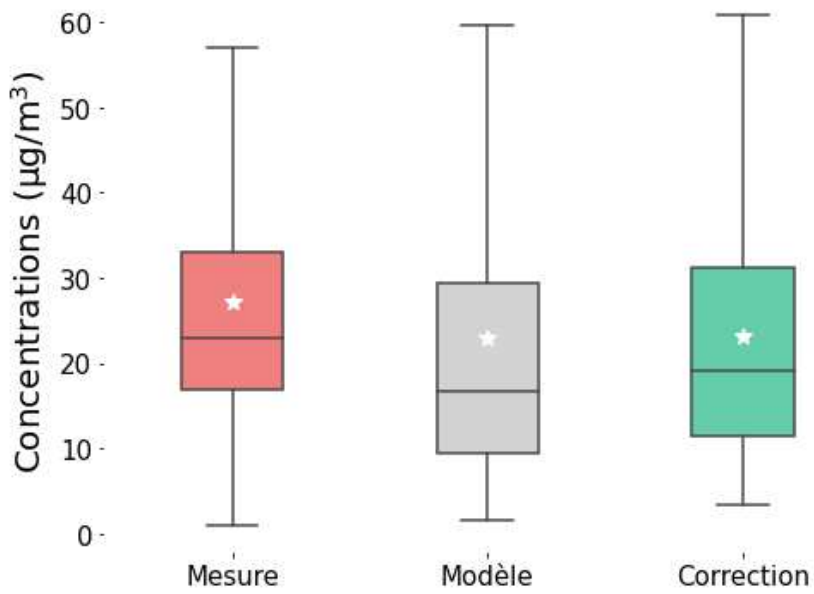
Comparaison des boxplots TREMB : PM₁₀, 2009



VITRY: Site de fond (NO₂, 2009)



Comparaison des boxplots VITRY : PM₁₀, 2009

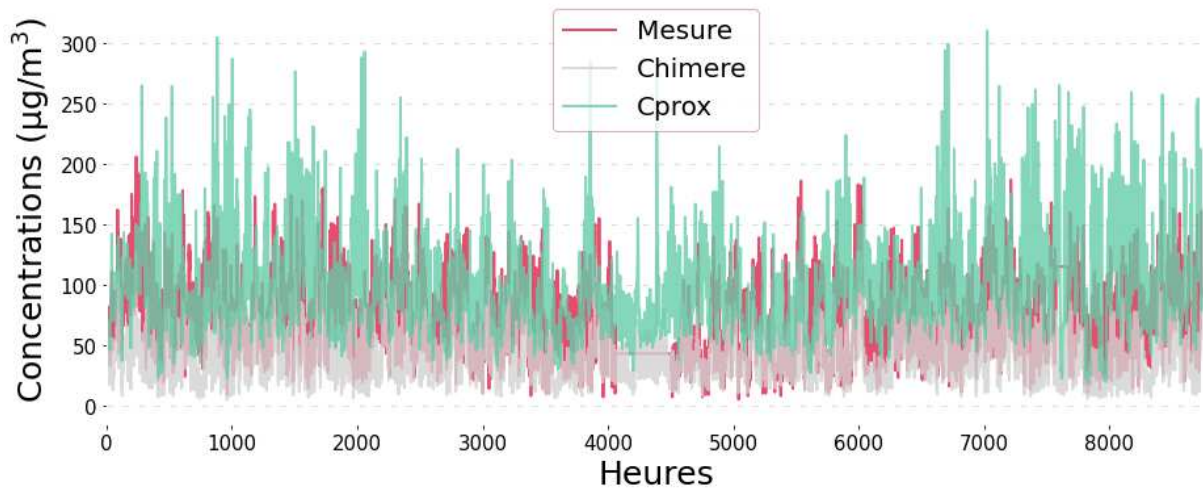


ANNEXE 4

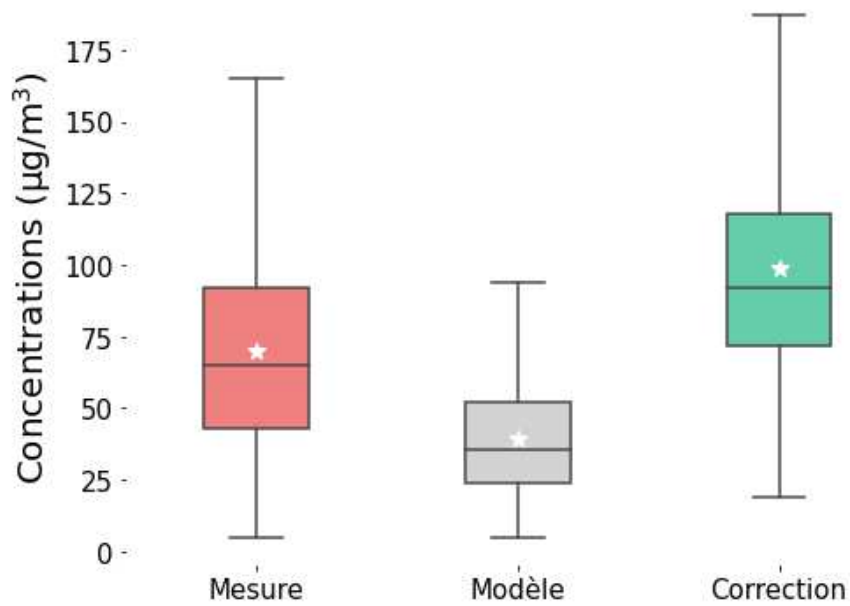
Présentation des résultats de la phase de validation de l'approche.

Séries temporelles et boxplots NO₂ – Sites de trafic.

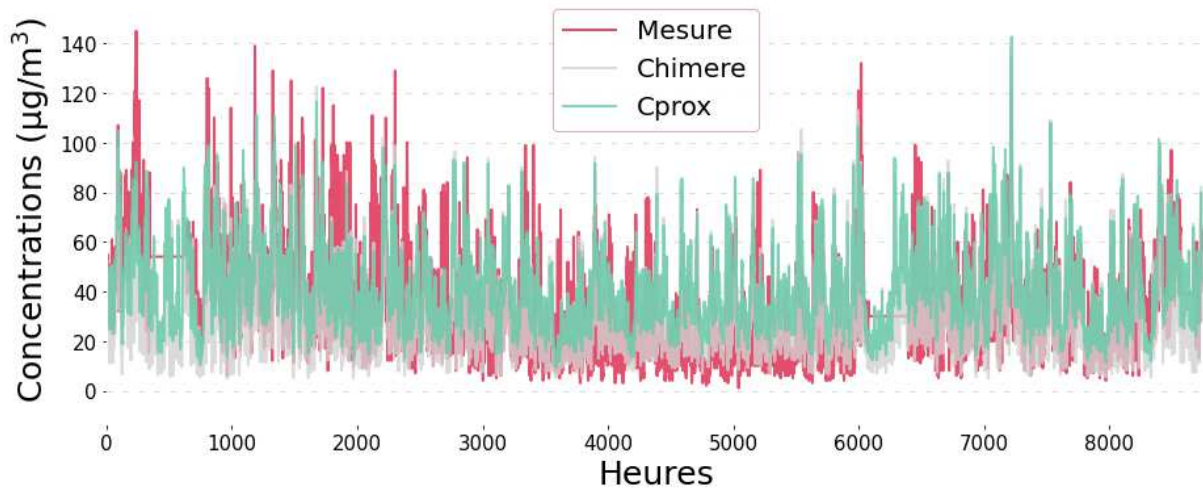
RN2: Site de trafic (NO₂, 2009)



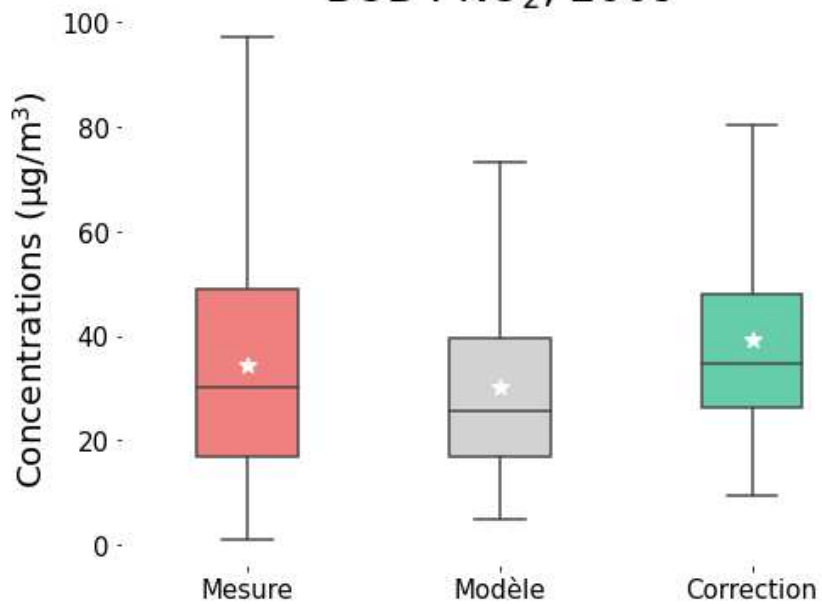
Comparaison des boxplots
RN2 : NO₂, 2009



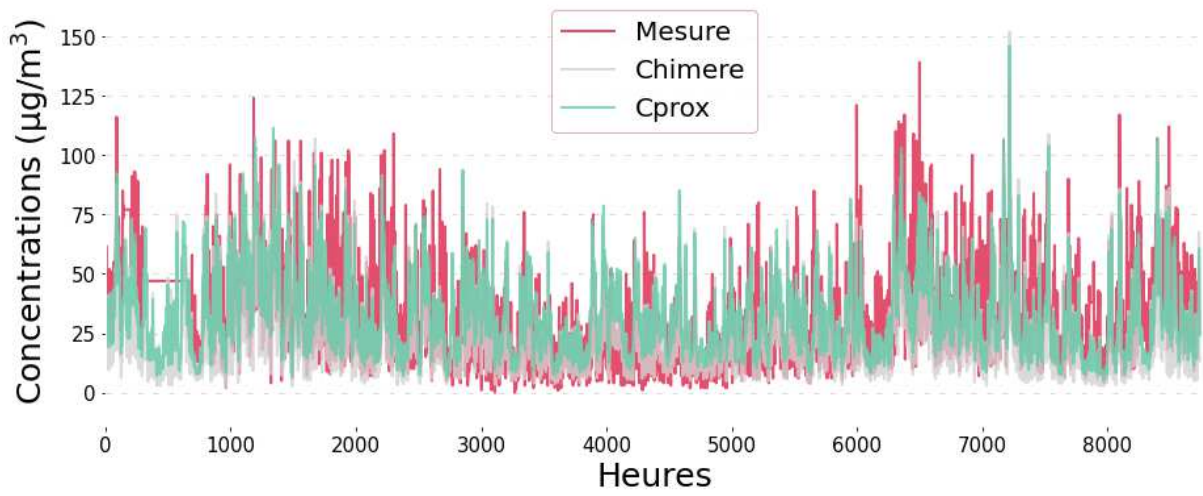
BOB: Site de fond (NO₂, 2009)



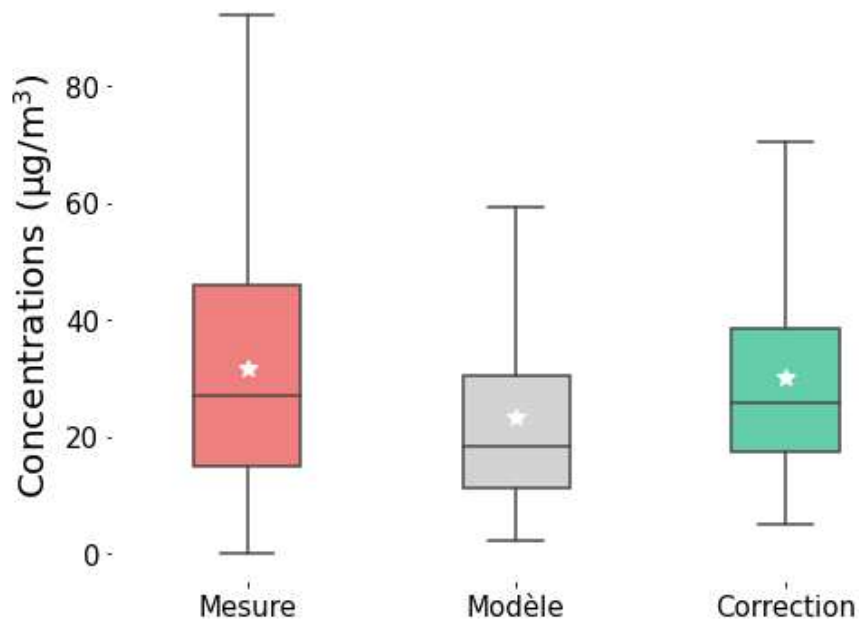
Comparaison des boxplots
BOB : NO₂, 2009



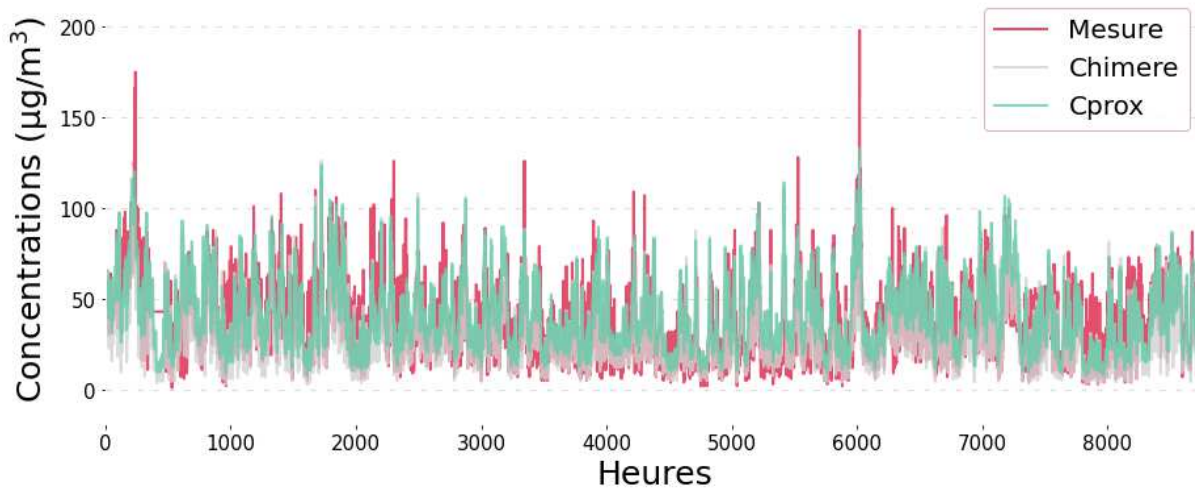
CHAMP: Site de fond (NO₂, 2009)



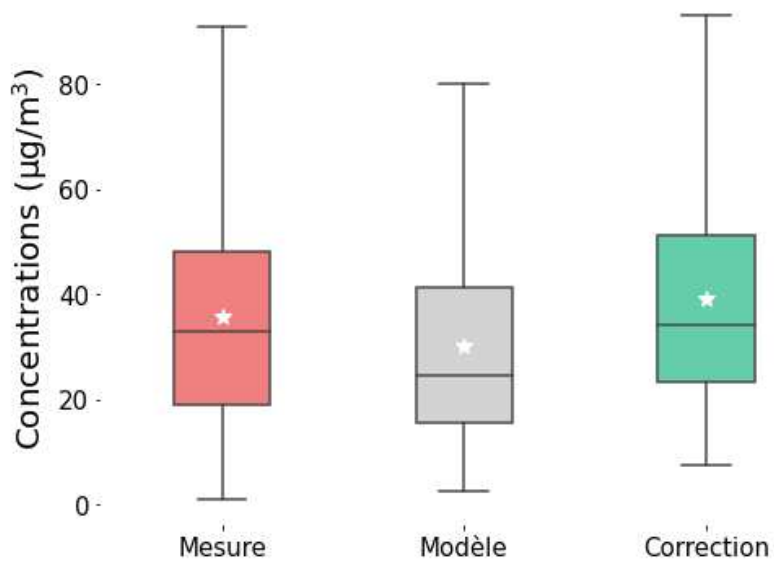
Comparaison des boxplots CHAMP : NO₂, 2009



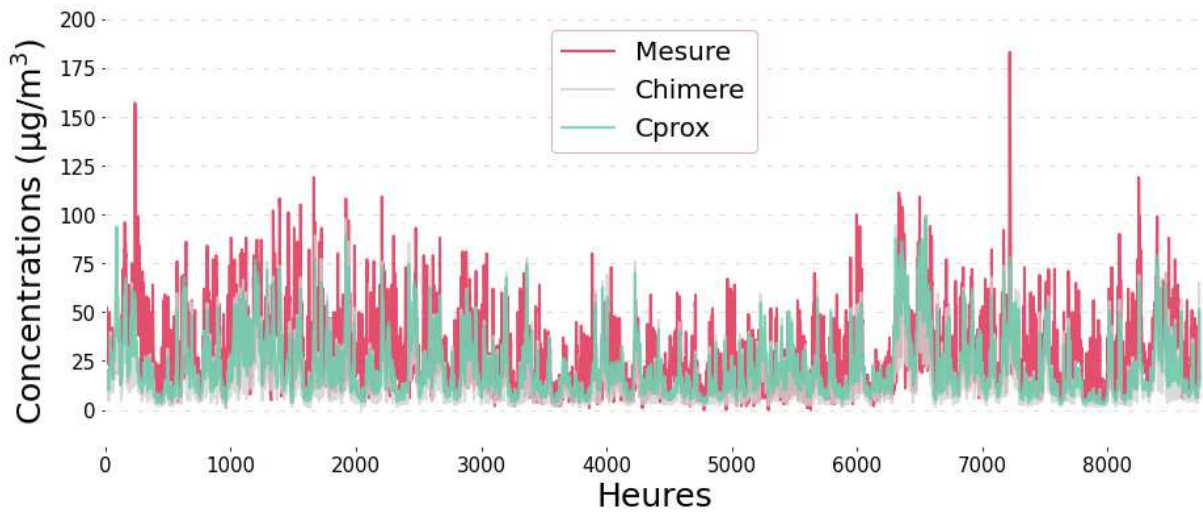
DEF: Site de fond (NO₂, 2009)



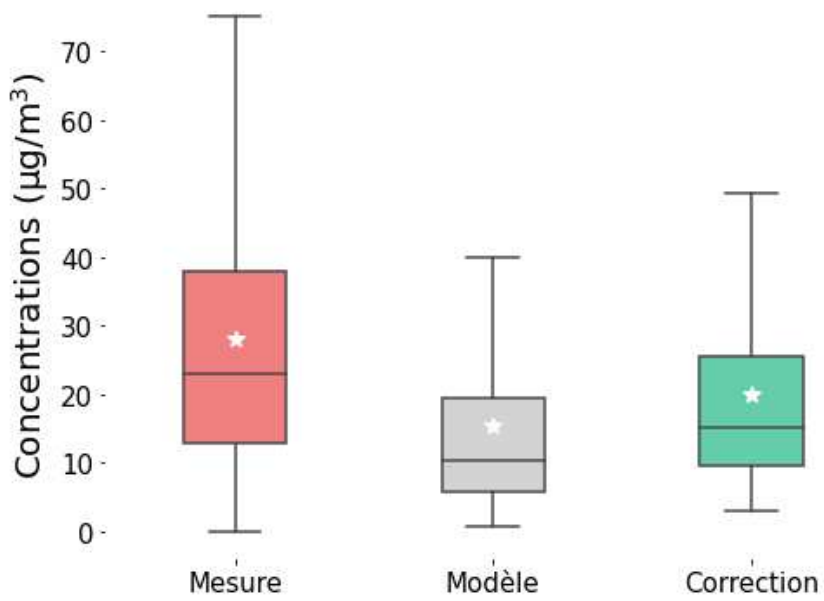
Comparaison des boxplots
DEF : NO₂, 2009



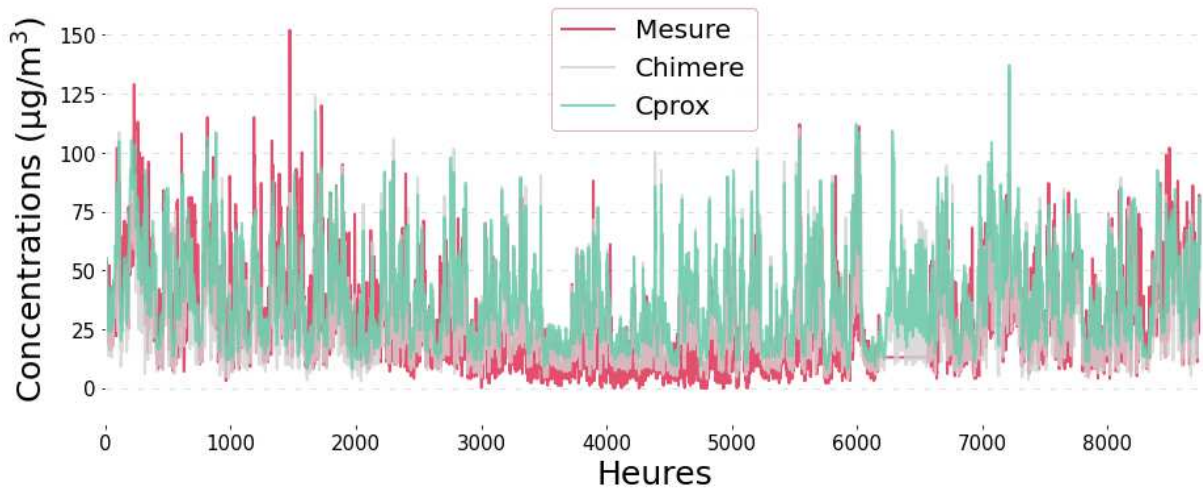
EVRY: Site de fond (NO₂, 2009)



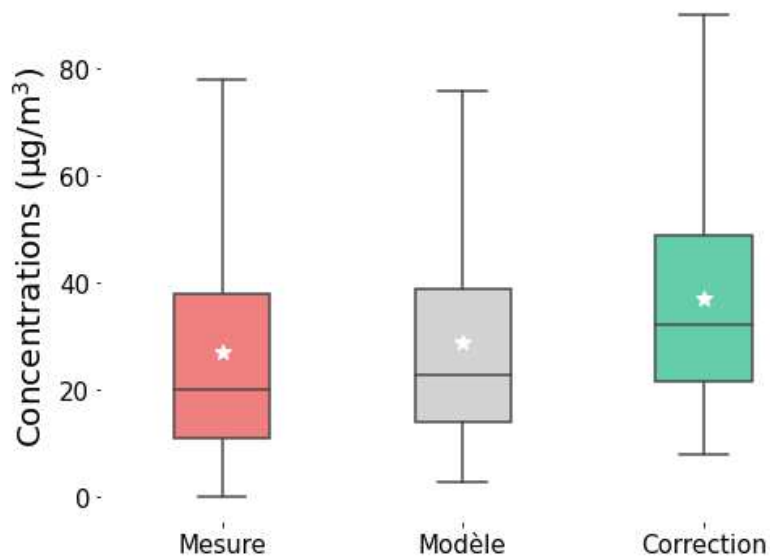
Comparaison des boxplots EVRY : NO₂, 2009



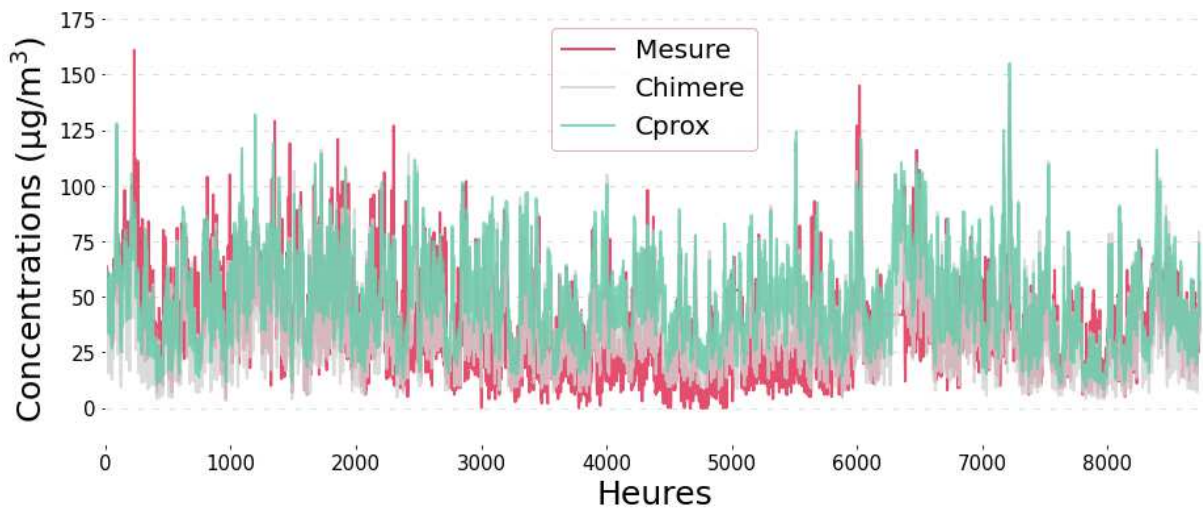
GON: Site de fond (NO₂, 2009)



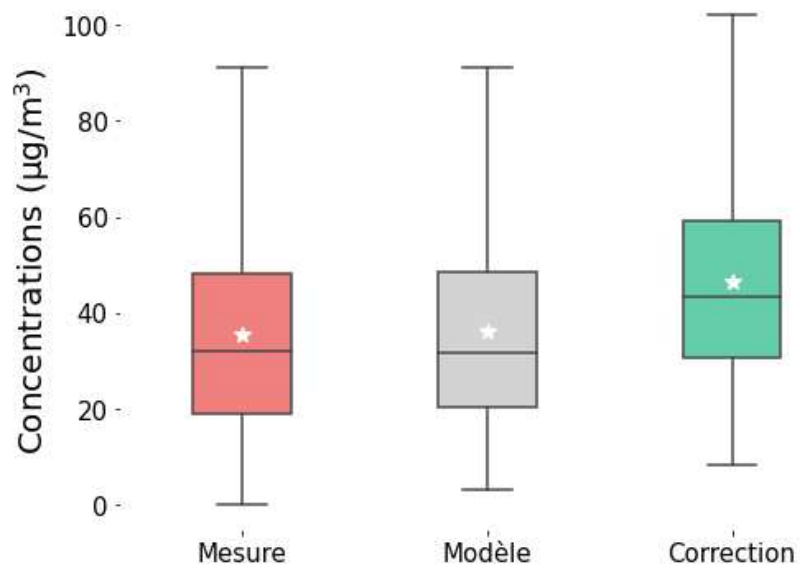
Comparaison des boxplots GON : NO₂, 2009



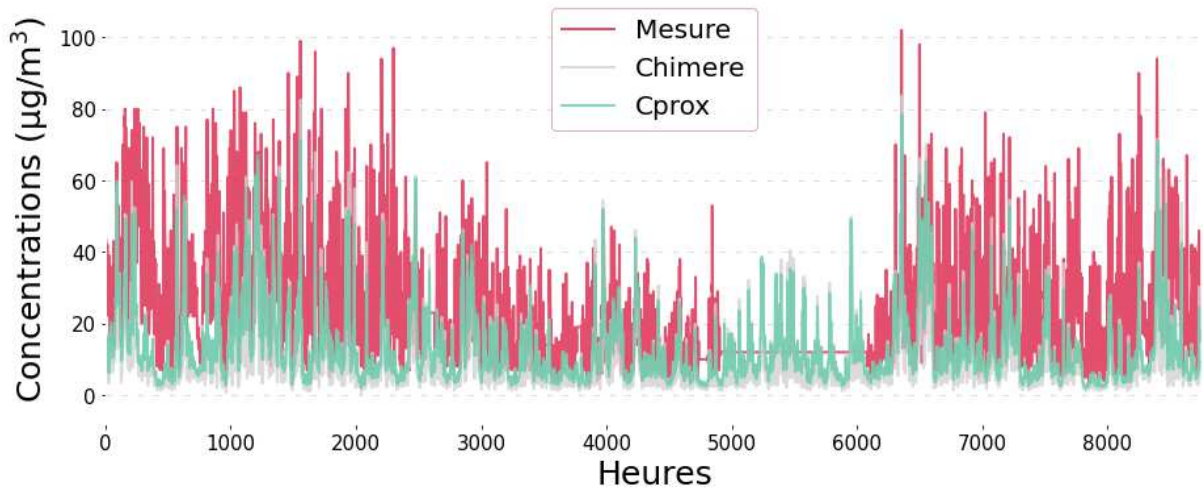
IVRY: Site de fond (NO₂, 2009)



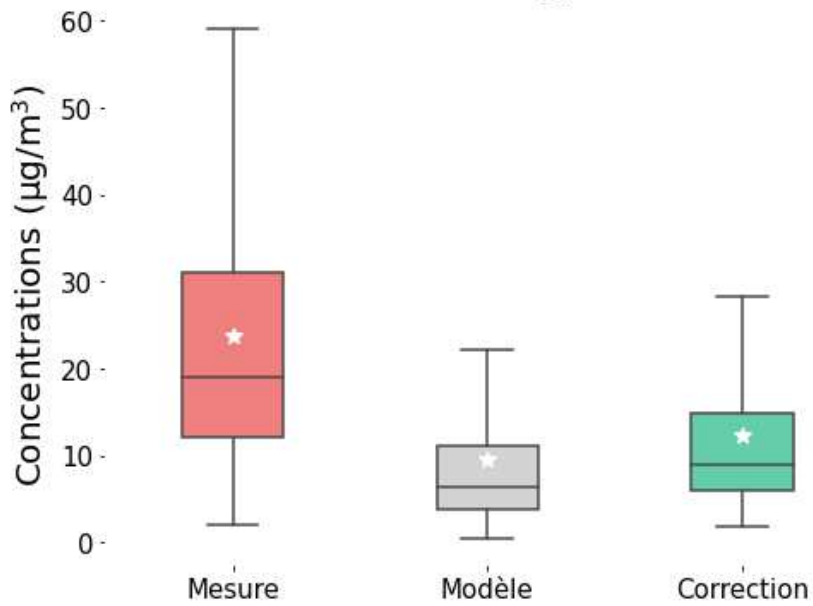
Comparaison des boxplots
IVRY : NO₂, 2009



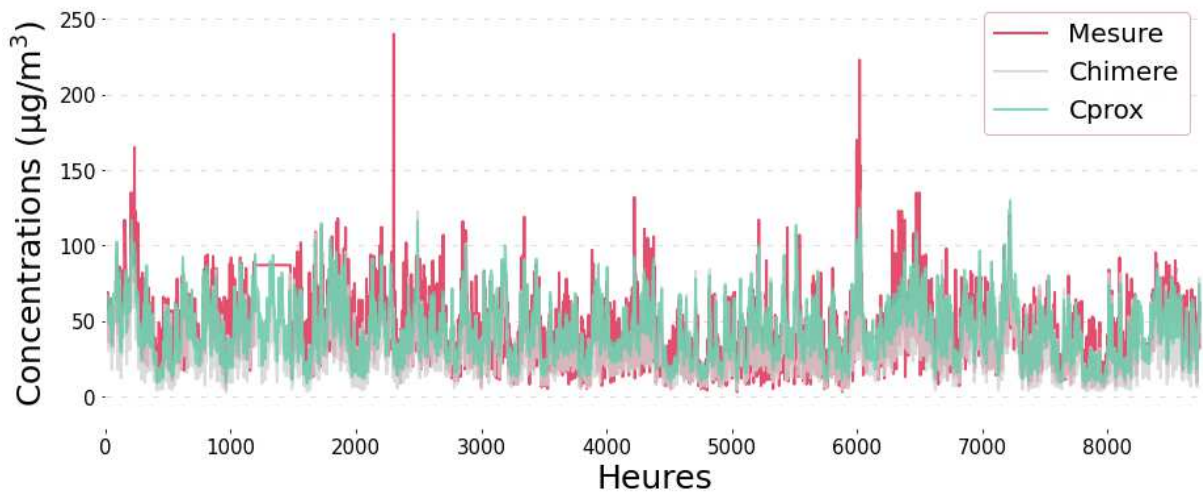
MELUN: Site de fond (NO₂, 2009)



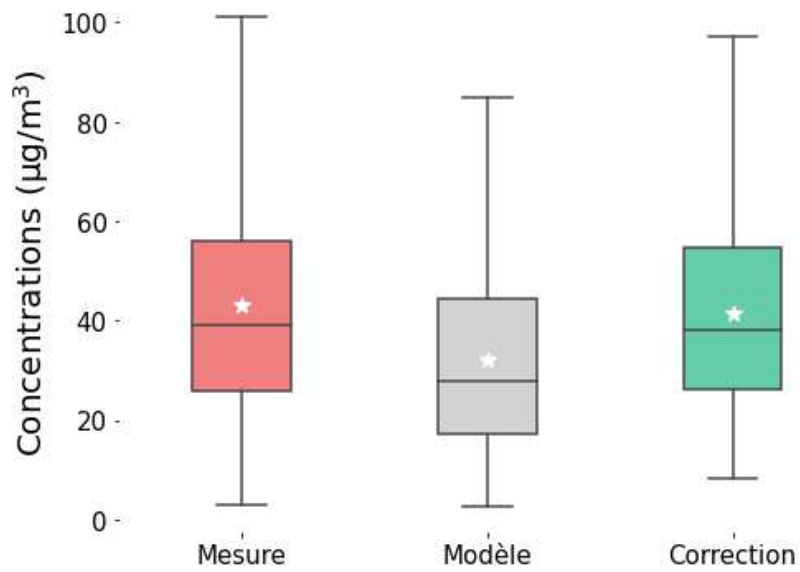
Comparaison des boxplots MELUN : NO₂, 2009



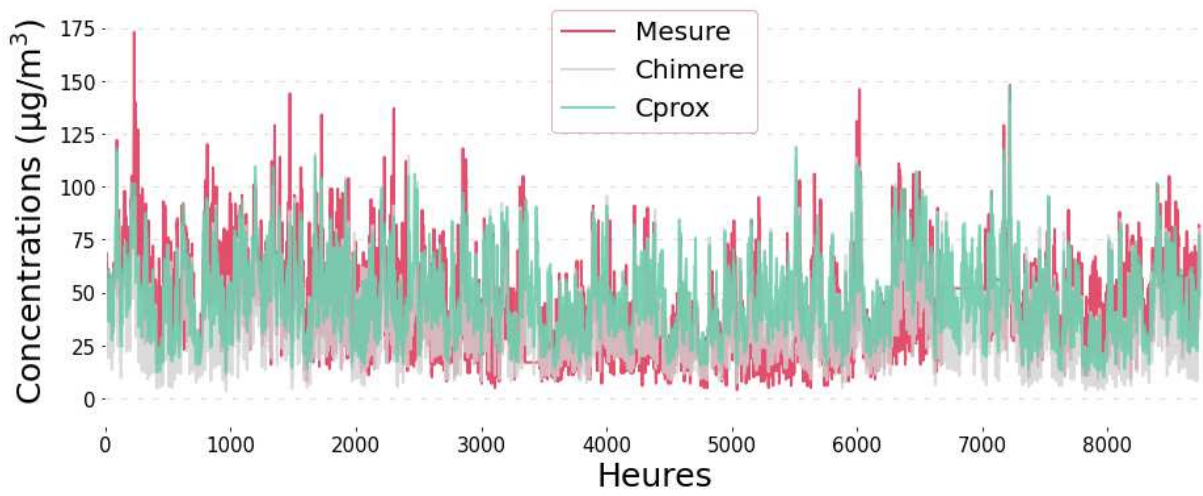
PA07: Site de fond (NO₂, 2009)



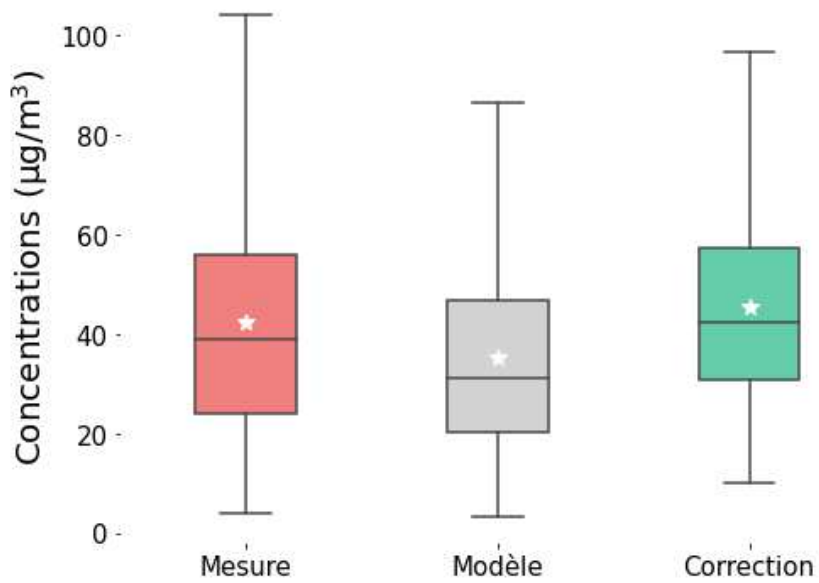
Comparaison des boxplots
PA07 : NO₂, 2009



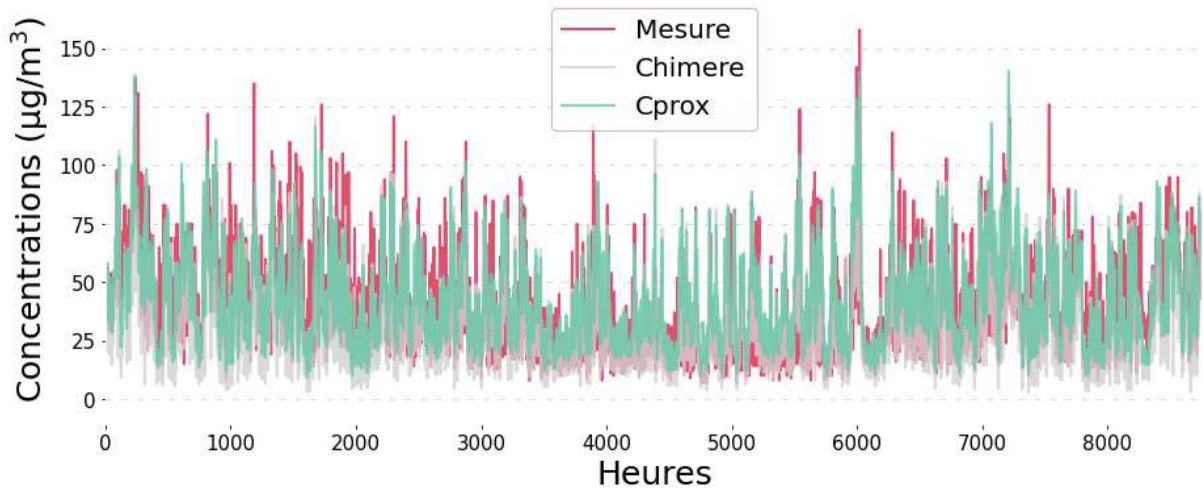
PA12: Site de fond (NO₂, 2009)



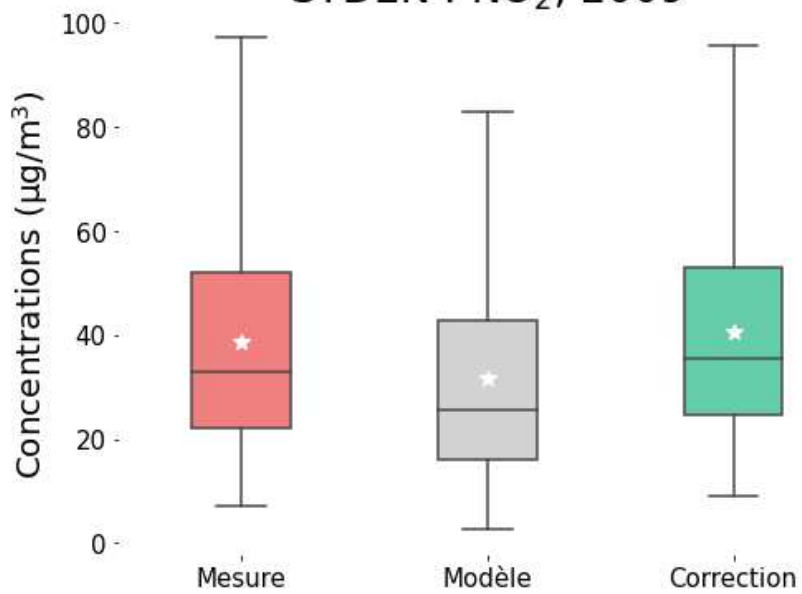
Comparaison des boxplots PA12 : NO₂, 2009



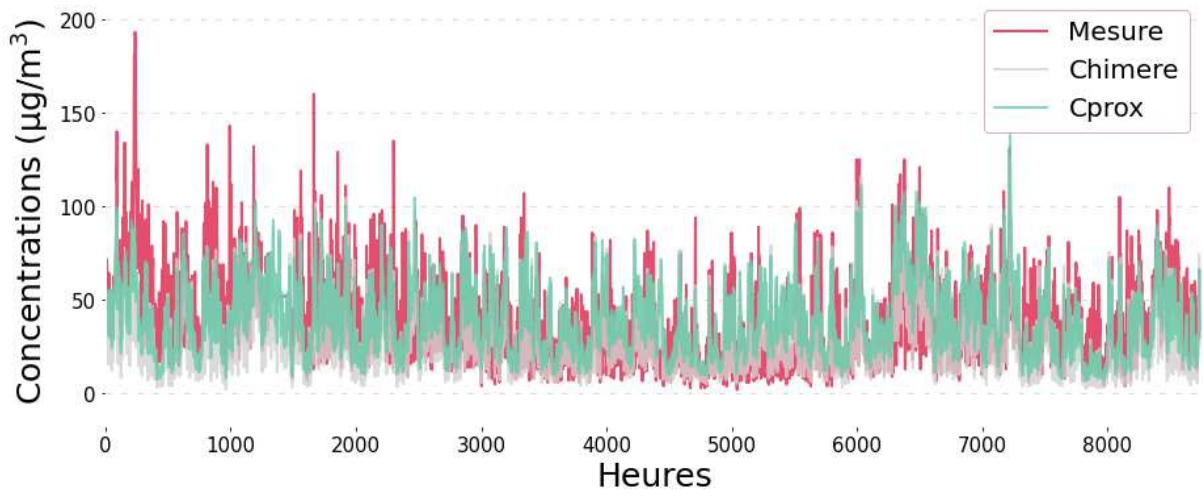
STDEN: Site de fond (NO₂, 2009)



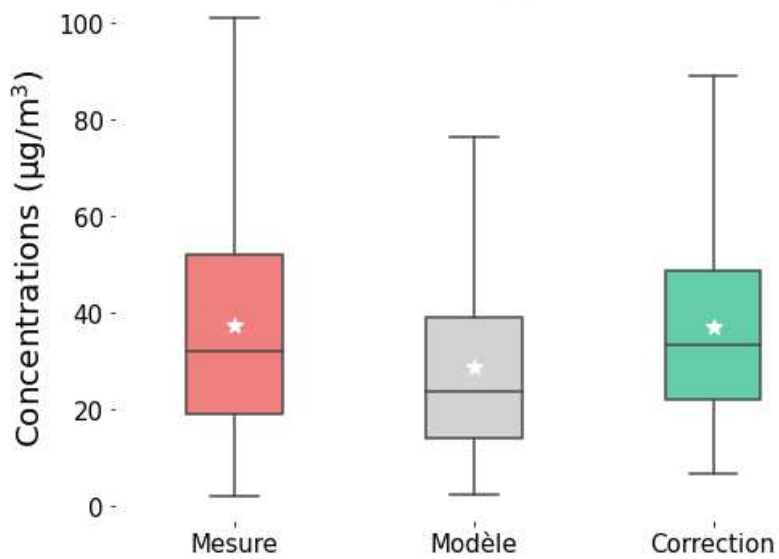
Comparaison des boxplots
STDEN : NO₂, 2009



VITRY: Site de fond (NO₂, 2009)

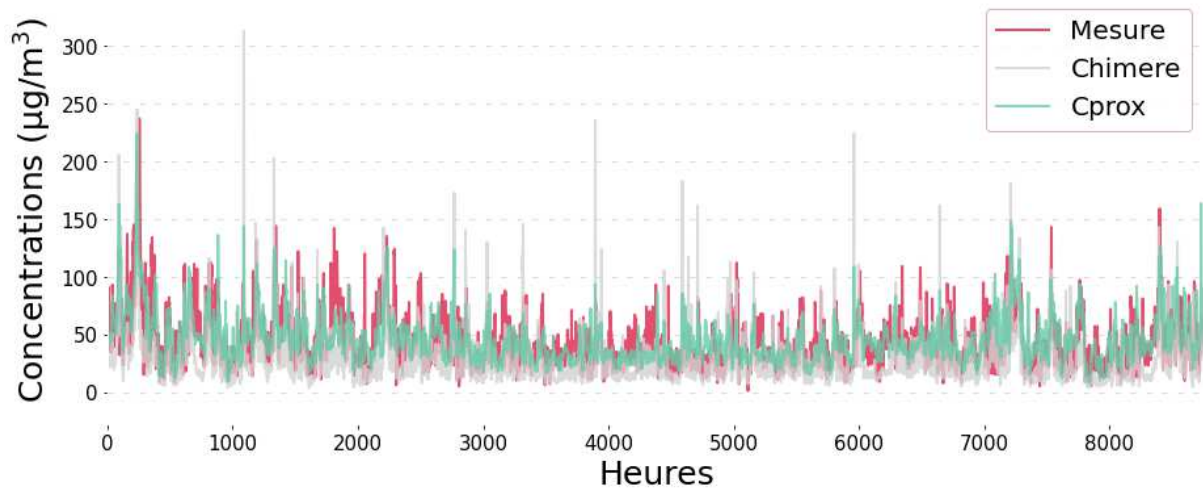


Comparaison des boxplots
VITRY : NO₂, 2009

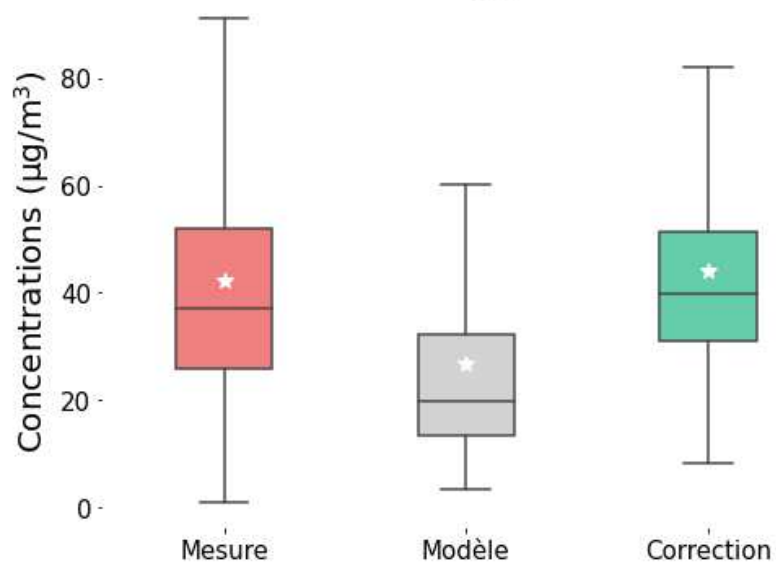


Séries temporelles et boxplots PM_{10} – Sites de trafic.

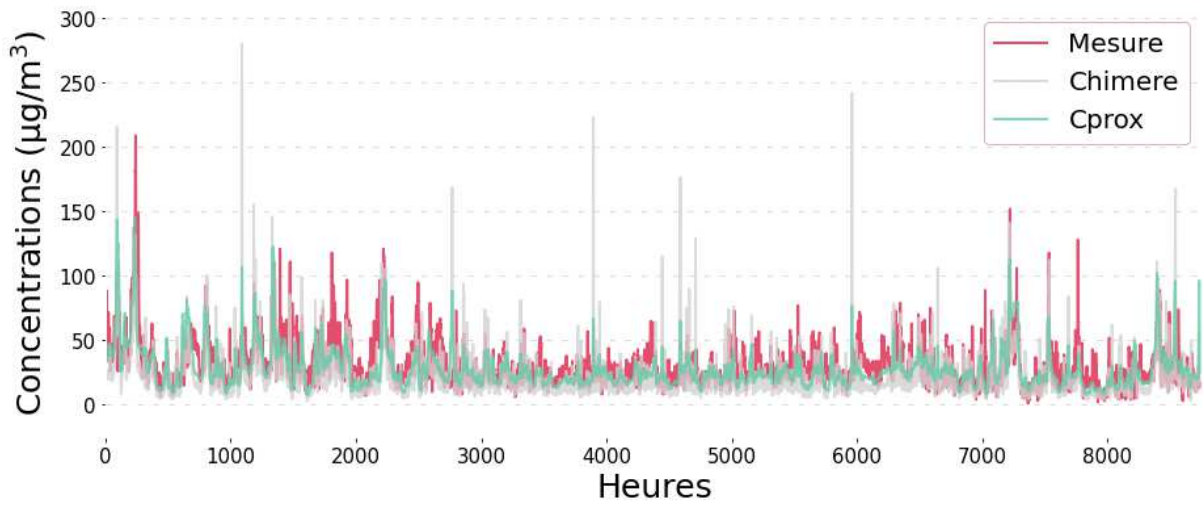
RN2: Site de trafic (NO_2 , 2009)



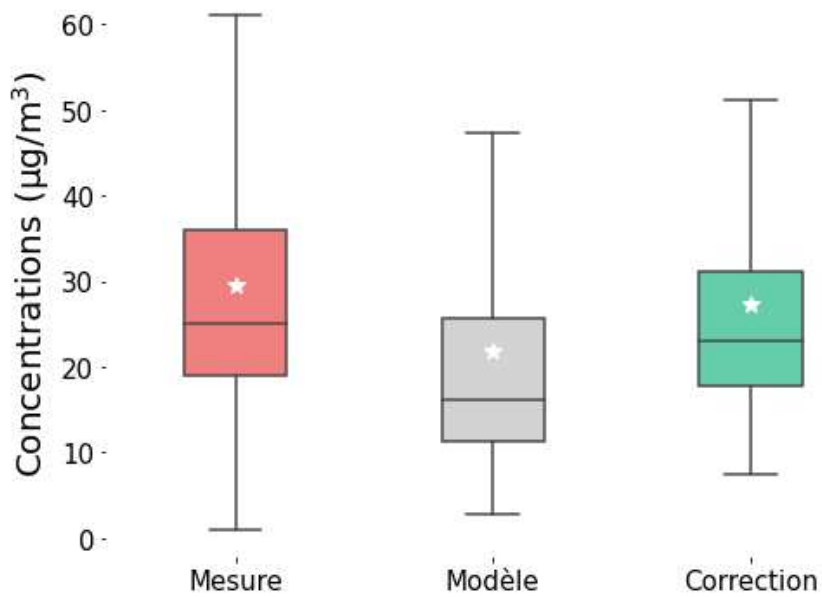
Comparaison des boxplots
RN2 : PM_{10} , 2009



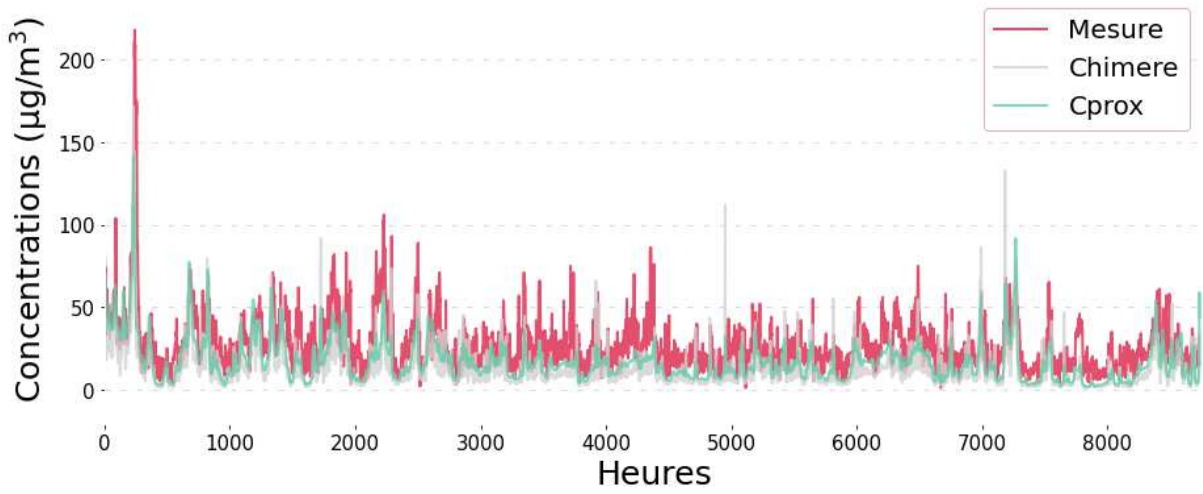
BOB: Site de fond (NO₂, 2009)



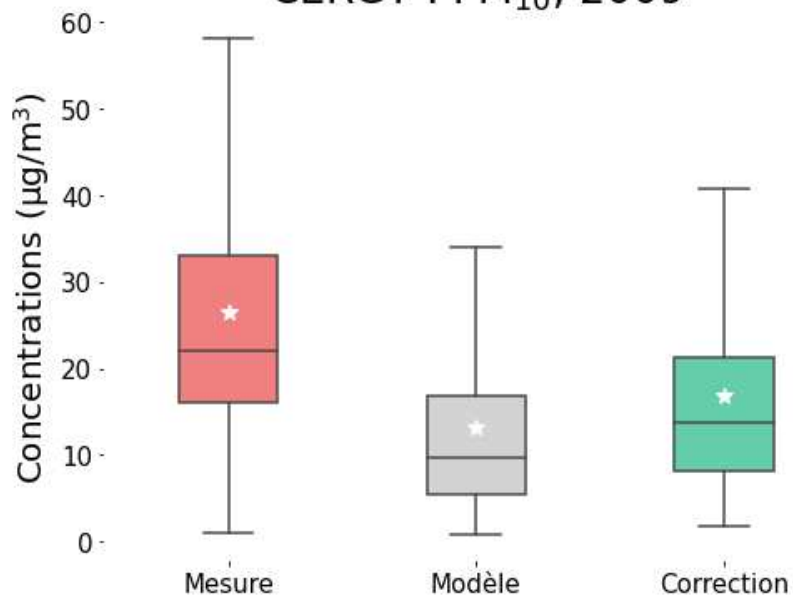
Comparaison des boxplots
BOB : PM₁₀, 2009



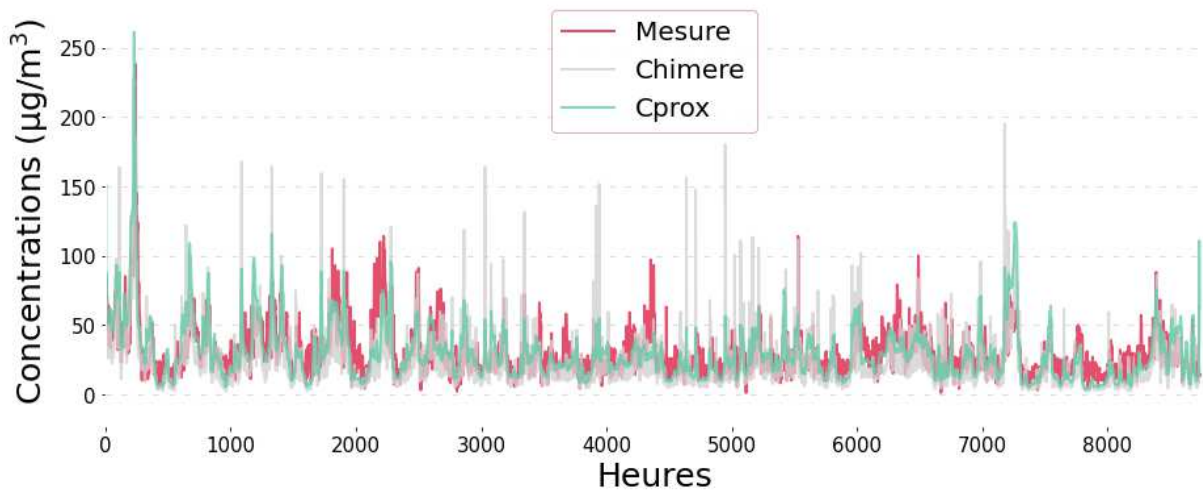
CERGY: Site de fond (NO₂, 2009)



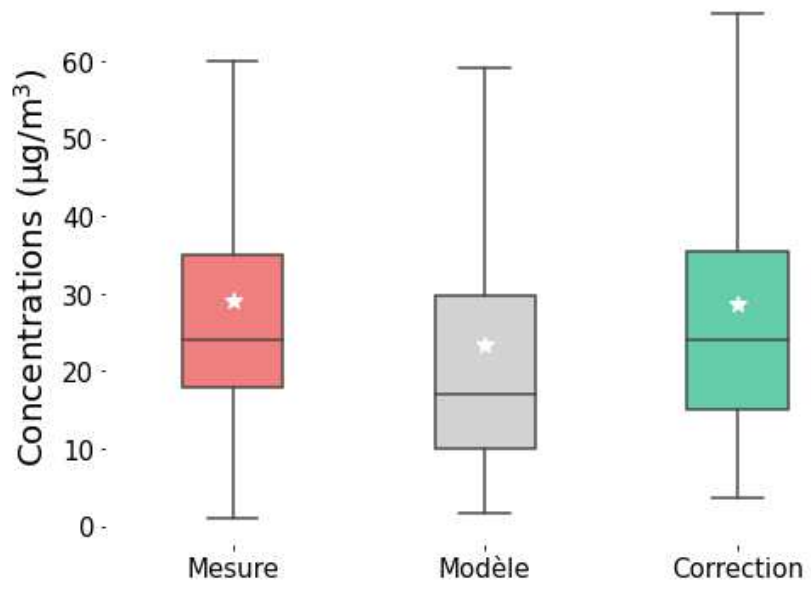
Comparaison des boxplots CERGY : PM₁₀, 2009



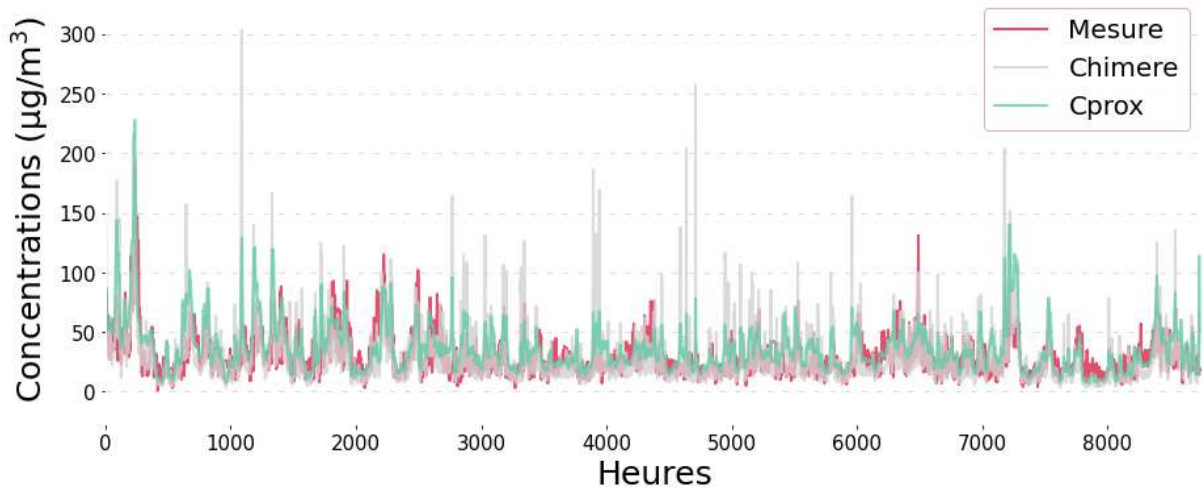
DEF: Site de fond (NO₂, 2009)



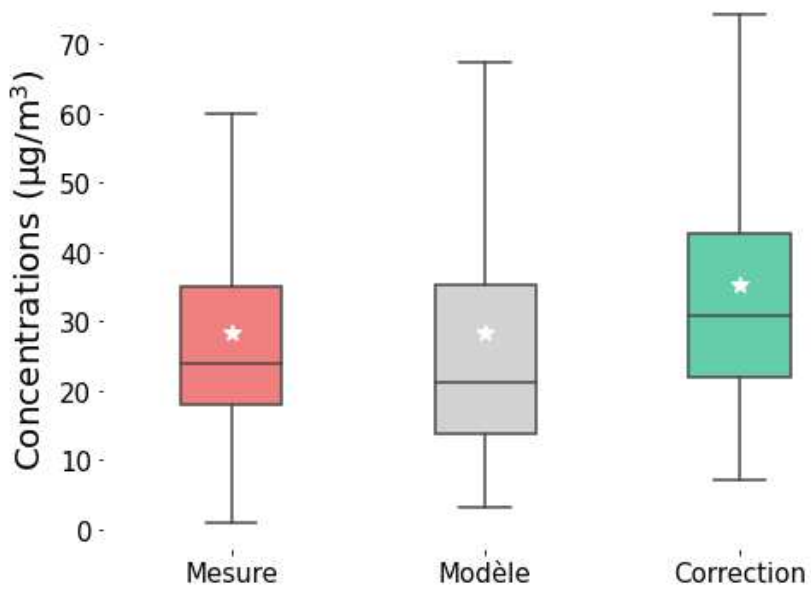
Comparaison des boxplots DEF : PM₁₀, 2009



PA18: Site de fond (NO₂, 2009)



Comparaison des boxplots PA18 : PM₁₀, 2009



ANNEXE 5

Résultats statistiques de la validation de l'approche pour les sites de validation de fond de l'année 2009.

Station	NO ₂									
	kfit	kfond	RMSE CHIM (µg/m ³)	RMSE Cprox (µg/m ³)	Moyenne CHIMERE (µg/m ³)	Moyenne Cprox (µg/m ³)	Moyenne Mesure (µg/m ³)	Variance Mesure (µg/m ³) ²	Variance CHIMERE (µg/m ³) ²	Variance Cprox (µg/m ³) ²
BOB	1,1	1,29	19,5	19,1	30,3	39	34,2	471,5	330,8	277,2
CHAMP	1,3	1,29	19,7	17,9	23,4	30,2	31,6	443,3	290,7	273,6
DEF	1,7	1,29	16,7	16,4	30,3	39,1	35,9	438,4	381,6	382,2
EVRY	1,1	1,29	20,2	18,1	15,4	19,9	28	377,5	200,2	205,2
GON	0,9	1,29	17,1	19,1	28,7	37	26,8	434,9	380,5	361,7
IVRY	1,6	1,29	19,6	22,2	36	46,4	35,5	461,7	423,8	390,1
MELUN	2,2	1,29	19,5	17,5	9,5	12,2	23,8	245,3	91,6	90,9
PA07	1,2	1,29	22,3	19,1	32,2	41,5	43,2	531,9	376,9	364,4
PA12	1,2	1,29	21,7	20,2	35,1	45,3	42,5	517,7	372	328,1
STDEN	1,3	1,29	18,1	16,3	31,5	40,6	38,7	473,3	403	382,6
VITRY	1,2	1,29	21,9	20,3	28,6	36,9	37,5	550,7	353,6	344,8

Station	PM ₁₀									
	kfit	kfond	RMSE CHIM (µg/m ³)	RMSE Cprox (µg/m ³)	Moyenne CHIMERE (µg/m ³)	Moyenne Cprox (µg/m ³)	Moyenne Mesure (µg/m ³)	Variance Mesure (µg/m ³) ²	Variance CHIMERE (µg/m ³) ²	Variance Cprox (µg/m ³) ²
BOB	2,1	3,4	19,2	14,4	21,9	27,3	29,6	294,2	347,1	262,7
CERGY	2,4	3,3	18,5	15,2	13,2	16,9	26,5	272,3	147	183,1
DEF	1,6	2,4	18,8	15,2	23,4	28,8	29,1	297,7	450	439,4
PA18	2,8	2	19,9	17,4	28,3	35,3	28,4	271,3	539,3	446,2

ANNEXE 6

Données comparées à la mesure pour les sites d'apprentissage de l'année 2017.

Tableaux de données de NO₂ – Les moyennes sont écrites en bleu dans les tableaux.

Stations de trafic										
Station	Coef. α	Coef. β	Coef. k	RMSE CHIMERE ($\mu\text{g}/\text{m}^3$)	RMSE Cprox ($\mu\text{g}/\text{m}^3$)	Moyenne CHIMERE ($\mu\text{g}/\text{m}^3$)	Moyenne Cprox ($\mu\text{g}/\text{m}^3$)	Moyenne Mesure ($\mu\text{g}/\text{m}^3$)	% α	% β
A1	1,5	0,3	1,8	66,7	49,2	21,1	45,6	83,3	83	17
AUT	2	0,6	2,6	78,2	52,5	17,7	55,6	89,9	77	23
BASCH	1,4	0,8	2,2	59,4	36,3	23	64,1	76,1	64	36
BONAP	0,9	0,5	1,4	28,4	21,4	27,3	48,3	50	64	36
CELES	1,4	0,4	1,8	41,6	32,6	26,5	58,4	59,4	78	22
ELYS	0,9	0,5	1,4	29,6	22,3	27,3	48,6	50,4	64	36
RN6	1,8	1	2,8	38,9	26,3	11,4	41	44,5	64	36
									71	29

Station de fond											
Station	Coef. α	Coef. β	Coef. k	RMSE CHIMERE ($\mu\text{g}/\text{m}^3$)	RMSE Cprox ($\mu\text{g}/\text{m}^3$)	Moyenne CHIMERE ($\mu\text{g}/\text{m}^3$)	Moyenne Cprox ($\mu\text{g}/\text{m}^3$)	Moyenne Mesure ($\mu\text{g}/\text{m}^3$)	% α	% β	
AUB	0,6	0,8	1,4	20	16,8	26,4	37	38,1	43	57	
LOGNES	0,8	0,9	1,7	18,5	15,9	15,5	26,4	27,1	47	53	
MANT	1,9	1,1	3	18,5	13,3	6,1	18,4	19,8	63	37	
MONTG	0,7	0,8	1,5	15,9	13,3	14,5	21,8	24,7	47	53	
NEUIL	0,8	0,7	1,5	19,7	16,1	23,9	35,9	36,1	53	47	
PA04C	0,7	0,6	1,3	18,6	15,1	26,5	34,4	36,8	54	46	
PA13	0,6	0,7	1,3	17	14,6	24,2	31,4	32,6	46	54	
PA15L	0,6	0,6	1,2	17,2	15,5	22,9	27,5	29,5	50	50	
PA18	0,7	0,6	1,3	19,4	16,1	28	36,4	38	54	46	
TREMB	0,4	0,8	1,2	15,7	14,8	21,8	26,2	27,9	33	67	
VERS	0,8	0,8	1,6	15,3	12,5	13,8	22	22,9	50	50	
VILLEM	0,7	0,8	1,5	18,7	16,7	20,1	30,1	30,8	47	53	
									1,54	49	51

Tableaux de données des PM₁₀ – Les moyennes sont écrites en bleu dans les tableaux.

Station de trafic										
Station	Coef. α	Coef. β	Coef. k	RMSE CHIMERE ($\mu\text{g}/\text{m}^3$)	RMSE Cprox ($\mu\text{g}/\text{m}^3$)	Moyenne CHIMERE ($\mu\text{g}/\text{m}^3$)	Moyenne Cprox ($\mu\text{g}/\text{m}^3$)	Moyenne Mesure ($\mu\text{g}/\text{m}^3$)	% α	% β
A1	1,8	0,7	2,5	32,4	25,4	15,1	25,1	43,1	72	28
AUT	1,8	0,7	2,5	25,6	19,8	13	22,1	34,5	72	28
BASCH	1,4	0,2	1,6	19,2	16,5	17	20,5	30,9	88	13
ELYS	1,4	0,5	1,9	22,7	19,4	17,9	23,1	33,2	74	26
RN6	2,1	0,8	2,9	20	15,4	11	22,7	25,9	72	28
									76	24

Station de fond										
Station	Coef. α	Coef. β	Coef. k	RMSE CHIMERE ($\mu\text{g}/\text{m}^3$)	RMSE Cprox ($\mu\text{g}/\text{m}^3$)	Moyenne CHIMERE ($\mu\text{g}/\text{m}^3$)	Moyenne Cprox ($\mu\text{g}/\text{m}^3$)	Moyenne Mesure ($\mu\text{g}/\text{m}^3$)	% α	% β
LOGNES	1,5	0,3	1,8	12,2	10,9	12,7	15,1	19,6	83	17
NOGENT	1,4	0,2	1,6	12,5	11,5	14,7	15,7	19,9	88	13
PA04C	1,1	0,2	1,3	13,8	12	18,2	15,4	21,1	85	15
PA15L	1,2	0,2	1,4	12,8	11,8	15,7	14,6	19,4	86	14
RAMBO	1,6	0,7	2,3	9,9	8,6	8,5	11,7	14,5	70	30
TREMB	1,7	0,2	1,9	12,9	11,3	14,7	18,8	20,2	89	11
VITRY	1,2	0,3	1,5	13	12,2	14,7	14	20,2	80	20
			1,69						83	17

Tableaux de données de NO₂ – Comparaison des variances des sites de trafic.

Station	Étape d'apprentissage			Étape de généralisation
	Variance Mesure ($\mu\text{g}/\text{m}^3$) ²	Variance Cprox ($\mu\text{g}/\text{m}^3$) ²	Variance CHIMERE ($\mu\text{g}/\text{m}^3$) ²	Variance Cprox ($\mu\text{g}/\text{m}^3$) ²
A1	658,4	640,9	317,4	542,9
AUT	1020,5	1015,7	262,4	465,6
BASCH	937,8	926,4	311,5	1377,9
BONAP	485,9	432,9	369,4	457,5
CELES	891,5	608,0	349,9	517,2
ELYS	494,9	453,5	358,5	617,1
RN6	487,5	484,6	106,2	328,9

Tableaux de données de NO₂ – Comparaison des variances des sites de fond.

Station	Étape d'apprentissage			Étape de généralisation
	Variance Mesure (µg/m ³) ²	Variance Cprox (µg/m ³) ²	Variance CHIMERE (µg/m ³) ²	Variance Cprox (µg/m ³) ²
AUB	486,3	403,1	326,6	390,9
LOGNES	345,0	299,8	173,6	204,7
MANT	229,5	224,3	45,6	56,8
MONTG	286,5	283,5	193,3	259,2
NEUIL	475,2	451,9	334,0	427,1
PA04C	422,1	408,4	295,9	418,1
PA13	357,9	348,1	253,0	399,2
PA15L	378,8	340,4	270,1	376,9
PA18	404,3	337,1	349,9	423,7
TREMB	391,4	353,3	322,3	287,9
VERS	397,3	340,2	321,9	230,6
VILLEM	407,8	265,0	305,5	306,3

Tableaux de données de PM₁₀ – Comparaison des variances des sites de trafic.

Station	Étape d'apprentissage			Étape de généralisation
	Variance Mesure (µg/m ³) ²	Variance Cprox (µg/m ³) ²	Variance CHIMERE (µg/m ³) ²	Variance Cprox (µg/m ³) ²
A1	343,8	333,9	202,8	175,2
AUT	327,8	327,8	168,0	211,7
BASCH	238,1	235,2	256,7	473,7
ELYS	268,2	267,8	276,6	257,1
RN6	255,8	255,4	91,2	168,4

Tableaux de données de PM₁₀ – Comparaison des variances des sites de fond.

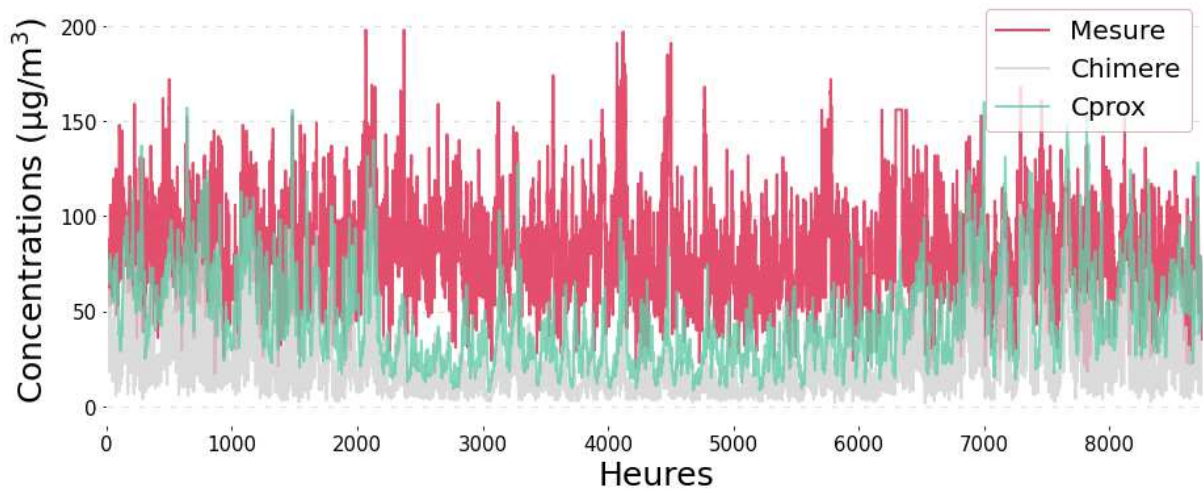
Station	Étape d'apprentissage			Étape de généralisation
	Variance Mesure ($\mu\text{g}/\text{m}^3$) ²	Variance Cprox ($\mu\text{g}/\text{m}^3$) ²	Variance CHIMERE ($\mu\text{g}/\text{m}^3$) ²	Variance Cprox ($\mu\text{g}/\text{m}^3$) ²
LOGNES	138,7	122,0	135,4	120,4
NOGENT	172,9	186,7	168,1	174,9
PA04C	165,6	279,0	162,1	257,7
PA15L	170,6	224,6	166,9	223,4
RAMBO	113,1	60,9	110,8	78,2
TREMB	201,2	162,1	195,2	145,8
VITRY	161,1	208,1	156,9	200,2

ANNEXE 7

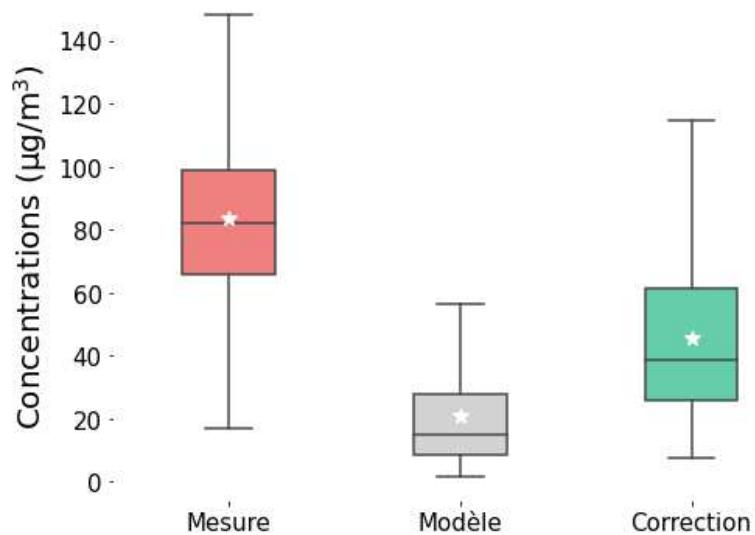
Données comparées à la mesure pour les sites d'apprentissage de l'année 2017.

Séries temporelles NO₂ – Sites de trafic.

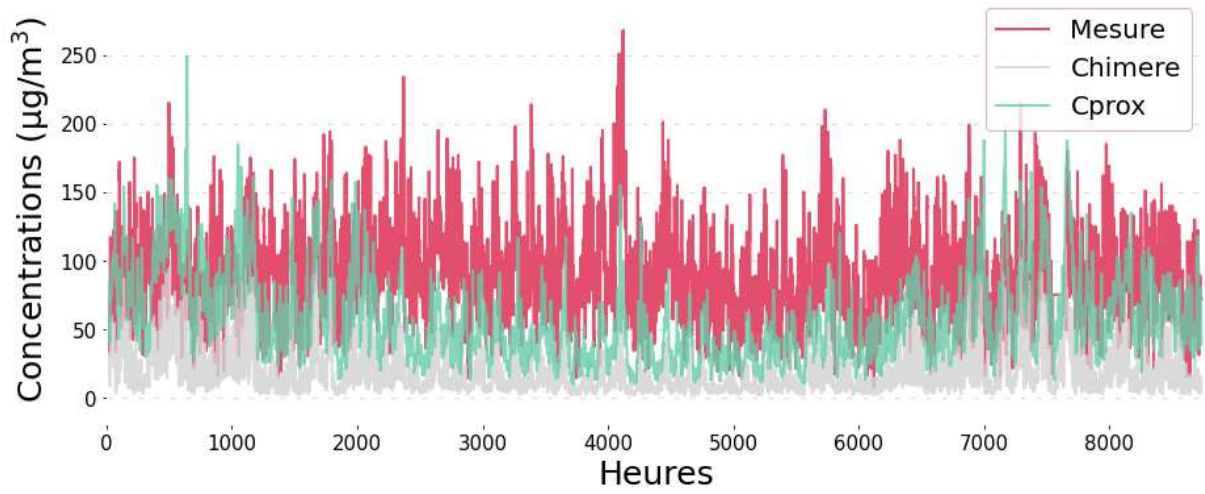
A1: Site de trafic (NO₂, 2017)



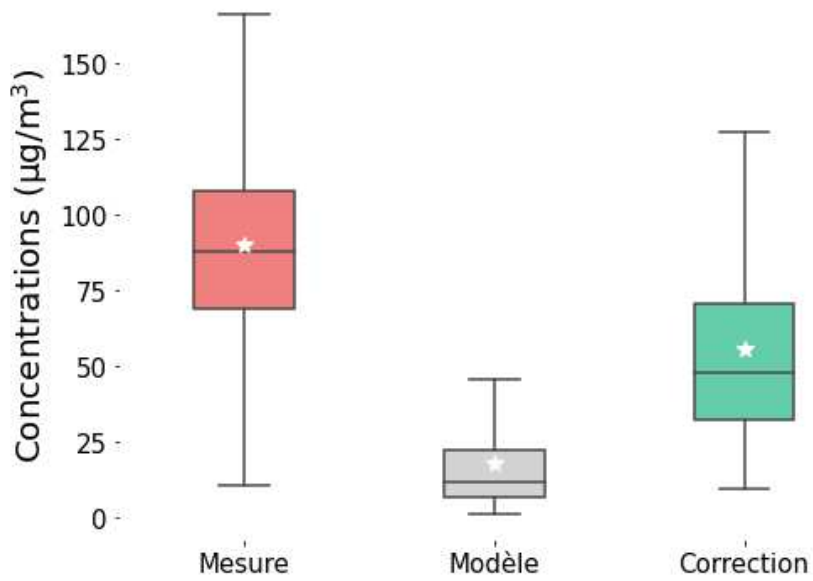
Comparaison des boxplots
A1 : NO₂, 2017



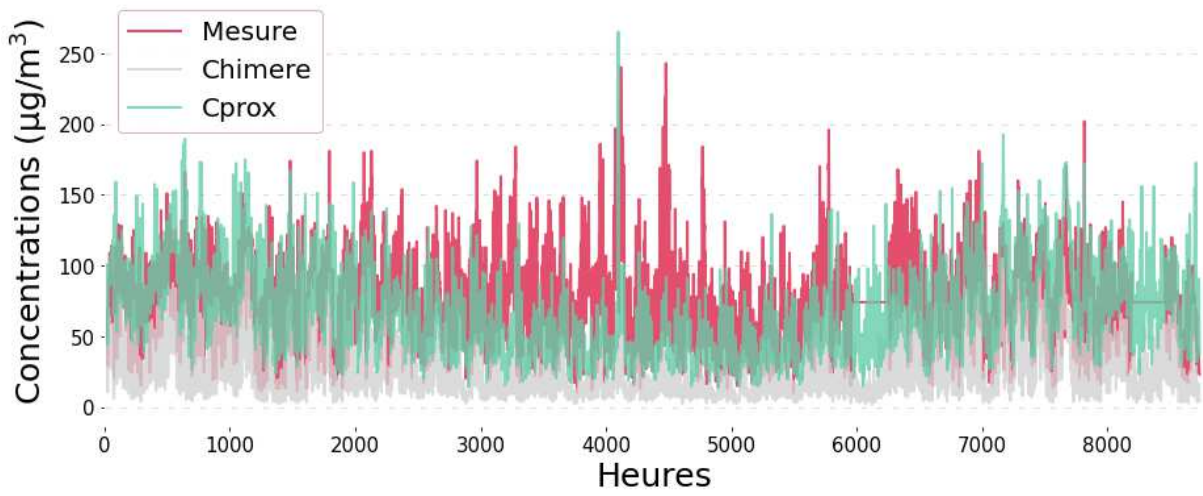
AUT: Site de trafic (NO₂, 2017)



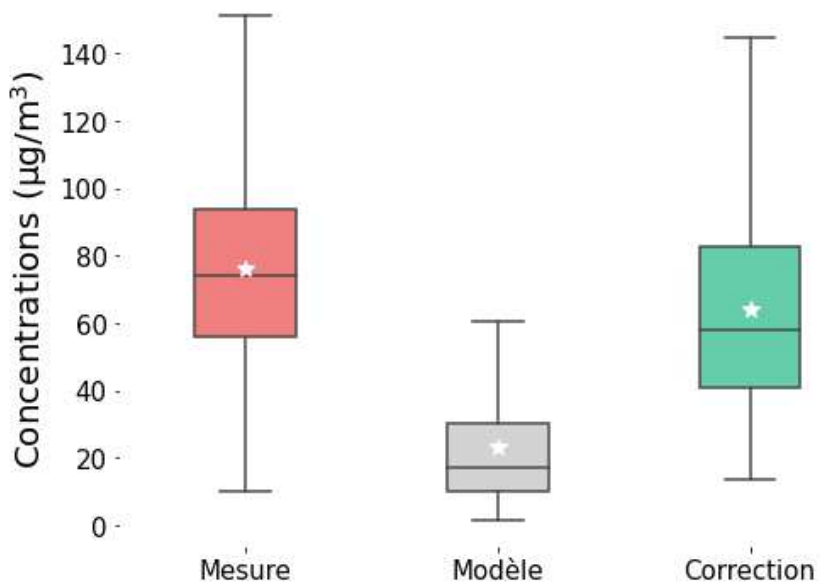
Comparaison des boxplots AUT : NO₂, 2017



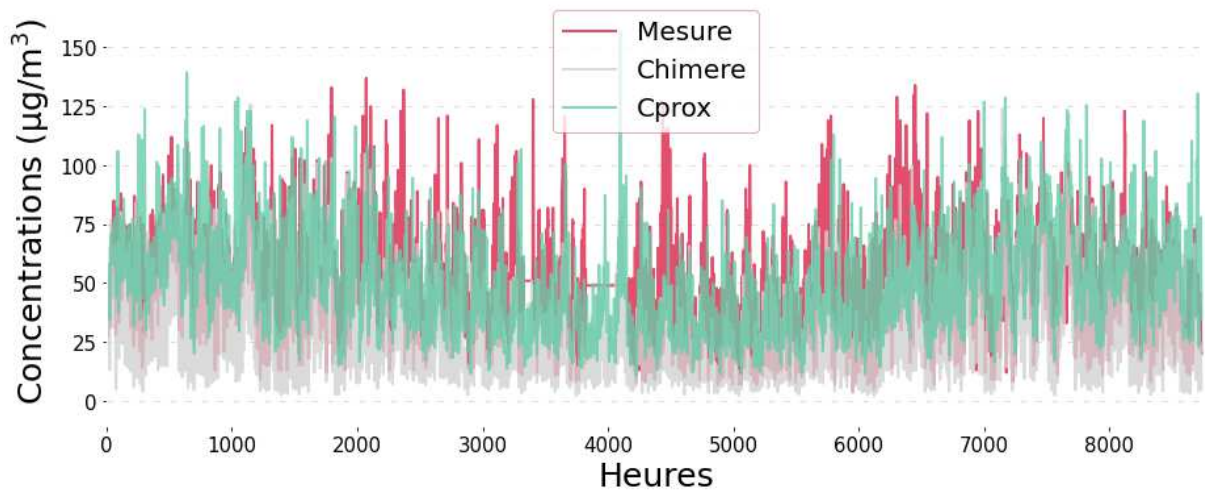
BASCH: Site de trafic (NO₂, 2017)



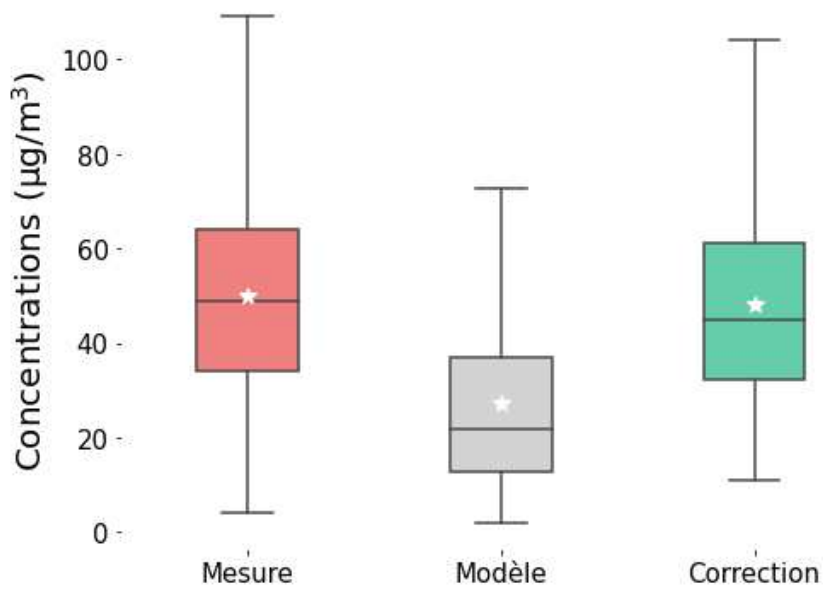
Comparaison des boxplots BASCH : NO₂, 2017



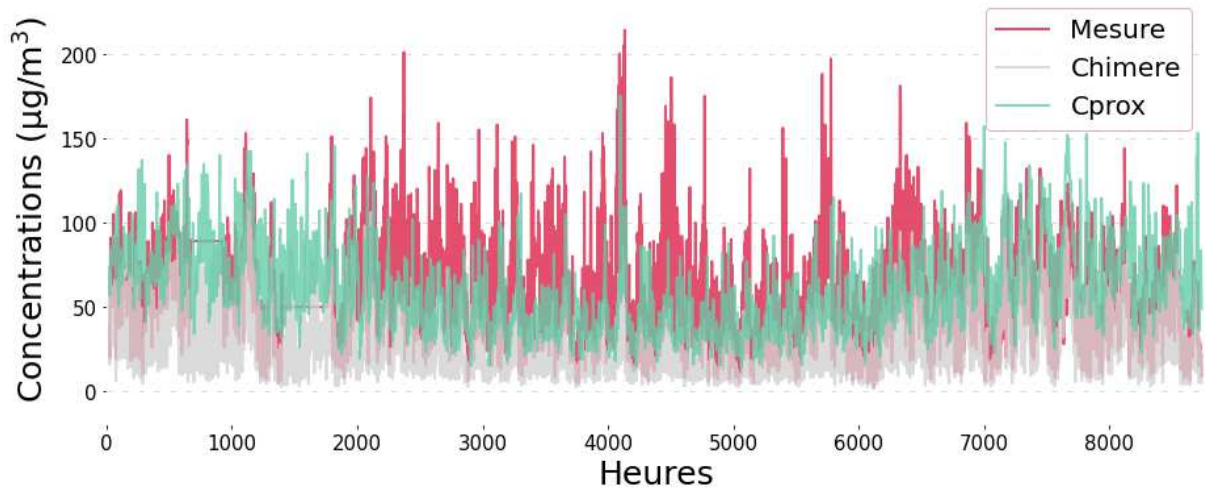
BONAP: Site de trafic (NO₂, 2017)



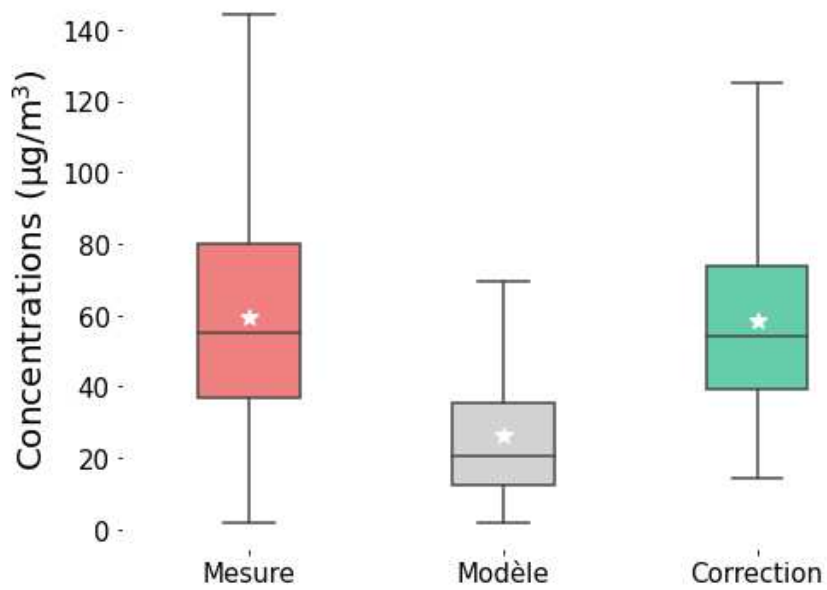
Comparaison des boxplots BONAP : NO₂, 2017



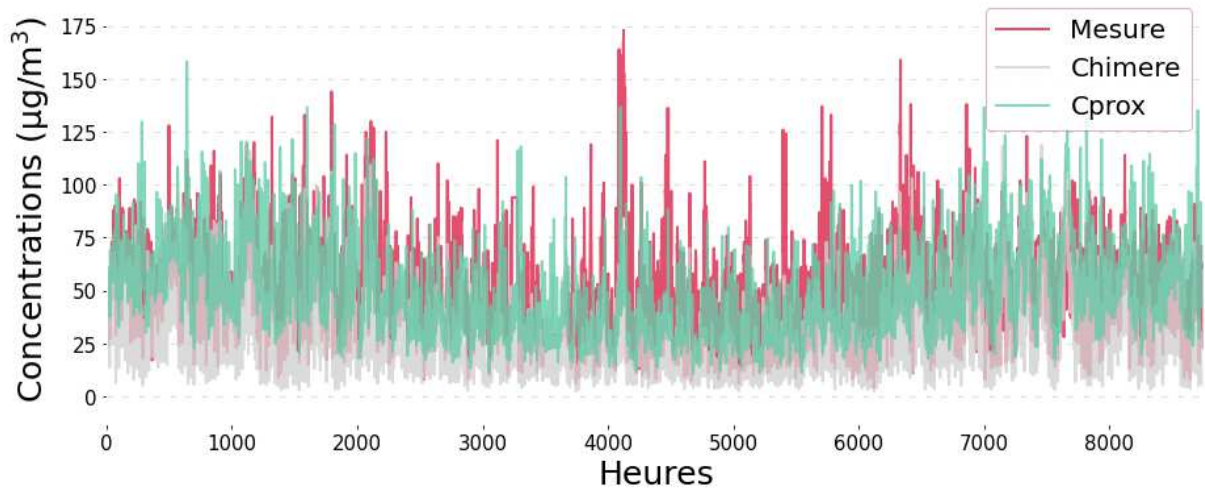
CELES: Site de trafic (NO₂, 2017)



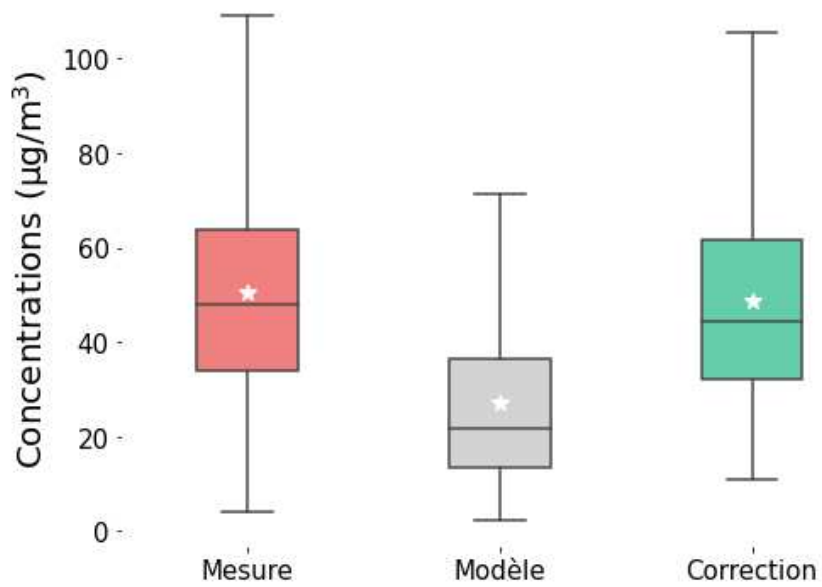
Comparaison des boxplots
CELES : NO₂, 2017



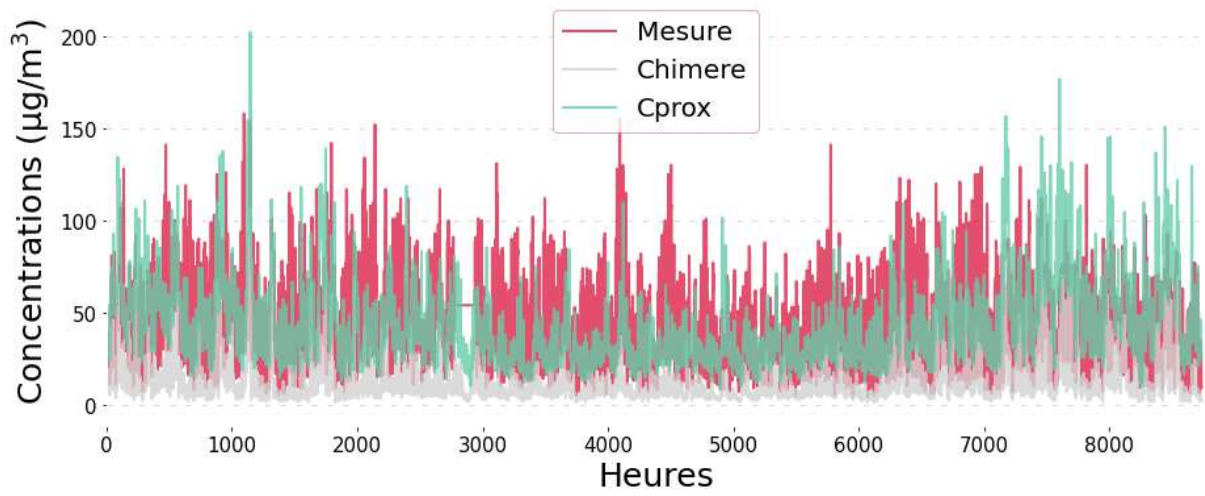
ELYS: Site de trafic (NO₂, 2017)



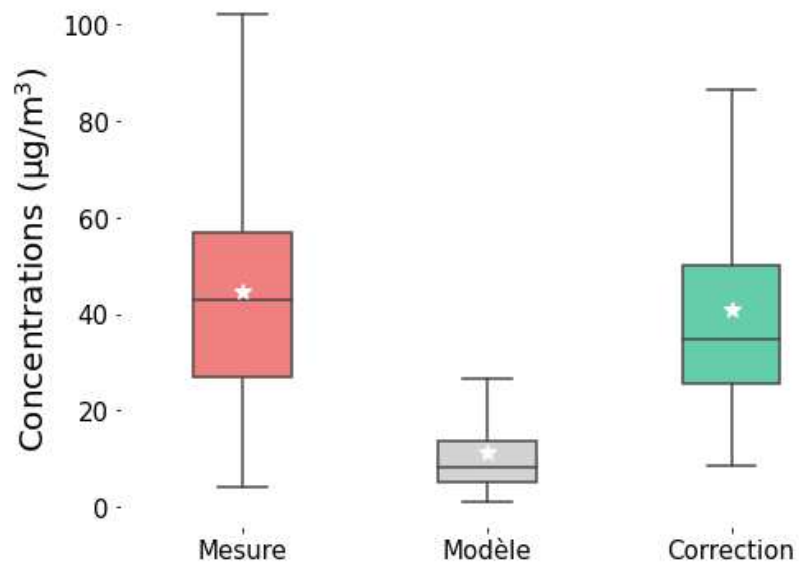
Comparaison des boxplots
ELYS : NO₂, 2017



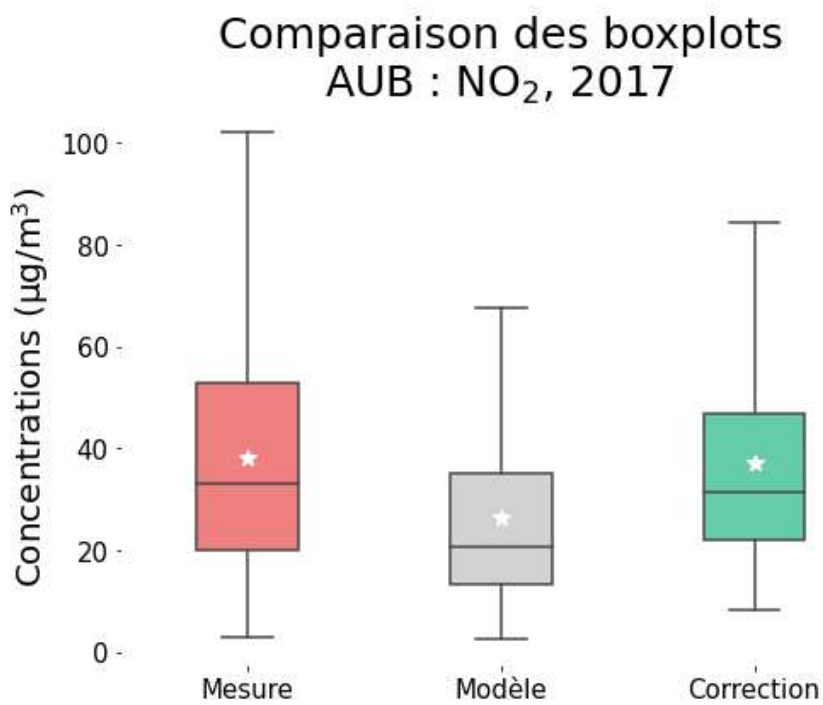
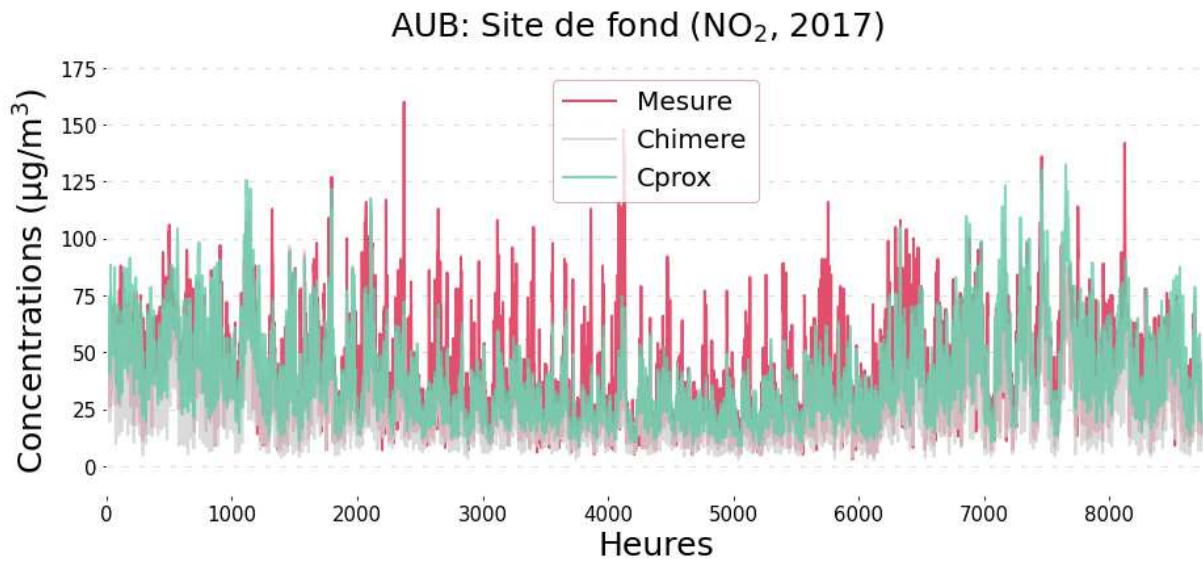
RN6: Site de trafic (NO₂, 2017)



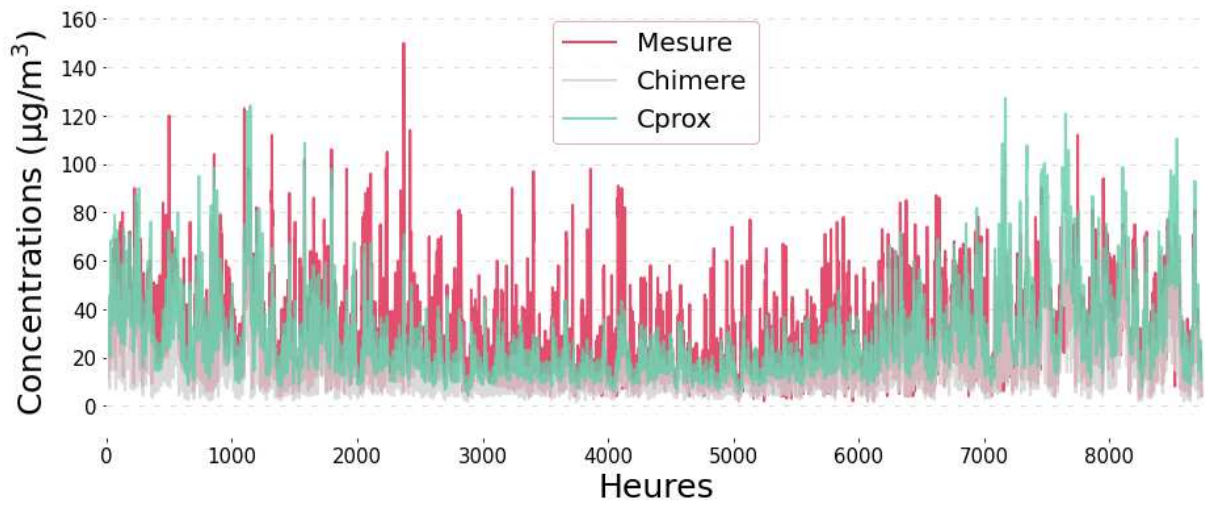
Comparaison des boxplots
RN6 : NO₂, 2017



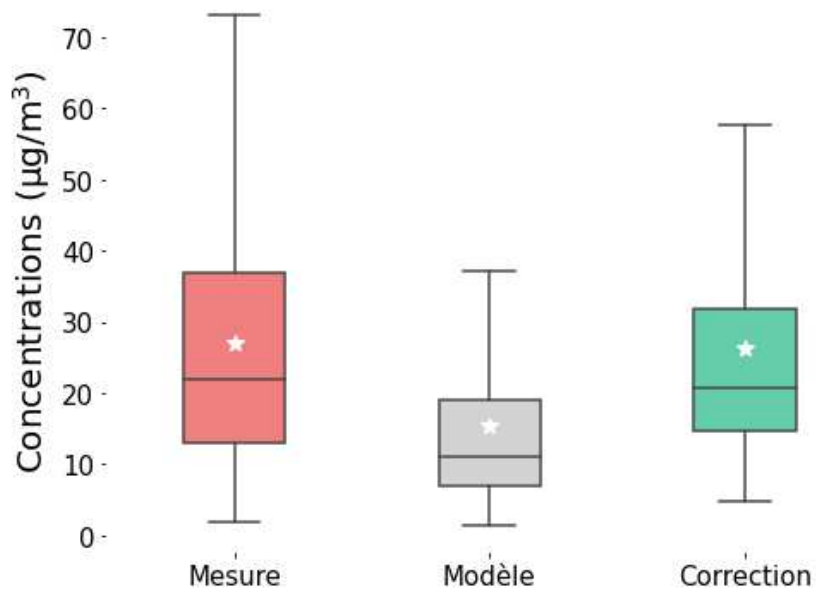
Séries temporelles NO₂ – Sites de fond.



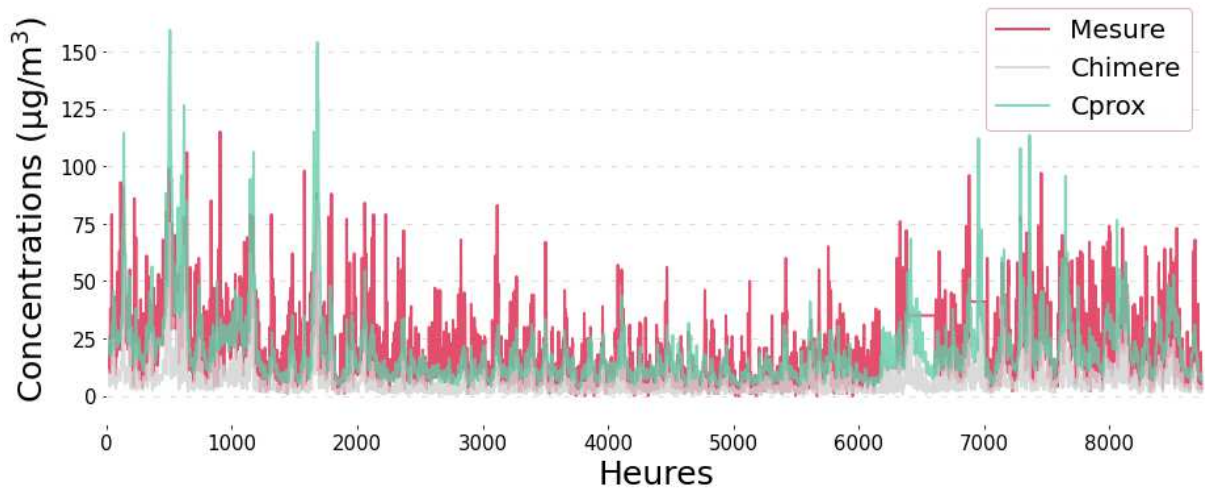
LOGNES: Site de fond (NO₂, 2017)



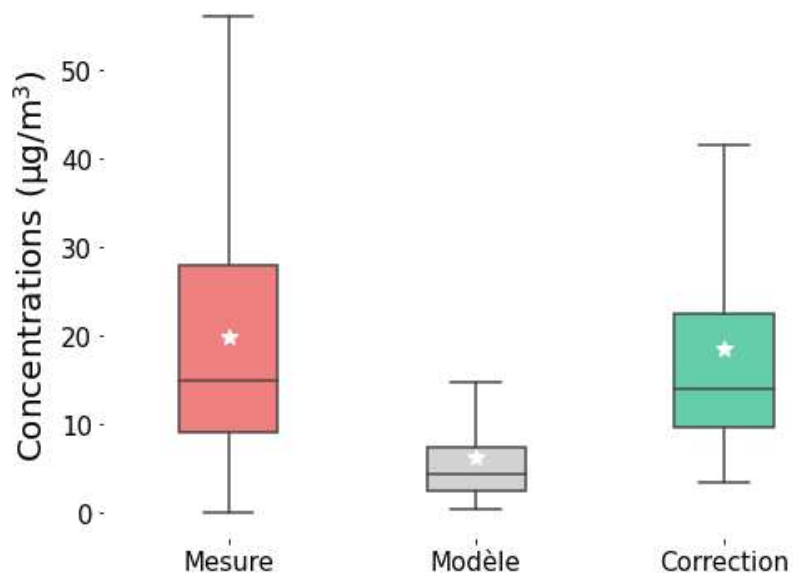
Comparaison des boxplots LOGNES : NO₂, 2017



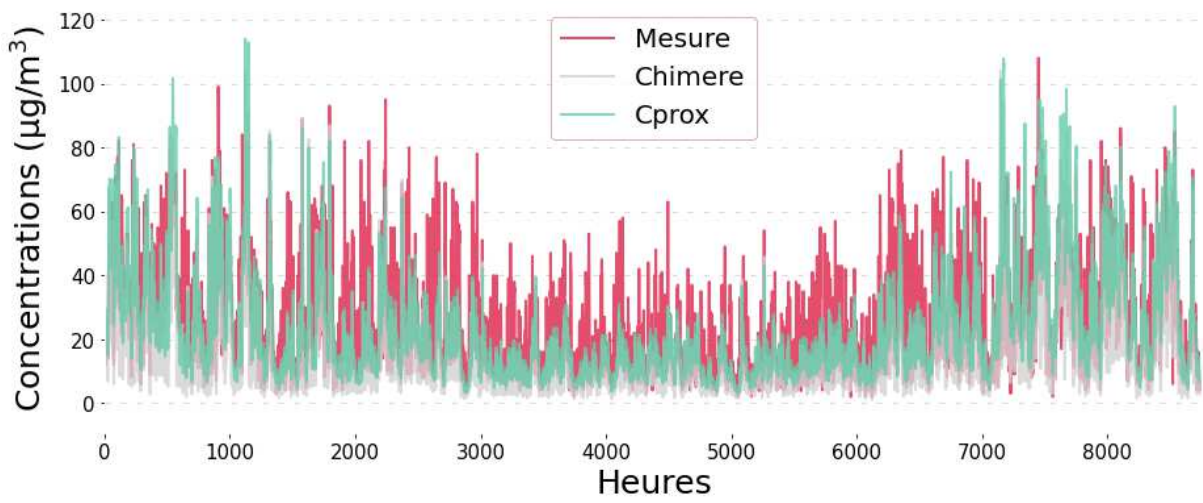
MANT: Site de fond (NO₂, 2017)



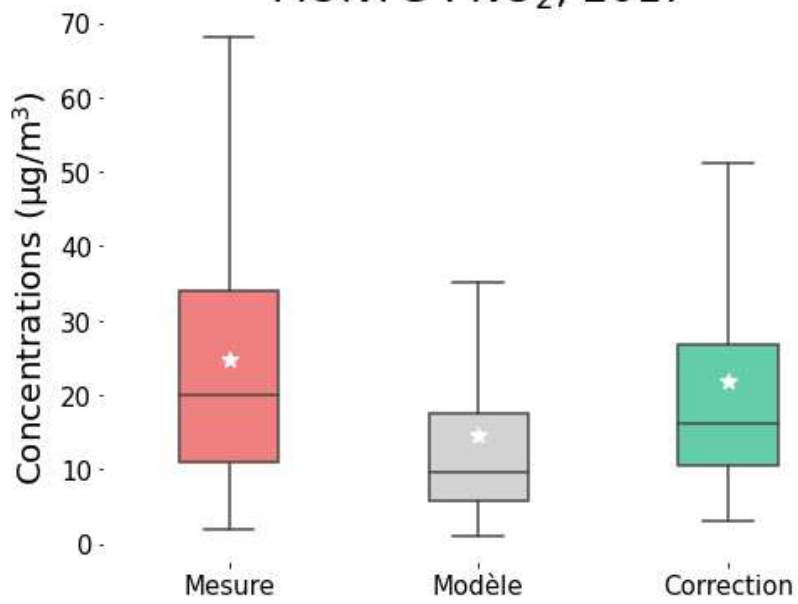
Comparaison des boxplots MANT : NO₂, 2017



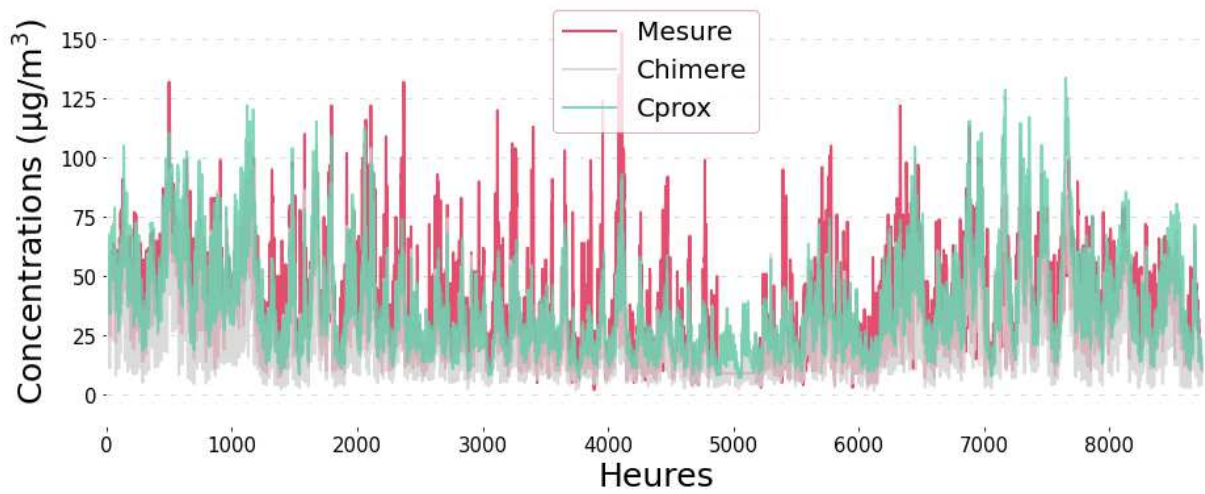
MONTG: Site de fond (NO₂, 2017)



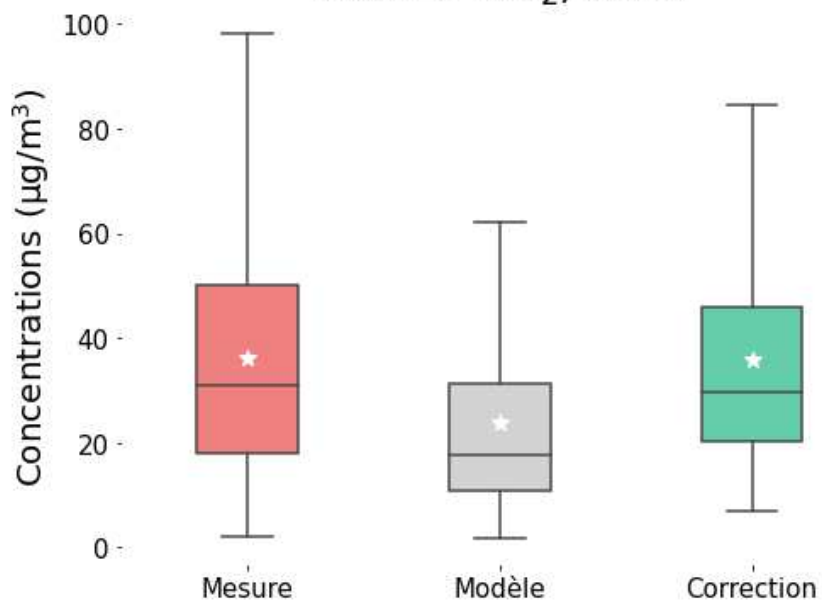
Comparaison des boxplots
MONTG : NO₂, 2017



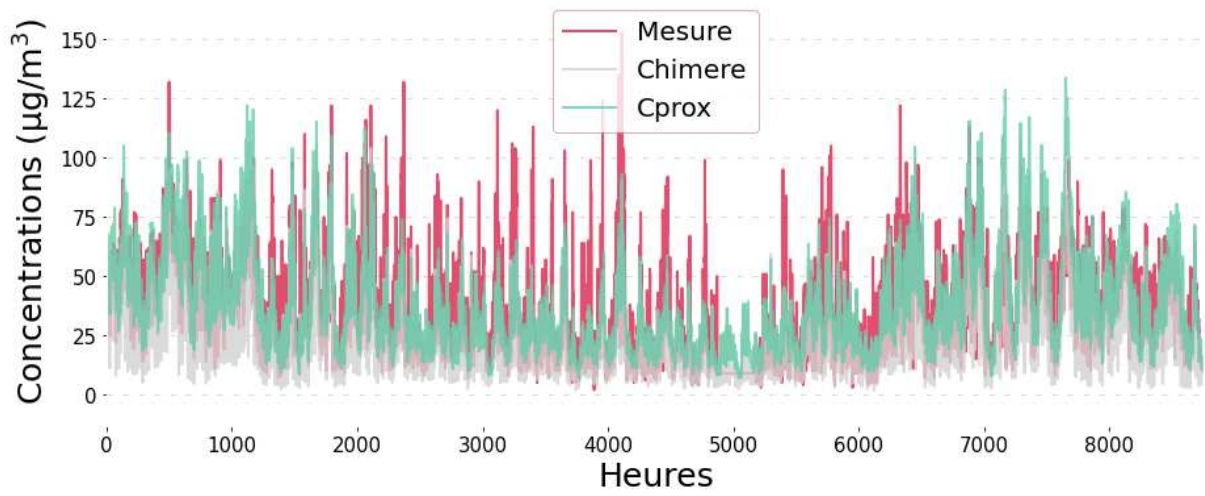
NEUIL: Site de fond (NO₂, 2017)



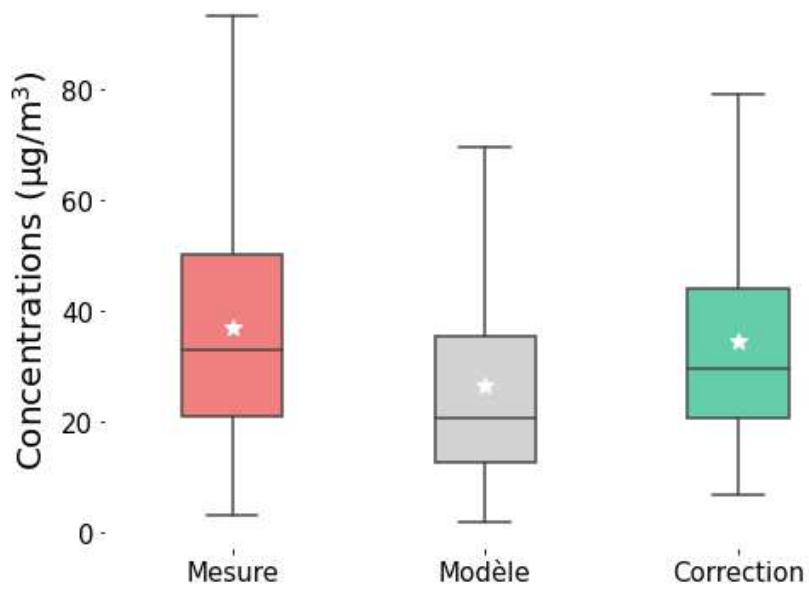
Comparaison des boxplots NEUIL : NO₂, 2017



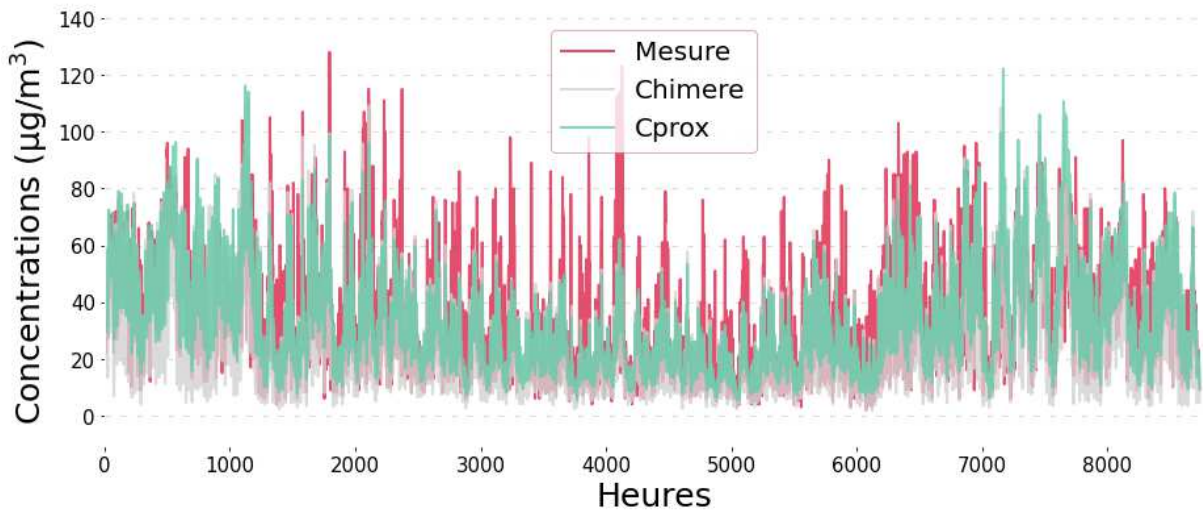
NEUIL: Site de fond (NO₂, 2017)



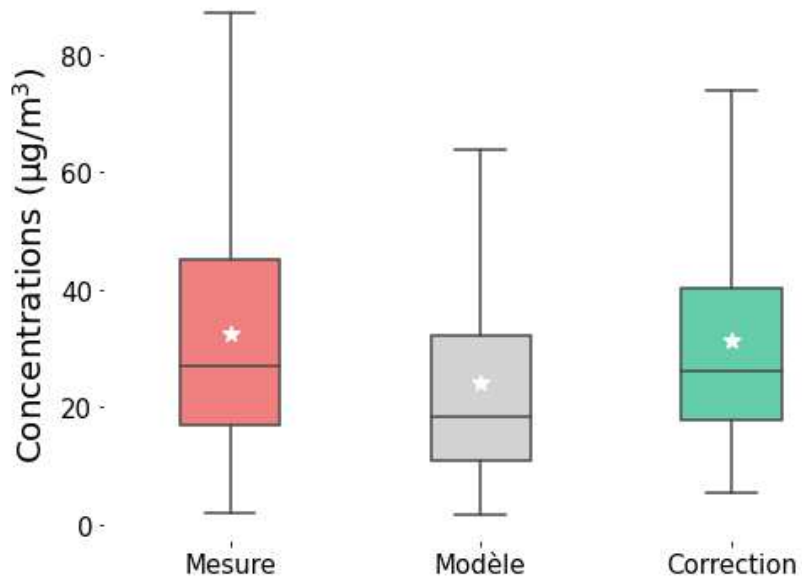
Comparaison des boxplots PA04C : NO₂, 2017



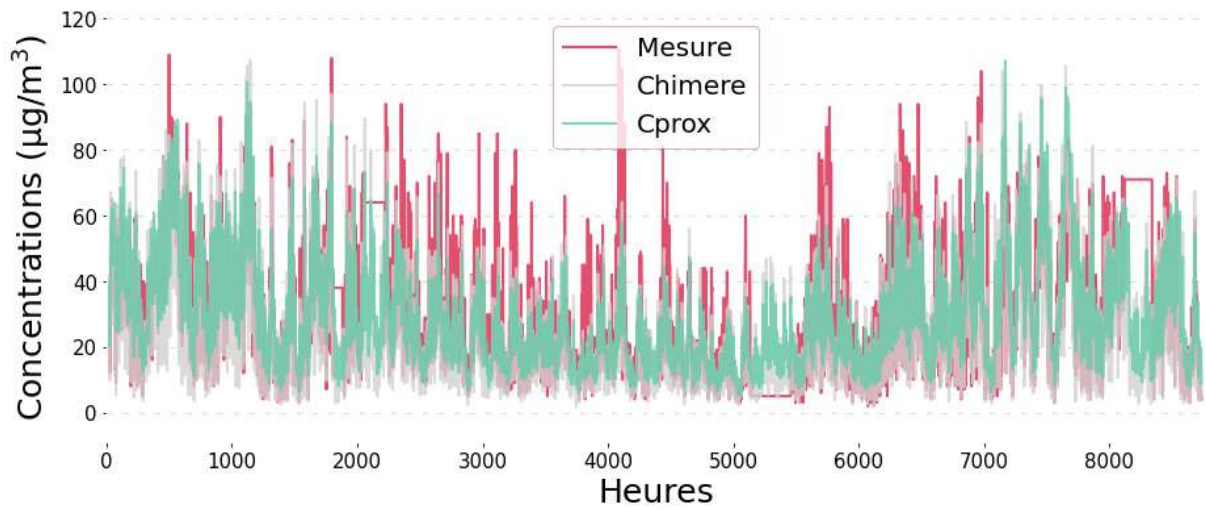
PA13: Site de fond (NO₂, 2017)



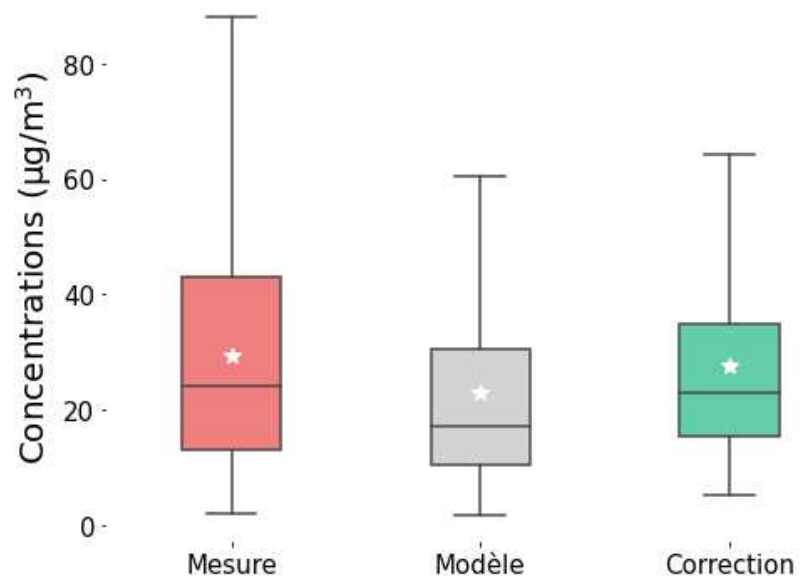
Comparaison des boxplots
PA13 : NO₂, 2017



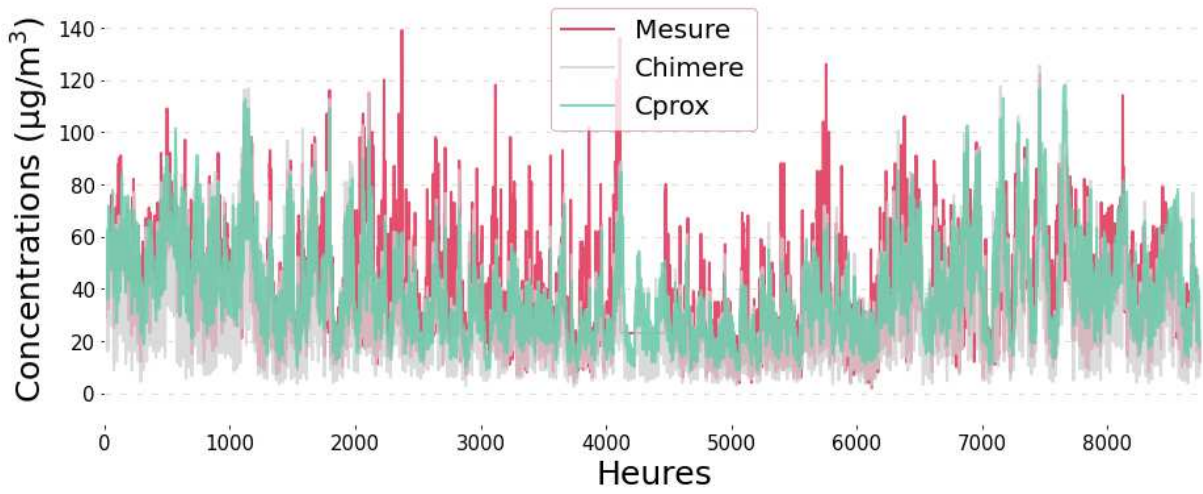
PA15L: Site de fond (NO₂, 2017)



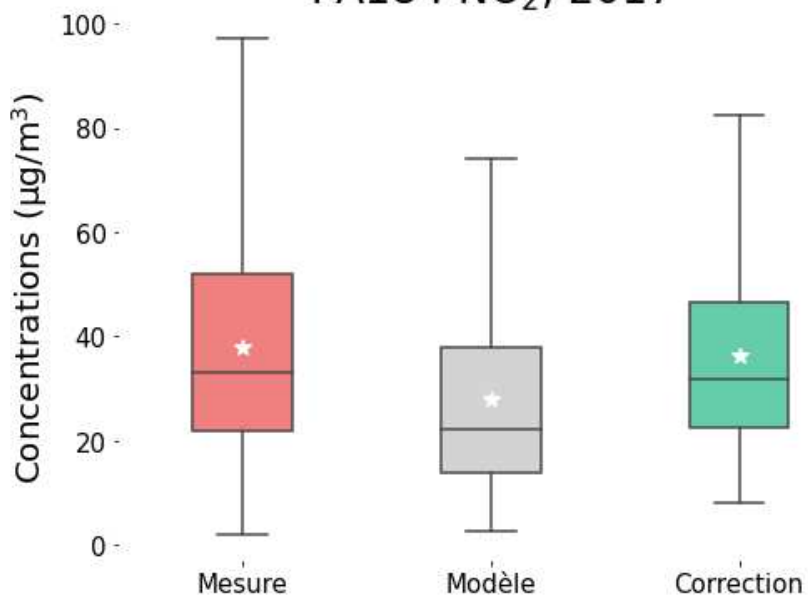
Comparaison des boxplots
PA15L : NO₂, 2017



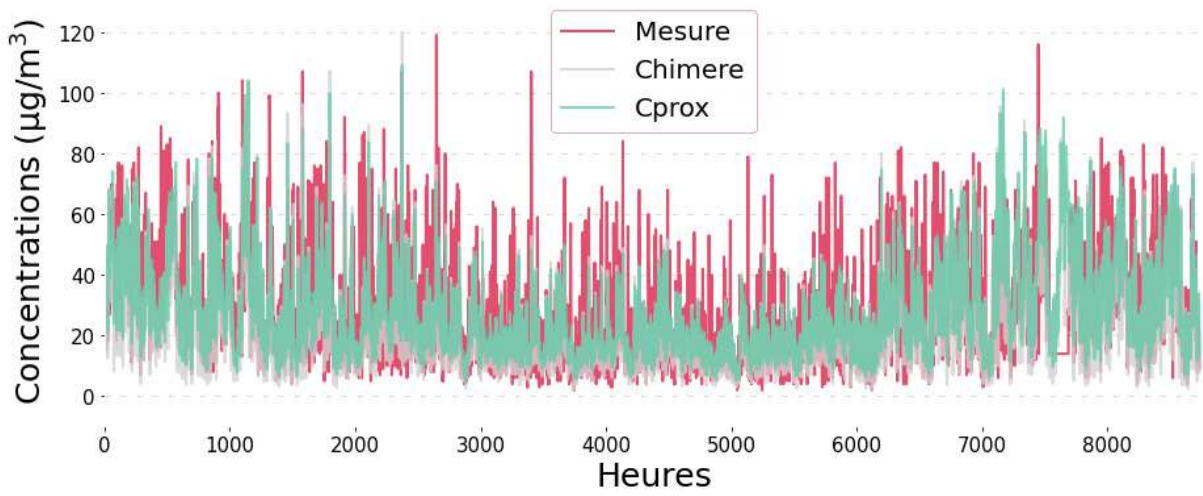
PA18: Site de fond (NO₂, 2017)



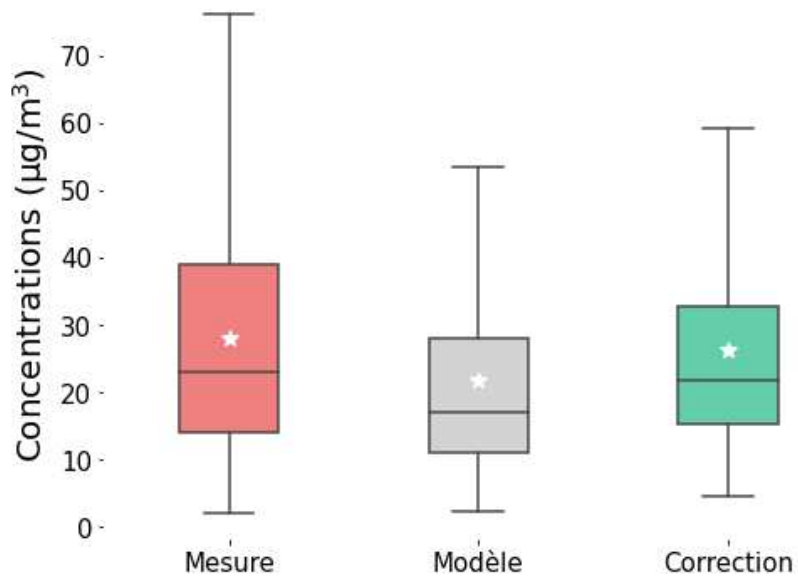
Comparaison des boxplots
PA18 : NO₂, 2017



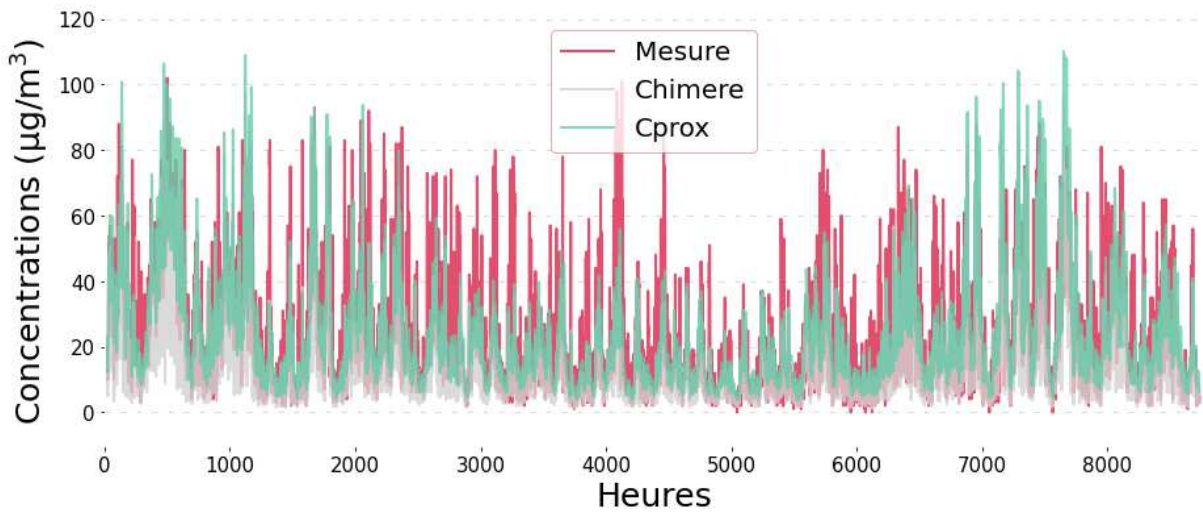
TREMB: Site de fond (NO₂, 2017)



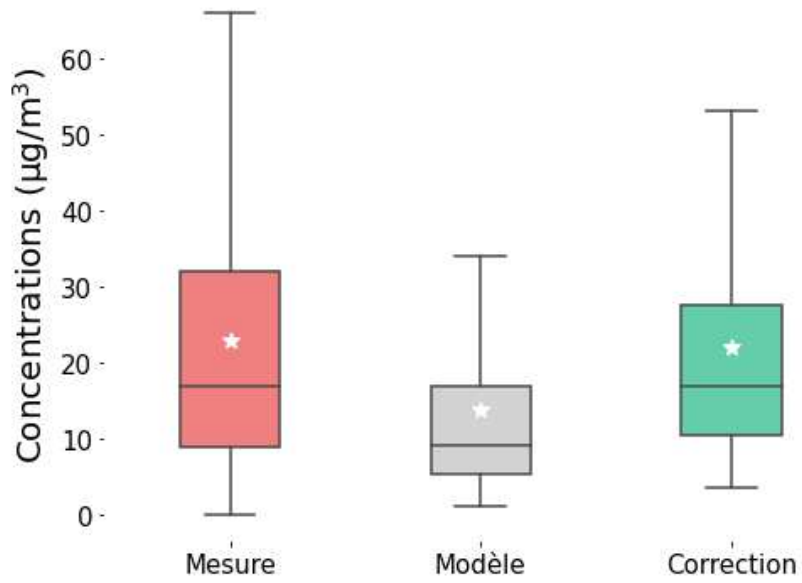
Comparaison des boxplots TREMB : NO₂, 2017



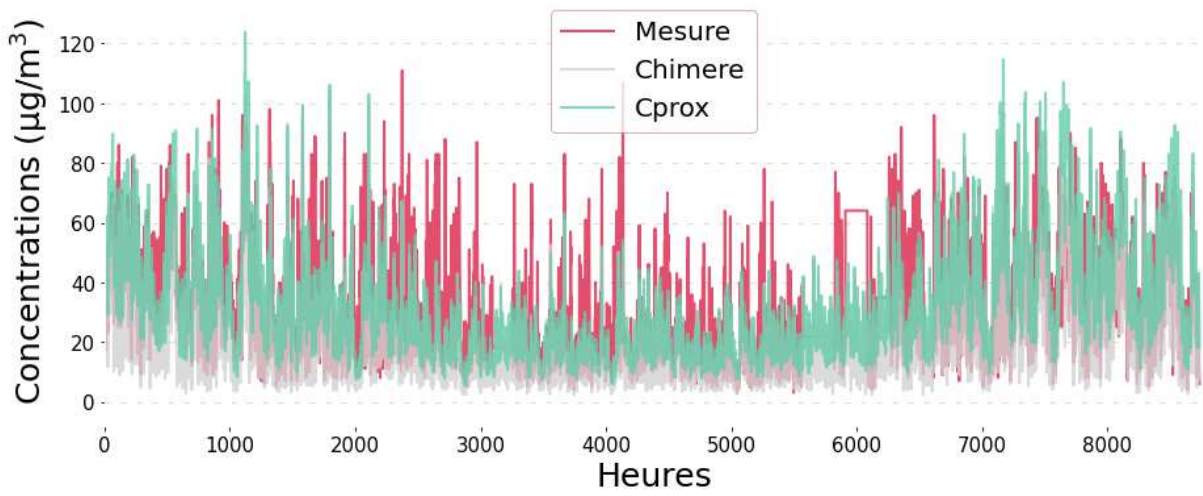
VERS: Site de fond (NO₂, 2017)



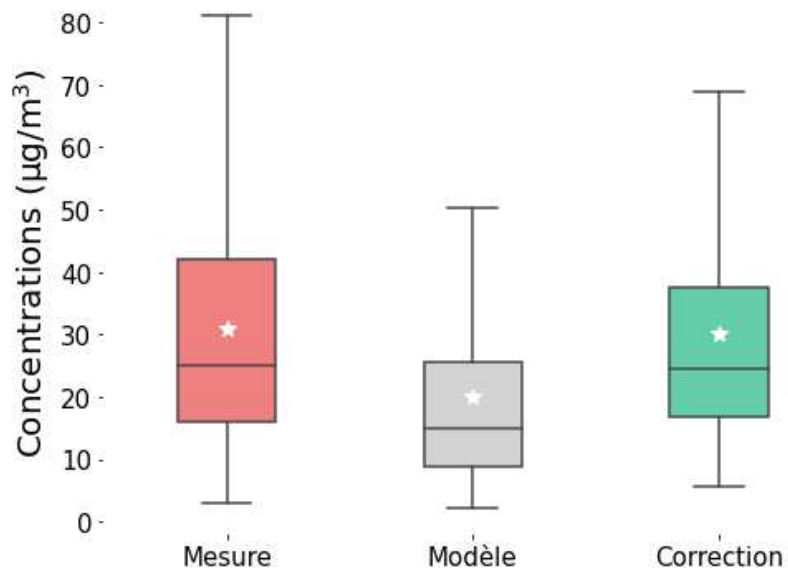
Comparaison des boxplots
VERS : NO₂, 2017



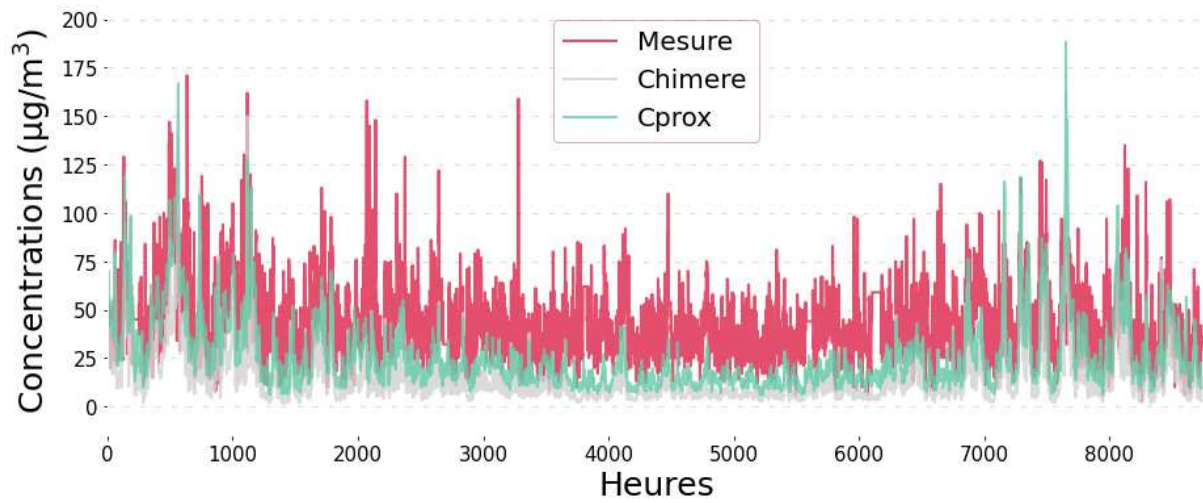
VILLEM: Site de fond (NO₂, 2017)



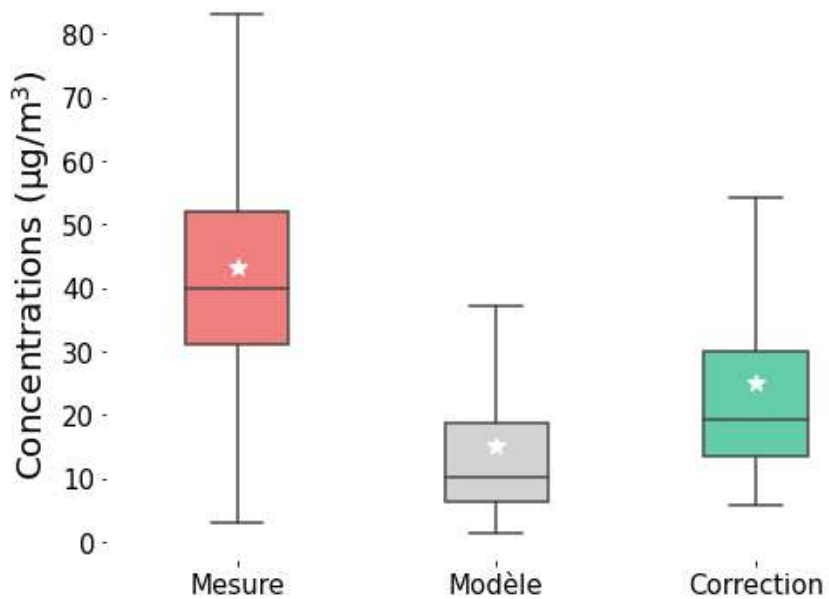
Comparaison des boxplots VILLEM : NO₂, 2017



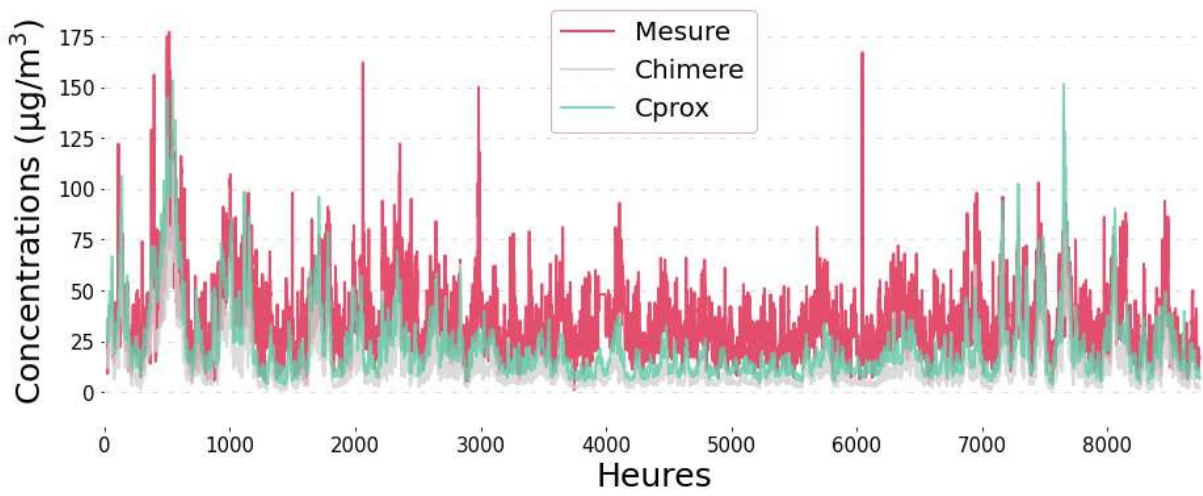
A1: Site de trafic (NO₂, 2017)



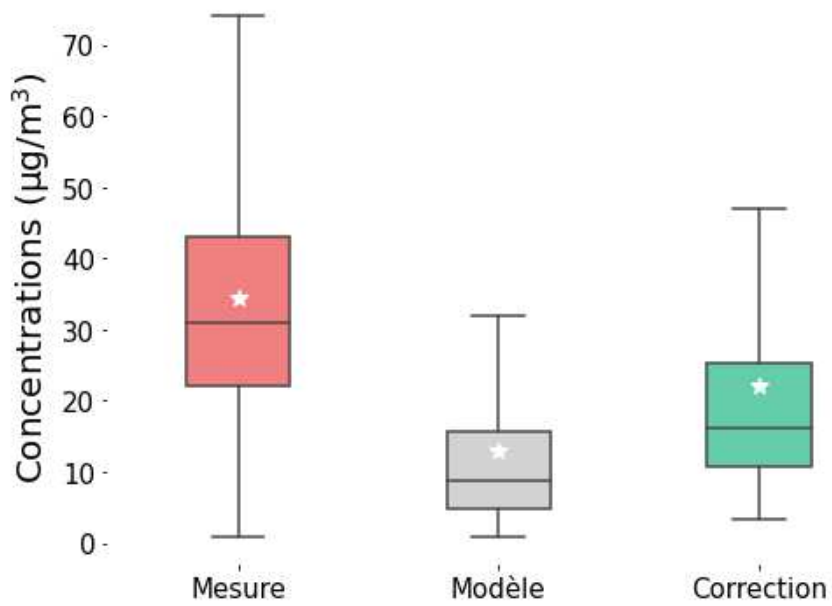
Comparaison des boxplots
A1 : PM₁₀, 2017



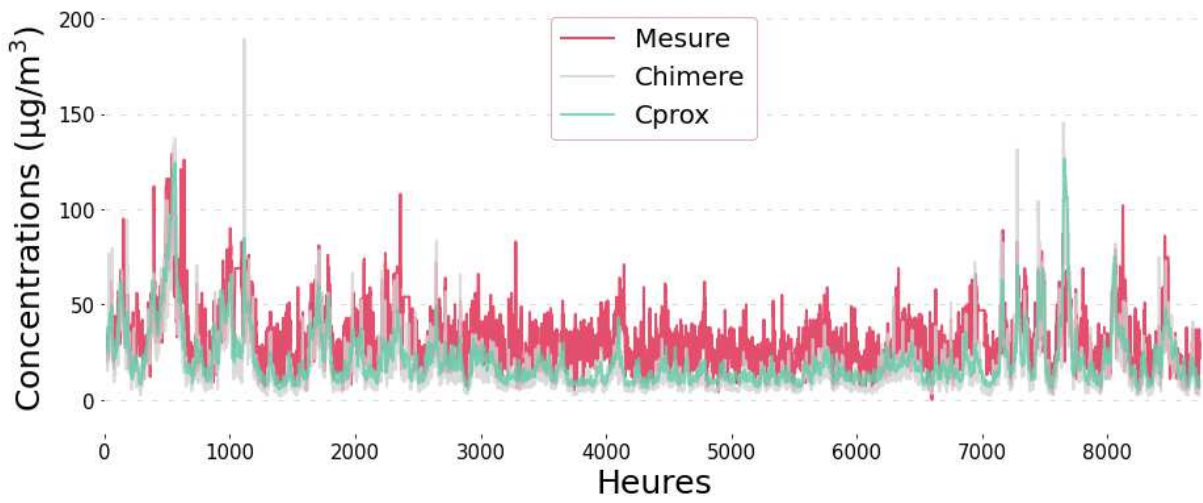
AUT: Site de trafic (NO₂, 2017)



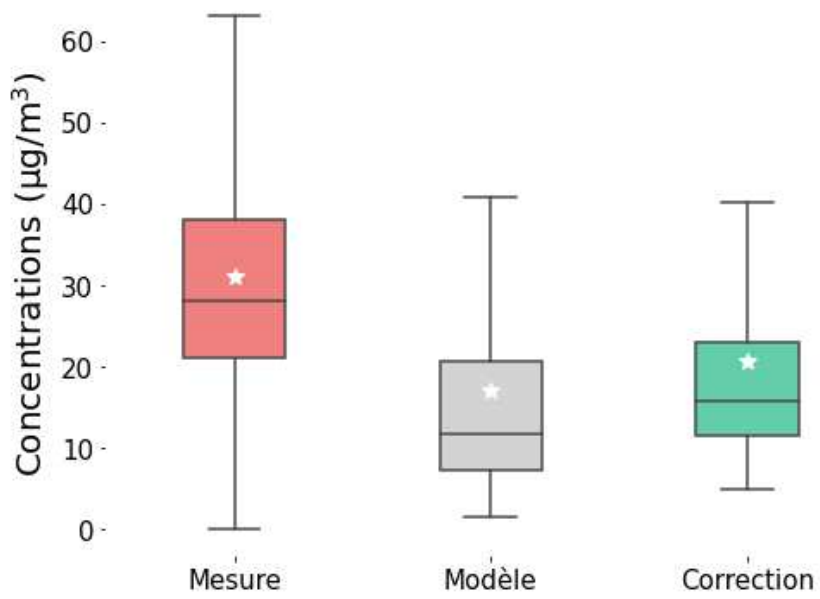
Comparaison des boxplots AUT : PM₁₀, 2017



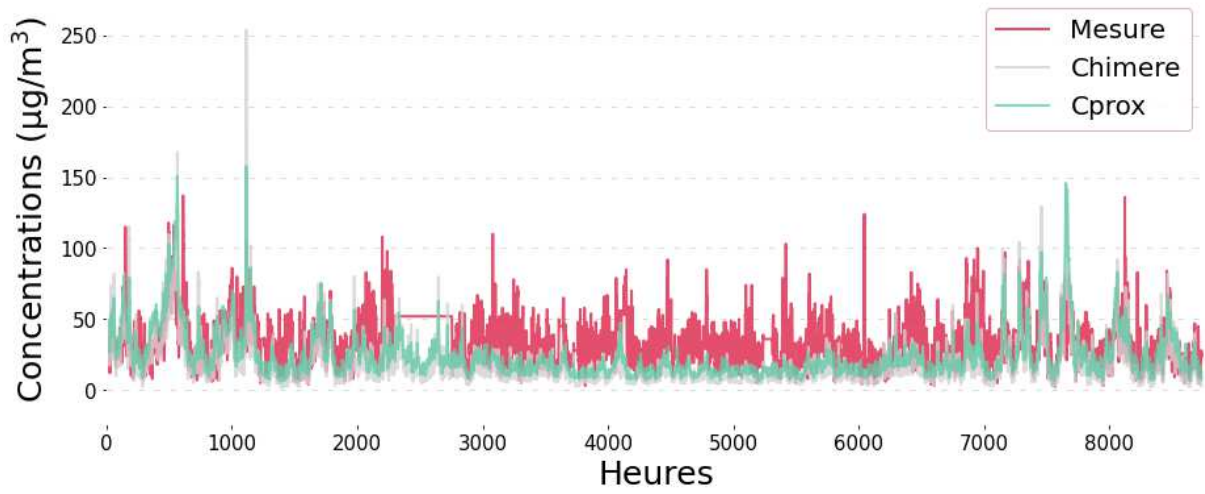
BASCH: Site de trafic (NO₂, 2017)



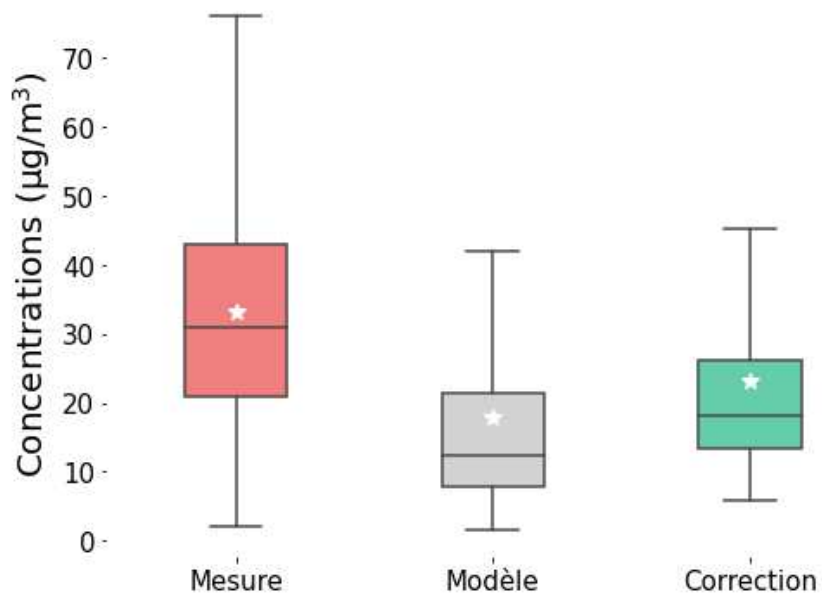
Comparaison des boxplots
BASCH : PM₁₀, 2017



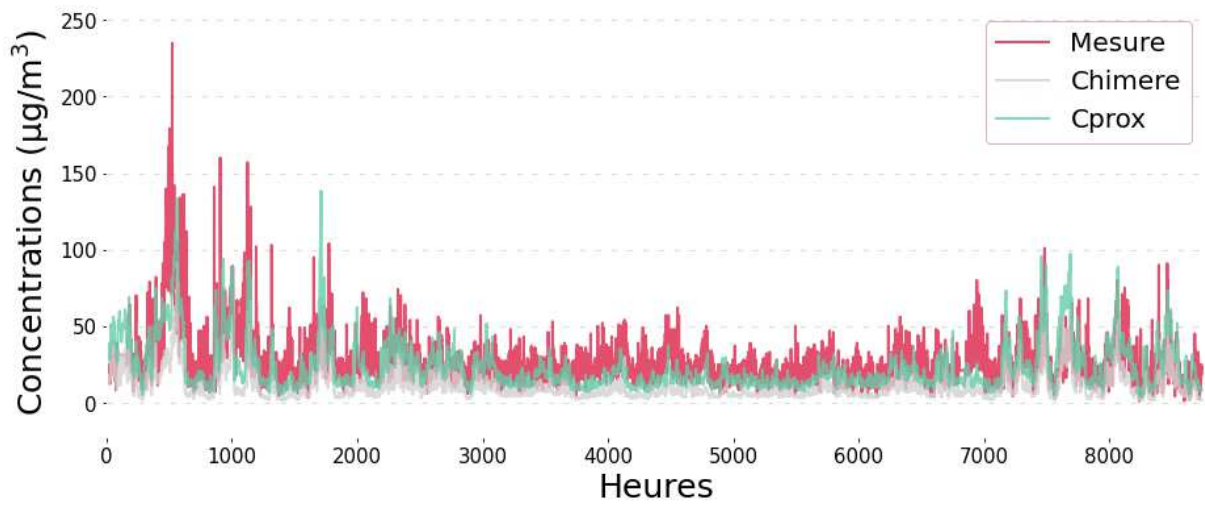
ELYS: Site de trafic (NO₂, 2017)



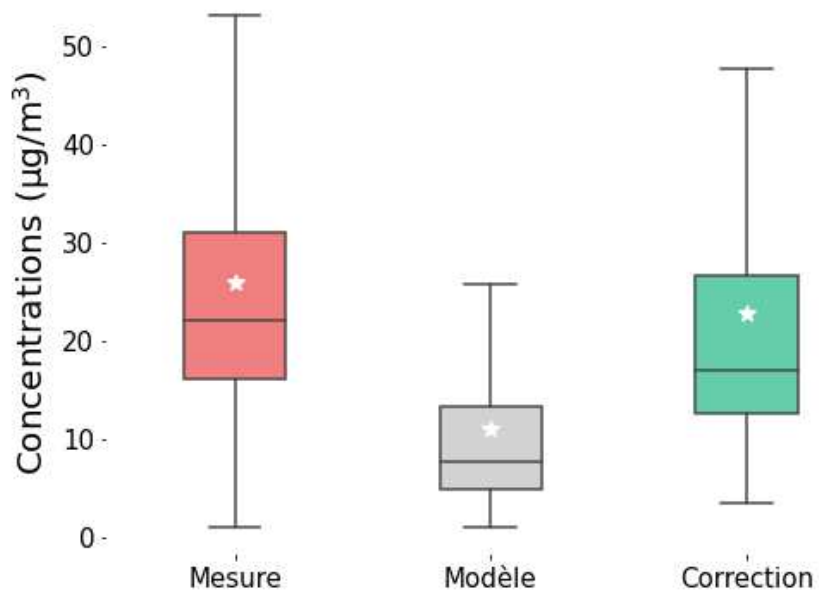
Comparaison des boxplots ELYS : PM₁₀, 2017



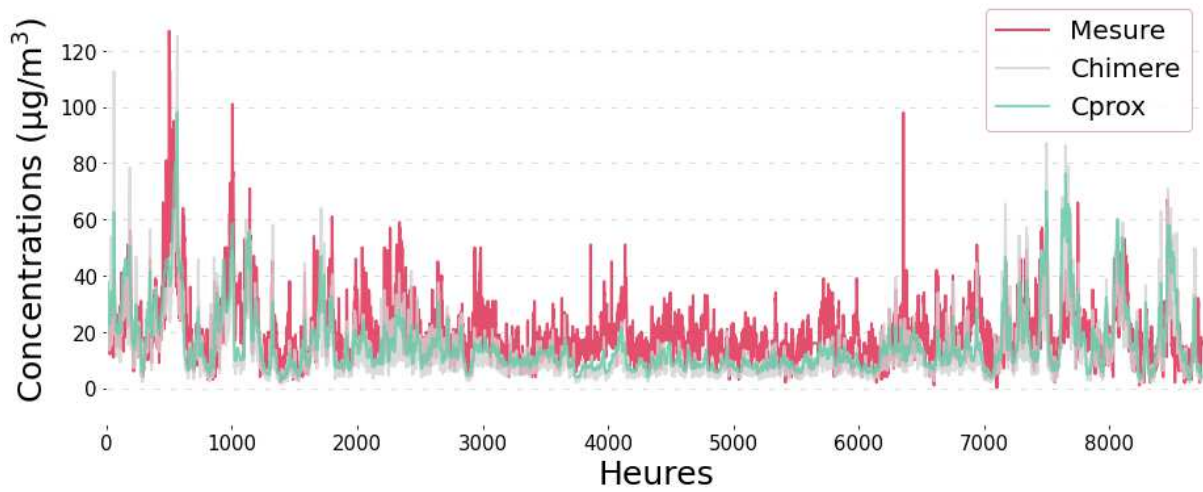
RN6: Site de trafic (NO₂, 2017)



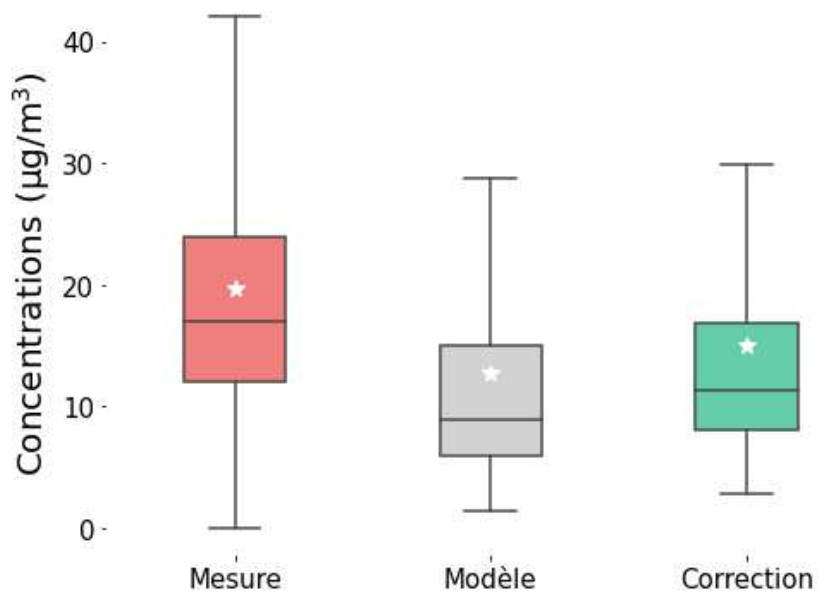
Comparaison des boxplots
RN6 : PM₁₀, 2017



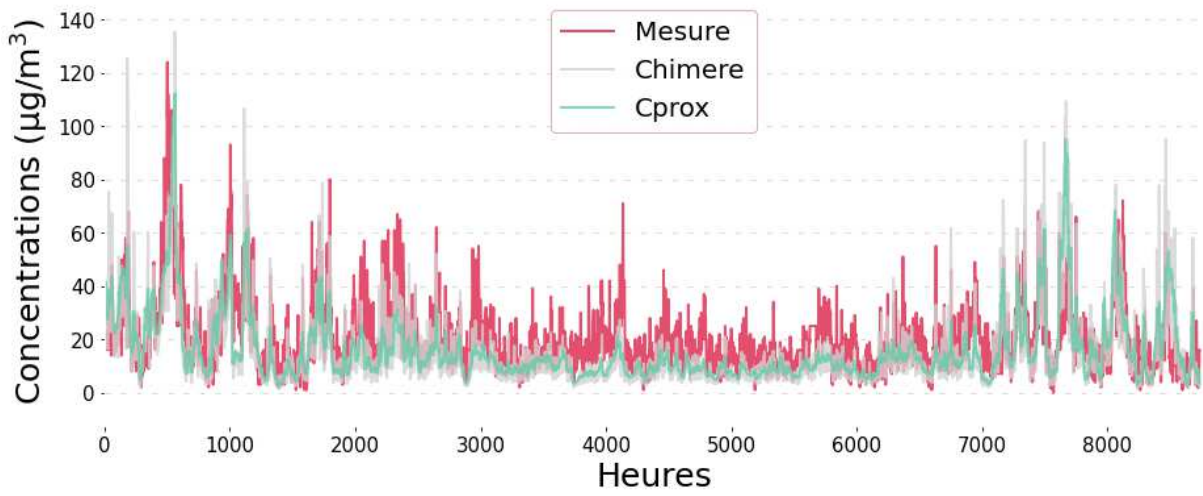
LOGNES: Site de fond (NO₂, 2017)



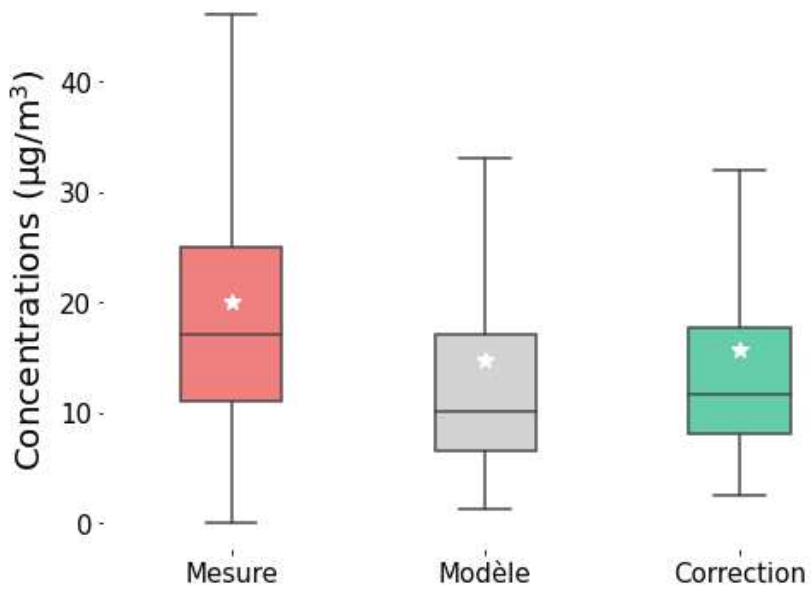
Comparaison des boxplots
LOGNES : PM₁₀, 2017



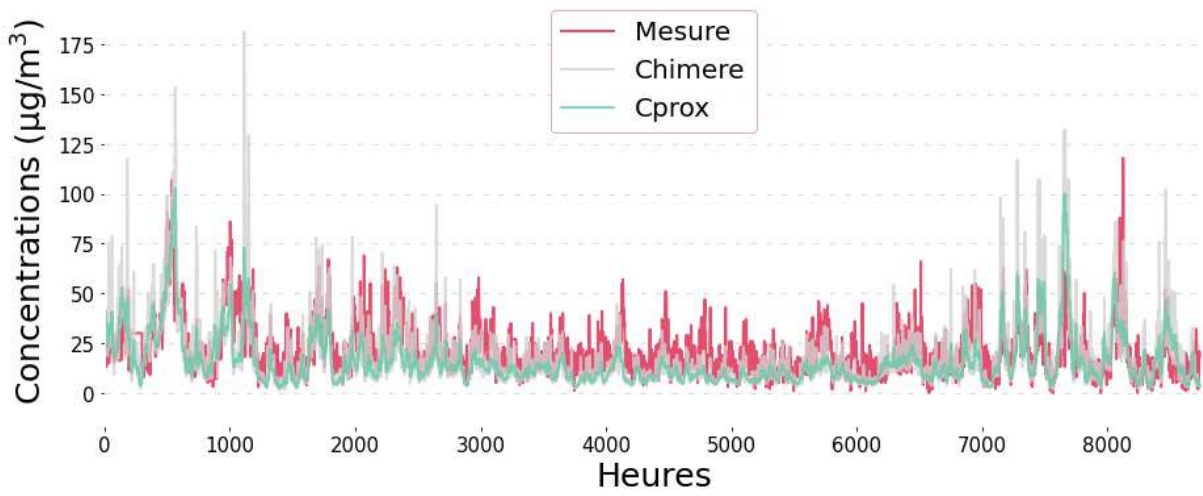
NOGENT: Site de fond (NO₂, 2017)



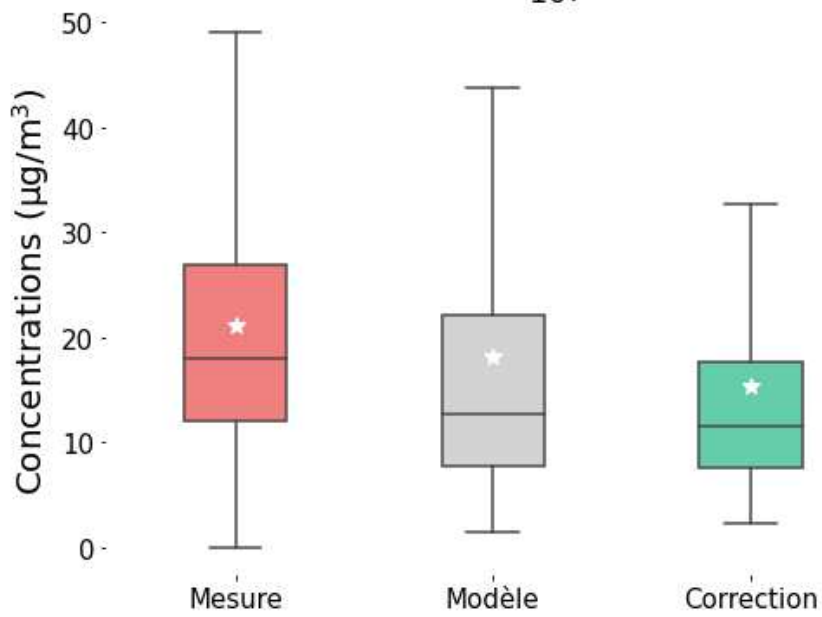
Comparaison des boxplots NOGENT : PM₁₀, 2017



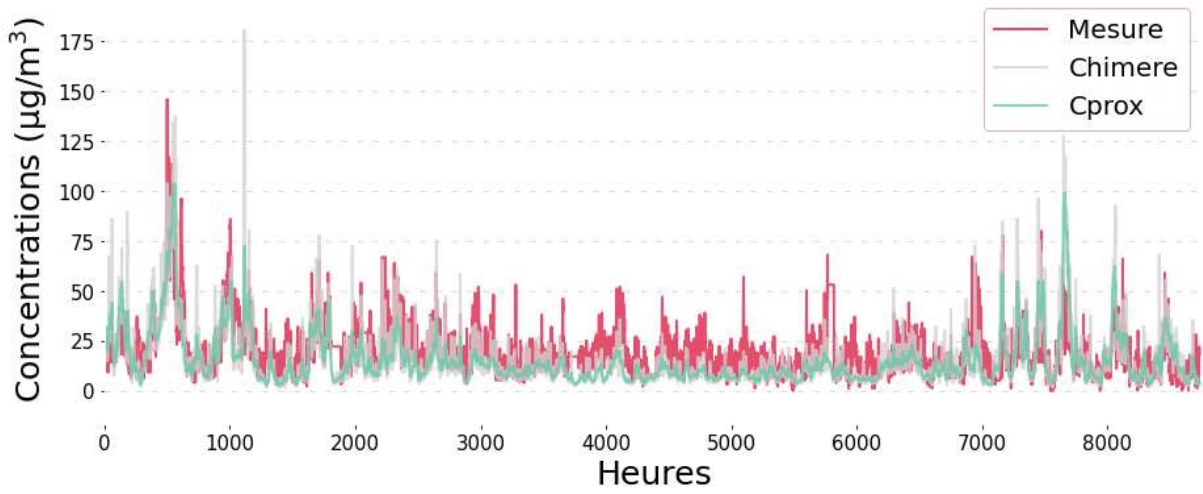
PA04C: Site de fond (NO₂, 2017)



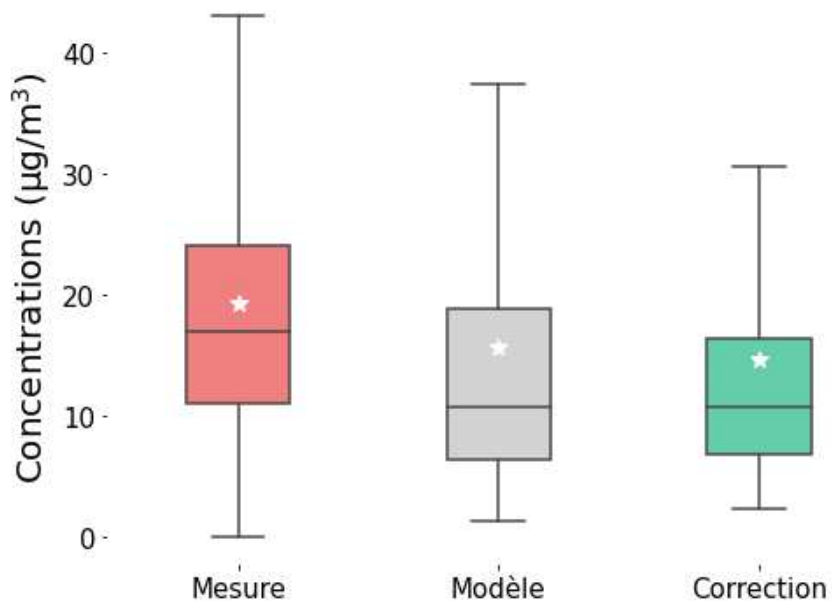
Comparaison des boxplots PA04C : PM₁₀, 2017



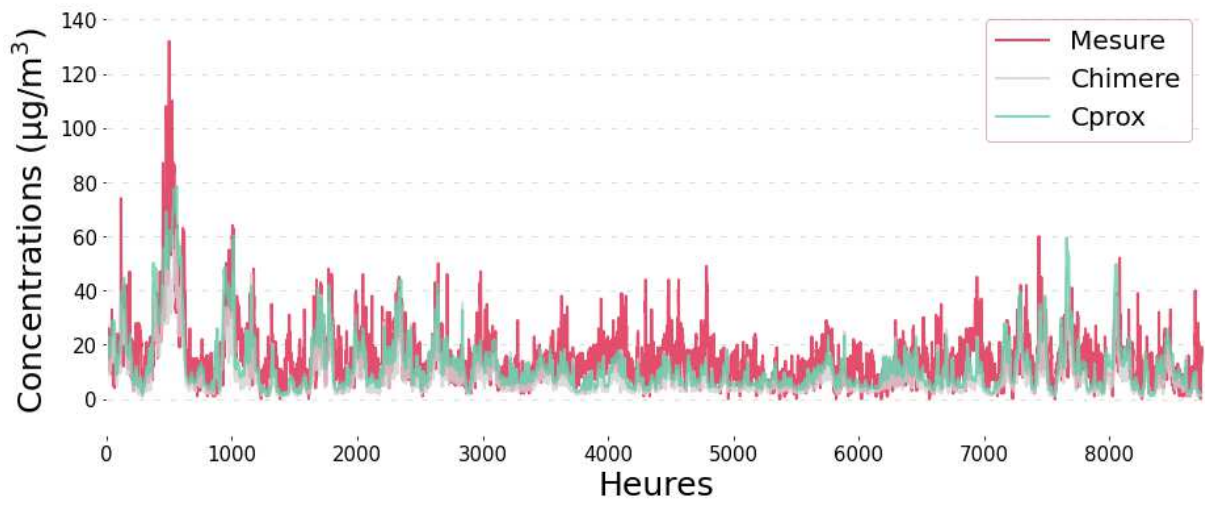
PA15L: Site de fond (NO₂, 2017)



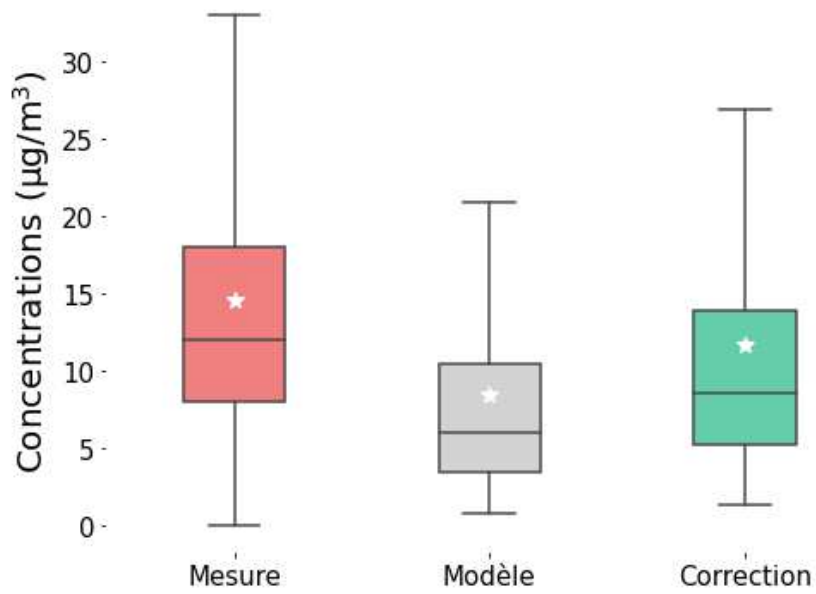
Comparaison des boxplots PA15L : PM₁₀, 2017



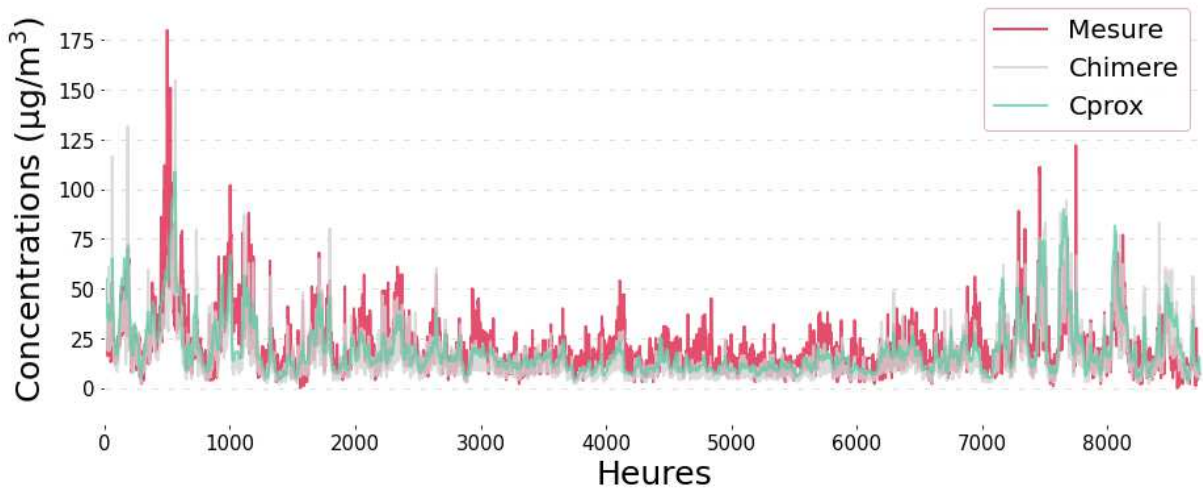
RAMBO: Site de fond (NO₂, 2017)



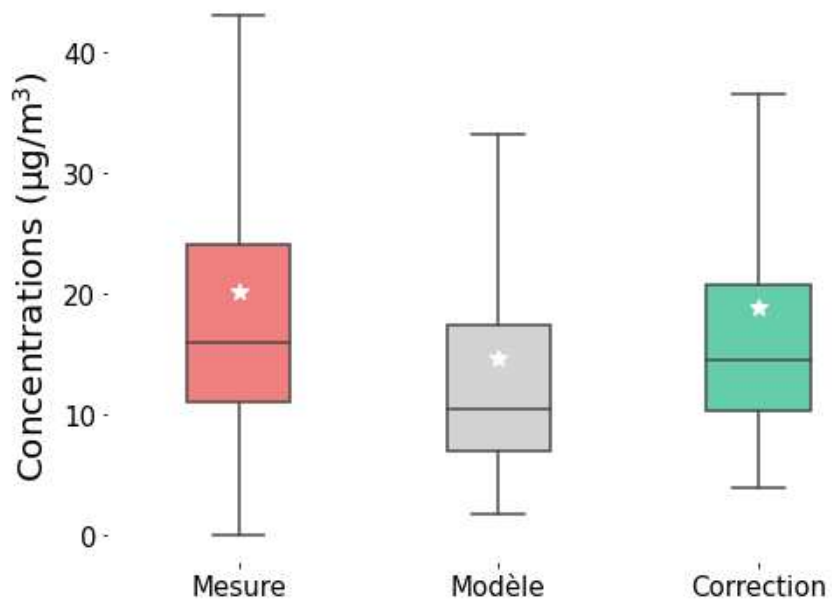
Comparaison des boxplots
RAMBO : PM₁₀, 2017



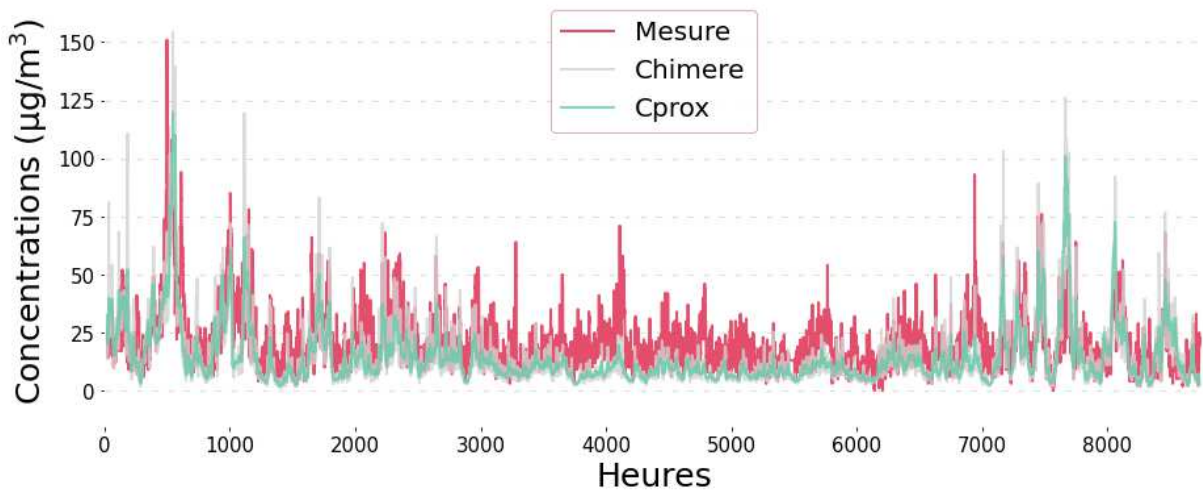
TREMB: Site de fond (NO₂, 2017)



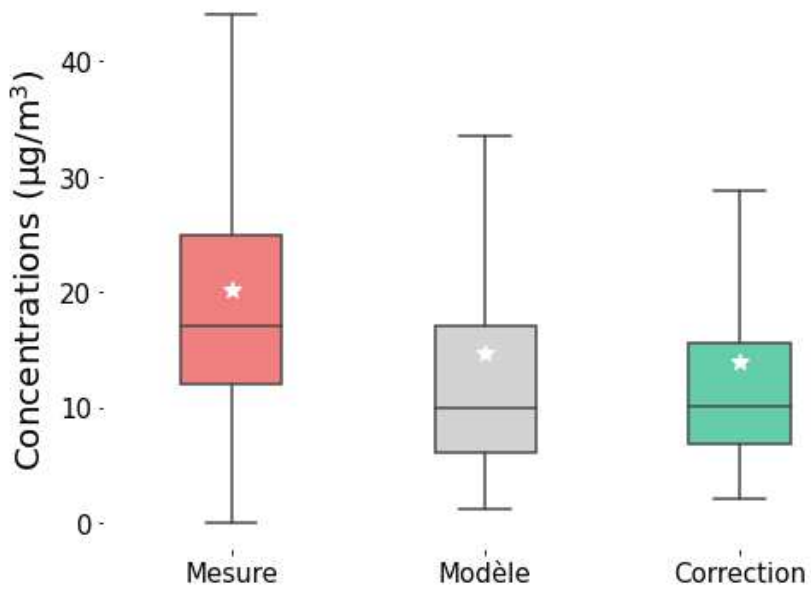
Comparaison des boxplots TREMB : PM₁₀, 2017



VITRY: Site de fond (NO₂, 2017)



Comparaison des boxplots VITRY : PM₁₀, 2017

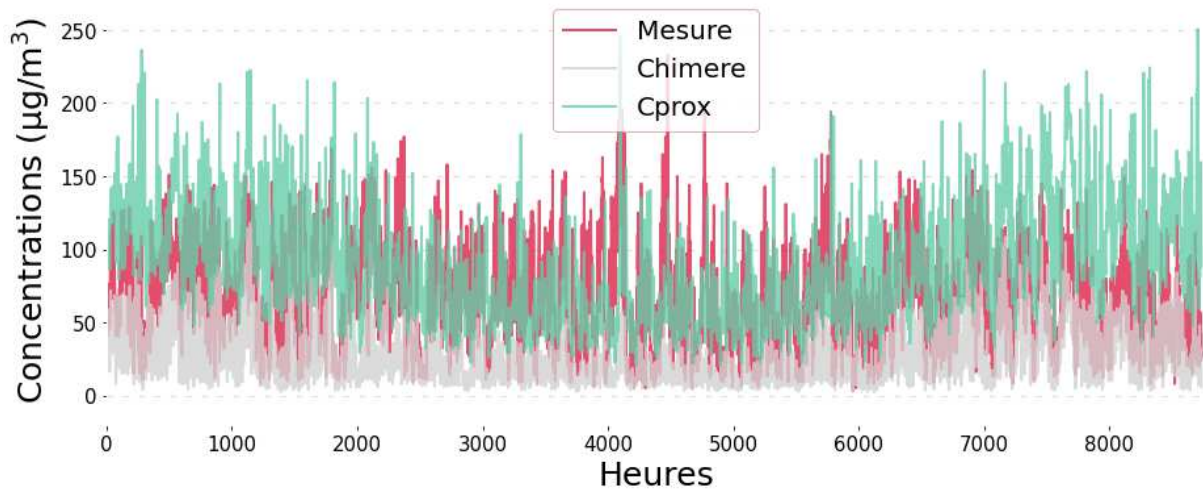


ANNEXE 8

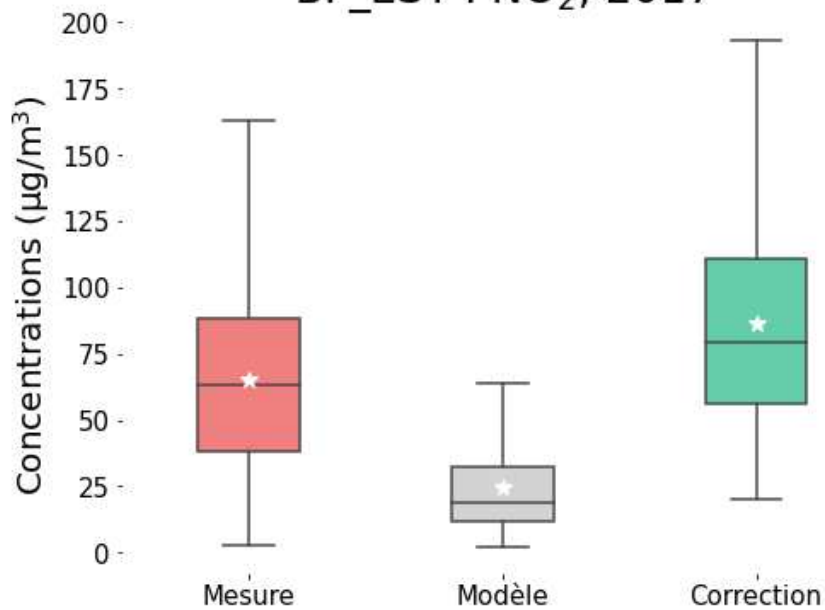
Données de l'universalisation appliquée aux sites de validation de 2017.

Séries temporelles NO₂ – Sites de trafic.

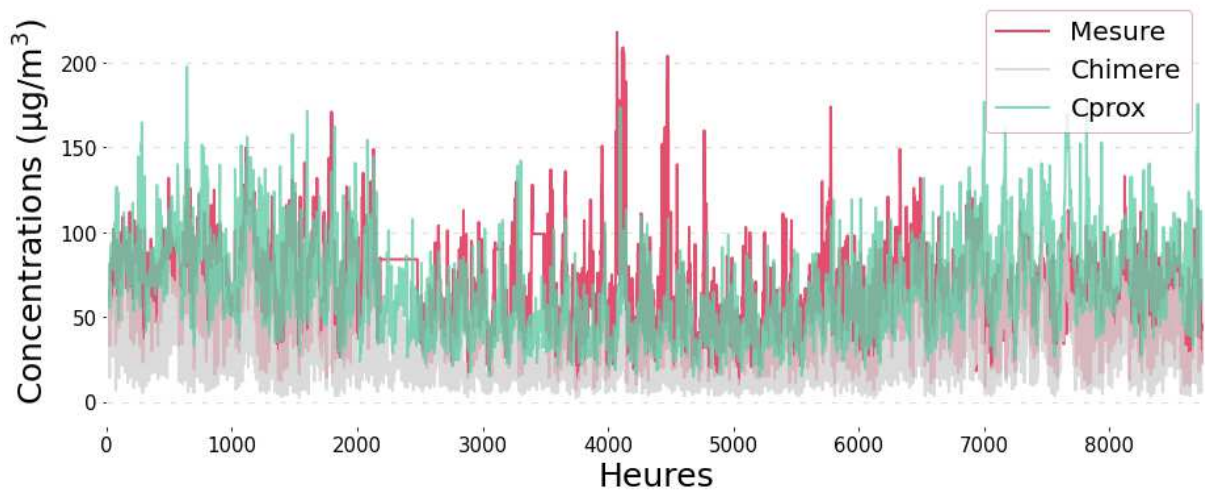
BP_EST: Site de trafic (NO₂, 2017)



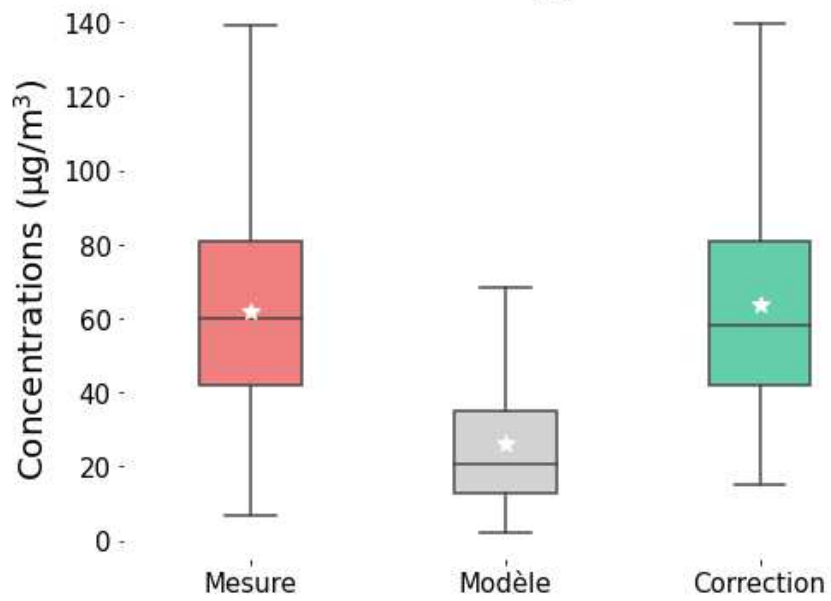
Comparaison des boxplots
BP_EST : NO₂, 2017



HAUS: Site de trafic (NO₂, 2017)



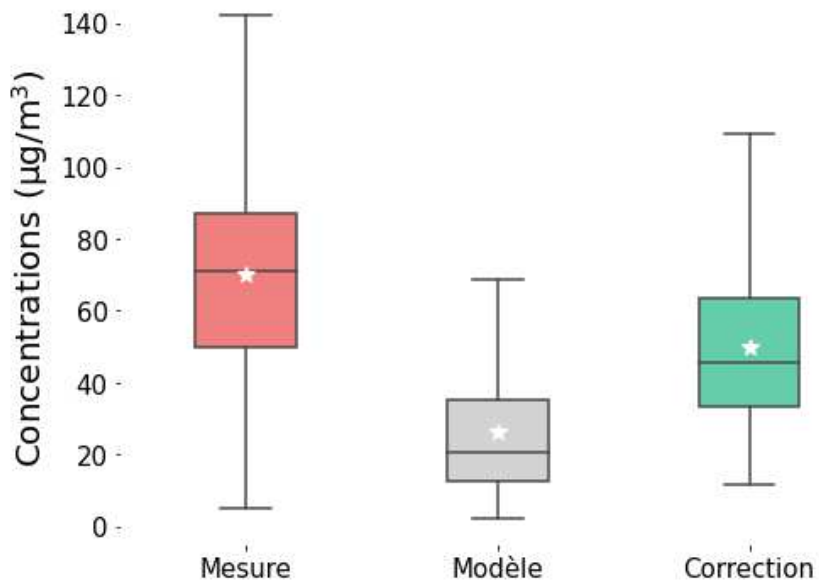
Comparaison des boxplots
HAUS : NO₂, 2017



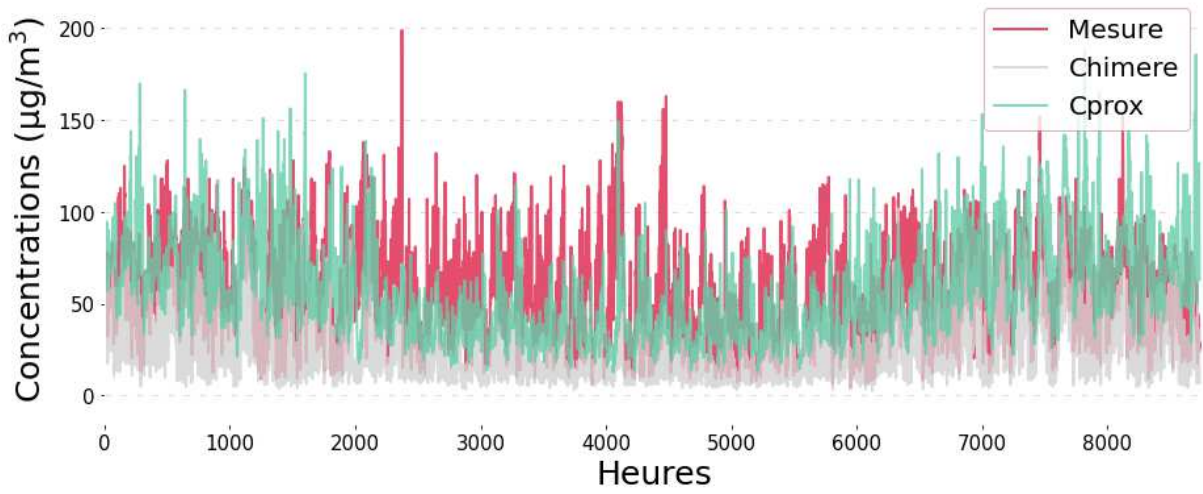
OPERA: Site de trafic (NO₂, 2017)



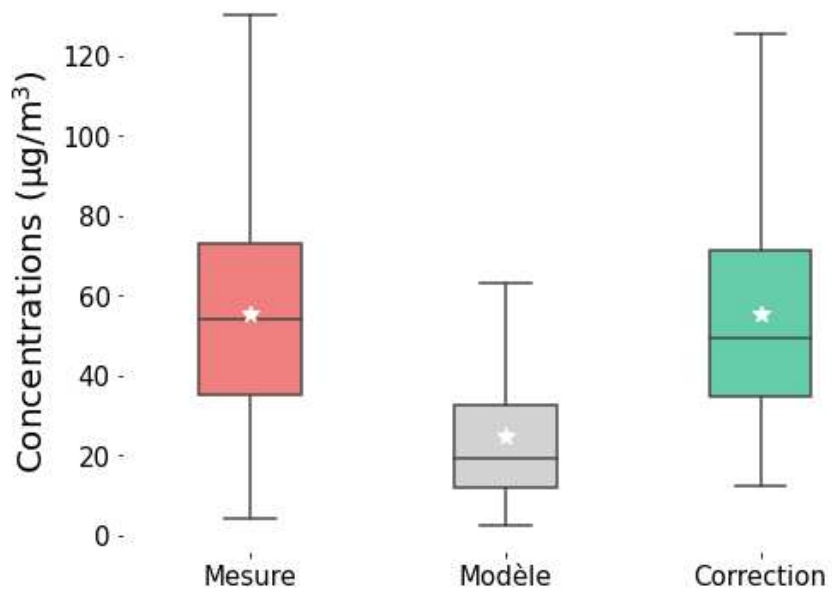
Comparaison des boxplots
OPERA : NO₂, 2017



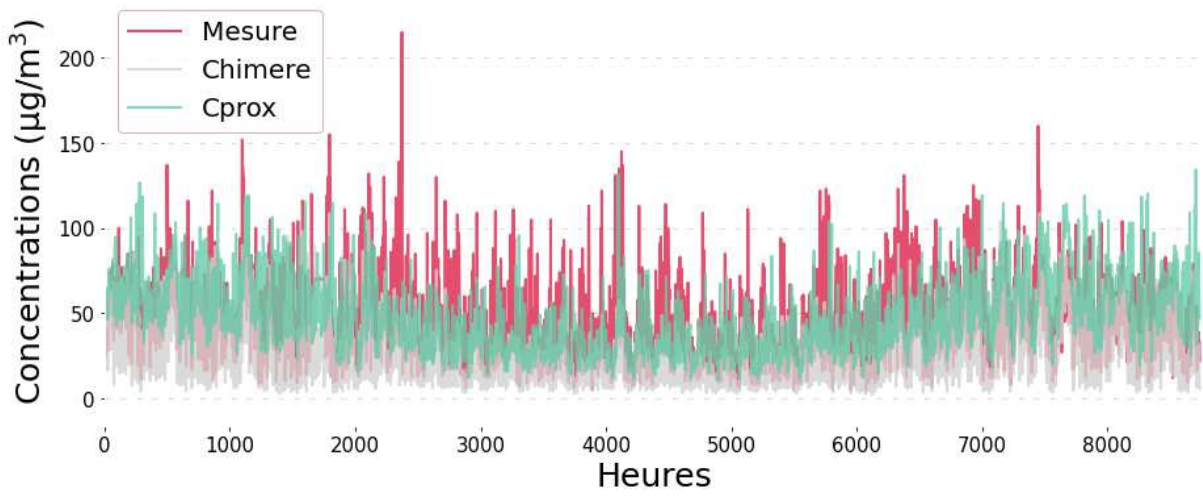
RN2: Site de trafic (NO₂, 2017)



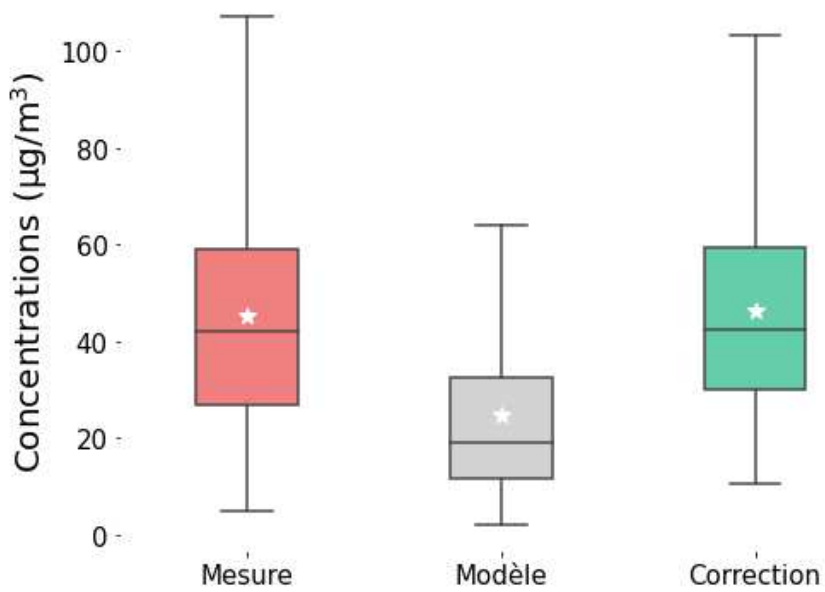
Comparaison des boxplots
RN2 : NO₂, 2017



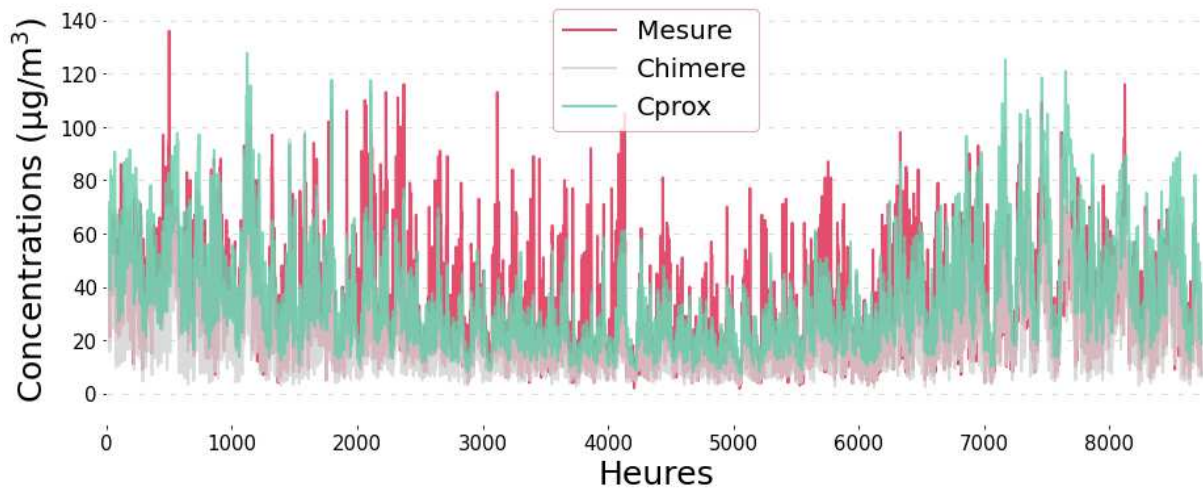
SOULT: Site de trafic (NO₂, 2017)



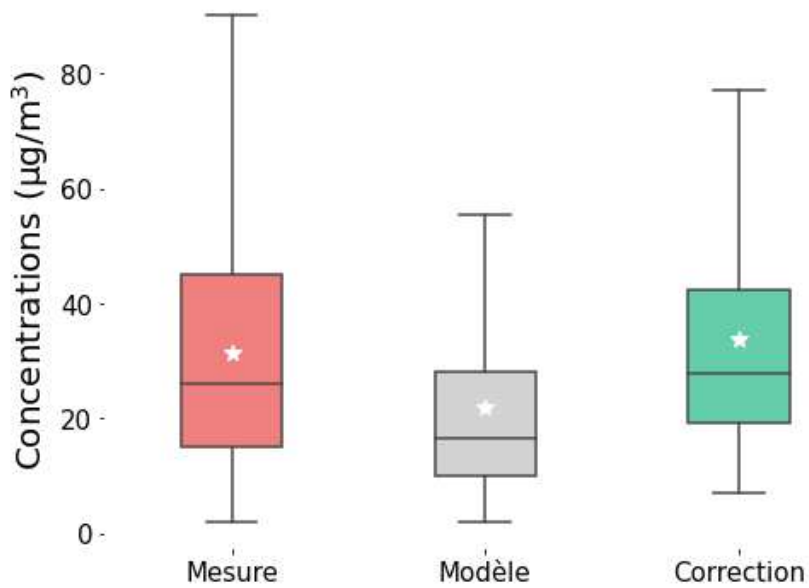
Comparaison des boxplots SOULT : NO₂, 2017



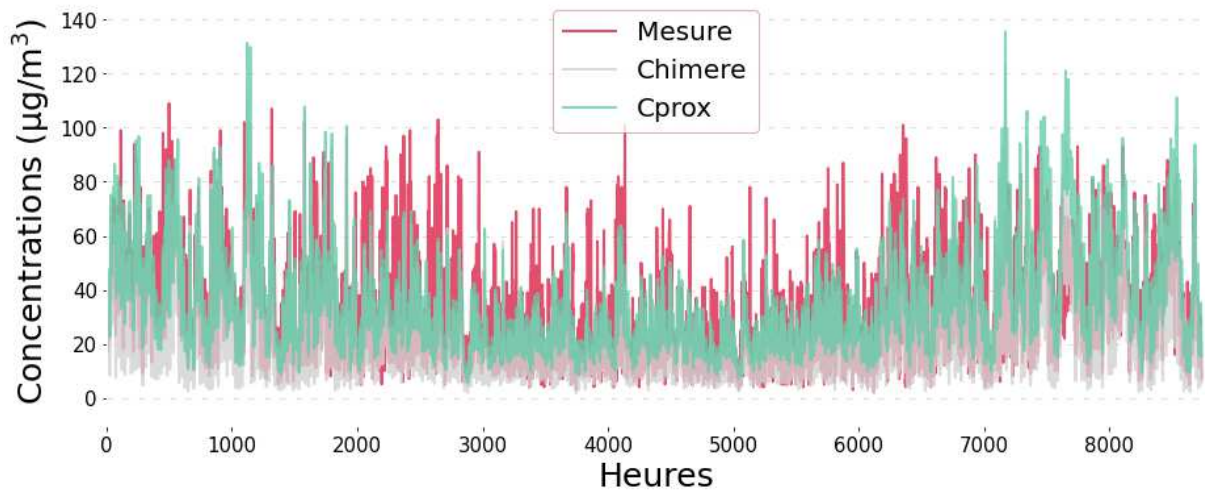
BOB: Site de fond (NO₂, 2017)



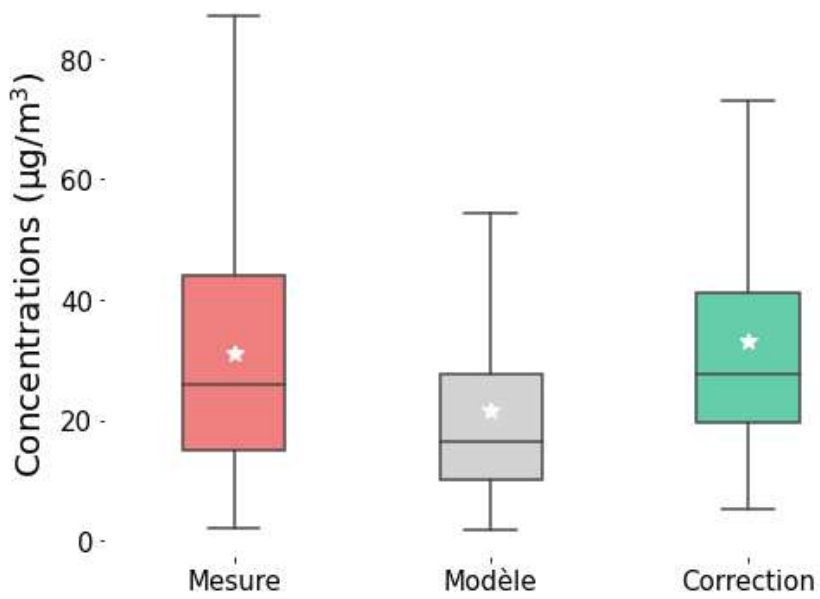
Comparaison des boxplots
BOB : NO₂, 2017



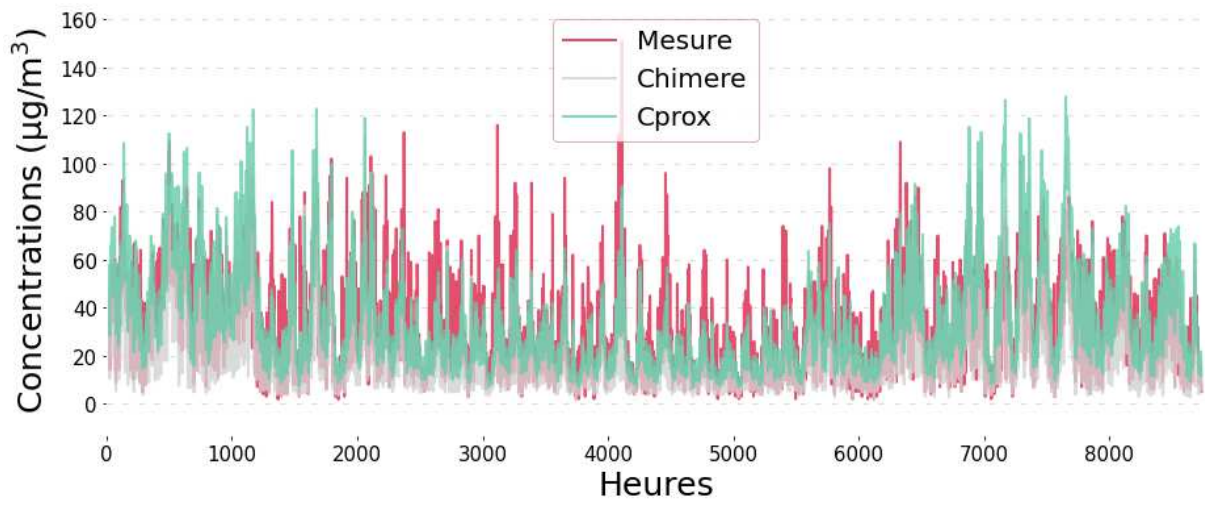
CHAMP: Site de fond (NO₂, 2017)



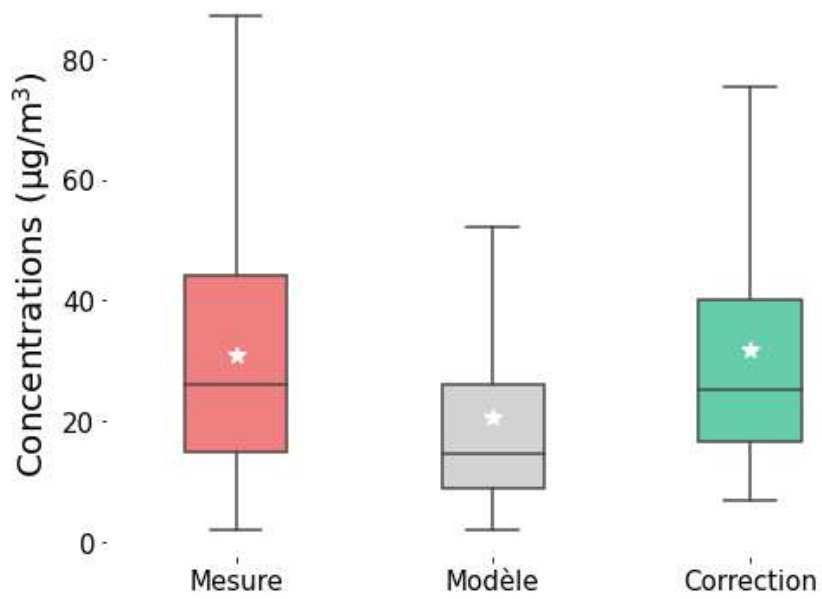
Comparaison des boxplots CHAMP : NO₂, 2017



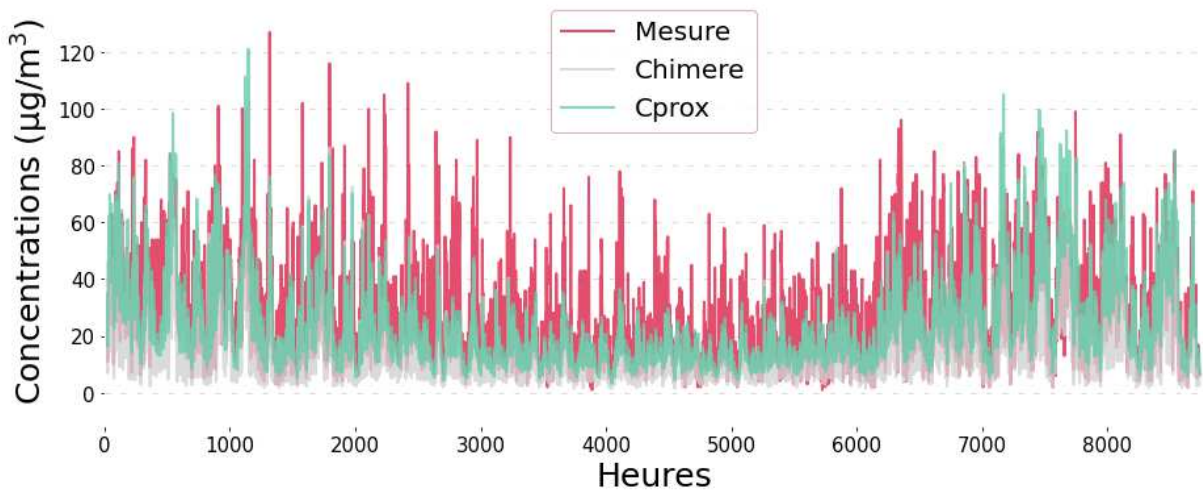
DEF: Site de fond (NO₂, 2017)



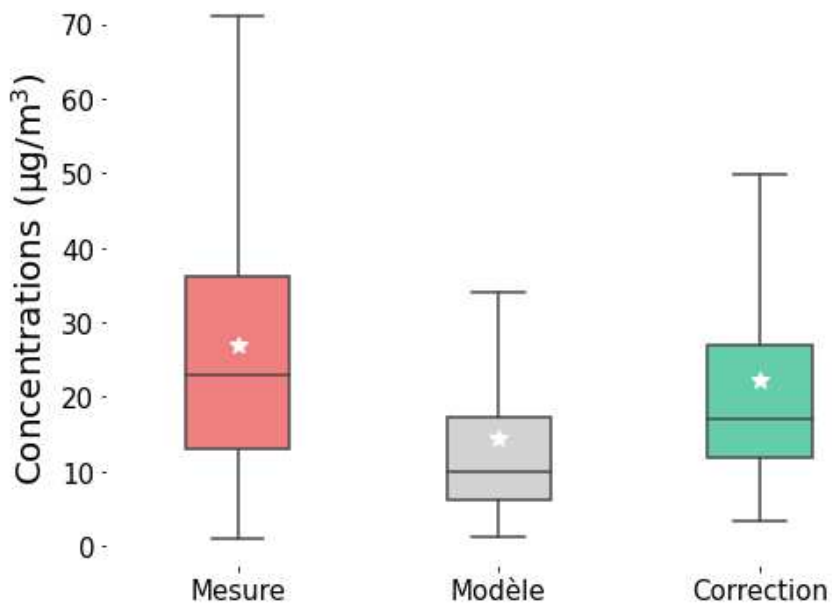
Comparaison des boxplots
DEF : NO₂, 2017



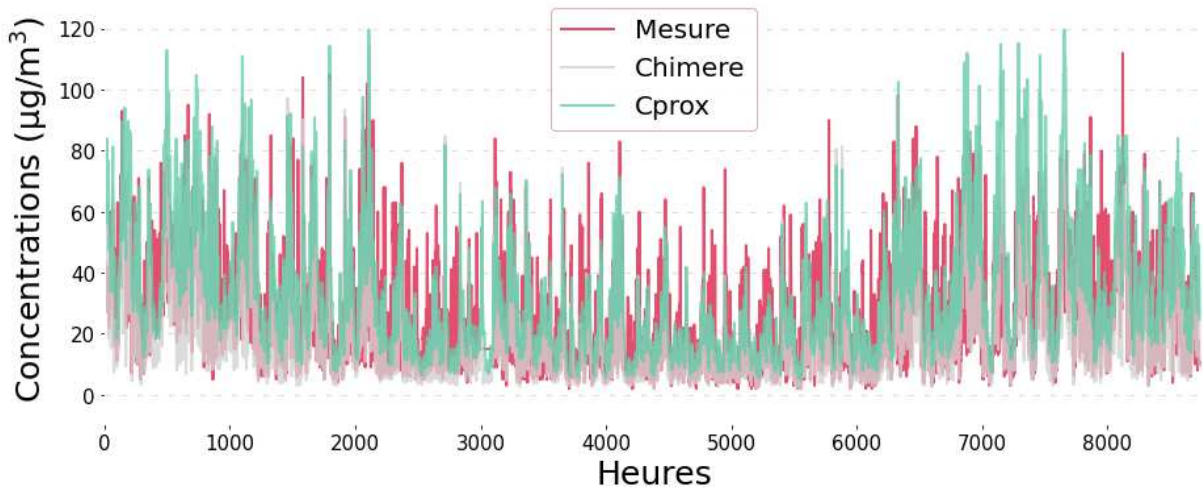
EVRY: Site de fond (NO₂, 2017)



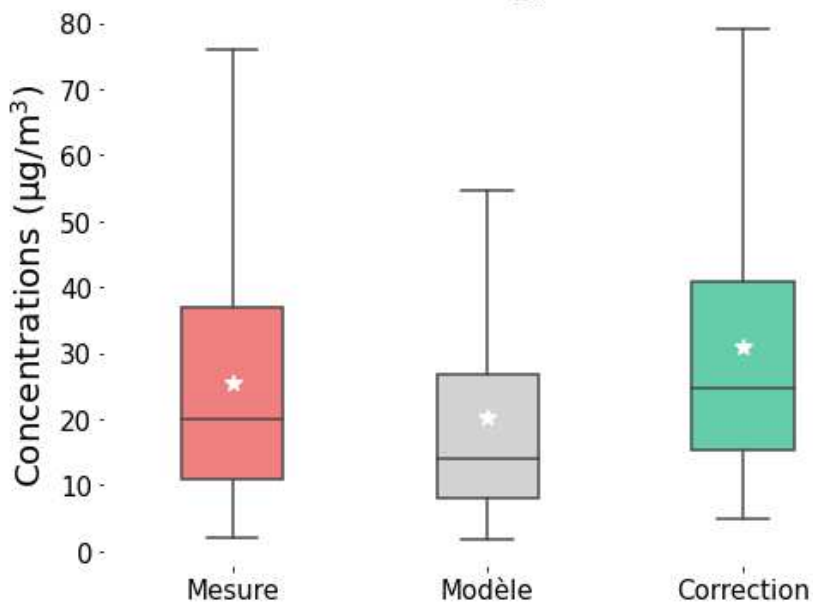
Comparaison des boxplots EVRY : NO₂, 2017



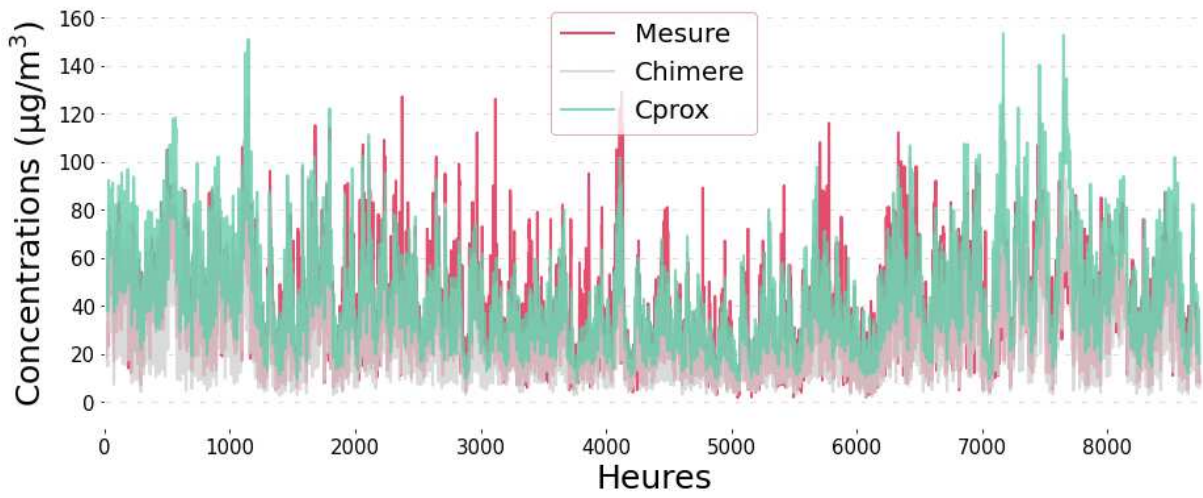
GON: Site de fond (NO₂, 2017)



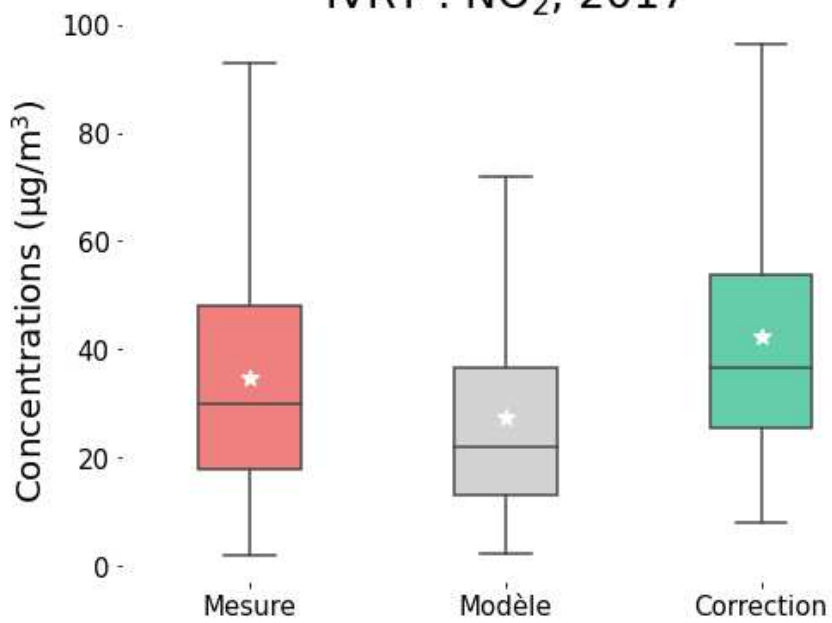
Comparaison des boxplots GON : NO₂, 2017



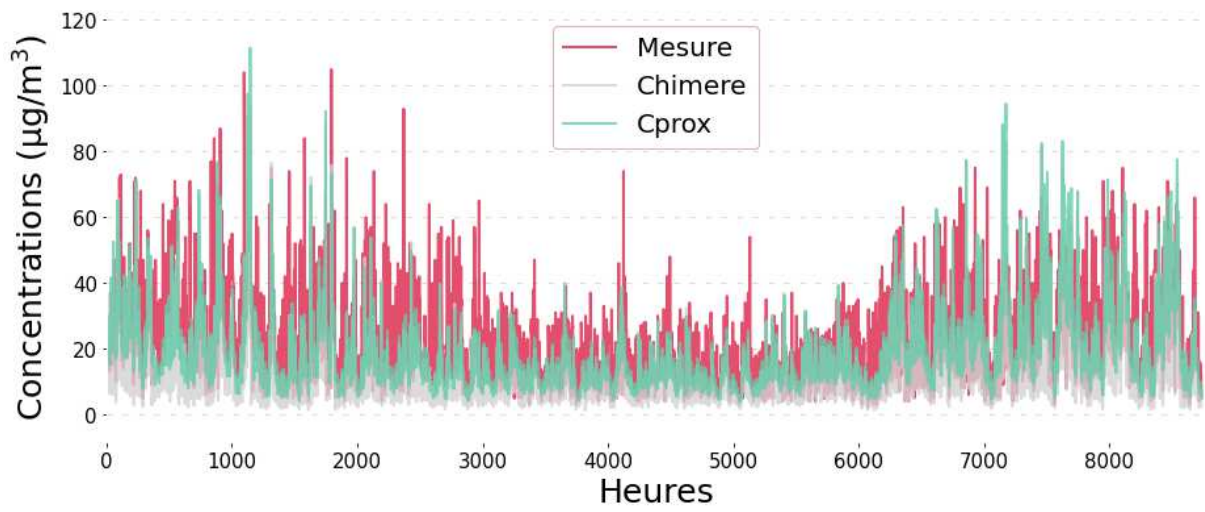
IVRY: Site de fond (NO₂, 2017)



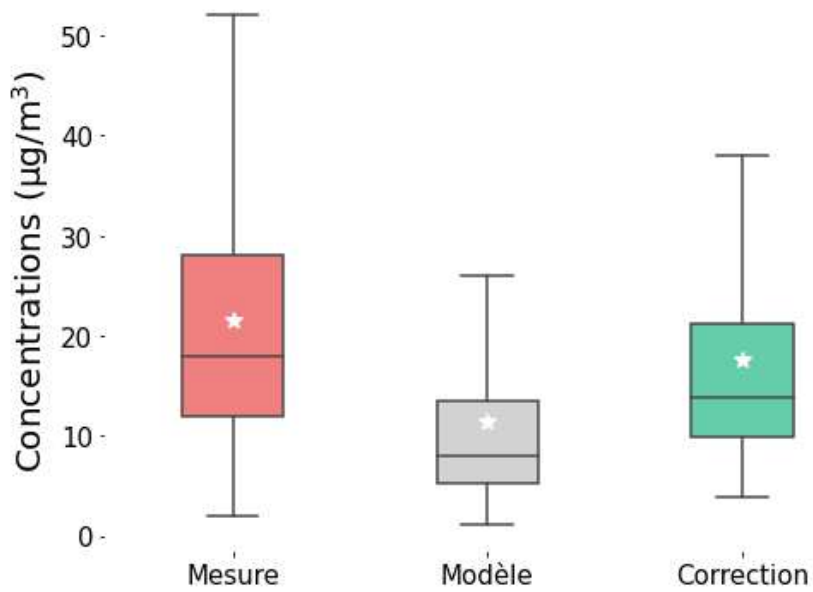
Comparaison des boxplots IVRY : NO₂, 2017



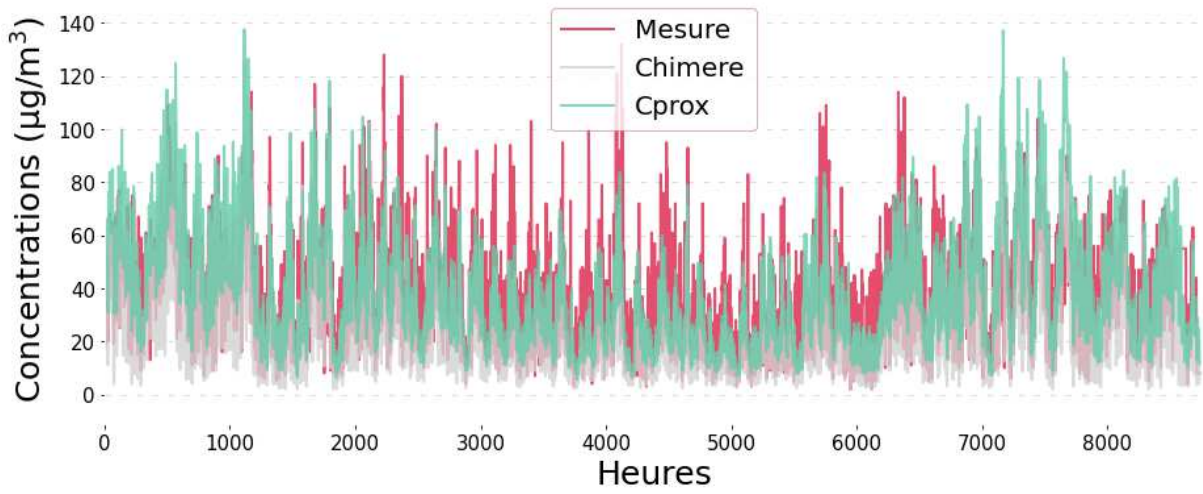
MELUN: Site de fond (NO₂, 2017)



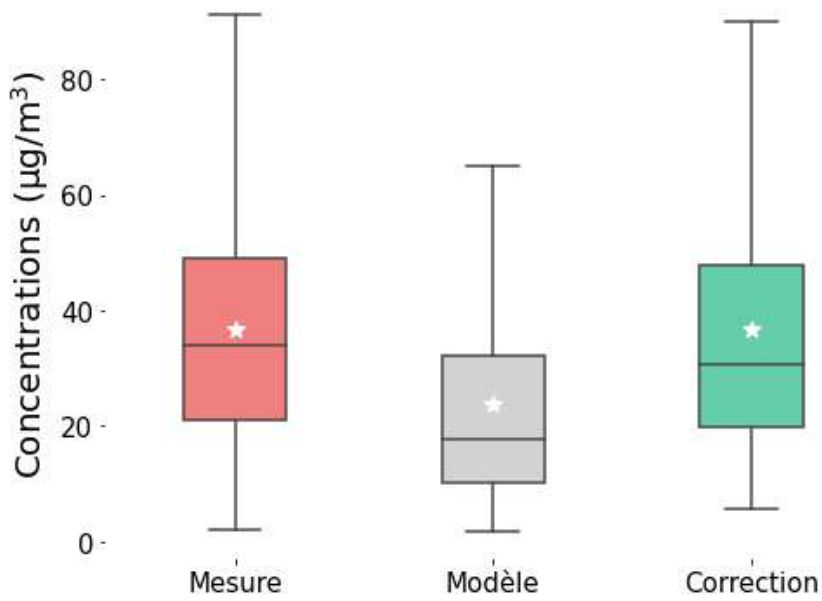
Comparaison des boxplots
MELUN : NO₂, 2017



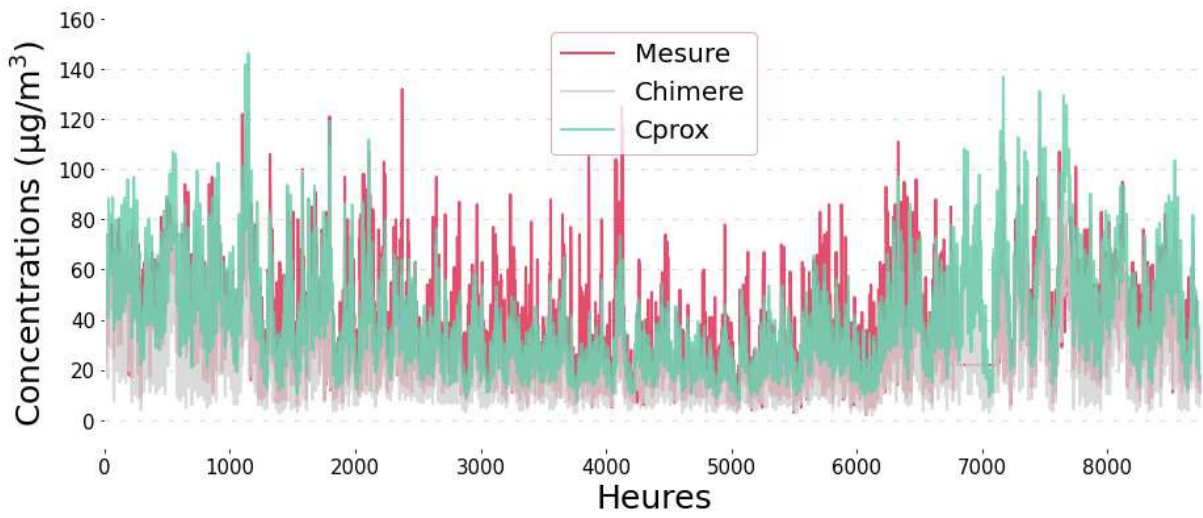
PA07: Site de fond (NO₂, 2017)



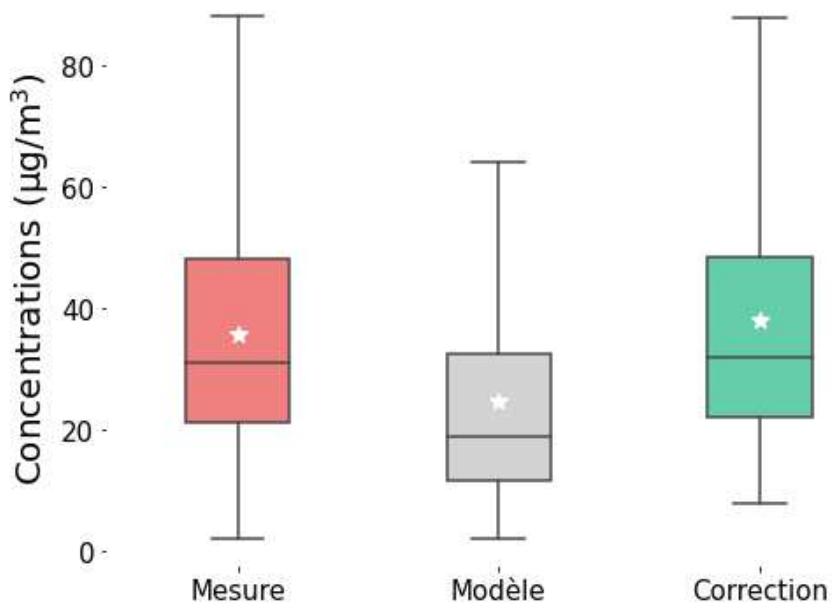
Comparaison des boxplots
PA07 : NO₂, 2017



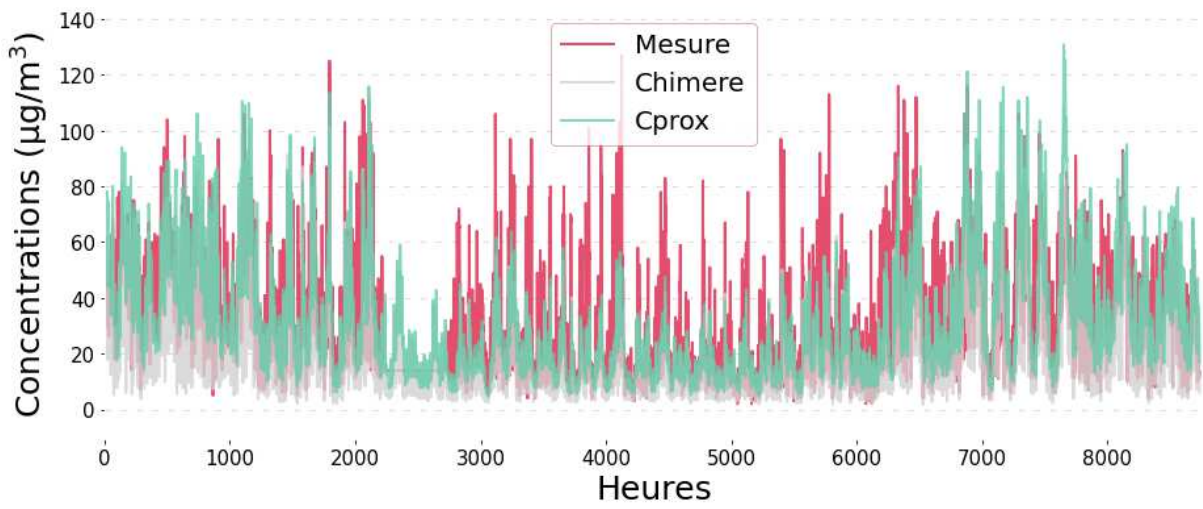
PA12: Site de fond (NO₂, 2017)



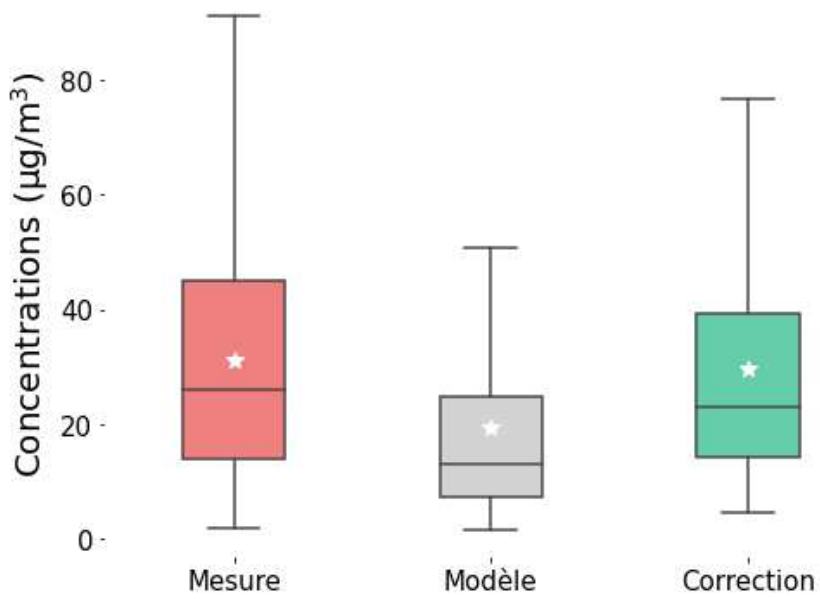
Comparaison des boxplots
PA12 : NO₂, 2017



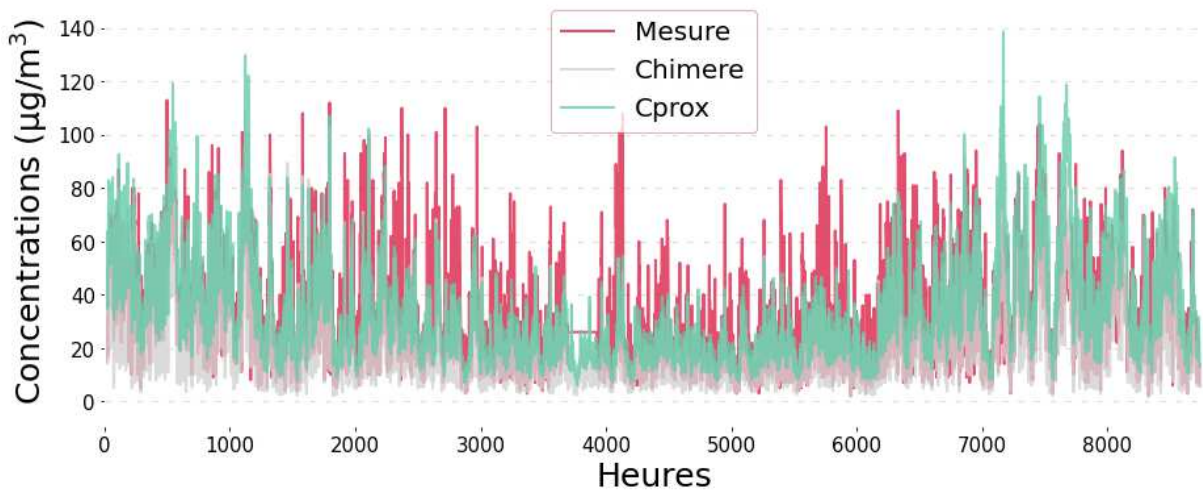
STDEN: Site de fond (NO₂, 2017)



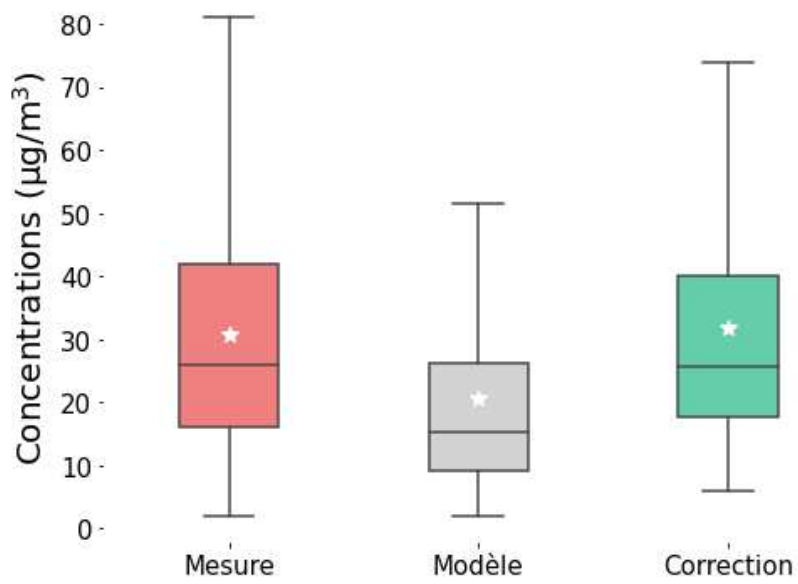
Comparaison des boxplots
STDEN : NO₂, 2017



VITRY: Site de fond (NO₂, 2017)

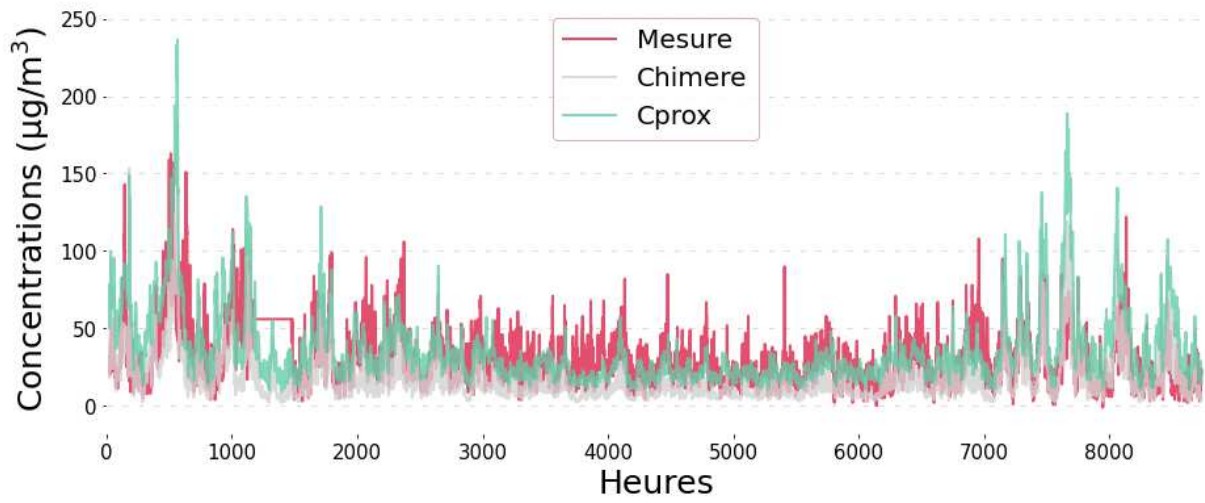


Comparaison des boxplots VITRY : NO₂, 2017

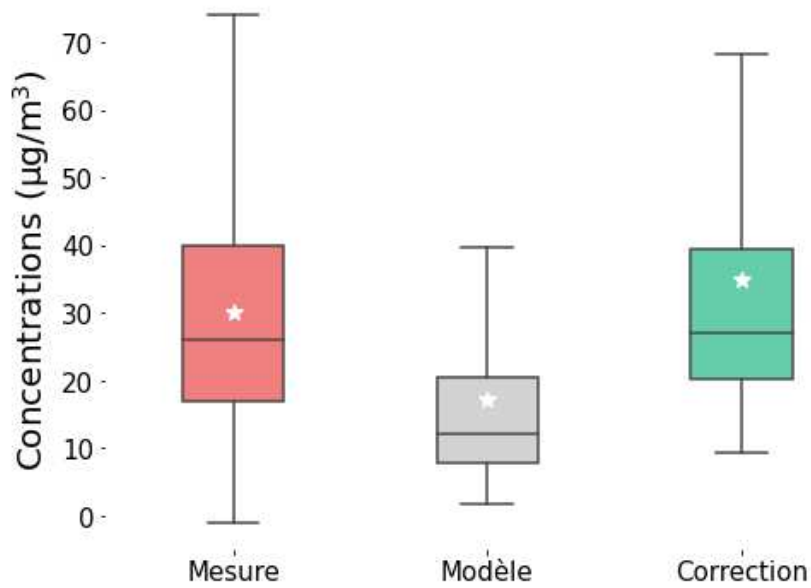


Séries temporelles PM₁₀ – Sites de trafic.

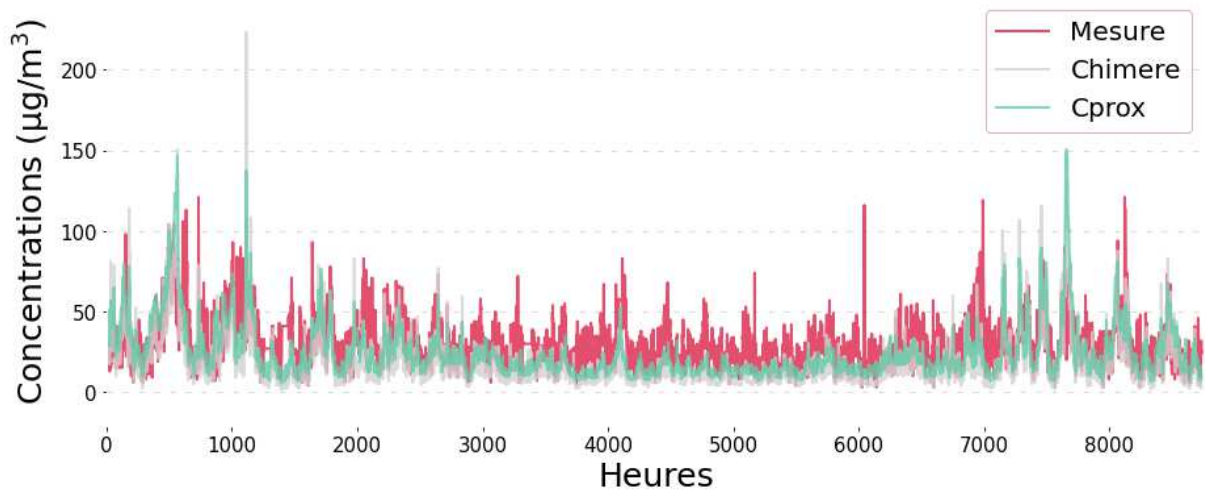
BP_EST: Site de trafic (NO₂, 2017)



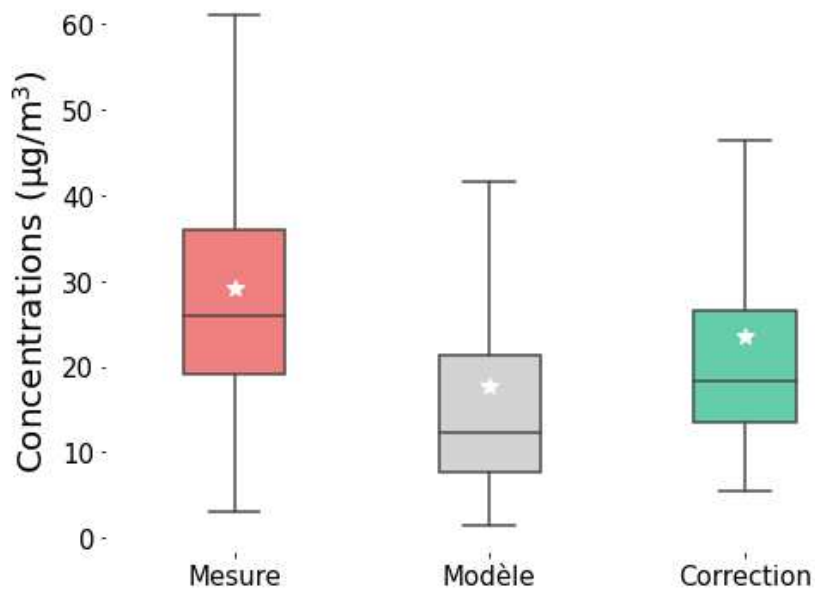
Comparaison des boxplots
BP_EST : PM₁₀, 2017



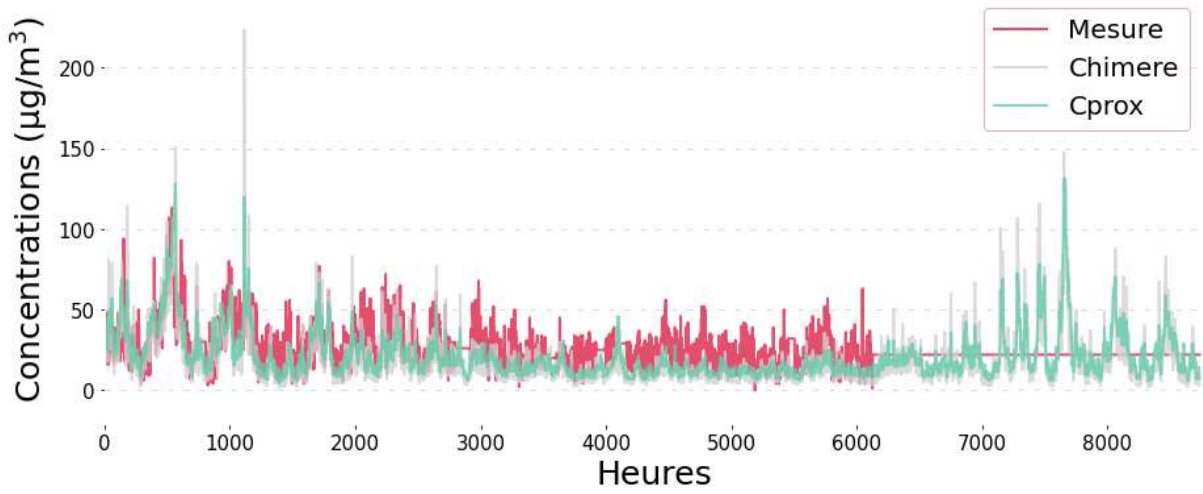
HAUS: Site de trafic (NO₂, 2017)



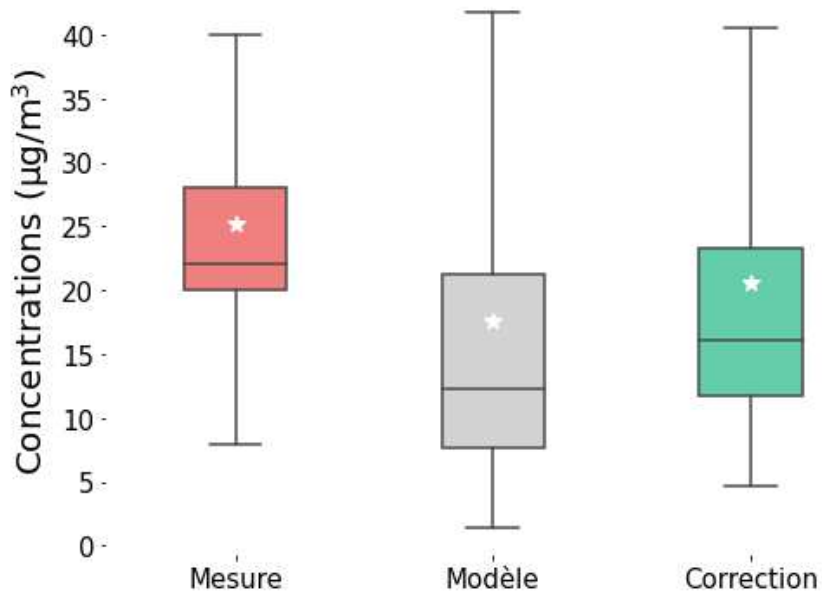
Comparaison des boxplots HAUS : PM₁₀, 2017



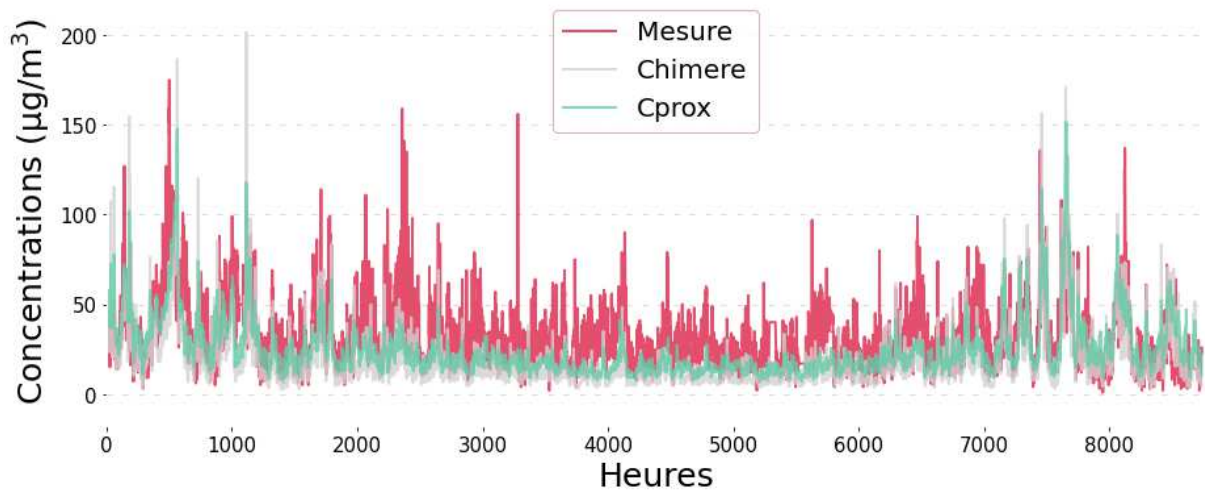
OPERA: Site de trafic (NO₂, 2017)



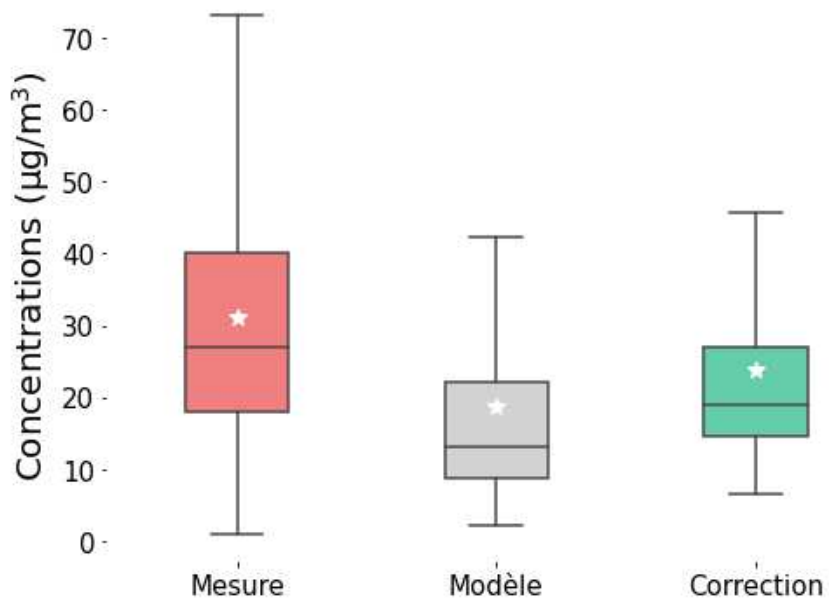
Comparaison des boxplots OPERA : PM₁₀, 2017



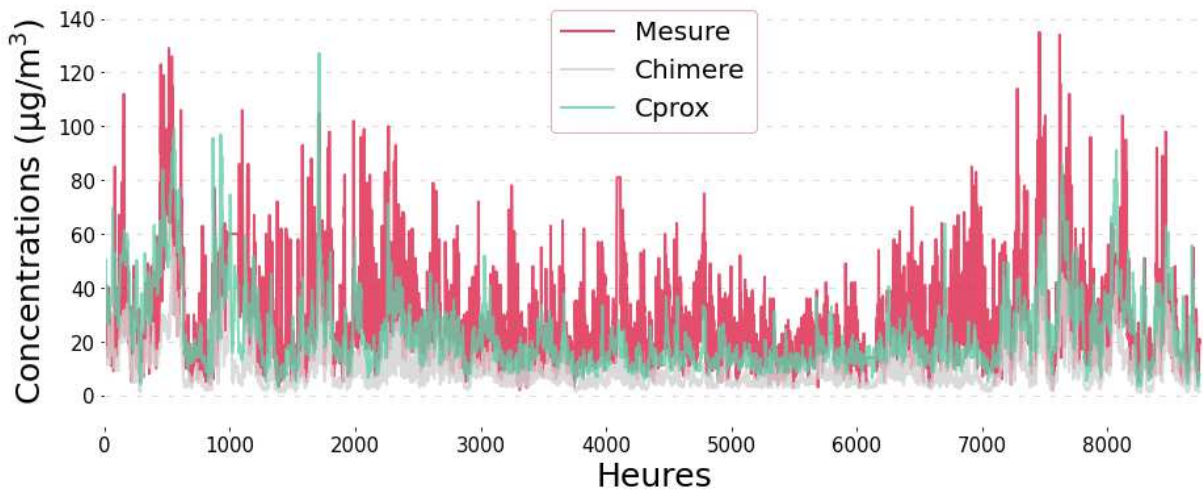
RN2: Site de trafic (NO₂, 2017)



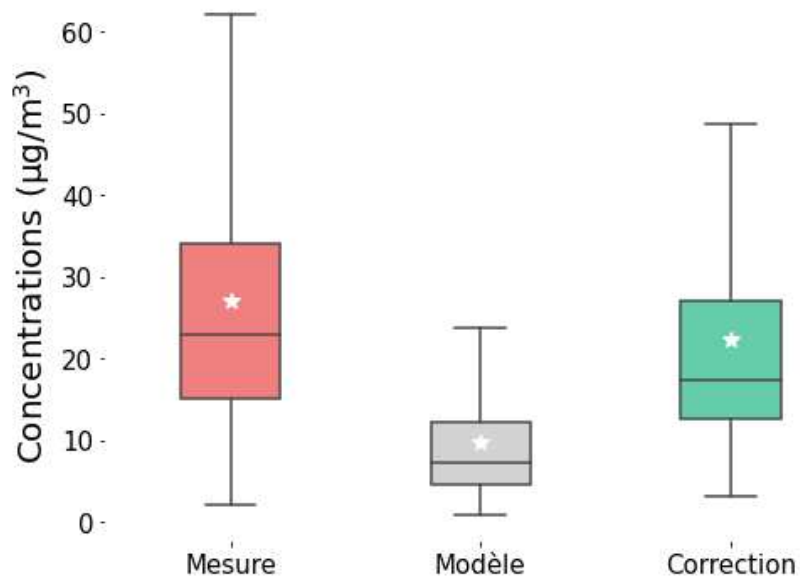
Comparaison des boxplots RN2 : PM₁₀, 2017



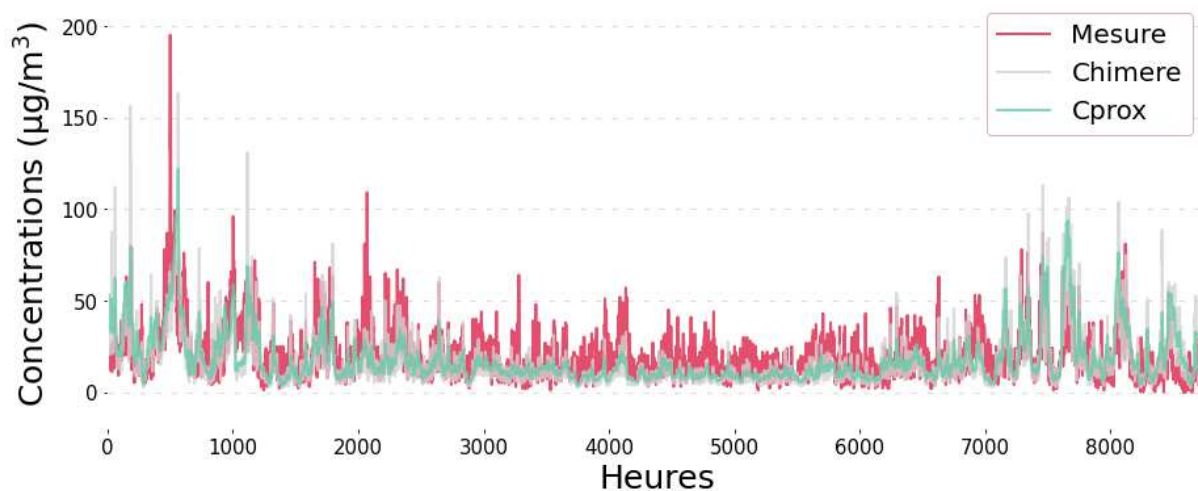
RD934: Site de trafic (NO₂, 2017)



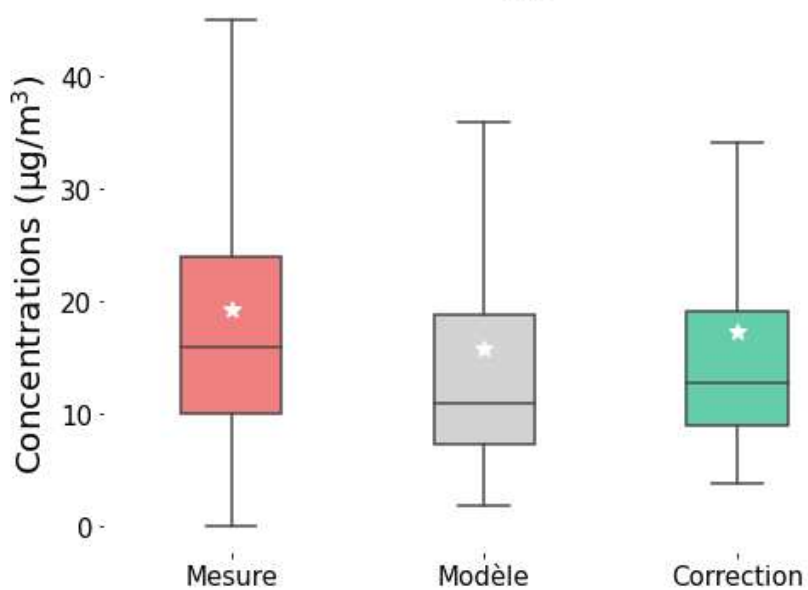
Comparaison des boxplots
RD934 : PM₁₀, 2017



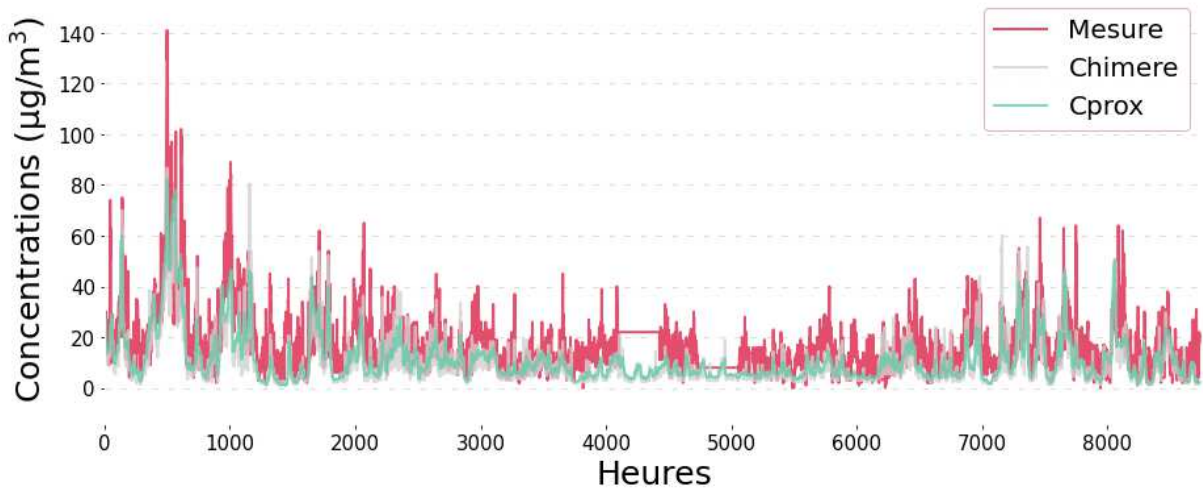
BOB: Site de fond (NO₂, 2017)



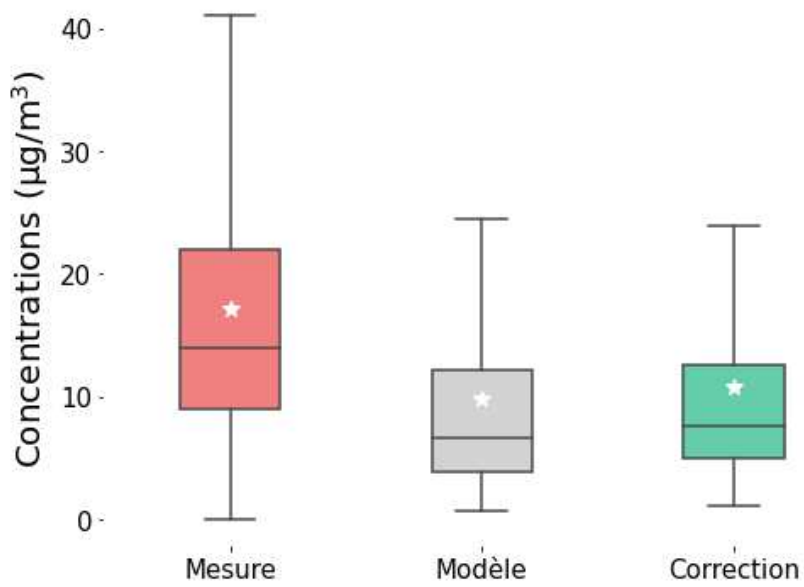
Comparaison des boxplots
BOB : PM₁₀, 2017



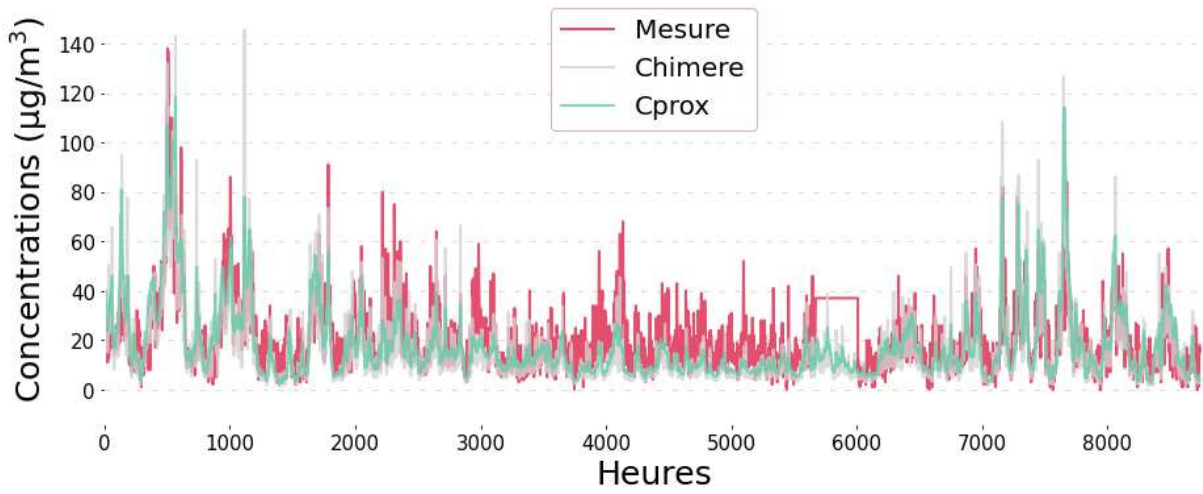
CERGY: Site de fond (NO₂, 2017)



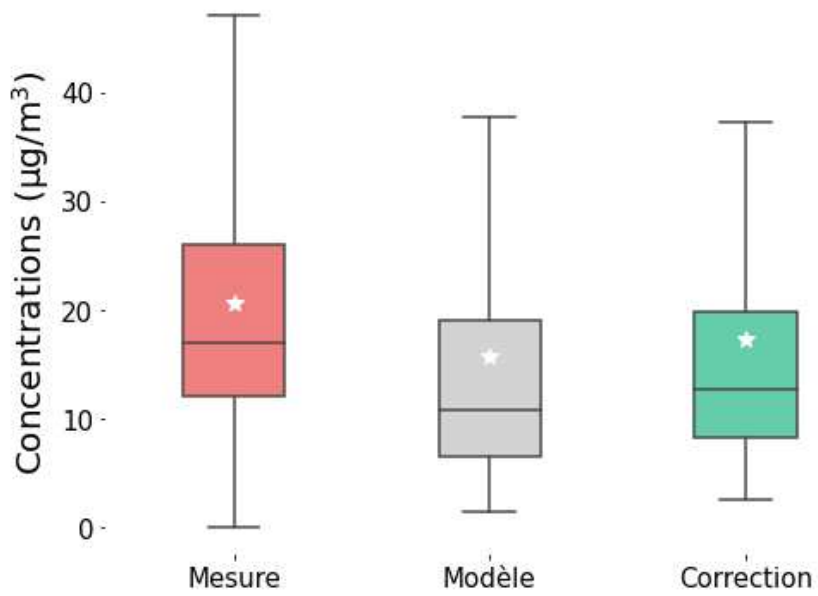
Comparaison des boxplots CERGY : PM₁₀, 2017



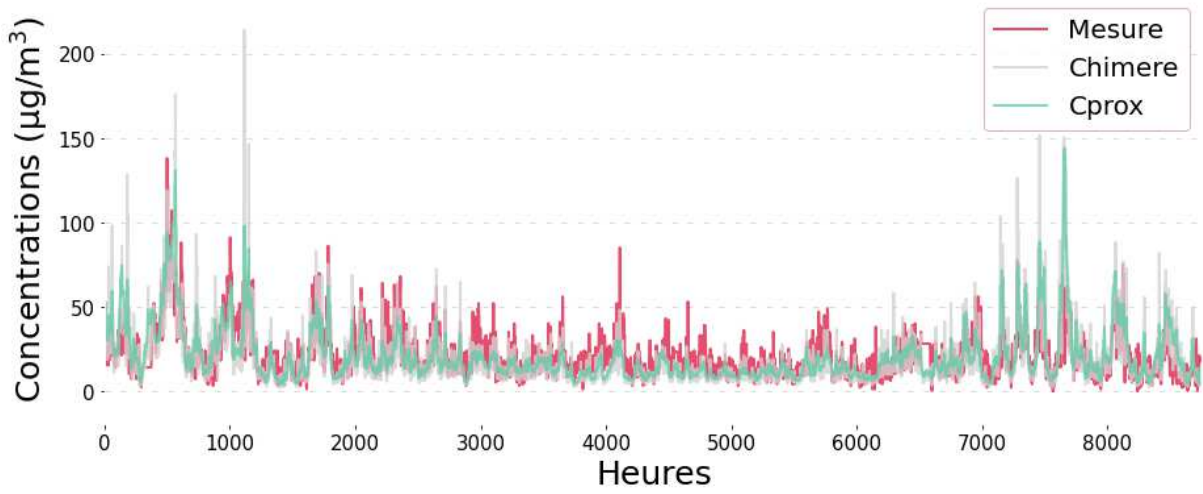
DEF: Site de fond (NO₂, 2017)



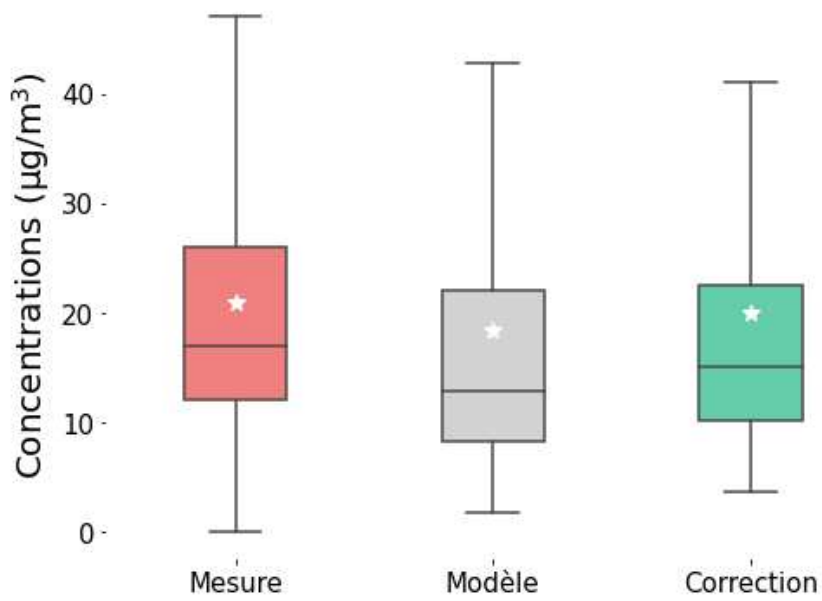
Comparaison des boxplots DEF : PM₁₀, 2017



PA18: Site de fond (NO₂, 2017)



Comparaison des boxplots PA18 : PM₁₀, 2017



ANNEXE 9

Article	Pays/Ville	Mode de transport	Polluant	Nombre de mesure	Moyenne (µg/m3)	Ecart-type	Min/Max	Commentaire	Article cité dans :
Int Panis et al., 2010	Belgique	Vélo	PM10	31	62	/	40-92		Karanasiou et al., 2014
		Voiture	PM10	9	47	/	nov-93		Karanasiou et al., 2014
Strak et al., 2010	Pays Bas, Utrecht	Vélo	PM10	14	Fort trafic : 44	29,4	16,7-118,7		Karanasiou et al., 2014
					Faible trafic : 45,7	28,4	14,2-109,3		Karanasiou et al., 2014
Zuurbier et al., 2010	Pays Bas, Arnhem	Vélo	PM10	15	Fort trafic : 38,8	14,1	/		Karanasiou et al., 2014
					Faible trafic : 37,2	11,6	/		Karanasiou et al., 2014
		Voiture	PM10	14	Diesel : 78,5	101,5	/		Karanasiou et al., 2014
					Pétrole : 58,7	34,6	/		Karanasiou et al., 2014
Berghmans et al., 2009	Belgique, Flanders	Bus	PM10	11	Diesel : 68,5	53,6	/		Karanasiou et al., 2014
					Electrique : 53,2	23,2	/		Karanasiou et al., 2014
Geiss et al., 2010	Italie, Ispra	Vélo	PM10	7	62,4	33,5	18,8-160		Karanasiou et al., 2014
		Voiture	PM10	18	48,6	30,7	0,9-263,7		Karanasiou et al., 2014

Briggs et al., 2008	GB, Londres	Voiture	PM10	46	5,9	3,1	/		Karanasiou et al., 2014
Diapouli et al., 2008	Grèce, Athènes	Marche à pied	PM10		27,6				Karanasiou et al., 2014
		Voiture	PM10	Fort trafic : 3 Faible trafic : 2	172,5 80	90 42	42-1060 20-157	Mesure moyenne sur 5 mins	Karanasiou et al., 2014 Karanasiou et al., 2014
Gulliver and Briggs, 2007	GB, Leicester	Voiture	PM10-2.5	33	15,1	5,1	8,6-28,3		Karanasiou et al., 2014
Gulliver and Briggs, 2004	GB, Northampton	Voiture	PM10	2	43,2	22,7	/		Karanasiou et al., 2014
Colombi et al., 2013	Italie, Milan	Métro	PM10	3	Quais : 103-184		58-299		Karanasiou et al., 2014
Querol et al., 2012	Espagne, Barcelone	Métro	PM10	40	quais Vieille ligne : 79 et 367	/	/		Karanasiou et al., 2014
					quais Nouvelle ligne : 45 et 193	/	/		
Raut et al., 2009	France, Paris	Métro	PM10	/	hall des stations Semaine : 183	/	/		Karanasiou et al., 2014
					hall des stations Weekend : 84	/	/		
Seaton et al., 2005	GB, Londres	Métro	PM10	3	Quais : 1200	33-160	/		Karanasiou et al., 2014
Johansson and Johansson, 2003	Suède, Stockholm	Métro	PM10	257	Quais, semaine : 469	/	212-722		Karanasiou et al., 2014
				65	Quais, weekend : 336	/	59,1-467		

Branis et al., 2006	Prague	Méto	PM10	108	Rame : 113	/	/	/	Karanasiou et al., 2014
Ripanucci et al., 2006	Italie, Rome	Méto	PM10	10	Quais : 251	/	39-2888	/	Karanasiou et al., 2014
Molle et al., 2013	France, Paris	Bus	NO2	32 mesures de 40 mins	203	55	/	Trajets en bus en plein paris ;	/
		Voiture	NO2	73	137		40-350	Figure 13	/
		Bus	NO2	31 mesures pour des itinéraires d'une heure	120	63,6	68-440	Mesures d'octobre 2007 à avril 2008 Valeurs moyennes = 120, 210 ; Indice Atmo des trajets en bus = 4 = bon	/
Delaunay et al., 2010	France	Tramway	NO2	16	30		45-89	Valeurs moyennes : 30	/
		Vélo	NO2	40	23	15,3	36-120	Valeurs moyennes : 10, 20, 40	/
		Marche à pied	NO2	67	72,5	13,2	34-155	Valeurs moyennes : 90, 65, 60, 75	/
		Méto	NO2	46	66	5,1	30-100	Valeurs moyennes : 70, 67, 60	/
		RER	NO2	31	35	7,1	17-84	Valeurs moyennes : 30, 40	/
		Méto	NO2	/	Quais : 60	/	/		/
		Méto	PM10	/	Quais : 60	/	/		/
Airparif (site web)	France, Ile-de-France	RER	PM10	/	Quais : 330	/	/		/
		Correspondances	PM10	/	50	/	/		/
		Bus	NO2	2	26	/	/		/
		Marche à pied	NO2	1	37	/	/		/
Fable et al., 2009 (projet INTERMODAL)	France, Paris	Méto	NO2	3	Rame : 17	/	/	Revue de la littérature effectuée dans le cadre du projet INTERMODAL sur 63 articles (dont 36 pour l'Europe et dont 13 pour la France)	/
		Méto	NO2	12	Quais : 20	/	/		/
		Train	NO2	1	81	/	/		/
		Vélo	NO2	1	22	/	/		/

ANNEXE 10

Nom de l'article	Auteur/Date	Pays/Ville	Type de bâtiment	Polluant	Nombre de mesure	Moyennes	Médianes	Ecart-types	Min/Max	Commentaire
OQAI - Campagne Ecoles	Derbez & Dassonville	France	Ecoles	NO2	300	/	5 µg/m3	/	/	
OQAI - Campagne Logements	Letinois & Malherbe, 2010	France	Logements	PM10	111 (Rural)	55,6 µg/m3	34,4 µg/m3	56,9 µg/m3	6,5 - 374 µg/m3	https://www.oqai.fr/media/download/221/1_1_LOG_CNLI_EXT.pdf
					90 (Péri-urbain)	52,5 µg/m3	31,35 µg/m3	75,2 µg/m3	1,6 - 522 µg/m3	
					96 (Urbain)	52,3 µg/m3	30,05 µg/m3	61,5 µg/m3	10,8 - 332 µg/m3	
Mesures de la qualité de l'air intérieur par autodiagnostic Collège Puget - Toulon	AtmoSud	France	Ecoles	NO2	5 sites mesurés avec des échantillonneurs passifs	24,8 µg/m3	/	/	/	https://www.atmosud.org/sites/paca/files/atoms/files/190716_note_technique_college_puget_toulon_vf.pdf

OOAI - CSTB Description des premiers résultats de la qualité de l'air intérieur et du confort de bâtiments d'habitation performants en énergie	Derbez et al, 2016	France	Logements	NO2	47	13,4 µg/m3	12,7 µg/m3	10,6 µg/m3	2,4 - 58,8 µg/m3	https://www.oqai.fr/media/download/253/5_BPE_1er-Etat-QAI-Confort.pdf						
Deuxième état descriptif de la qualité de l'air intérieur et du confort de bâtiments d'habitation performants en énergie	Derbez et al, 2016	France	Logements	NO2	62	16,6 µg/m3	16,2 µg/m3	9 µg/m3	max = 52 µg/m3	https://www.oqai.fr/media/download/256/5_BPE-2eme-Etat-QAI-Confort.pdf						
Indoor-Outdoor air pollution relationships in ten commercial buildings: PM2.5 and NO2	Challoner & Gill, 2014	Irlande, Dublin	Magasin	NO2	7	17,1	20	8,6								
											PM2.5	6	28,2	28	10,1	14 - 41 µg/m3
											NO2	11	16,4	20	9	
			Bureaux	PM2.5	9	14,7	13	6,4								

Human exposure to NO2 in school and office indoor environments	Salonen et al., 2019	REVUE	Ecoles	NO2	963	30,1 µg/m3	26,1 µg/m3	/	6 - 68,5 µg/m3	Revue de la littérature : 47 articles ont examinés + de 963 écoles
				Bureaux	2760	25,1 µg/m3	22,7 µg/m3	/	3 - 56,5 µg/m3	42 articles ont examinés 2760 localisations
Indoor air quality in energy-sufficient dwellings: Levels and sources of pollutants	Derbez et al., 2017	France	Logements	NO2	70	19,5 µg/m3	/	10,7 µg/m3	/	Mesures PM2.5 en automne-hiver (chauffage), mesure de NO2 aussi (la moyenne diffère de 2 µg/m3 sans le chauffage)
				PM2.5	45	25,4 µg/m3	/	29,5 µg/m3	/	
Indoor environmental quality in French dwellings and building characteristics	Langer et al., 2016	France	Logements	PM10	297	54 µg/m3	31 µg/m3	64 µg/m3	1,6 - 523 µg/m3	
				NO2	112	16 µg/m3	16 µg/m3	5,1 µg/m3	2,7 - 29 µg/m3	Campagne d'été, les valeurs de la campagne d'hiver sont de 18 µg/m3 pour la moyenne et 18 µg/m3 pour la médiane sur 128 mesures
Assessment of indoor air quality in office buildings across Europe - The OFFICAIR study	Mandin et al., 2017	Europe	Bureaux	PM2.5	16	9,7 µg/m3	9,2 µg/m3	4,6 µg/m3	2,7 - 17 µg/m3	Campagne d'été
					22	15 µg/m3	16 µg/m3	8,4 µg/m3	3,4 - 32 µg/m3	Campagne d'hiver

Particulate matter and health effects in offices - a review	Nezis et al., 2019	REVUE	Bureaux	PM10	Entre 1 et 121 mesures	76,4	57	74,7	11,3 - 333 µg/m3		
School air quality related to dry cough, rhinitis and nasal patency in children	Simoni et al., 2010	6 pays	Ecoles	PM10	21 écoles, 45 classes	112 µg/m3	106 µg/m3	69 µg/m3	14 - 260 µg/m3		
Determinants of perceived air pollution annoyance and association between annoyance scores and air pollution (PM2.5, NO2) concentrations in the European EXPOLIS study	Rotko et al., 2002	6 pays	Loisirs	NO2	176	30 µg/m3					Ils donnent les valeurs moyennes par pays ici.
				PM2.5	201	25 µg/m3					