MODELISATION SPATIO-TEMPORELLE DES CONCENTRATIONS DE PARTICULES ULTRAFINES PAR APPRENTISSAGE AUTOMATIQUE : APPLICATION A BORDEAUX METROPOLE

Agnès Hulin*[1], Olivier Le Bihan[2], Cécilia Samieri[3], Fleur Delva[4,5], Anne-Claire Devanne[1], Sabyne Audignon-Durand[4,5]

[1] Atmo Nouvelle-Aquitaine, 33690 Mérignac, France ;
[2]LB Environnement, 35200 RENNES, France ;
[3]Université de Bordeaux, INSERM, UMR1219, BPH, équipe ELEANOR, Bordeaux, F-33000, France;
[4]Université de Bordeaux, INSERM, UMR1219, BPH, équipe EPICENE, Bordeaux, F-33000, France;
[5]Centre Hospitalier Universitaire de Bordeaux, Service Santé Travail Environnement, Bordeaux, F-33000, France;
*Courriel de l'orateur : ahulin@atmo-na.org

TITLE

Spatio-temporal modeling of ultrafine particle concentrations using machine learning: application to the Bordeaux metropolitan area

RESUME

Dans le cadre d'une étude épidémiologique, les concentrations moyennes journalières de particules ultrafines (PUF) ont été estimées par apprentissage automatique à l'échelle kilométrique sur Bordeaux Métropole pour les années 2016 à 2024. Le modèle, construit à partir de 34 variables d'entrées (polluants réglementés, paramètres météorologiques, occupation du sol, mesures satellitaires), présente de bonnes performances (corrélation de 0.82 en validation croisée k-folds, k = 10). En 2024, les concentrations estimées varient de 3 800 à 9 700 particules·cm⁻³, avec des niveaux plus élevés à proximité des axes routiers.

ABSTRACT

As part of an epidemiological study, daily average concentrations of ultrafine particles (UFP) were estimated using machine learning at a one-kilometer resolution over the Bordeaux metropolitan area for the years 2016 to 2024. The model, based on 34 variables (regulated pollutants, meteorological parameters, land use, satellite measurements), shows good performance (correlation of 0.82 under cross-validation, k-folds k = 10). In 2024, estimated concentrations range from 3 800 to 9 700 particles·cm⁻³, with higher levels observed near major roadways.

MOTS-CLES : particules ultrafines, Machine Learning, air, cartes / **KEYWORDS:** ultrafine particles, Machine Learning, air, maps

1. INTRODUCTION ET OBJECTIFS

Les particules ultrafines (ou PUF, de diamètre ≤ 100 nm) dominent en nombre la distribution des particules atmosphériques tout en ne représentant qu'une faible part de leur masse totale. Elles peuvent traverser les barrières biologiques et pénétrer dans l'organisme, avec des effets potentiels sur la santé cardiovasculaire, respiratoire et neurologique (ANSES, 2019).

Pour évaluer l'impact sanitaire des PUF dans l'air ambiant par des études épidémiologiques, tâche rendue difficile par la forte variabilité spatiale et temporelle de ce polluant (Morawska, 2019), il est nécessaire de disposer de données d'exposition spatialisées à une résolution supérieure à celle des réseaux de mesure existants. Alors que la mesure des particules fines (PM2.5) en air ambiant, conduite depuis plus de vingt ans, a permis d'établir leur impact majeur sur la santé, celle des PUF demeure beaucoup plus récente et fragmentaire.

Les récents progrès du machine learning permettent désormais d'intégrer un grand nombre de variables environnementales complexes pour prédire les concentrations de polluants émergents, sans les contraintes propres aux modèles déterministes ((Ahmad Makhdoomi, 2025); (Umesh Kumar Lilhore, 2025), (Petrić, 2024)). Ainsi, une étude récente menée à Taïwan (Chau-Ren Jung, 2023) propose un modèle de machine learning pour la prédiction à l'échelle kilométrique des concentrations des PUF.

Lauréat du PNR-EST de l'ANSES et piloté par l'équipe EPICENE de l'Inserm, le projet B cube-PUF vise à étudier les liens entre l'exposition aux PUF et les processus de neurodégénérescence, à partir d'une cohorte de 2 000 jeunes séniors (cohorte B cube, 55-80 ans vivant en métropole bordelaise, PI C Samieri). Dans ce cadre, un modèle spatio-temporel de machine learning a été développé en s'inspirant des travaux de l'équipe

DOI : 10.25576/ASFERA-CFA2026-50134

de Chau-Ren Jung (Chau-Ren Jung, 2023). Ce modèle a permis de produire des estimations de l'exposition aux PUF dans l'air ambiant sur la métropole bordelaise, à l'échelle kilométrique, pour la période 2016-2024. Ces données d'exposition viendront compléter celles relatives à l'exposition professionnelle et à l'air intérieur produites par les autres partenaires du projet.

2. METHODES

Les algorithmes d'apprentissage automatique génèrent, à l'issue de la phase d'apprentissage, un modèle prédictif des concentrations de PUF en nombre (la variable cible), à partir des variables d'entrées, selon la forme :

PUF = f[DOY(Day Of the Year) + Aerosol (MODIS) + NH3 (IASI) + NO2, O3, PM10, PM2.5 (PREVAIR)+ Météorologie (ERA5)+ occupation du sol (routier et bâti, IGN et INSEE)]

L'apprentissage consiste à entraîner l'algorithme à reconnaître des relations entre les données cibles et d'entrées afin de prédire de nouvelles observations. Le modèle une fois validé est utilisé pour prédire les concentrations de PUF sur Bordeaux Métropole. Il s'appuie sur des algorithmes de type XGBoost. Son apprentissage et la prédiction des concentrations de PUF ont été réalisées à l'échelle journalière, sur un maillage de résolution kilométrique couvrant le territoire administratif de Bordeaux Métropole ainsi que les stations de mesure des PUF du territoire national. La résolution kilométrique a été retenue pour être cohérente avec celle les données d'entrée.
Les concentrations moyennes journalières de PUF proviennent des mesures réalisées dans la gamme granulométrique [7 nm-1 µm] par une sélection de compteurs CPC et UFP 3031 répartis sur 15 sites en France métropolitaine et en Corse.

Les variables d'entrée sont les suivantes :

- **Polluants atmosphériques** : dioxyde d'azote (NO_2), ozone (O_3), particules fines PM10 et PM2.5. Les données proviennent de la plate-forme PREV'AIR, basée sur le modèle de chimie-transport Chimère et développée par l'INERIS.
- **Météorologie :** vent, pression, hauteur de couche limite, température, point de rosée, couche nuageuse, rayonnement solaire et précipitations. Les données proviennent du modèle de données réanalysées Européen ERA5 (ECMWF Reanalysis v5) produit par l'ECMWF.
- **Réseau routier :** sa représentation s'appuie sur la BD TOPO® de l'IGN. Seul le réseau principal a été conservé. A défaut d'une donnée trafic disponible sur l'ensemble du réseau, c'est la largeur de la voie qui a été utilisée pour caractériser l'importance du brin routier. Ainsi à chaque maille kilométrique est associée la somme des surfaces des routes qu'elle contient (longueur du tronçon × largeur de la voie).
- **Chauffage au bois résidentiel :** il est représenté à partir de deux sources de données, l'enquête « Détail Logement » de l'INSEE, qui indique la part des logements utilisant le bois comme combustible principal et la surface bâtie résidentielle issue de la BD TOPO de l'IGN. A chaque maille de la grille kilométrique est associée la surface du bâtie résidentiel pondérée par la part de chauffage au bois.
- **Mesures satellitaires :** les concentrations d'ammoniac (NH_3) sont issues de IASI (AERIS (Clarisse, 2023). Les données pour les aérosols sont issues de la filière MODIS 3 km (Moderate Resolution Imaging Spectroradiometer, NASA (L. A. Remer, 2013)).

3. RESULTATS

Les cartes kilométriques produites à l'échelle journalière pour les années 2016 à 2024 (cf. Figure 1) montrent en 2024 des concentrations comprises entre 3 800 et 9 700 particules·cm^{-3}, et entre 3 700 et 13 200 particules·cm^{-3} pour l'année 2016. Les maxima se situent sur la rocade et les communes de l'intra-rocade. Les concentrations annuelles décroissent sur la période étudiée (cf. Figure 2), en cohérence avec la baisse observée sur la station de mesure des PUF de Talence.

La validation 10-folds donne une corrélation moyenne de 0.82 (Racine de l'erreur quadratique moyenne ou RMSE ≈ 2 214 particules·cm^{-3}) avec un léger biais de sous-estimation des valeurs élevées. La corrélation moyenne pour la validation spatiale (leave-one-station-out ou exclusion successive de chaque site) est de 0.63, avec de meilleurs résultats sur les stations de fond (Talence) que sur les sites trafic (Montpellier Liberté, Lille Leeds)

Les cinq variables les plus influentes sur la prédiction des PUF sont le NO_2, l'occupation du sol avec le bâti et le réseau routier, l'ozone et les PM2.5. Viennent ensuite les variables météo. Le vent moyen, le rayonnement solaire incident et le rayonnement UV ressortent comme facteurs majeurs, reflétant la dispersion et les

DOI : 10.25576/ASFERA-CFA2026-50134

processus photochimiques de formation de particules secondaires. Les données satellites (MODIS, IASI) apportent une amélioration marginale.

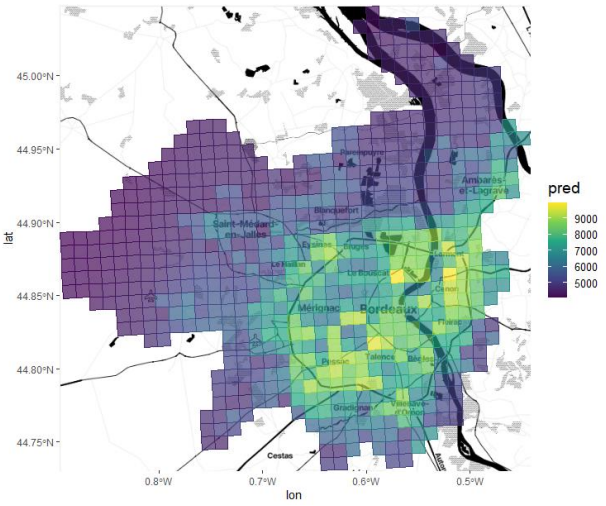

Figure 1. Concentrations moyennes annuelle prédites par le modèle pour l'année 2023 sur Bordeaux métropole en nombre de particules par m3 (grille kilométrique)

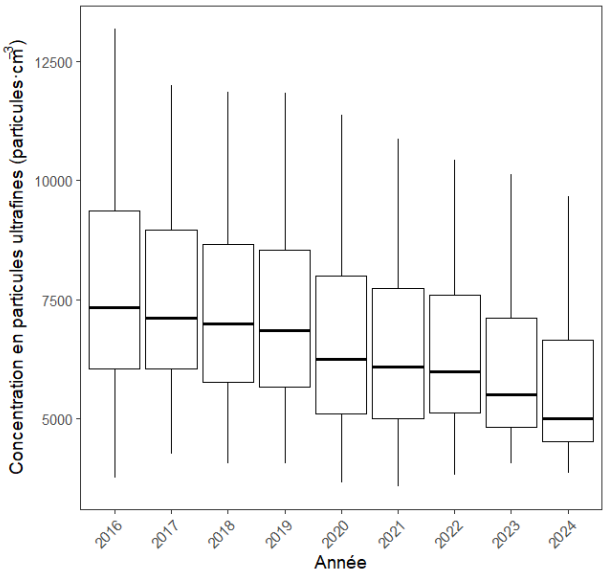

Figure 2. Répartition des concentrations moyennes annuelles estimées par le modèle sur la grille kilométrique de Bordeaux Métropole (min, 1er quart, med, 3eme quart, max).

4. DISCUSSION

Ce travail constitue une avancée significative car il offre un jeu de données inédit sur les concentrations de particules ultrafines (PUF) sur Bordeaux Métropole. L'intégration des données kilométriques de NO_2, d'ozone et de PM2.5 issues du modèle PREVAIR a sensiblement amélioré les performances et la cohérence spatiale du modèle. Malgré ces résultats prometteurs, plusieurs défis demeurent, en raison notamment de l'hétérogénéité et de la disponibilité limitée des mesures de PUF. Une base de données harmonisée et des données de trafic sur l'ensemble du réseau routier national permettraient de renforcer la précision du modèle et de mieux représenter l'exposition au trafic. Pour finir, la résolution kilométrique représente mal l'exposition de proximité au trafic, comme en témoignent les performances plus faibles observées sur les stations trafic.

5. CONCLUSION

L'approche fondée sur l'apprentissage automatique a offert une solution pertinente en l'absence de modèle déterministe dédié et face au nombre limité de stations de mesure des PUF sur le territoire. Les performances obtenues se révèlent robustes sur le plan temporel et prometteuses sur le plan spatial. Bien que la demande initiale du projet B cube-PUF portait sur la production d'une moyenne annuelle des concentrations, les données ont été générées à l'échelle journalière afin de répondre aux besoins de construction du modèle, offrant ainsi un jeu de données plus riche et potentiellement réutilisable pour d'autres projets.
L'analyse de l'importance des variables confirme le rôle prépondérant du trafic routier, ainsi que l'influence, plus modérée, du chauffage au bois, soulignant la nécessité de politiques publiques ciblant la mobilité et le chauffage résidentiel. L'importance des variables d'entrées, tels les maxima d'ozone, le rayonnement solaire, met en évidence le rôle des processus photochimiques, suggérant un renforcement potentiel de ces phénomènes dans un contexte de changement climatique.
Ce travail ouvre ainsi la voie à une meilleure estimation de l'exposition chronique des populations urbaines aux PUF, enjeu majeur pour la recherche épidémiologique et la santé publique. À terme, l'approche pourrait être étendue à d'autres territoires, enrichie par de nouvelles données de mesure et intégrée à des outils opérationnels d'aide à la décision pour la gestion de la qualité de l'air.

Remerciements : ANSES (contrat n°23-EST-166) LCSQA/ INERIS (PREV'AIR et GEOD'AIR) ; les AASQA (données de mesure des PUF), AERIS (IASI-NH3), NASA (données MODIS), IGN, INSEE.

Références principales

Ahmad Makhdoomi, M. S. (2025). PM2.5 concentration prediction using machine learning algorithms: an approach to virtual monitoring stations. *Scientific Reports volume 15*, Article number: 8076 .

DOI : 10.25576/ASFERA-CFA2026-50134

ANSES. (2019). *Particules de l'air ambiant extérieur - Effets sanitaires des particules de l'air ambiant extérieur selon les composés, les sources et la granulométrie.*

Chau-Ren Jung, W.-T. C.-H.-C. (2023). A hybrid model for estimating the number concentration of ultrafine particles based on machine learning algorithms in central Taiwan. *Environment International 175* .

Clarisse, L. F.-L. (2023). The IASI NH3 version 4 product: averaging kernels and improved consistency. *Atmos. Meas. Tech.*

L. A. Remer, S. M. (2013). MODIS 3 km aerosol product: algorithm and global perspective. *Atmos. Meas. Tech., 6*, 1829-1844.

Morawska, L. &.-K. (2019). *Ambient ultrafine particles: evidence for policy makers. A report prepared by the 'Thinking outside the box' team.*

Petrić, V. H. (2024). Ensemble Machine Learning, Deep Learning, and Time Series Forecasting: Improving Prediction Accuracy for Hourly Concentrations of Ambient Air Pollutants. *Aerosol Air Qual. Res.* .

Umesh Kumar Lilhore, S. S. (2025). Advanced air quality prediction using multimodal data and dynamic modeling techniques. *Scientific Reports volume 15*, Article number: 27867 .

DOI : 10.25576/ASFERA-CFA2026-50134