THÈSE DE DOCTORAT

Préparée à l'Université de Lille par

Valisoa Miadana RAKOTONIRINJANAHARY

Soutenue publiquement le 16 avril 2025 devant la Commission d'Examen

En vue de l'obtention du grade de Docteur en Chimie Théorique, Physique et Analytique

Évaluation et optimisation de la qualité de l'air dans les enceintes ferroviaires souterraines : analyse des concentrations en particules et efficacité des systèmes de filtration

Assessment and optimization of air quality in underground railway stations: analysis of PM concentrations and the effectiveness of filtration systems

Composition du Jury :

RIFFAULT Véronique	Professeure, CERI EE, IMT Nord Europe	Présidente du Jury
BLONDEAU Patrice	Maître de conférences, LASIE, La Rochelle Université	Rapporteur
LAVERGE Jelle	Professeur associé, Ghent University	Rapporteur
COLL Isabelle	Professeure, LISA, Université de Paris Est Créteil	Examinatrice
ZEITOUNI Karine	Professeure, DAVID, Université Versailles Saint-Quentin	Examinatrice
MELLOUKI Wahid	Professeur, Université Mohammed VI Polytechnique, Ben Guerir, Maroc	Examinateur
HANOUNE Benjamin	Chargé de recherche CNRS, PC2A, Université de Lille	Directeur de thèse
CRUMEYROLLE Suzanne	Maîtresse de conférences, LOA, Université de Lille	Co-encadrante
VIMARD Anne-Marie	Ingénieure, DSG, SNCF	Invitée, encadrante
BOGDAN Mateusz	Docteur, AREP L'hypercube	Invité, encadrant

« Cela semble toujours impossible jusqu'à ce que ce soit fait »

Nelson Mandela

REMERCIEMENTS

Je tiens tout d'abord à exprimer ma plus sincère gratitude envers mon directeur de thèse, Benjamin Hanoune, directeur du laboratoire PhysicoChimie des Processus de Combustion et de l'Atmosphère (PC2A), pour sa bienveillance, son aide précieuse, sa patience, ses orientations éclairées, son soutien enthousiaste, ses retours d'expertise et ses corrections linguistiques de mes articles et de ma thèse. Il m'a offert l'opportunité de rejoindre son équipe de recherche, et je lui en suis profondément reconnaissant. Au cours de ces trois années, il m'a confié des missions passionnantes et m'a permis de participer à des conférences scientifiques, m'exposant ainsi aux avancées les plus récentes dans notre domaine.

Je remercie également ma co-encadrante de thèse, Suzanne Crumeyrolle, du Laboratoire d'Optique Atmosphérique (LOA), pour ses conseils avisés et sa capacité à me pousser au-delà de mes limites, depuis le début de ma thèse jusqu'à la rédaction de ce manuscrit. Son accompagnement a été essentiel pour enrichir ma culture scientifique et m'a permis de progresser dans ma carrière de chercheur.

Au sein de la SNCF, je tiens à exprimer ma profonde gratitude envers mes référents en entreprise qui se sont succédé pour encadrer ma thèse. Je remercie tout particulièrement Frédéric Villot, qui m'a offert l'opportunité de réaliser cette thèse au sein de la SNCF et m'a fait confiance dès le départ. Merci également à Isabelle Delobel pour ses conseils et sa gentillesse, ainsi qu'à Anne-Marie Vimard, mon encadrante actuelle, pour ses idées géniales, nos échanges hebdomadaires et son soutien constant.

Je souhaite également remercier Mateusz Bogdan, de chez AREP, qui a été bien plus qu'un encadrant : un véritable grand frère. Merci pour tes conseils, ta disponibilité pour répondre à mes questions, et nos échanges enrichissants, tant professionnels que personnels, à la gare du Nord.

Je tiens à remercier Patrice Blondeau, enseignant-chercheur à l'Université de La Rochelle, et Jelle Laverge, professeur associé à l'Université de Gand en Belgique, pour avoir accepté de rapporter mes travaux de thèse. Mes remerciements vont également aux examinateurs : Véronique Riffault, professeure à l'IMT Nord Europe ; Isabelle Coll, professeure à l'Université Paris-Est Créteil ; Karine Zeitouni, professeure à l'Université de Versailles Saint-Quentin ; et Wahid Mellouki, professeur à l'Université Mohammed VI Polytechnique de Ben Guerir au Maroc.

Je remercie également les membres de mon Comité de Suivi Individuel : Isabelle Coll, Marc Abadie, enseignant-chercheur à l'Université de La Rochelle, et Luc Dauchet, maître de conférences à l'Université de Lille. Vos remarques lors de nos réunions ont permis d'enrichir mes travaux en leur ouvrant de nouvelles perspectives.

Je tiens à remercier la direction DRASS (Direction des Risques, Audits, Sécurité et Sûreté), dirigée par Jean-Claude Larrieu, ainsi que Yohanna François, responsable de la DSG (Direction Sécurité Groupe) au sein de la DRASS, pour m'avoir accueilli au sein de la DSG. Un grand merci également à mon équipe, l'équipe SST (Santé Sécurité au Travail), dirigée par Dominique Latard, et à ses membres : Hugues Biseuil, Michel Lacombe, Laurence Mai, Patricia Bon Mardion, Catherine Mergault, Anne-Sophie Bergounhon, Laurent Leonardi et Ilyana Valière. Vous n'étiez pas seulement mes collègues, mais ma deuxième famille. Nos échanges autour d'un café et nos sorties d'équipe resteront des souvenirs inoubliables qui ont rendu mon passage à la SNCF particulièrement agréable.

Je remercie également les membres des groupes de travail COPROJ, COTECH, GT Qualité de l'Air et GT Intelligence Artificielle au sein de la SNCF, notamment Anaïs Watbled, Cora Cremezi, Laurent Dupont, Fayez Ghozzi et Théo Schwartz, pour m'avoir facilité mon intégration dans l'entreprise et m'avoir permis de me familiariser avec les enjeux de la qualité de l'air dans les enceintes ferroviaires souterraines.

Je tiens également à remercier mes interlocutrices chez AIRPARIF, Géraldine Le Nir et Heidi Carrilho de l'AEF, pour leurs conseils et les échanges de données qui ont été précieux pour mes recherches.

Je souhaite exprimer ma gratitude à mes collègues du PC2A, en particulier Laure, Guillaume, Sonia, Amaury et Maxence, qui ont rendu mon quotidien au laboratoire si agréable. Je n'oublie pas non plus mes camarades doctorants, notamment Jeyan, Hisham, Zainab, Jérôme et Luna, qui sont devenus au fil du temps de véritables amis.

Je voudrais adresser mes remerciements les plus sincères à mes parents, Clément et Fanja, qui m'ont guidé, soutenu et encouragé tout au long de ma vie. Leurs efforts et leur confiance en moi dépassent tout ce que les mots peuvent exprimer.

Je suis également très reconnaissant envers mes deux sœurs, Miantra et Felana, pour leur aide, leur patience et leurs encouragements constants. Vous êtes les meilleures.

Enfin, je tiens à remercier du fond du cœur Hafatra, ma conjointe, qui m'a toujours soutenu et encouragé. Sans elle, je n'aurais jamais pu mener ce travail à bien. Merci infiniment pour ta patience, ta compréhension, tes sacrifices et ton amour indéfectible au cours de ces années. J'espère que moi et mes particules ne t'ont pas trop importunée durant ces trois années !

SOMMAIRE

REMERCIEMENTS 5
SOMMAIRE 7
TABLE DES FIGURES 10
TABLE DES TABLEAUX 14
NOMENCLATURE 16
INTRODUCTION 19
CHAPITRE 1. ÉTAT DE L'ART 23
1.1 Sources de particules en EFS 27
1.2 Variabilité et facteurs d'influence des concentrations en particules 30
1.2.1 Variabilité spatiale 30
1.2.2 Variabilité temporelle des particules au niveau du quai 35
1.3 Solutions pour améliorer la qualité de l'air dans les EFS 40
1.3.1 Stratégies passives 40
1.3.2 Stratégies actives 42
1.3.2.1 Les systèmes de ventilation 42
1.3.2.2 Equipements de purifications de l'air 44
1.4 Conclusion du chapitre 45
CHAPITRE 2. DISPOSITIFS DE MESURE ET SITES D'ÉTUDE 47
2.1 Instrumentation 47
2.1.1 Instruments de mesures de la qualité de l'air en EFS 47
2.1.1.1 Analyseur automatique de particules 48
2.1.1.2 Analyseur de CO_2 par corrélation infrarouge 51
2.1.1.3 Capteur de température et d'humidité relative 52
2.1.1.4 Contrôle qualité des données 53
2.1.2 Instruments de filtration des particules 54
2.1.2.1 Par ionisation 54
2.1.2.2 Par voie humide 55
2.1.2.3 Par filtre classique 55
2.2 Les EFS instrumentées 55
2.2.1 Gares avec mesures de courte durée 56
2.2.2 Gares avec mesures de longue durée 58
2.2.2.1 Gare d'Avenue Foch 58
2.2.2.2 Gare de Sevran Beaudottes 61
2.2.2.3 Gare de Magenta 63

2.2.3 Station de mesures de l'air extérieur 65
2.3 Conclusion du chapitre 67
CHAPITRE 3. MÉTHODOLOGIES D'ANALYSE DES DONNÉES 69
3.1 Prétraitement et nettoyage des données 69
3.2 Méthode basée sur le Coefficient d'Amplitude Journalier (DAC) 72
3.2.1 Calcul de profils moyens initiaux (jours ouvrés et weekends séparés) 73
3.2.2 Ajustement linéaire et coefficient d'amplitude journalier (DAC) 74
3.2.3 Elimination des profils atypiques 75
3.2.4 Application aux gares avec des mesures courte durée 77
3.2.5 Evaluation de la réduction des PM pour les systèmes de filtration 79
3.3 Modèle additif généralisé (GAM) 80
3.3.1 Critères de sélection 81
3.3.2 Identification et traitement des variables catégorielles 83
3.3.3 Identification et traitement des variables continues 83
3.3.4 Équation finale 84
3.3.5 Evaluation de la réduction des PM à l'aide du modèle GAM 85
3.4 Réseau de neurones récurrents à mémoire à long et court terme (LSTM) 86
3.4.1 Architecture et fonctionnement interne d'une cellule LSTM 86
3.4.2 Entraînement et validation du modèle 87
3.4.3 Hyperparamètres et optimisation du modèle 87
3.4.4 Métriques d'évaluation des prédictions 88
3.4.5 Précision des prévisions des concentrations de PM sur différents horizons temporels . 89
3.4.6 Evaluation de la réduction des PM à l'aide du modèle LSTM 89
3.5 Conclusion du chapitre 90
CHAPITRE 4. ANALYSE DES CONCENTRATIONS EN PARTICULES DANS LES EFS 91
4.1 Variabilité et tendances à long terme des concentrations en particules les gares de références 91
4.2 Profils journaliers dans les 3 EFS avec des mesures de longue durée 100
4.3 Profils journaliers dans les gares avec des mesures de courte durée 108
4.4 Incertitude des profils en fonction de la période de mesure 111
4.5 Conclusion du chapitre 113
CHAPITRE 5. ANALYSE DES FACTEURS D'INFLUENCE SUR LES CONCENTRATIONS EN PARTICULES 115
5.1 Analyse des facteurs d'influence avec le modèle GAM 115
5.1.1 Cas d'Avenue Foch 115
5.1.2 Cas de Sevran Beaudottes 120
5.1.3 Cas de Magenta 124
5.2 Evaluation de l'efficacité des systèmes de filtration 128

5.2.1 Evaluation des deux technologies de filtration à Avenue Foch 128
5.2.1.1 Evaluation avec la méthode DAC 128
5.2.1.2 Evaluation avec le modèle LSTM 129
5.2.2 Évaluation de la technologie de filtration à Sevran Beaudottes 131
5.3 Conclusion du chapitre 133
CONCLUSION GÉNÉRALE ET PERSPECTIVES POUR DE FUTURS TRAVAUX 135
BIBLIOGRAPHIE 141
ANNEXES I
Annexe 1 : A novel method for establishing typical profile in underground railway stations I
Annexe 2 : Comparison of three methods for evaluating filtration system performance in underground railway stations XIII
Annexe 3 : Périodes notables et élimination des profils atypiques dans les gares avec mesures longue durée XLIII
Annexe 4 : En EFS – Profils journaliers types du lundi au vendredi (jours ouvrés) et du samedi au dimanche (weekends) et analyse du temps de déposition ($PM_{2.5}$ – $PM_{2.5-10}$) XLVII
Annexe 5 : En extérieur – Profils journaliers types du lundi au vendredi (jours ouvrés) et du samedi au dimanche (weekends) LI
RÉSUMÉ LV
ABSTRACT LVI

TABLE DES FIGURES

Figure 1.1 : Pourcentage d'études sur la qualité de l'air dans les EFS, réparties par continents, de 2000 à 2024 (Wang et al., 2024). 23
Figure 1.2 : Concentrations en particules dans les EFS dans différentes villes du monde (adapté et mis à jour de Passi et al. 2021) 24
Figure 1.3 : Cartographie des niveaux de pollution dans les gares et stations d'Île-de-France (https://data.iledefrance-mobilites.fr/explore/dataset/qualite-de-lair-dans-le-reseau-de-transport-francilien/custom/)..... 26
Figure 1.4 : Différentes sources de particules dans les EFS (adapté de PIXLR générateur d'images IA) 27
Figure 1.5 : Taux de déposition en fonction du diamètre de particule (Walther et al., 2017)..... 29
Figure 1.6 : Concentrations des PM_{10} et $PM_{2.5}$ entre les zones de stations de métro souterraines à Séoul (adapté de Kim et al., 2008) 30
Figure 1.7 : Profils des concentrations de $PM_{2,5}$ ($\mu g/m^3$) sur le quai de la station de métro Syntagma à Athènes (adapté de Martins et al. 2016). 31
Figure 1.8 : Simulation de la concentration moyenne de PM_{10} sur un quai de station de Métro à Tianjin, Chine (Yang et al., 2024)..... 32
Figure 1.9 : Relation entre la profondeur de la station de métro et la concentration moyenne journalière de PM_{10} aux gares Aéroport, Aksaray, 4Levent, Sisli et Taksim à Istanbul, Turquie (adapté de Şahin et al., 2012)..... 32
Figure 1.10 : Concentrations de $PM_{2.5}$ dans les lignes en profondeur du réseau ferroviaire londonien, classées selon la proximité avec l'entrée de tunnel (adapté de Saunders et al., 2019)..... 34
Figure 1.11 : Concentrations de particules (PM_{10}, $PM_{2.5}$, PM_1) et de gaz (CO_2, CO) dans différentes stations de la ligne L2 du métro de Barcelone (Moreno et al., 2014)..... 35
Figure 1.12 : Evolution des concentrations en PM_{10} avec la cadence d'exploitation dans la station d'Arlanda C à Stockholm (adapté de Gustafsson et al., 2012 par Durand, 2021)..... 36
Figure 1.13 : Répartition mensuelle des concentrations de $PM_{2.5}$ à l'intérieur et à l'extérieur de l'ensemble des stations de métro de Séoul (Park et al., 2024) 37
Figure 1.14 : Concentrations de PM_{10}, $PM_{2.5}$ et PM_1 sur le quai de la station de métro Joanic à Barcelone à 4 sites différents (P1, P2, P3 et P4) (adapté de Martins et al. 2015)..... 38
Figure 1.15 : Relation entre les concentrations de $PM_{2,5}$ ($\mu g/m^3$) sur les quais de métro et l'air ambient extérieur à a) Barcelone (40 m de profondeur), b) Athènes (20 m de profondeur) et à c) Porto (Peu profonde) (adapté de Martins et al. 2016) 38
Figure 1.16: Différentes solutions pour améliorer la qualité de l'air dans les EFS..... 40
Figure 1.17 : Porte palières en EFS : (a) PSDs, (b) FPBDs, (c) HPBDs et (d) AVPDs (Yang et al., 2023) 41
Figure 1.18 : Roues en caoutchouc utilisées dans le métro parisien (https://www.theautopian.com/heres-why-some-trains-have-rubber-tires/)..... 42
Figure 1.19 : Fonctionnement du système UPE (adapté de Izadi et al. (2022) 43
Figure 1.20 : Système de ventilation dans une EFS (adapté de Wen et al. 2020)..... 43
Figure 1.21 : Modèle de précipitateur électrostatique. (https://www.iqsdirectory.com/articles/air-pollution-control/electrostatic-precipitators.html)..... 44
Figure 2.1 : Baie de mesure de la qualité de l'air installée dans la gare de Musée d'Orsay lors de la campagne de mesure du 10 avril au 26 avril 2018 (AIRPARIF, 2018). 48
Figure 2.2 : Principe du TEOM (adapté de Su et al. 2020) 49
Figure 2.3 : Analyseurs automatiques (a) TEOM 1405-D et (b) TEOM 1405-F (https://www.thermofisher.com/)..... 50

Figure 2.4 : (a) Analyseur de CO_2 modèle 410i, (b) Schéma fonctionnel illustrant son principe de fonctionnement (adapté de https://www.thermofisher.com/)..... 52
Figure 2.5 : Capteur de température et d'humidité relative (https://www.atal.nl/)..... 52
Figure 2.6 : Principe de fonctionnement de la technologie d'ionisation électrostatique positive 54
Figure 2.7 : Principe de fonctionnement de la technologie par voie humide 55
Figure 2.8 : Localisation des gares avec mesures de courte durée en Île-de-France..... 56
Figure 2.9 : Illustration de la gare d'Avenue Foch : (a) vue extérieure, (b) quai, et (c) plan simplifié de la gare 58
Figure 2.10 : Disposition des modules de filtration Air Liquide et localisation du TEOM..... 60
Figure 2.11 : Disposition des modules de filtration Starklab et localisation du TEOM 60
Figure 2.12 : Illustration de la gare de Sevran Beaudottes : (a) vue extérieure, (b) quai et (c) plan simplifié de la gare 61
Figure 2.13 : Concentrations journalières de PM_{10} et $PM_{2.5}$ à Sevran-Beaudottes, mettant en évidence l'impact du changement d'instrumentation le 22 décembre 2021. 62
Figure 2.14 : Illustration de la gare de Magenta : (a) vue extérieure, (b) quai, et (c) plan simplifié de la gare 64
Figure 2.15 : Concentrations journalières de PM_{10} et $PM_{2.5}$ à Magenta, mettant en évidence l'impact du changement d'instrumentation le 22 décembre 2021..... 65
Figure 2.16 : Localisation des trois gares de référence (Avenue Foch, Sevran Beaudottes, Magenta) et des stations de mesures de l'air extérieur correspondantes (Porte d'Auteuil, Bobigny, Paris 18è) 66
Figure 3.1 : Partitionnement des points dans une distribution gaussienne pour l'isolation des anomalies (Liu et al., 2008)..... 70
Figure 3.2 : Anomalies détectés à Avenue Foch..... 71
Figure 3.3 : Diagramme pour l'établissement de profile type journalier des concentrations en particules (adapté de Rakotonirinjanahary et al. 2024)..... 73
Figure 3.4 : Profils moyens journaliers initiaux de la concentration de PM_{10} pour les jours ouvrés (à gauche) et les weekends (à droite) à Avenue Foch..... 74
Figure 3.5 : (a) Comparaison des profils : profil moyen initial (ligne rouge pointillée) et profil du 10 novembre 2022 (ligne noire pointillée), (b) Modèle de régression linéaire, basé sur chaque mesure de PM du profil journalier, pour le profil moyen de PM_{10} et le profil de PM_{10} de ce jour-là (Rakotonirinjanahary et al., 2024)..... 75
Figure 3.6 : Variation temporelle du coefficient de détermination (R^2) pour les PM_{10} (en gris) et les $PM_{2.5}$ (en vert) à la gare d'Avenue Foch : (a) jours ouvrés pour PM_{10}, (b) weekends pour PM_{10}, (c) jours ouvrés pour $PM_{2.5}$, et (d) weekends pour $PM_{2.5}$ 76
Figure 3.7 : (a) Comparaison des profils : profil moyen initial (ligne rouge pointillée) et profil du 2 novembre 2022 (ligne noire pointillée), (b) Modèle de régression linéaire, basé sur chaque mesure de PM du profil journalier, pour le profil moyen de PM_{10} et le profil de PM_{10} de ce jour-là..... 76
Figure 3.8 : Profil journalier type obtenu à la gare de Grigny Centre durant les jours ouvrés..... 78
Figure 3.9 : Illustration du principe de la fenêtre glissante pour l'échantillonnage progressif des données (adapté de Polanco-Martínez 2019)..... 79
Figure 3.10 : Architecture interne d'une cellule LSTM (Fan et al., 2020)..... 86
Figure 3.11 : Évaluation de la performance du modèle LSTM : (a) Précision évaluée par R^2, (b) RMSE ($\mu g/m^3$). Les lignes bleues indiquent la performance sur les ensembles d'entraînement, tandis que les lignes vertes représentent la performance sur les ensembles de test (Article 2, en annexe) 89
Figure 4.1 : Évolution temporelle du DAC (points bleus) et des concentrations originales de PM_{10} (points gris) mesurées à la gare de (a) Avenue Foch (2018-2022), (b) Sevran Beaudottes (2018-2021) et (c) Magenta (2017-2021)..... 92
Figure 4.2 : Évolution temporelle du DAC (points bleus) et des concentrations originales de $PM_{2.5}$ (points verts) mesurées à la gare de (a) Avenue Foch (2018-2022), (b) Sevran Beaudottes (2018-2021) et (c) Magenta (2017-2021)..... 93

Figure 4.3 : Évolution mensuelle du DAC moyen des PM_{10} (gris) et $PM_{2.5}$ (vert) dans les gares d'Avenue Foch (a, b), Sevran Beaudottes (c, d) et Magenta (e, f), comparée à celle des PM_{10} et $PM_{2.5}$ (violet) de leurs stations extérieures respectives : Porte d'Auteuil (a, b), Bobigny (c, d) et Paris 18ème (e, f)..... 95
Figure 4.4 : (a) Évolution temporelle du ratio $PM_{2.5}/PM_{10}$ à la gare d'Avenue Foch et sa distribution selon différentes échelles temporelles à Avenue Foch : (b) par année, (c) par mois, (d) par jour de la semaine et (e) par période de la journée 97
Figure 4.5 : (a) Évolution temporelle du ratio $PM_{2.5}/PM_{10}$ à la gare de Sevran Beaudottes et sa distribution selon différentes échelles temporelles à Sevran Beaudottes : (b) par année, (c) par mois, (d) par jour de la semaine et (e) par période de la journée 98
Figure 4.6 : (a) Évolution temporelle du ratio $PM_{2.5}/PM_{10}$ à la gare de Magenta et sa distribution selon différentes échelles temporelles à Magenta : (b) par année, (c) par mois, (d) par jour de la semaine et (e) par période de la journée 99
Figure 4.7 : Analyse du temps de déposition nocturne des concentrations de PM_{10} (en gris) et $PM_{2.5}$ (en vert) dans les gares de (a) Avenue Foch, (b) Sevran Beaudottes et (c) Magenta. Les lignes pointillées représentent les ajustements exponentiels des courbes de concentration. 100
Figure 4.8 : Profil temporel des concentrations en PM_{10} sur une semaine (18/04/2022 – 25/04/2022) à la gare d'Avenue Foch..... 101
Figure 4.9 : Profils types des PM_{10} pour les jours de semaine (à gauche) et les weekends (à droite) dans les gares de référence : (a) Avenue Foch, (b) Sevran Beaudottes et (c) Magenta. La zone hachurée indique la période d'exploitation des trains. Les zones ombrées représentent un écart type (σ=1) 102
Figure 4.10 : Profils types des $PM_{2.5}$ pour les jours de semaine (à gauche) et les weekends (à droite) dans les gares de référence : (a) Avenue Foch, (b) Sevran Beaudottes et (c) Magenta. La zone hachurée indique la période d'exploitation des trains. Les zones ombrées représentent un écart type (σ=1) 103
Figure 4.11 : Profils types des PM_{10} pour les jours de semaine (à gauche) et les weekends (à droite) dans les stations de mesure extérieur : (a) Porte d'Auteuil, (b) Bobigny et (c) Paris 18ème. Les zones ombrées représentent un écart type (σ=1)..... 106
Figure 4.12 : Comparaison des concentrations moyennes de PM_{10} et $PM_{2.5}$ entre les gares de référence et leurs stations extérieures associées : (a) Avenue Foch et Porte d'Auteuil, (b) Sevran Beaudottes et Bobigny, et (c) Magenta et Paris 18ème, pour les jours ouvrés (à gauche) et les weekends (à droite). 107
Figure 4.13 : Profils journaliers et écarts-types des concentrations en PM_{10} pour les jours ouvrés et les weekends dans les stations temporaires (campagnes de 2 à 3 semaines) 108
Figure 4.14 : Concentrations normalisées en PM_{10} (jours ouvrés et weekends) pour les différentes EFS 110
Figure 4.15 : Moyenne glissante de l'erreur absolue moyenne (points en bleu, avec des barres d'erreur correspondant à un écart-type) accompagnée de l'évolution glissante de l'écart-type (ligne continue en olive).....112
Figure 4.16 : Variation des profils types pour (a) les jours ouvrés et (b) les weekends au fil du temps. Les zones ombrées représentent un écart-type (σ=1). 113
Figure 5.1 : Effet des facteurs d'influence sur les concentrations de PM_{10} dans le modèle GAM à la gare d'Avenue Foch 117
Figure 5.2 : Effet des facteurs d'influence sur les concentrations de $PM_{2.5}$ dans le modèle GAM à la gare d'Avenue Foch 119
Figure 5.3 : Effet des facteurs d'influence sur les concentrations de PM_{10} dans le modèle GAM à la gare de Sevran Beaudottes 121
Figure 5.4 : Effet des facteurs d'influence sur les concentrations de $PM_{2.5}$ dans le modèle GAM à la gare de Sevran Beaudottes 123
Figure 5.5 : Effet des facteurs d'influence sur les concentrations de PM_{10} dans le GAM à la gare de Magenta..... 125
Figure 5.6 : Effet des facteurs d'influence sur les concentrations de $PM_{2.5}$ dans le GAM à la gare de Magenta..... 127

Figure 5.7 : Évolution des ratios DAC pour PM_{10} et $PM_{2.5}$: (a) PM_{10}, expérimentation 1 ; (b) PM_{10}, expérimentation 2 ; (c) $PM_{2.5}$, expérimentation 1 ; et (d) $PM_{2.5}$, expérimentation 2..... 129
Figure 5.8 : Comparaison des concentrations mesurées de PM (ligne bleue) et prédites par le modèle LSTM (ligne rouge) : (a) PM_{10}, expérimentation 1 ; (b) PM_{10}, expérimentation 2 ; (c) $PM_{2.5}$, expérimentation 1 ; (d) $PM_{2.5}$, expérimentation 2. 130
Figure 5.9 : Evolution temporelle des PM_{10} (points gris) et des PM2.5 (puces vertes) 131
Figure 5.10 : Évolution des ratios DAC pour (a) PM_{10} et (b) $PM_{2.5}$ entre 2018-2019 et 2020-2021 .. 132

Figure A. 1 : Évolution temporelle des concentrations en PM_{10} et $PM_{2.5}$ mesurées à la gare de (a) Avenue Foch entre 2018 et 2022, (b) Sevran Beaudottes entre 2018 et 2021 et (c) Magenta entre 2017 et 2021 mettant en évidence les périodes notables.....XLIV
Figure A. 2 : Variation temporelle de R^2 pour les PM_{10} (en gris) et les $PM_{2.5}$ (en vert) à la gare de Sevran-Beaudottes : (a) jours ouvrés pour PM_{10}, (b) weekends pour PM_{10}, (c) jours ouvrés pour $PM_{2.5}$, et (d) weekends pour $PM_{2.5}$.....XLV
Figure A. 3 : Variation temporelle de R^2 pour les PM_{10} (en gris) et les $PM_{2.5}$ (en vert) à la gare de Magenta : (a) jours ouvrés pour PM_{10}, (b) weekends pour PM_{10}, (c) jours ouvrés pour $PM_{2.5}$, et (d) weekends pour $PM_{2.5}$.....XLV
Figure A. 4 : Profils journaliers types des concentrations de PM_{10} du lundi au vendredi (à gauche) et du samedi au dimanche (à droite) dans les gares avec mesures longue durée : (a) Avenue Foch, (b) Sevran Beaudottes et (c) Magenta. La zone hachurée indique la période d'exploitation des trains. Les zones ombrées représentent un écart type (σ=1) sur les différents jours.....XLVII
Figure A. 5 : Profils typiques des concentrations de $PM_{2.5}$ du lundi au vendredi (à gauche) et du samedi au dimanche (à droite) dans les gares avec mesures longue durée : (a) Avenue Foch, (b) Sevran Beaudottes et (c) Magenta. La zone hachurée indique la période d'exploitation des trains. Les zones ombrées représentent un écart type (σ=1) sur les différents jours.....XLVIII
Figure A. 6 : Profils journaliers et écarts-types des concentrations en $PM_{2.5}$ pour les jours ouvrés et les weekends dans les stations temporaires (campagnes de 2 à 3 semaines).....XLIX
Figure A. 7 : Analyse du temps de déposition nocturne des concentrations de $PM_{2.5\text{-}10}$ (en gris) et $PM_{0\text{-}2.5}$ (en vert) dans les gares de (a) Avenue Foch, (b) Sevran Beaudottes et (c) Magenta. Les lignes pointillées représentent les ajustements exponentiels des courbes de concentration XLIX
Figure A. 8 : Profils typiques des concentrations de PM_{10} du lundi au vendredi (à gauche) et du samedi au dimanche (à droite) dans les stations de mesure extérieur : (a) Porte d'Auteuil, (b) Bobigny et (c) Paris 18ème. Les zones ombrées représentent un écart type (σ=1) LI
Figure A. 9 : Profils types des $PM_{2.5}$ pour les jours de semaine (à gauche) et les weekends (à droite) dans les stations de mesure extérieur : (a) Porte d'Auteuil et (b) Bobigny. Les zones ombrées représentent un écart type (σ=1) LII
Figure A. 10 : Profils types des $PM_{2.5}$ pour les jours de semaine (à gauche) et les weekends (à droite) dans les stations de mesure extérieur : (a) Porte d'Auteuil et (b) Bobigny. Les zones ombrées représentent un écart type (σ=1) LIII

TABLE DES TABLEAUX

Tableau 1.1 : Concentrations dans l'air en EFS recommandées a minima (ANSES, 2022) 25
Tableau 1.2 : Résumé des principales sources de particules dans les EFS 30
Tableau 1.3 : Principaux facteurs d'influences régissant les concentrations en particules souterraines (d'après Chang et al. 2021) 39
Tableau 2.1 : Spécifications techniques des analyseurs TEOM (https://www.thermofisher.com/) 51
Tableau 2.2 : Spécifications techniques de l'analyseur de CO_2 (https://www.thermofisher.com/) 51
Tableau 2.3 : Spécification technique du capteur analogique de température et d'humidité relative ... 53
Tableau 2.4 : Caractéristiques des gares instrumentées pour les campagnes de mesures de courte durée. 57
Tableau 2.5 : Caractéristiques principales de la gare d'Avenue Foch 59
Tableau 2.6 : Statistiques descriptives des données collectées à Avenue Foch durant la période étudiée 59
Tableau 2.7 : Caractéristiques principales de la gare de Sevran-Beaudottes 62
Tableau 2.8 : Statistiques descriptives des données collectées à Sevran-Beaudottes durant la période étudiée 62
Tableau 2.9 : Caractéristiques principales de la gare de Magenta 64
Tableau 2.10 : Statistiques descriptives des données collectées à Magenta durant la période étudiée.. 65
Tableau 2.11 : Synthèse des dispositifs de mesure et des paramètres mesurés dans les gares de référence 67
Tableau 2.12 : Synthèse des paramètres mesurés dans les stations de surveillance de l'air extérieur... 67
Tableau 3.1 : Résumé du nettoyage des données et proportion de points supprimés aux gares Avenue Foch, Sevran Beaudottes et Magenta 72
Tableau 3.2 : Périodes de mesure et nombre de profils journaliers retenus pour les jours ouvrés et les weekends dans les gares avec des mesures de courte durée. 78
Tableau 3.3 : VIF des variables explicatives dans le GAM pour Avenue Foch (Annexe 2) 82
Tableau 3.4 : VIF des variables explicatives dans le GAM pour Sevran-Beaudottes 82
Tableau 3.5 : VIF des variables explicatives dans le GAM pour Magenta 83
Tableau 4.1 : DAC mensuel à la gare d'Avenue Foch (Rakotonirinjanahary et al., 2024).....110
Tableau 5.1 : Résultats des termes paramétriques (linéaires) du modèle GAM pour les concentrations de PM_{10} à Avenue Foch 116
Tableau 5.2 : Résultats des termes lisses (non linéaires) du modèle GAM pour les concentrations de PM_{10} à Avenue Foch 116
Tableau 5.3 : Résultats des termes paramétriques (linéaires) du modèle GAM pour les concentrations de $PM_{2.5}$ à Avenue Foch 118
Tableau 5.4 : Résultats des termes lisses (non linéaires) du modèle GAM pour les concentrations de $PM_{2.5}$ à Avenue Foch 119
Tableau 5.5 : Résultats des termes paramétriques (linéaires) du modèle GAM pour les concentrations de PM_{10} à Sevran-Beaudottes 120
Tableau 5.6 : Résultats des termes lisses (non linéaires) du modèle GAM pour les concentrations de PM_{10} à Sevran-Beaudottes 121
Tableau 5.7 : Résultats des termes paramétriques (linéaires) du modèle GAM pour les concentrations de $PM_{2.5}$ à Sevran Beaudottes 122
Tableau 5.8 : Résultats des termes lisses (non linéaires) du modèle GAM pour les concentrations de $PM_{2.5}$ à Sevran Beaudottes 123
Tableau 5.9 : Résultats des termes paramétriques (linéaires) du modèle GAM pour les concentrations de PM_{10} à Magenta 124

Tableau 5.10 : Résultats des termes lisses (non linéaires) du modèle GAM pour les concentrations de PM_{10} à Magenta 125
Tableau 5.11 : Résultats des termes paramétriques (linéaires) du modèle GAM pour les concentrations de $PM_{2.5}$ à Magenta 126
Tableau 5.12 : Résultats des termes lisses (non linéaires) du modèle GAM pour les concentrations de $PM_{2.5}$ à Magenta 127
Tableau 5.13 : Taux de réductions de PM_{10} et $PM_{2.5}$ (%) évalués par les trois méthodes 131
Tableau 5.14 : Taux de réductions de PM_{10} et $PM_{2.5}$ (%) évalués par les deux méthodes : DAC et GAM 133

NOMENCLATURE

A	
AEF	Agence d'Essai Ferroviaire
ANSES	Agence Nationale Sécurité Sanitaire Alimentaire Nationale
AVPDs	Adjustable Vent Platform Doors
B	
BTEX	Benzène Toluène Ethylbenzène Xylènes
C	
CFD	Computational Fluid Dynamics
CLA	Couche Limite Atmosphérique
COV	Composés Organiques Volatils
D	
DAC	Daily Amplitude Coefficient
DOW	Day Of Week
DSG	Direction Sécurité Groupe
E	
EDF	Effective Degrees of Freedom
EFS	Enceintes Ferroviaires Souterraines
EOLE	Est Ouest Liaison Express
F	
FDMS	Filter Dynamic Measurement System
FPBDs	Full-height Platform Bailout Doors
G	
GAM	Generalized Additive Models
H	
HAP	Hydrocarbures Aromatiques Polycycliques
HEPA	High Efficiency Particulate Air
HPBDs	Half-height Platform Bailout Doors
I	
IA	Intelligence Artificielle
INERIS	Institut National de l'Environnement Industriel et des RISques
ISO	International Organization for Standardization
J	
JO	Jours ouvrés
L	
LCD	Liquid Crystal Display
LSTM	Long short-term memory
M	
MAE	Mean Absolute Error
MI	Matériel d'Interconnexion
N	
NF	Norme Française
O	
OMS	Organisation Mondiale de la Santé
OPE	Over-Platform Exhaust
P	
PC	Personal Computer
PM	Particulate Matter

PMF	Positive Matrix Factorization
PSDs	Platform Screen Doors
R	
RATP	Régie Autonome des Transports Parisiens
RER	Réseau Express Régional
RH	Relative Humidity
RMSE	Root Mean Square Error
S	
SNCF	Société Nationale des Chemins de Fer Français
T	
TEI	Thermo Environmental Instruments
TEOM	Tapered Element Oscillating Microbalance
U	
UPE	Under-Platform Exhaust
URS	Underground Railway Stations
UTC	Universal Time Coordinated
V	
VIF	Variance Inflation Factor
W	
WE	Weekend
WHO	World Health Organization
WMAPE	Weighted Mean Absolute Percentage Error

INTRODUCTION

Les trains figurent aujourd'hui parmi les modes de transport les plus utilisés et s'imposent progressivement comme un élément central des déplacements en milieu urbain. Avec l'augmentation de leur fréquentation, les usagers passent désormais davantage de temps dans les infrastructures ferroviaires, notamment dans les enceintes ferroviaires souterraines (EFS). Selon l'ANSES (2022)[1], ces espaces désignent l'ensemble des espaces couverts situés en dessous du niveau du sol, connectés à une voie de transport ferroviaire souterraine, et fréquentés par les usagers du transport ferroviaire en souterrain. Ils regroupent des stations, des gares, des couloirs, des rames et des locaux commerciaux. Au cours des trente dernières années, la qualité de l'air dans les EFS a fait l'objet d'une attention croissante, en raison de leurs caractéristiques structurelles et des sources internes de pollution. Les concentrations de particules y sont souvent bien supérieures à celles mesurées en extérieur, pouvant être 3 à 10 fois plus élevées, voire davantage (Cartenì et al., 2015; Mohsen et al., 2018; Tu & Olofsson, 2021, 2024; Karim et al., 2024). Ces niveaux dépassent fréquemment les seuils recommandés par l'Organisation mondiale de la santé (OMS, ou WHO en anglais pour World Health Organization, 2021), bien que ces valeurs de référence s'appliquent principalement à l'air ambiant extérieur. Bien que les passagers ne restent que peu de temps dans ces espaces, l'exposition de courte durée (en moyenne 7 minutes selon Marcus, 2011) à des niveaux élevés de polluants atmosphériques peut avoir des impacts sur la santé. Ces effets concernent notamment le système cardiovasculaire, avec des risques accrus d'infarctus du myocarde, d'insuffisance cardiaque (Renzi et al., 2020; Luo et al., 2021b; Wyatt et al., 2022), ainsi que le système respiratoire, pouvant entraîner des arythmies, des thromboses veineuses profondes, des accidents vasculaires cérébraux, de l'asthme, des bronchiolites et d'autres infections respiratoires (Lei et al., 2019; Zhang et al., 2020; Chen et al., 2021; Pini et al., 2021; Sun et al., 2021). Cependant, il n'existe pas encore de consensus scientifique sur la dangerosité spécifique des particules présentes dans les EFS, ni de réglementation clairement définie pour encadrer la qualité de l'air dans ces environnements. Les pays se réfèrent souvent aux recommandations de l'OMS, bien que celles-ci ne soient pas spécifiquement adaptées à ces espaces confinés.

Face à ces enjeux, la SNCF (Société Nationale des Chemins de Fer Français), en collaboration avec AIRPARIF[2] et AEF[3] (Agence d'Essai Ferroviaire), a lancé en 2016 un programme ambitieux pour évaluer et surveiller la qualité de l'air dans 24 gares souterraines. Ces gares ont été instrumentées pendant quelques semaines, et trois d'entre elles ont été équipées d'instruments de mesure en continu, fournissant ainsi une base de données exceptionnelle couvrant environ cinq ans, une durée inédite dans la littérature. Ce programme vise à documenter finement les niveaux de particules dans les gares franciliennes exploitées par la SNCF, afin de guider l'élaboration de plans d'amélioration et de prioriser les travaux nécessaires. Parallèlement, des expérimentations ont été menées pour tester l'efficacité de différents systèmes de filtration de l'air.

Dans le cadre de cette thèse, nous cherchons à analyser ces données pour caractériser les concentrations de particules à court et long terme en EFS, ce qui approfondira notre compréhension des variations temporelles des concentrations de particules en développant une méthodologie d'analyse de données rigoureuse et robuste.

[1] ANSES : Agence Nationale de Sécurité Sanitaire de l'alimentation, de l'environnement et du travail

[2] AIRPARIF : Association agréée pour la surveillance de la qualité de l'air en Île-de-France. Membre des AASQA (Associations Agréées Surveillance Qualité de l'Air), organismes indépendants chargés de surveiller et d'informer sur la qualité de l'air à l'échelle régionale, conformément aux réglementations nationales et européennes.

[3] AEF : Laboratoire de la SNCF situé à Vitry-sur-Seine, spécialisé dans les essais et expertises sur le matériel roulant. Créée en 1999, AEF hérite d'une tradition débutée en 1933 avec le « banc des locomotives », devenu ensuite le centre d'essai de Vitry.

En effet, les concentrations de particules dans les EFS présentent des variations significatives, à la fois d'une gare à l'autre et d'un jour à l'autre (Kam et al., 2011; Barmparesos et al., 2016). Les études portant sur la qualité de l'air dans ces espaces se limitent généralement à des périodes allant de quelques semaines à une année. Par conséquent, les variations à long terme des concentrations de particules restent mal documentées. En ce qui concerne les variations à court terme, un profil diurne distinct a été identifié, avec des différences marquées entre les jours ouvrés et les weekends, contrastant fortement avec les tendances observées en extérieur. Birenzvige et al. (2003), Raut et al. (2009), Querol et al. (2012a) et Walther and Bogdan (2017) ont mis en évidence des cycles diurnes différents entre les jours de semaine et les weekends. Ils ont également montré que le profil journalier des concentrations de particules et la temporalité du cycle diurne sont remarquablement cohérents d'un jour à l'autre, malgré des fluctuations quotidiennes des niveaux de concentration. À notre connaissance, aucune étude ne s'est penchée sur l'établissement d'un profil journalier de référence des concentrations de particules reflétant les tendances des jours ouvrés et des weekends. La littérature se contente souvent de calculer des moyennes, sans approfondir l'analyse des variations temporelles. Pour combler cette lacune, nous avons développé une approche méthodologique permettant d'établir des profils journaliers types des concentrations de particules, tout en examinant les facteurs influençant ces concentrations dans les EFS franciliennes, y compris l'impact des systèmes de filtration.

Ce manuscrit s'articule autour de cinq chapitres, chacun apportant une contribution essentielle à la compréhension des dynamiques de pollution particulaire dans les EFS.

Le premier chapitre dresse un état de l'art complet sur les sources de particules dans les EFS, les facteurs influençant leurs concentrations et les solutions existantes pour améliorer la qualité de l'air. Il explore la variabilité spatiale et temporelle des particules, ainsi que les stratégies passives et actives, telles que la purification de l'air et les systèmes de ventilation. Ce chapitre pose les bases théoriques nécessaires pour les analyses empiriques et méthodologiques développées dans les chapitres suivants.

Le deuxième chapitre présente les dispositifs de mesure et les sites d'étude utilisés pour évaluer la qualité de l'air dans les EFS. Il décrit en détail les instruments de mesure des particules, du CO_2, de la température et de l'humidité relative, ainsi que les technologies de filtration testées. Les gares étudiées, incluant les gares de référence (Avenue Foch, Sevran Beaudottes, Magenta) et les stations de mesure de l'air extérieur, sont également présentées. Ce chapitre fournit les fondements expérimentaux qui sont analysés dans ce travail de thèse.

Le troisième chapitre est consacré aux méthodologies d'analyse des données. Il commence par exposer le prétraitement et le nettoyage des données, avant de présenter trois approches distinctes. La première est la méthode basée sur le Coefficient d'Amplitude Journalier (DAC), que nous avons développée, et qui permet de quantifier les variations intra-journalières tout en évaluant les performances des systèmes de filtration. La deuxième approche est le Modèle Additif Généralisé (GAM), qui permet de modéliser les relations complexes entre les concentrations de particules et les facteurs influençant leur variation. Enfin, la troisième méthode est le Réseau de Neurones Récurrents à Mémoire à Long et Court Terme (LSTM), utilisé pour la prédiction à court terme des concentrations de particules et pour l'évaluation en temps réel des systèmes de filtration.

Le quatrième chapitre applique ces méthodologies pour analyser les concentrations de particules dans les EFS. Il explore la variabilité et les tendances à long terme dans les gares de référence, les profils journaliers des concentrations de particules, ainsi que les spécificités des gares avec des mesures de courte durée. Une attention particulière est portée à l'incertitude des profils journaliers en fonction de la durée des mesures, fournissant des recommandations pour la collecte de données futures.

Le cinquième chapitre approfondit l'analyse en identifiant les principaux facteurs influençant les concentrations de particules dans les EFS, à l'aide du modèle GAM. Il évalue également l'efficacité des systèmes de filtration testés dans les gares d'Avenue Foch et de Sevran Beaudottes, en utilisant les

méthodes DAC et LSTM. Les résultats mettent en lumière les performances des différentes technologies de filtration et leurs variations en fonction des conditions locales.

Enfin, la conclusion générale synthétise les principaux résultats de l'étude et met en évidence les contributions scientifiques et pratiques. Elle formule des recommandations pour améliorer la qualité de l'air dans les EFS et ouvre des perspectives pour de futurs travaux de recherche, notamment l'exploration de nouvelles technologies de filtration, l'étude des particules ultrafines, l'identification de facteurs non abordés dans ce travail de thèse, ainsi que l'intégration de modèles d'intelligence artificielle avancés.

Ce projet de thèse CIFRE (ANRT n°2021/0856), en collaboration entre la SNCF et l'Université de Lille, bénéficie du soutien d'AIRPARIF et d'AEF pour la fourniture des données, ainsi que de l'expertise d'AREP en matière de traitement de données et de qualité de l'air en EFS.

CHAPITRE 1. ÉTAT DE L'ART

Au cours des trois dernières décennies, l'intérêt pour la recherche sur la qualité de l'air dans les enceintes ferroviaires souterraines (EFS) s'est nettement intensifié. Cette attention accrue s'explique par les caractéristiques particulières de ces espaces : confinement, faible renouvellement de l'air et forte fréquentation, avec des affluences pouvant atteindre jusqu'à 4 millions de passagers par an (AEF, 2022b). Une revue récente par Wang et al. (2024) montre une tendance croissante depuis les années 2000 (Figure 1.1), avec une augmentation du nombre d'études annuelles consacrées à ce sujet de 2000 à 2024. Ce domaine suscite particulièrement l'intérêt en Asie, en Europe et en Amérique du Nord, qui dominent les contributions scientifiques sur la qualité de l'air dans les EFS.

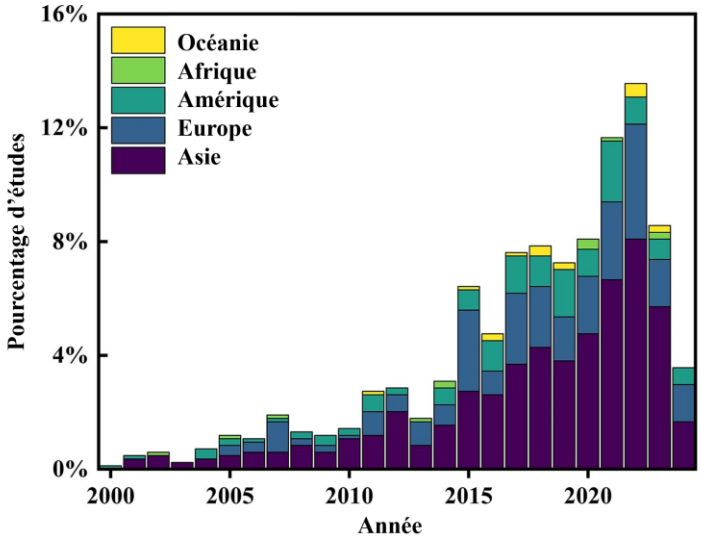


Figure 1.1 : Pourcentage d'études sur la qualité de l'air dans les EFS, réparties par continents, de 2000 à 2024 (Wang et al., 2024).

Selon Passi et al. (2021) et Yang et al. (2023), les principaux polluants générés à l'intérieur des EFS sont les particules en suspension (PM, pour Particulate Matter en anglais), le dioxyde de carbone (CO_2), l'ozone (O_3), les oxydes d'azote (NOx), le monoxyde de carbone (CO), les composés organiques volatils (COV) et les bioaérosols. Ces derniers sont des particules biologiques en suspension, composées de poussières, de microbes et de leurs fragments (Humbal et al., 2018). Ces polluants, bien que communs à d'autres environnements, présentent des dynamiques particulières dans les EFS en raison des interactions complexes entre les activités ferroviaires, les infrastructures souterraines et la ventilation souvent limitée.

De nombreuses études ont examiné la qualité de l'air dans les EFS, mettant particulièrement l'accent sur les PM. Ce travail de thèse s'inscrit dans cette lignée et se focalisera spécifiquement sur l'étude des PM. La Figure 1.2 illustre les concentrations de particules mesurées dans différentes villes du monde de 1999 à 2024. Les études portant sur la qualité de l'air dans les EFS se concentrent principalement sur les fractions PM_{10} et $PM_{2.5}$, tandis que peu de travaux s'intéressent à la fraction PM_1. En effet, selon l'ANSES (2015), moins de 5 % des études prennent en compte la fraction PM_1. Comme l'indique la Figure 1.2, les concentrations en PM_{10} et $PM_{2.5}$ sont souvent supérieures aux seuils recommandés par l'Organisation Mondiale de la Santé (WHO, 2021). Bien que ces seuils concernent principalement l'air extérieur, ils offrent une référence utile pour évaluer les niveaux de pollution dans les EFS.

Les comparaisons des différentes études présentées sur la Figure 1.2 sur plusieurs continents indiquent que les concentrations de particules les plus élevées sont observées en Europe, notamment dans des villes comme Stockholm, Londres et Athènes, où les climats présentent des étés chauds et des

hivers modérés. En Asie, des niveaux élevés sont également mesurés dans des villes telles que Pékin et Séoul, qui connaissent des étés chauds et humides, suivis d'hivers froids et secs. À l'inverse, les niveaux les plus bas ont été relevés en Amérique, en particulier à Buenos Aires, caractérisée par un climat subtropical (Passi et al., 2021a).

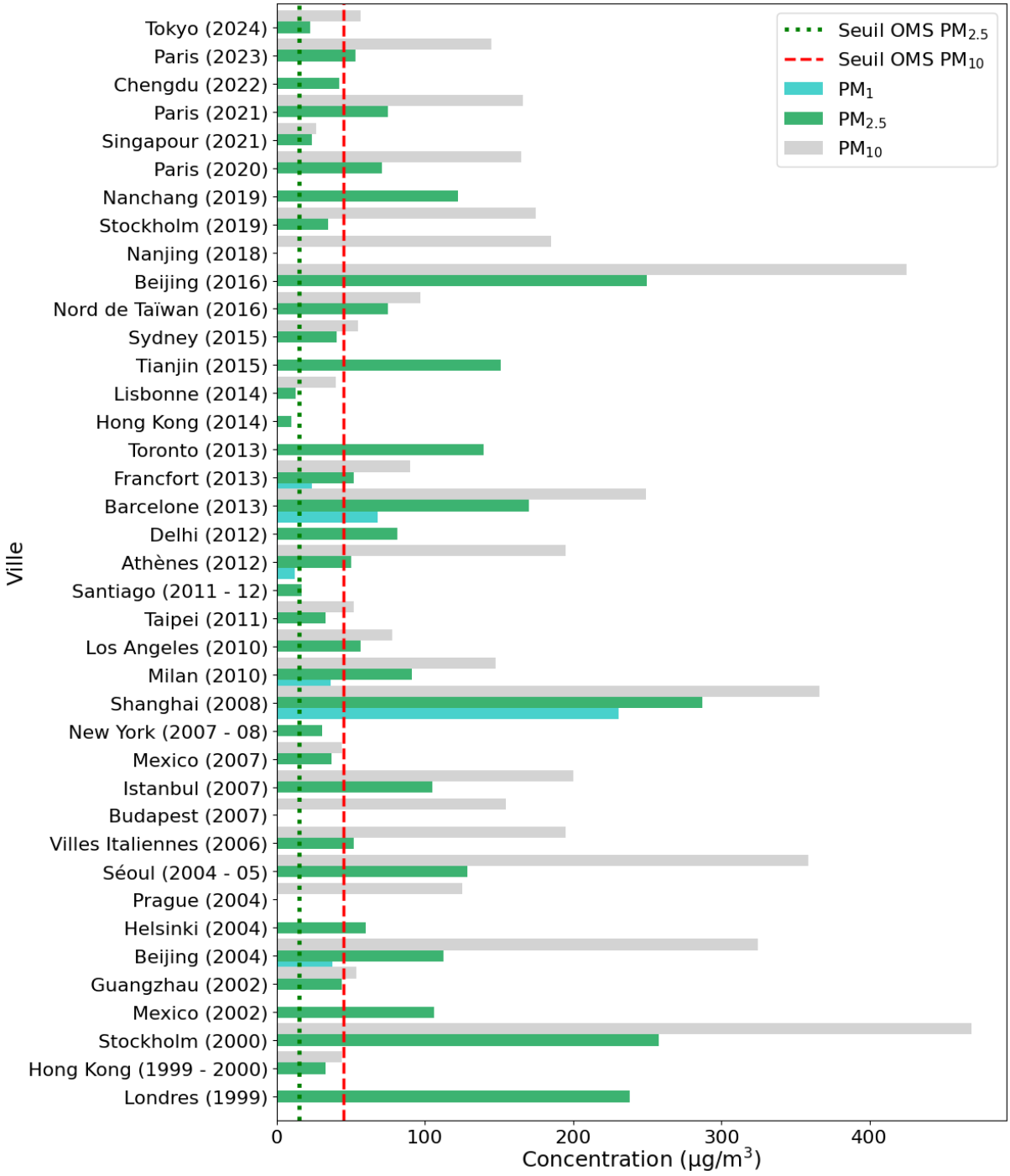


Figure 1.2 : Concentrations en particules dans les EFS dans différentes villes du monde (adapté et mis à jour de Passi et al. 2021)

La plupart de ces études ont révélé que les concentrations de particules dans les EFS peuvent être de 3 à 10 fois supérieures à celles observées à l'extérieur (Cartenì et al., 2015; Mohsen et al., 2018; Tu & Olofsson, 2021, 2024; Karim et al., 2024), voire atteignant des niveaux 5 à 30 fois supérieurs selon les travaux menés par Raut et al. (2009) à Paris. Par ailleurs, la majorité des recherches, réalisées notamment à Nanjing, Chine (Mao et al., 2019); Athènes, Grèce (Barmparesos et al., 2016); Taipei,

Taïwan (Cheng et al., 2012); Séoul, Corée du Sud (Kim et al., 2008; Lee et al., 2018); Los Angeles, États-Unis (Kam et al., 2011); Toronto, Canada (Van Ryswyk et al., 2017); Naples, Italie (Cartenì et al., 2015); et Sydney, Australie (Mohsen et al., 2018), indiquent que les concentrations de particules sont généralement plus élevées sur les quais que dans les trains ou métros eux-mêmes. Cependant, il y a des exceptions, comme les études menées à Prague, République Tchèque (Braniš, 2006); Séoul, Corée du Sud (Kwon et al., 2010); et Lisbonne, Portugal (Ramos et al., 2015). Passi et al. (2021) et Wang et al. (2024) proposent une revue exhaustive des études portant sur les concentrations en particules mesurées sur les quais des EFS à travers le monde, offrant ainsi une vue d'ensemble des niveaux de pollution particulaire dans ces environnements spécifiques.

À ce jour, en France, aucun texte réglementaire ne régit spécifiquement la surveillance de la qualité de l'air dans les EFS. Cependant, en s'appuyant sur les valeurs limites de la directive 2008/50/CE et les recommandations de l'OMS (WHO, 2021), l'ANSES (2022) a formulé des recommandations pour limiter l'exposition des usagers. Dans son rapport de 2022, le groupe de travail de l'ANSES préconise de poursuivre les efforts pour améliorer la qualité de l'air intérieur dans les EFS et de veiller à ne pas dépasser les valeurs indiquées dans le Tableau 1.1. Les durées maximales journalières en EFS ont été fixées à 30 minutes, 1 heure, 1 heure 30 ou 2 heures, sur la base de scénarios de fréquentation des usagers, établis à partir des données disponibles des réseaux ferroviaires souterrains en Île-de-France, Toulouse et Rennes (ANSES, 2022).

Tableau 1.1 : Concentrations dans l'air en EFS recommandées a minima (ANSES, 2022)

Polluant	Durée cumulée de fréquentation de l'EFS sur une journée	Concentration calculée à partir de la valeur limite journalière de la directive européenne 2008/50/CE (µg/m^3)	Concentration calculée à partir de la valeur guide journalière de l'OMS (µg/m^3)
PM_{10}	2h	260	80
	1h30	330	100
	1h	480	140
	30min	940	250
$PM_{2.5}$	2h	Sans objet	50
	1h30		60
	1h		80
	30min		140

En 2024, s'appuyant sur les recommandations de l'ANSES, AIRPARIF a publié une cartographie résumant les niveaux de concentration de particules dans différentes EFS d'Île-de-France, illustrée dans la Figure 1.3. Une version interactive est disponible sur les sites d'AIRPARIF (Qualité de l'air dans les métros, RER et Transiliens | Airparif) et d'Île-de-France Mobilités (Qualité de l'Air dans le réseau de transport francilien — Open Data Île-de-France Mobilités (iledefrance-mobilites.fr)). Cette cartographie vise à permettre à Île-de-France Mobilités et aux opérateurs (RATP et SNCF) de prioriser leurs actions pour améliorer la qualité de l'air dans ces espaces confinés. Elle repose sur un modèle statistique utilisant une technique de machine learning développée par AIRPARIF pour évaluer la pollution particulaire en fonction des valeurs de référence de l'ANSES. Le modèle compare les concentrations mesurées durant une heure, qui représentent les 5 % des niveaux les plus élevés de particules, aux normes recommandées par l'ANSES (Tableau 1.1). Cette classification s'appuie sur des campagnes de mesure de courte durée et des mesures continues effectuées entre 2015 et 2022 sur 44 quais de gares et stations (AIRPARIF, 2024), dont les gares gérées par la SNCF qui seront étudiées dans ce travail de thèse.

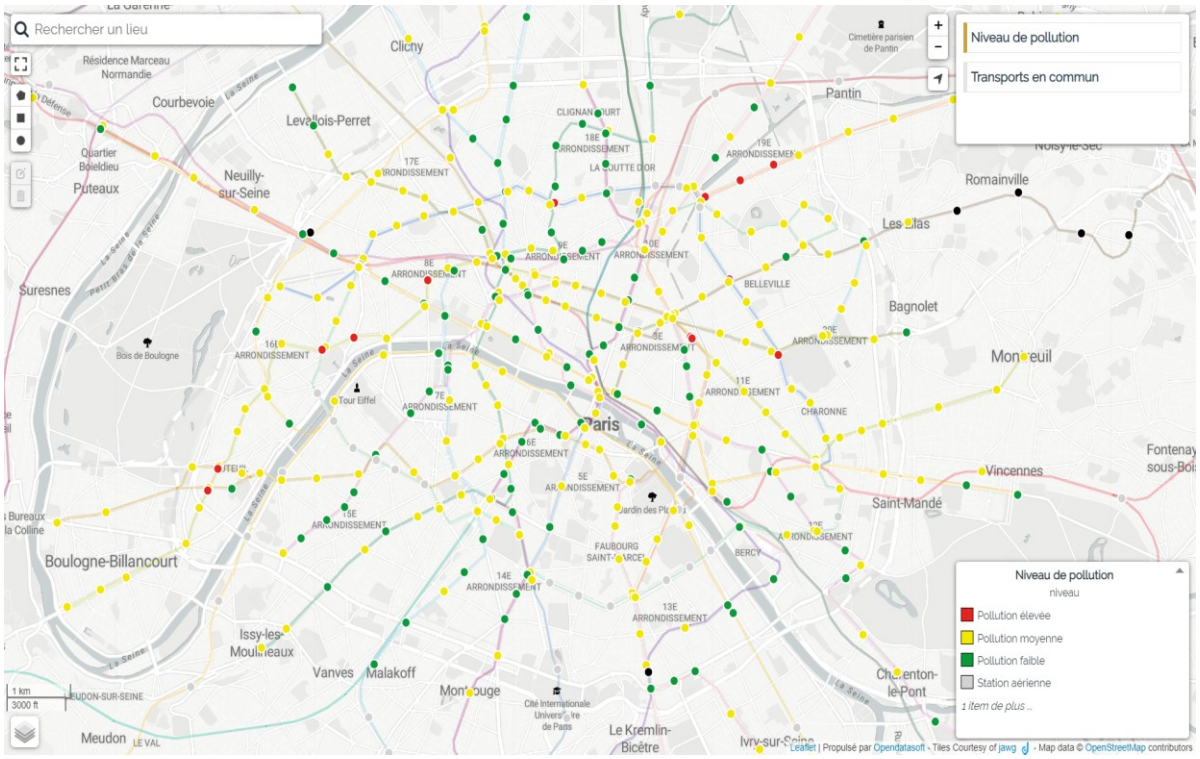


Figure 1.3 : Cartographie des niveaux de pollution dans les gares et stations d'Île-de-France (https://data.iledefrance-mobilites.fr/explore/dataset/qualite-de-lair-dans-le-reseau-de-transport-francilien/custom/)

Parmi les 426 quais de gares et stations mesurés, les niveaux de pollution aux particules, en fonction des valeurs de référence recommandées par l'ANSES, se répartissent comme suit :

- 13 quais de gares et stations présentent des niveaux de pollution élevés (supérieurs à 480 µg/m³ sur 1 heure),
- 276 quais de gares et stations présentent des niveaux modérés (compris entre 140 µg/m³ et 480 µg/m³ sur 1 heure),
- 123 quais de gares et stations présentent des niveaux faibles (inférieurs à 140 µg/m³ sur 1 heure),
- 14 quais de gares et stations n'ont pas pu être classés en raison d'un manque de documentation.

La suite de ce chapitre explore les multiples aspects liés à la pollution atmosphérique dans les EFS, en abordant à la fois les sources de polluants, leur variabilité, et les solutions pour améliorer la qualité de l'air. La première section est consacrée à l'identification des sources de polluants dans les EFS, avec un focus particulier sur les particules en suspension, dont la composition chimique et l'origine diffèrent considérablement de celles des particules issues de l'air extérieur. Les facteurs influençant les concentrations en particules des EFS sont ensuite analysés. Ces facteurs incluent la fréquence des trains, l'effet piston généré par leur passage, la saisonnalité, la profondeur des stations, l'environnement extérieur, ainsi que les systèmes de ventilation. Par ailleurs, la variabilité des concentrations de particules, tant dans l'espace que dans le temps (variabilité temporelle au niveau des quais), est examinée pour fournir des perspectives sur la dynamique de la pollution dans ces milieux confinés. Enfin, le chapitre se conclut par une revue des solutions actuelles visant à améliorer la qualité de l'air dans ces enceintes, telles que l'utilisation de matériaux non métalliques dans les systèmes de freinage, l'optimisation des systèmes de ventilation, l'installation de portes palières et de systèmes de filtration.

1.1 Sources de particules en EFS

Les particules en suspension dans les EFS se caractérisent par plusieurs paramètres physicochimiques, tels que leur taille, leur concentration, leur composition chimique et leur densité. La taille des particules peut varier entre 1,3 nm et 20 µm (Gustafsson et al., 2012; Cha & Olofsson, 2018), offrant une grande hétérogénéité de dimensions. Ces particules proviennent de multiples sources, tant internes qu'externes, qui contribuent de manière significative à la qualité de l'air en milieu clos (Figure 1.4).

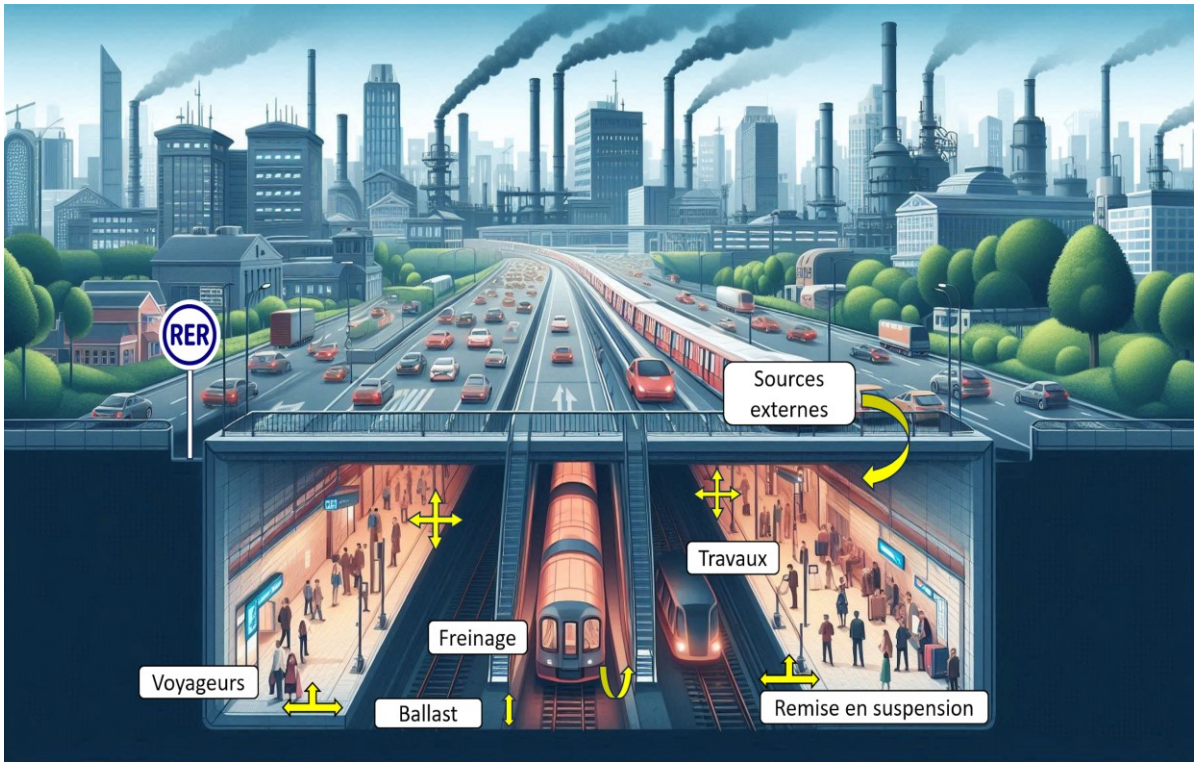


Figure 1.4 : Différentes sources de particules dans les EFS (adapté de PIXLR générateur d'images IA)

Les sources intérieures sont principalement liées aux opérations ferroviaires et aux activités humaines (Salma et al., 2007), comme illustré dans la Figure 1.4. Les activités ferroviaires, telles que le contact entre les roues et les rails (Ma et al., 2012; Colombi et al., 2013; Olofsson et al., 2013; Font et al., 2019), l'usure des freins (Moreno et al., 2015a; Namgung et al., 2017; Durand et al., 2021), ainsi que le contact entre les pantographes et les caténaires (Cha et al., 2016; Lee et al., 2018) constituent des contributeurs majeurs à la pollution particulaire. Ces activités génèrent des particules principalement par friction et usure.

Les émissions associées au contact roue-rail proviennent principalement de la friction mécanique, représentant entre 0,3 et 5,6 g/km d'émissions de PM_{10}, dont 87,5% sont des $PM_{2.5}$ (Fruhwirt et al., 2023b). Deux types de mouvements relatifs contribuent à l'usure : le roulement de la roue sur le rail et le glissement, ce dernier pouvant se produire dans les directions longitudinale et transversale. Cette friction mécanique entraîne l'usure des surfaces de contact et la libération de particules, en particulier des particules métalliques issues des matériaux constitutifs des rails et des roues (Olofsson et al., 2013).

En ce qui concerne les émissions liées aux systèmes de freinage, ces derniers subissent une dégradation importante lors des freinages. Ce processus libère une quantité significative de particules dans l'air ambiant, avec des tailles et des compositions chimiques variables selon les types de freins utilisés (Olofsson, 2011), ce qui explique leurs différences chimiques significatives. Selon les études de Fruhwirt et al. (2024), qui ont testé deux combinaisons différentes de disques de frein et de matériaux

de plaquettes de frein — une plaquette de frein frittée avec un disque de frein en acier et une plaquette de frein organique avec un disque de frein en fonte —, les émissions des plaquettes de frein organiques varient de 0,1 g/km à 4,2 g/km de PM_{10}, dont 56 % sont des $PM_{2.5}$. Pour les plaquettes de frein frittées, les émissions se situent entre 0,2 g/km et 1,0 g/km, avec 70 % de $PM_{2.5}$. Les travaux de Moreno et al. (2015a) ont révélé que les particules émises lors du freinage pneumatique se distinguent clairement par leur composition chimique et leur texture. Ces particules sont généralement carbonées et présentent une forte teneur en baryum, reflétant les spécificités des matériaux utilisés dans les plaquettes de frein. En complément, l'étude de Font et al. (2019) sur deux types de matériaux de freinage – les plaquettes frontales et les garnitures latérales – a révélé que l'usure des plaquettes frontales est une source majeure de Baryum (Ba) et de Zinc (Zn) dans les PM. Par ailleurs, l'usure des garnitures latérales contribue au rejet d'Antimoine (Sb), d'Arsenic (As), de Zinc (Zn), de Molybdène (Mo), de Mercure (Hg) et de Plomb (Pb), soulignant ainsi la diversité des éléments métalliques libérés en fonction des matériaux utilisés. Selon Fruhwirt et al. (2024), dès qu'une température de surface critique des systèmes de freinage est dépassée, une augmentation exponentielle du nombre de particules est observée. Les tests qu'ils ont menés avec des matériaux frittés ont révélé que cette température critique atteignait environ 450 °C pour le disque de frein. En revanche, avec des plaquettes de frein organiques, la température critique s'est avérée nettement plus basse, se situant autour de 310 °C. Ces valeurs, bien que spécifiques aux matériaux testés, mettent en évidence l'influence cruciale de la composition des matériaux sur le comportement thermique des systèmes de freinage. En été, la température sur les quais est similaire à celle de l'extérieur, ce qui accentue l'usure des freins et, par conséquent, contribue à augmenter l'émission de particules. En hiver, les températures plus basses limitent l'élévation thermique des composants de freinage, réduisant ainsi leur usure et les émissions de particules associées. De plus, Walia et al. (2019) soulignent que l'énergie de freinage des blocs de frein en fonte est généralement plus faible en hiver qu'en été.

En ce qui concerne les émissions générées par friction mécanique entre le pantographe et la caténaire, elles s'élèvent entre 0,1 et 0,6 g/km pour les PM_{10}, parmi lesquelles 87 % sont des $PM_{2.5}$, selon Fruhwirt et al. (2023a).

Une étude récente menée par Ngo et al. (2023) dans une EFS parisienne a révélé, à l'aide du modèle de factorisation de matrices positives (PMF, pour Positive Matrix Factorization), que les activités ferroviaires contribuent à environ 80 % des émissions de PM_{10}. Cependant, les auteurs soulignent que, en raison de la simultanéité des émissions d'origine ferroviaire et de leur composition chimique similaire, il reste difficile de distinguer plus finement ces sources avec le modèle PMF et les données actuellement disponibles. Ces résultats s'inscrivent dans la continuité des travaux de Park et al. (2014), qui ont montré que l'abrasion des roues, des rails et des freins est responsable de 60 % des PM_{10} dans le métro de Séoul. Les 40 % restants proviennent de la combustion du pétrole (17 %), des aérosols secondaires (10 %), de l'usure des câbles électriques (8 %) et des poussières de sol et de route (5 %). Par ailleurs, Sitzmann et al. (1999) ont estimé une contribution de 68 % des activités ferroviaires à la pollution particulaire dans les gares souterraines londoniennes.

Outre ces opérations ferroviaires, d'autres activités contribuent aux émissions de particules dans les EFS. Par exemple, les travaux de maintenance, le découpage de rails et le soudage (Querol et al., 2012b) génèrent des particules en suspension dans l'air. Par ailleurs, les mouvements des voyageurs jouent également un rôle. Ces déplacements contribuent à la remise en suspension des particules déposées sur le quai (Johansson & Johansson, 2003). Pan et al. (2019) ont observé que le mouvement des voyageurs dans les EFS pouvait entraîner des fluctuations des concentrations de $PM_{2.5}$, bien que ces fluctuations n'affectent pas de manière significative les concentrations moyennes globales de $PM_{2.5}$, qui restent relativement stables.

Les sources extérieures, quant à elles, sont principalement dues à l'air ambiant qui pénètre dans les stations par les ouvertures et le réseau de ventilation (Fortain, 2008; Lee et al., 2018; Pan et al., 2019; Son et al., 2021). Ce phénomène permet aux particules issues de la pollution extérieure, comme celles

provenant du trafic routier et des activités industrielles, d'infiltrer les espaces souterrains. Selon Ngo et al. (2023), ces sources extérieures représentent environ 20 % des émissions totales de PM_{10} au niveau des quais.

En termes de composition chimique, les particules présentes dans les EFS se distinguent donc de celles de l'air extérieur principalement en raison de leur teneur plus riche en métaux (Ben Rayana et al., 2022). Parmi les métaux les plus courants figurent le Fer (Fe), le Manganèse (Mn), le Molybdène (Mo), le Cuivre (Cu), le Nickel (Ni), le Chrome (Cr), le Calcium (Ca), l'Aluminium (Al), le Sodium (Na), le Baryum (Ba), le Zinc (Zn), le Cobalt (Co) et le Plomb (Pb). Le Fer en particulier, est le métal le plus abondant dans les particules des EFS (Furuya et al., 2001; Aarnio et al., 2005; Salma et al., 2009; Colombi et al., 2013; Van Ryswyk et al., 2017; Mohsen et al., 2018; AEF, 2022a; Russell et al., 2022; INERIS, 2023), se présentant principalement sous forme d'Oxyde de Fer (Fe_2O_3) (Salma et al., 2009).

L'analyse de la composition élémentaire permet de remonter aux sources spécifiques de pollution, telles que l'usure des freins (Fe, C, Cu, Sr, Si, Zr, S, Cu, Mo, Ba, Zn, Ti, Sb, As, Pb), le contact roue-rail (Fe, Mn, C, Si, Al), et le contact pantographe-caténaire (C, Cu) (Sitzmann et al., 1999; Fortain, 2008; Moreno et al., 2015a; Lee et al., 2018; Durand, 2021; ANSES, 2022). Cette approche contribue ainsi à l'identification des sources principales de particules dans les EFS et permet une meilleure évaluation de leurs impacts potentiels sur la qualité de l'air intérieur et la santé.

Enfin, les particules souterraines, avec une densité relativement élevée estimée à environ 4 g/cm³ en raison de leur composition métallique, présentent un comportement de dépôt et de dispersion distinct comparé aux particules extérieures, dont la densité est généralement d'environ 2,5 g/cm³ (Salma et al., 2007; Fridell et al., 2010; Cha & Olofsson, 2018). Cette différence de densité se traduit par une vitesse de sédimentation plus rapide dans les EFS. En effet, celle-ci est plus importante pour les particules de diamètre important (Figure 1.5).

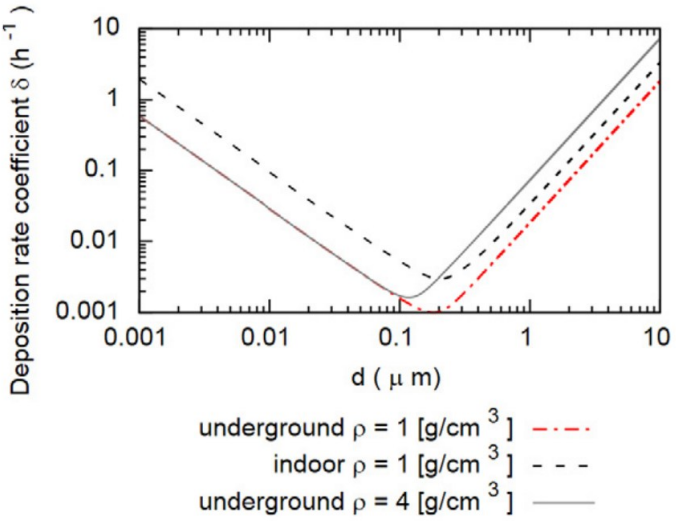


Figure 1.5 : Taux de dépôt en fonction du diamètre de particule (Walther et al., 2017)

Le dépôt gravitationnel domine pour les particules dont le diamètre dépasse 1 µm. Cependant, les particules ultrafines, de diamètre inférieur à 0,1 µm, présentent également un dépôt significatif, dû au mouvement brownien qui favorise leur diffusion et leur adhésion aux surfaces. En revanche, pour les particules de taille intermédiaire, leur masse est insuffisante pour une sédimentation gravitationnelle efficace, et elles sont trop grosses pour être fortement influencées par le mouvement brownien. Ainsi, les courants de convection naturelle de l'air suffisent à les maintenir en suspension, limitant leur dépôt (Walther & Bogdan, 2017; Walther et al., 2017).

En résumé, le Tableau 1.2 présente les principales sources de particules dans les EFS.

Tableau 1.2 : Résumé des principales sources de particules dans les EFS

Types		Sources
Intérieures	Activités humaines	Remise en suspension des particules déposées
	Activités ferroviaires	Contact roues/rails Usure des freins Contact pantographes/caténaires Travaux de maintenance Découpage de rails Creusement de tunnels Soudage Nettoyage
Extérieures	Air extérieur	Infiltration des particules extérieures

1.2 Variabilité et facteurs d'influence des concentrations en particules

Les concentrations de particules dans les EFS sont soumises à des variations et fluctuations importantes en fonction de multiples facteurs. Cette variabilité, qu'elle soit spatiale ou temporelle, est le résultat de processus internes, tels que le mouvement des trains, l'effet piston et la présence des passagers, ainsi que de facteurs externes comme l'environnement extérieur. Comprendre cette variabilité est essentiel pour l'établissement de stratégies efficaces de gestion de la qualité de l'air dans les EFS.

1.2.1 Variabilité spatiale

La variabilité spatiale des concentrations en particules dans les EFS est, entre autres, influencée par leur proximité aux sources de pollution. L'étude de Kim et al. (2008), menée à Séoul, a révélé que les niveaux de PM sont plus élevés sur les quais, en raison de leur proximité immédiate avec les trains, tandis que les zones plus éloignées, comme les bureaux et guichets, affichent des concentrations plus faibles. La Figure 1.6 illustre les concentrations de PM_{10} et $PM_{2.5}$ mesurées dans différents espaces – quais, bureaux, et guichets – dans 12 EFS et 10 gares en surface.

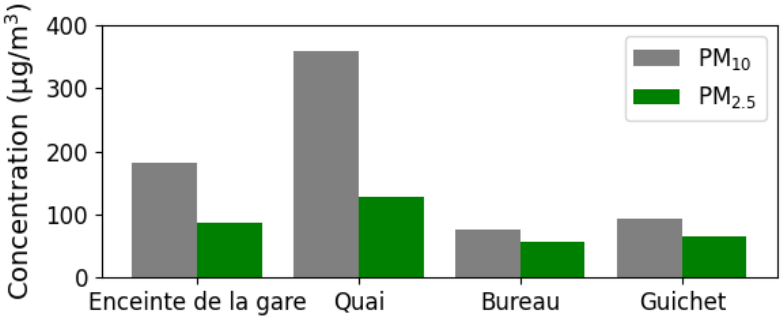


Figure 1.6 : Concentrations des PM_{10} et $PM_{2.5}$ entre les zones de stations de métro souterraines à Séoul (adapté de Kim et al., 2008)

De plus, Ben Rayana et al. (2022) ont étudié l'exposition aux particules ultrafines chez trois catégories de personnel (agents de station, conducteurs, agents de sécurité) dans les EFS, en explorant les déterminants de sa variabilité. Ils ont observé que l'exposition aux particules ultrafines (diamètre < 100 nm) diminue à mesure que l'on s'éloigne des voies ferrées, suggérant que les sources de particules ultrafines dans les EFS proviennent principalement de sources internes. Ces observations rejoignent ainsi les conclusions de Kim et al. (2008) avec les PM_{10} et les $PM_{2.5}$.

Sur les quais, la variabilité spatiale des concentrations en particules est principalement liée au phénomène dit d'effet piston. Ce phénomène fait référence à l'effet créé par les trains lorsqu'ils se déplacent dans le tunnel. Quand le train approche de la station, un gradient de pression est généré le long du train, et l'air est poussé en avant du train et aspiré par derrière, produisant ainsi un flux d'air. De même, quand le train quitte le quai, la station devient un espace de pression négative, et les particules présentes dans la station sont aspirées dans le tunnel au départ du train (Martins et al., 2016; Chang et al., 2021; Yang et al., 2024). De ce fait l'air est aspiré vers le tunnel, ce qui peut réduire la concentration des PM dans la station. La concentration des PM dans la gare change donc périodiquement, au rythme de l'intervalle entre les trains.

L'étude de Martins et al. (2016) menée à Athènes et Porto a montré que les concentrations de $PM_{2.5}$ augmentent lorsque le train entre en gare, l'effet piston poussant l'air pollué du tunnel vers le quai, en y ajoutant des particules en suspension. Inversement, au départ du train, l'effet piston inverse évacue une partie de l'air pollué du quai vers le tunnel, renouvelant ainsi l'air ambiant sur le quai. La Figure 1.7 illustre ce phénomène à la station de métro Syntagma à Athènes où les profils des concentrations de $PM_{2.5}$ sont mesurés en quatre positions distinctes (P1, P2, P3, P4). Les emplacements des accès des passagers aux quais sont représentés par des flèches (▼) situées soit entre deux positions (P), soit à l'extrême d'une position. En se positionnant par exemple à la position P1 du quai, les concentrations en $PM_{2.5}$ augmentent significativement lors de l'arrivée d'un train (●) et diminuent lors de son départ (○). Les concentrations les plus élevées de $PM_{2.5}$ sont souvent mesurées aux extrémités d'accès des trains, là où l'air pollué du tunnel est poussé par l'effet piston. Toutefois, même si l'accent est mis sur la variabilité spatiale, des variations notables peuvent également être observées sur des intervalles de temps très brefs.

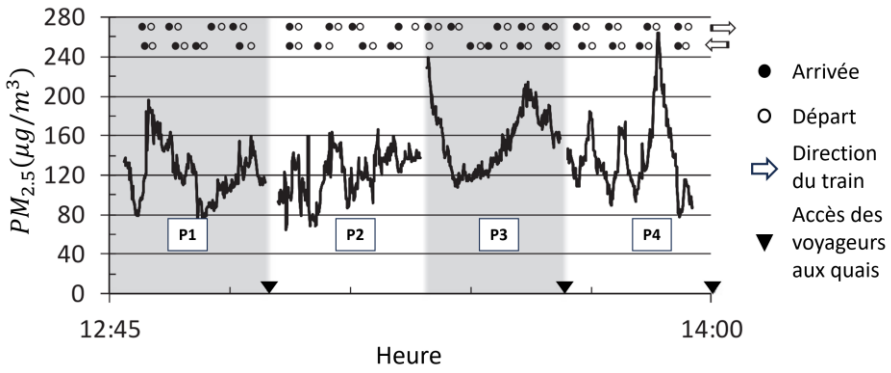


Figure 1.7 : Profils des concentrations de $PM_{2.5}$ ($\mu g/m^3$) sur le quai de la station de métro Syntagma à Athènes (adapté de Martins et al. 2016).

Une étude récente menée par Bendl et al. (2023) à Munich a révélé des résultats similaires. Ils ont démontré qu'une augmentation notable des concentrations de PM_{10} se produit également lorsque deux trains quittent la gare simultanément mais dans des directions opposées, provoquant une élévation rapide des niveaux de particules pendant un court laps de temps.

Plus récemment, Yang et al. (2024) ont mené une étude dans une station de métro à Tianjin, en Chine, simulant la dynamique des PM_{10}. La Figure 1.8 illustre la variation moyenne des concentrations de PM_{10} simulées lors de cette étude. Initialement fixée à 93 µg/m³, correspondant à la moyenne annuelle de PM_{10} à l'extérieur, la concentration sur le quai commence à augmenter légèrement lorsque le train circule dans le tunnel pour une durée de 30 secondes. Cette accumulation se poursuit pendant 70 à 100 secondes, contaminant plus de la moitié de la surface du quai lorsque le train s'arrête. Même après le départ du train, alors que celui-ci n'a pas encore quitté complètement le quai, les particules continuent de s'accumuler à l'arrière du quai pendant environ 10 secondes supplémentaires. Une fois passé 110 secondes, le train s'éloignant progressivement, la concentration de PM_{10} diminue grâce à la dilution par l'air neuf extérieur et le dépôt des particules. Les résultats de cette simulation indiquent une augmentation de 2,2 µg/m³ de PM_{10} après le passage du train.

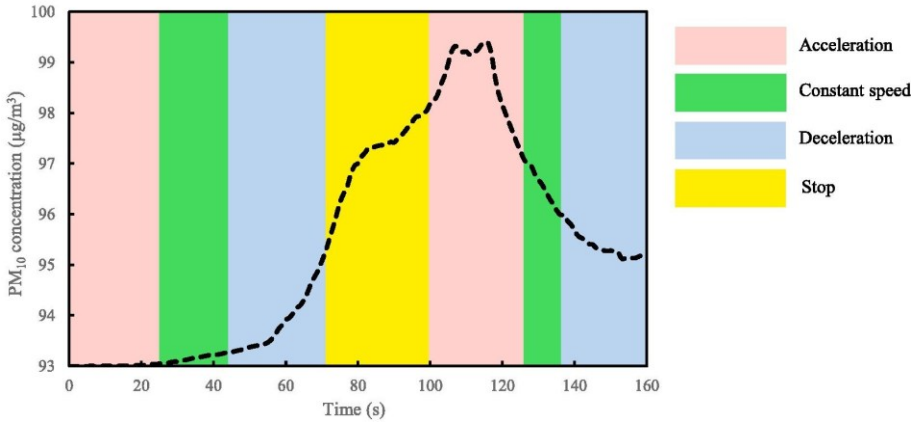


Figure 1.8 : Simulation de la concentration moyenne de PM_{10} sur un quai de station de Métro à Tianjin, Chine (Yang et al., 2024)

Des observations comparables ont été rapportées par Agarwal et al. (2025) dans la station de Shibuya à Tokyo. Ils ont constaté que, sur un même quai, les niveaux de PM variaient jusqu'à 63,72 % pour les $PM_{2.5}$ et 120,23 % pour les PM_{10}, avec des concentrations plus élevées aux extrémités du quai qu'au centre. Les auteurs attribuent principalement ces variations à l'effet piston, généré par l'entrée et le départ des trains. Ces variations montrent que la position sur le quai influence fortement l'exposition aux particules. D'après Chang et al. (2021), cet effet piston favorise l'échange entre l'air intérieur et les milieux extérieurs (tunnel, environnement extérieur) de la station et augmente la corrélation entre les concentrations de PM dans les EFS et l'air ambiant.

En plus de la proximité au quai et de l'effet piston, la profondeur de la station ainsi que les systèmes de ventilation jouent également un rôle dans cette variabilité spatiale. Les quais situés à des profondeurs plus faibles, proches de la surface, interagissent davantage avec l'air extérieur, favorisant un meilleur renouvellement de l'air. En revanche, les stations situées à des profondeurs plus importantes sont principalement exposées aux polluants générés à l'intérieur des tunnels, qui représentent leur principale source de pollution (Passi et al., 2021a). Des travaux, notamment ceux de Vilcassim et al. (2014) et de Figueroa-Lara et al. (2019), ont démontré que les stations profondes ont tendance à accumuler davantage de polluants, en grande partie à cause d'une ventilation insuffisante.

Globalement, les recherches suggèrent une relation directe entre la profondeur des stations et les concentrations de PM : plus la station est profonde, plus les niveaux de PM sont élevés (Şahin et al., 2012; Colombi et al., 2013; Saunders et al., 2019; Chang et al., 2021). Par exemple, dans leur étude à Istanbul, Şahin et al. (2012) rapportent une corrélation élevée (0,87) entre la profondeur des EFS et les concentrations moyennes de PM_{10}, indiquant que les niveaux de PM_{10} augmentent significativement avec la profondeur des stations (Figure 1.9).

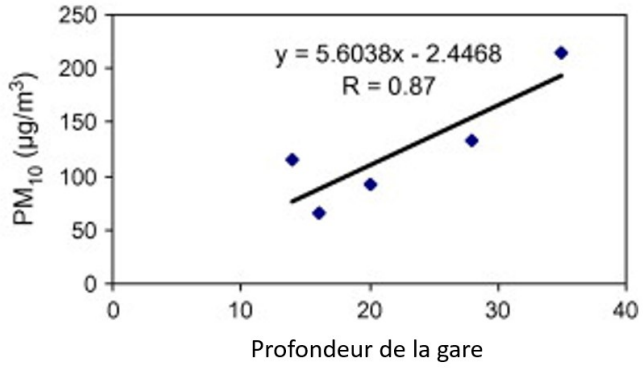


Figure 1.9 : Relation entre la profondeur de la station de métro et la concentration moyenne journalière de PM_{10} aux gares Aéroport, Aksaray, 4Levent, Sisli et Taksim à Istanbul, Turquie (adapté de Şahin et al., 2012)

Kamani et al. (2014) ont mesuré les concentrations de $PM_{2.5}$ dans des EFS de Mexico aux profondeurs différentes (10 m pour la ligne L9, 25 m pour L1 et 40 m pour L7). Les résultats montrent des concentrations croissantes de $PM_{2.5}$ avec la profondeur, atteignant respectivement 57 et 66 µg/m³, 90 et 111 µg/m³, et 104 et 122 µg/m³ en fonction des saisons (froides et chaudes). Enfin, l'étude de Van Ryswyk et al. (2017) portant sur les EFS de Toronto, Montréal et Vancouver met également en évidence cette tendance. Ils ont constaté des taux d'augmentation avec la profondeur de 11,1 %/10 m pour Toronto, 2,1 %/10 m pour Montréal, et 32,7 %/10 m pour Vancouver.

A l'inverse, Pan et al. (2019) ont rapporté une tendance contrastée à Beijing, où les concentrations en particules (PM_{10} et $PM_{2.5}$) variaient en fonction de la profondeur. Ils ont comparé les concentrations mesurées à trois niveaux : à l'extérieur, dans le hall et sur le quai de la station. Par exemple, les concentrations relevées étaient de 578 µg/m³ et 371 µg/m³ à l'extérieur, 561 µg/m³ et 317 µg/m³ au niveau du hall, et 533 µg/m³ et 291 µg/m³ au quai, respectivement pour les PM_{10} et les $PM_{2.5}$. Ces résultats montrent que, dans les EFS profondes, les concentrations de PM étaient inférieures à celles mesurées dans les zones moins profondes de la même station. Dans leur étude, ils ont émis l'hypothèse que cette différence pourrait être liée au mode de fonctionnement des systèmes de filtration et de ventilation, qui jouent un rôle clé dans les variations des concentrations de PM_{10} et $PM_{2.5}$. Ces observations soulignent l'importance de considérer d'autres facteurs, tels que les systèmes de ventilation, en plus de la profondeur de la station, pour comprendre les variations des concentrations de particules.

Agarwal et al. (2025) ont étudié les concentrations de particules sur plusieurs lignes de métro dans la station de Shibuya à Tokyo, en comparant des stations peu profondes dépourvues de système de ventilation (L3 et L4) à des stations plus profondes équipées de systèmes de ventilation (L5 et L6). Leur étude a révélé que les quais plus profonds (L5 et L6), dotés de systèmes de ventilation efficaces, présentaient des concentrations de particules nettement inférieures à celles des quais moins profonds dépourvus de ventilation. Cette différence pourrait également s'expliquer par le fait que les lignes 5 et 6 se trouvent dans des tunnels plus larges disposant de quatre quais, tandis que les lignes 3 et 4 sont situées dans des tunnels plus étroits avec seulement deux quais.

Par ailleurs, Saunders et al. (2019) ont mis en évidence la relation entre la concentration de particules et la distance par rapport à l'entrée du tunnel.

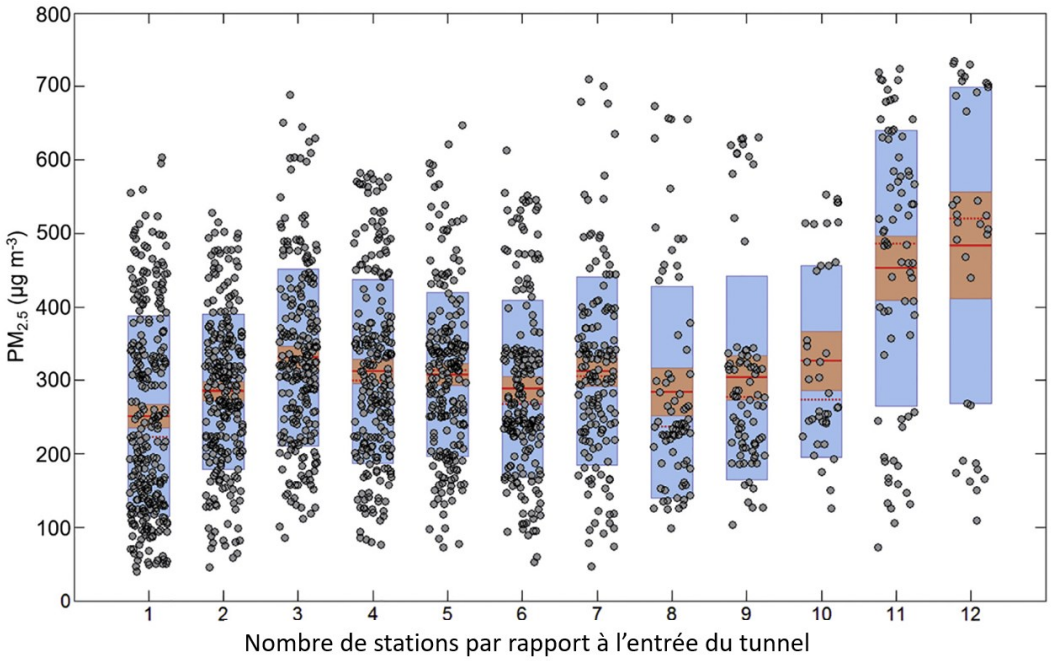


Figure 1.10 : Concentrations de $PM_{2.5}$ dans les lignes en profondeur du réseau ferroviaire londonien, classées selon la proximité avec l'entrée de tunnel (adapté de Saunders et al., 2019)

Plus l'on s'éloigne de l'entrée du tunnel, plus les concentrations augmentent, confirmant ainsi que les particules fines sont principalement produites et concentrées dans les sections souterraines des lignes (Figure 1.10).

Enfin, une étude menée par Moreno et al. (2014) a mis en évidence, outre la profondeur du quai et la proximité par rapport à l'entrée du tunnel, l'impact de la ventilation mécanique forcée des tunnels sur les concentrations en particules. La Figure 1.11 illustre les concentrations de particules (PM_{10}, $PM_{2.5}$, PM_1) et de gaz (CO_2, CO) mesurées dans les stations de la ligne L2 du métro de Barcelone, à quatre positions distinctes du quai (P1, P2, P3, P4), présentant des caractéristiques variées, notamment en termes de profondeur, taille du tunnel, nombre de voies et de présence ou d'absence de système de ventilation. Les accès aux quais sont indiqués par de petites flèches sur le schéma P1-P4. Les concentrations de particules, surlignées en gris, atteignent des niveaux exceptionnellement élevés (>50 µg/m³ pour PM_1, >100 µg/m³ pour PM_3, et >200 µg/m³ pour PM_{10}). Les concentrations les plus élevées sur chaque quai sont mises en évidence en gras. Selon leurs résultats, dans les gares équipées de tunnels à voie unique, la qualité de l'air se détériore significativement lorsque le système de ventilation est désactivé, en particulier pour les particules fines. En revanche, cet effet n'a pas été observé dans les tunnels à double voie, où les concentrations de PM restent similaires, que la ventilation forcée soit activée ou non. Ces résultats suggèrent que l'efficacité des systèmes de ventilation dépend fortement de la configuration et de la taille de l'espace intérieur (gare et tunnel).

Ainsi, l'influence des systèmes de ventilation sur la concentration de PM dans les EFS dépend des conditions locales, telles que la configuration des tunnels et la géométrie de l'espace. Ces éléments doivent être soigneusement pris en compte lors de l'évaluation de l'efficacité de la ventilation dans la gestion des concentrations de particules en milieu souterrain.

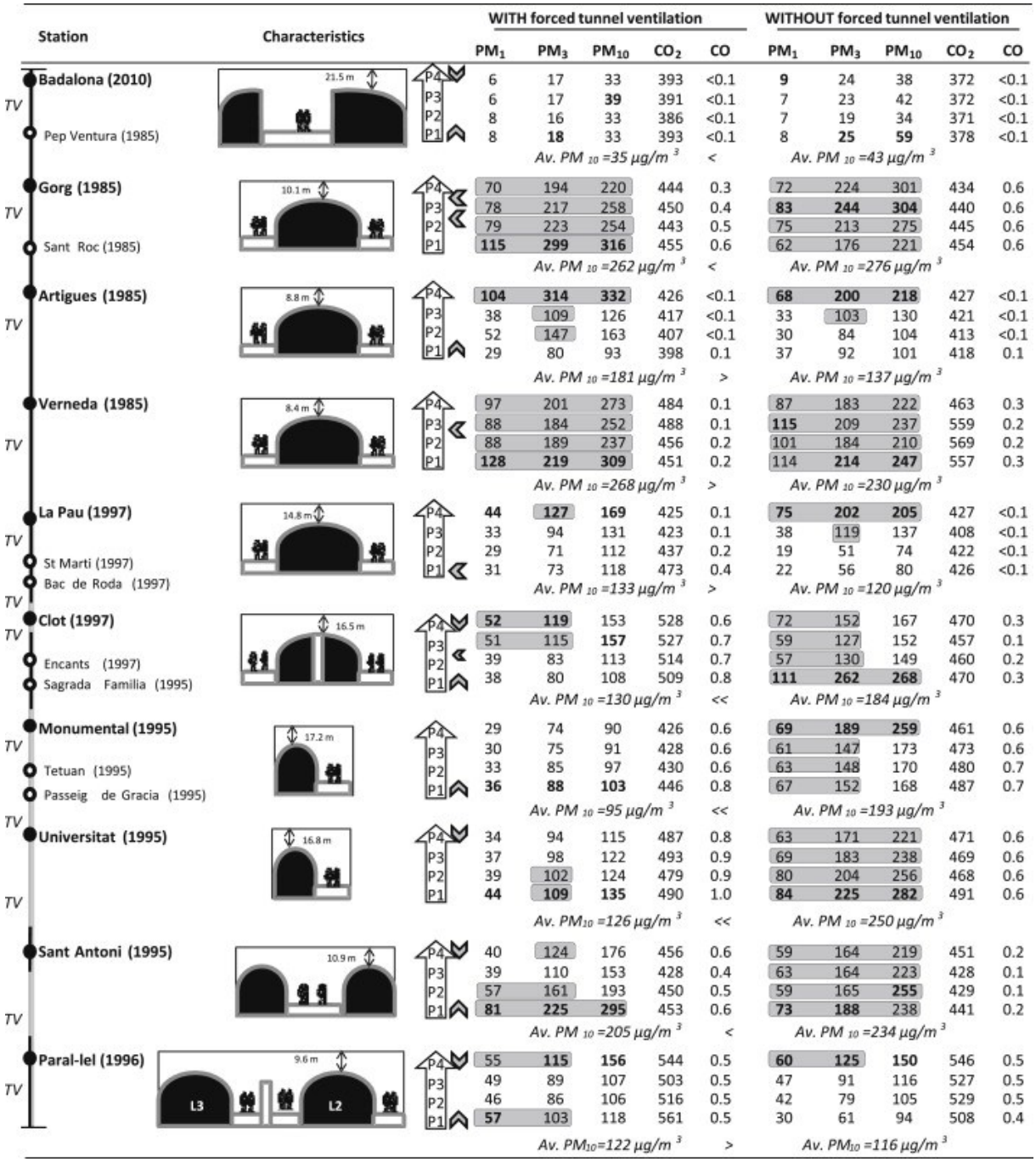


Figure 1.11 : Concentrations de particules (PM_{10}, $PM_{2.5}$, PM_1) et de gaz (CO_2, CO) dans différentes stations de la ligne L2 du métro de Barcelone (Moreno et al., 2014)

1.2.2 Variabilité temporelle des particules au niveau du quai

Tout comme la variabilité spatiale, la variabilité temporelle des particules est également influencée par plusieurs facteurs, tels que la fréquence des trains, entre autres.

Le passage fréquent des trains est l'un des principaux contributeurs à l'augmentation des concentrations de particules dans les EFS. La friction entre les roues des trains et les rails, ainsi que l'usure des freins, génèrent une grande quantité de particules métalliques. De plus, chaque mouvement de train provoque une turbulence qui soulève les particules déjà présentes dans l'air ou sur les surfaces environnantes.

En fonction de l'affluence des voyageurs durant les jours ouvrés, la période de fonctionnement des trains peut être divisée en heures de pointe et heures creuses. En effet, il existe deux périodes de pointe chaque jour, correspondant aux déplacements domicile-travail matinaux et aux retours en fin de journée, le reste du temps étant considéré comme période creuse. Dans la plupart des EFS, les heures de pointe se situent entre 07h00 et 09h00 le matin, puis entre 18h00 et 21h00 le soir. Pendant ces périodes, la fréquence des trains est augmentée pour accueillir un plus grand nombre de passagers, tandis qu'elle diminue durant les heures creuses, traduisant une baisse du flux de voyageurs. Par exemple, à la gare d'Avenue Foch (Paris), l'intervalle entre les trains est de 8 minutes en heure de pointe et de 15 minutes en période creuse.

Une augmentation en concentration de particules est notée durant la période des heures de pointe le matin (Birenzvige et al., 2003; Raut et al., 2009; Şahin et al., 2012; Querol et al., 2012b; Kwon et al., 2015; Walther & Bogdan, 2017; Durand, 2021), ce qui est attribuable non seulement à l'afflux de passagers, mais aussi à la fréquence plus élevée des trains ; le mouvement des trains favorise la remise en suspension des particules et leur génération due à l'abrasion des voies ferrées, des roues, des plaquettes de frein et des matériaux d'alimentation en courant, et les mouvements des voyageurs conduisent principalement à la remise en suspension des particules (Figure 1.12).

Pendant la nuit, une diminution progressive des concentrations de particules est observée en raison de l'interruption du service de transport (Martins et al., 2016). Cette tendance a également été relevée dans les travaux d'AIRPARIF, 2017; Fortain, 2008; Gustafsson et al., 2012; Walther and Bogdan, 2017 ainsi que Son et al. (2013), illustrant le lien direct entre l'intensité d'exploitation des transports et les niveaux de pollution mesurés dans les EFS. Pendant le weekend, les concentrations de particules fluctuent davantage tout au long de la journée, affichant des niveaux généralement inférieurs à ceux observés en semaine. Ces observations concordent avec les travaux de Raut et al. (2009) et de Birenzvige et al. (2003), qui ont montré que les concentrations de particules suivent étroitement l'intensité du trafic ferroviaire, mais avec un décalage d'une à deux heures entre le début du trafic et l'augmentation des concentrations de particules. Cette latence est attribuée à la dispersion progressive des particules par la resuspension, provoquée par l'intensité croissante du trafic et l'afflux de passagers, un phénomène qui semble fortement lié à la cadence de l'exploitation ferroviaire.

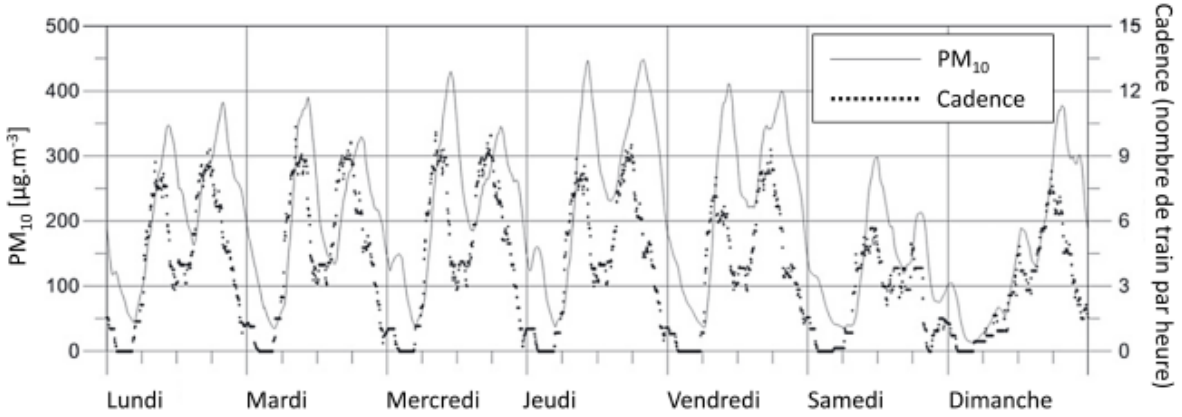


Figure 1.12 : Evolution des concentrations en PM_{10} avec la cadence d'exploitation dans la station d'Arlanda C à Stockholm (adapté de Gustafsson et al., 2012 par Durand, 2021)

Cependant, il convient de souligner les résultats de l'étude d'Agarwal et al. (2025) à la station de métro de Shibuya à Tokyo, qui montrent une dynamique différente. Contrairement aux tendances généralement observées, les concentrations moyennes de PM durant les weekends ne sont pas systématiquement inférieures à celles des jours ouvrés. Pour les $PM_{2.5}$, les concentrations étaient en moyenne de 36,7 µg/m³ durant les jours ouvrés contre 38,8 µg/m³ les weekends. En revanche, pour les PM_{10}, les concentrations étaient légèrement plus élevées les jours ouvrés, avec 58,6 µg/m³, contre 50,3 µg/m³ les weekends. Les auteurs n'ont cependant pas fourni d'explications détaillées pour ces

observations, laissant suggérer que des facteurs contextuels spécifiques pourraient influencer cette variabilité.

Outre la fréquence des trains, la saisonnalité joue également un rôle clé dans les concentrations de particules dans les EFS. De nombreuses études montrent que, de manière générale, les concentrations de PM sont plus élevées pendant la saison froide (Braniš, 2006; Chang et al., 2021; Furuya et al., 2001; Klepczynska Nyström, 2012; Kwon et al., 2015; Park et al., 2024; Van Ryswyk et al., 2017). Par exemple, à Prague, Braniš (2006) a observé que les niveaux de PM atteignent des pics hivernaux d'environ 120,2 µg/m³, contre 102,7 µg/m³ en été. Cette tendance est également visible au Canada, où, à Montréal, les PM_{10} passent de 26,3 µg/m³ en été à 46,9 µg/m³ en hiver, tandis qu'à Vancouver, bien que les niveaux absolus soient plus bas, une augmentation similaire est observée, passant de 11,3 µg/m³ en été à 24,4 µg/m³ en hiver (Van Ryswyk et al., 2017). La Figure 1.13, issue des travaux de Park et al. (2024) à Séoul, illustre clairement cette tendance. Les concentrations moyennes de $PM_{2.5}$, observées dans l'ensemble des stations de métro de Séoul, sont plus élevées en hiver et diminuent significativement durant les mois d'été, où les niveaux restent plus faibles, aussi bien à l'intérieur qu'à l'extérieur.

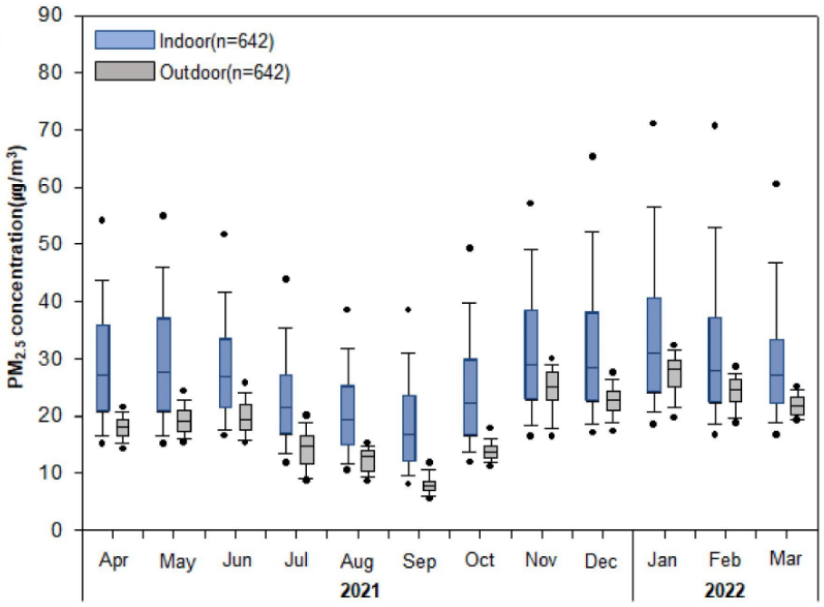


Figure 1.13 : Répartition mensuelle des concentrations de $PM_{2.5}$ à l'intérieur et à l'extérieur de l'ensemble des stations de métro de Séoul (Park et al., 2024)

Les concentrations élevées durant la période froide sont principalement attribuables à la modification saisonnière des opérations des systèmes de ventilation du métro (Martins et al., 2015) ainsi qu'aux niveaux de particules dans l'air extérieur (Hernández-Castillo et al., 2014). En effet, Martins et al. (2015) ont comparé les concentrations de PM_{10}, $PM_{2.5}$ et PM_1 durant les saisons chaude et froide dans différentes stations de métro. La Figure 1.14 illustre, par exemple, le cas de la station de métro Joanic à Barcelone pour les trois classes de particules. Les concentrations de particules sont plus faibles durant les périodes chaudes par rapport aux périodes froides. Selon les auteurs, cette observation s'explique par une utilisation accrue des systèmes de ventilation dans la station pendant l'été, contribuant ainsi à la dilution des particules. Chang et al. (2021) soulignent dans leur revue que l'utilisation des systèmes de climatisation peut renforcer la ventilation et augmenter l'efficacité de la filtration dans les stations de métro, contribuant ainsi à réduire les concentrations de particules.

Par ailleurs, les facteurs extérieurs, comme la température et la vitesse du vent, diffèrent entre les saisons et influencent la dispersion des particules en extérieur. En hiver, des températures plus froides et une faible vitesse du vent ralentissent la dispersion, favorisant ainsi leur accumulation. Par

conséquent, ces particules peuvent s'infiltrer dans les EFS par les entrées utilisées par les passagers, ainsi que par les tunnels et les bouches de ventilation.

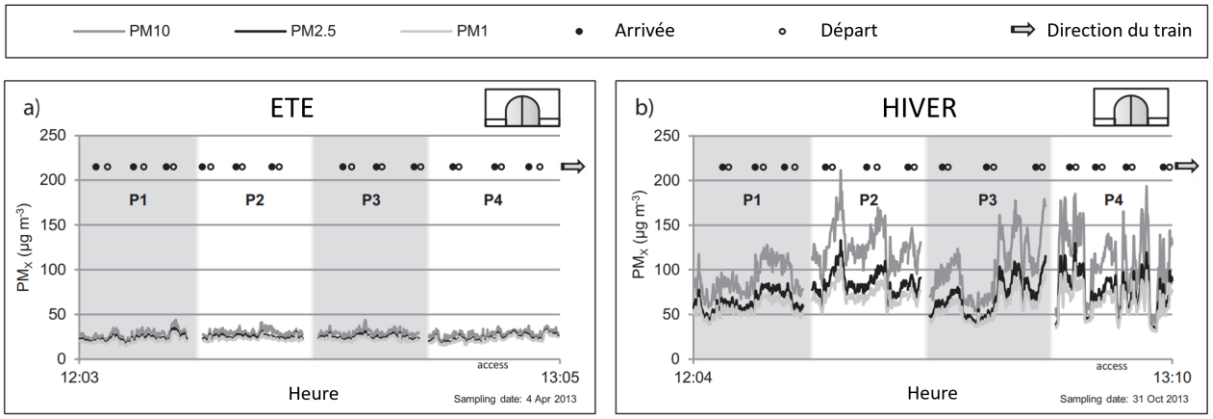


Figure 1.14 : Concentrations de PM_{10}, $PM_{2.5}$ et PM_1 sur le quai de la station de métro Joanic à Barcelone à 4 sites différents (P1, P2, P3 et P4) (adapté de Martins et al. 2015)

En plus de la fréquence des trains et de la saisonnalité, l'air extérieur est un facteur majeur de pollution dans les EFS, influençant la variabilité temporelle des concentrations de particules. Plusieurs études (Aarnio et al., 2005; Cheng et al., 2008; Kam et al., 2011; Kim et al., 2012; Kwon et al., 2015; Lee et al., 2018; Martins et al., 2015b, 2016; Mohsen et al., 2018; Pan et al., 2019; Wang et al., 2016; L. Zhao et al., 2017) ont conclu que les concentrations de particules dans les EFS et dans l'air extérieur sont positivement corrélées, bien que peu d'entre elles aient quantifié précisément cette relation. Par exemple, Martins et al. (2016) ont observé que les concentrations de $PM_{2.5}$ dans les stations de métro étaient fortement corrélées avec celles de l'air extérieur à Porto, au Portugal, et à Athènes, en Grèce avec des coefficients de détermination de Pearson de $R^2 = 0,74$ et de $R^2 = 0,60$, respectivement (voir Figure 1.15). Cependant, une exception notable est la gare de Barcelone, où la corrélation est beaucoup plus faible ($R^2 = 0.08$). Ces résultats montrent que les équations de régression linéaire reliant les concentrations sur les quais à celles de l'extérieur ne sont pas nécessairement applicables à toutes les EFS, en raison des variations de concentration de fond ambiant et des systèmes de contrôle environnemental spécifiques à chaque EFS (Chang et al., 2021).

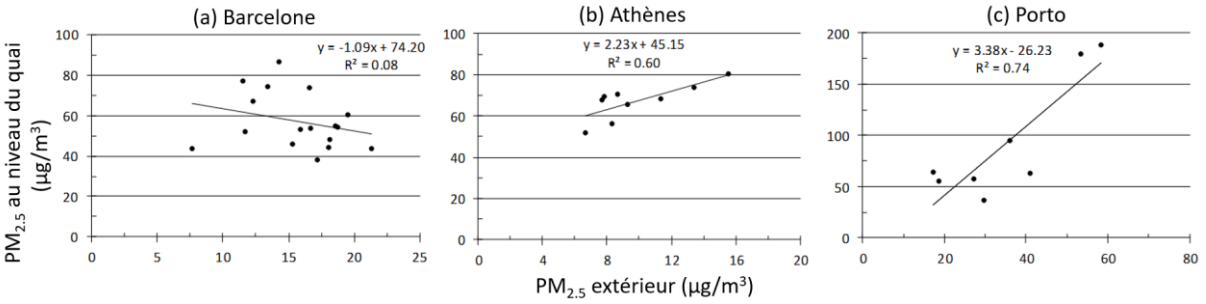


Figure 1.15 : Relation entre les concentrations de $PM_{2.5}$ (µg/m³) sur les quais de métro et l'air ambiant extérieur à a) Barcelone (40 m de profondeur), b) Athènes (20 m de profondeur) et à c) Porto (Peu profonde) (adapté de Martins et al. 2016)

Compte tenu des facteurs discutés précédemment, Chang et al. (2021) ont dressé une liste des principaux facteurs influençant la qualité de l'air dans les EFS, présentée dans le Tableau 1.3. La communauté scientifique accorde une attention particulière aux six premiers facteurs mentionnés dans cette liste, comme souligné par Chang et al. (2021).

Tableau 1.3 : Principaux facteurs d'influences régissant les concentrations en particules souterraines (d'après Chang et al. 2021)

Facteurs d'influence	Références
Concentration extérieure des PM	Aarnio et al., 2005; Cheng et al., 2008; Kam et al., 2011; Kim et al., 2012; Kwon et al., 2015; Lee et al., 2018; Martins et al., 2015b, 2016; Mohsen et al., 2018; Pan et al., 2019; Wang et al., 2016; L. Zhao et al., 2017
Fréquence des trains	Aarnio et al. 2005; Park and Ha 2008; Raut et al. 2009; Colombi et al. 2013; Kwon et al. 2015; Martins et al. 2016; Cartenì and Cascetta 2018
Effet piston	Moreno et al. 2014, 2015b; Martins et al. 2015; Wang et al. 2016; He et al. 2018
Saisonnalité	Braniš 2006; Klepczynska Nyström 2012; Martins et al. 2015; Kwon et al. 2015; Van Ryswyk et al. 2017; Minguillón et al. 2018
Profondeur de la gare	Park and Ha 2008; Cheng et al. 2008, p. 10; Şahin et al. 2012; Colombi et al. 2013; Kamani et al. 2014; Hernández-Castillo et al. 2014; Kwon et al. 2015; Van Ryswyk et al. 2017; Pan et al. 2019
Systèmes de ventilation	Aarnio et al. 2005; Park and Ha 2008; Cheng et al. 2008; Klepczynska Nyström 2012; Moreno et al. 2014; Hernández-Castillo et al. 2014
Âge de la gare	Cheng et al. 2008; Midander et al. 2012; Guo et al. 2014
Temps	Chan et al. 2002; Park and Ha 2008; Martins et al. 2016
Système de freinage	Aarnio et al. 2005; Park and Ha 2008; Cheng et al. 2008; Midander et al. 2012
Densité de passagers	Park and Ha 2008
Portes palières (PSDs)	Lee et al. 2009; Kim et al. 2010, 2012; Son et al. 2014b; Qiao et al. 2015; Martins et al. 2015; He et al. 2018; Minguillón et al. 2018
Conception de la gare	Chan et al. 2002; Aarnio et al. 2005; Park and Ha 2008; Cheng et al. 2008; Kam et al. 2011; Moreno et al. 2014
Circulation extérieure	Boudia et al. 2006

En France, AIRPARIF (2022b), en partenariat avec la SNCF, a conduit une étude sur les facteurs d'influence des concentrations en particules dans les gares souterraines franciliennes de la SNCF. Cette étude a permis d'établir une liste des facteurs susceptibles d'influencer les concentrations en particules dans les EFS. En complément de la liste dressée par Chang et al. (2021), cette étude intègre les dernières avancées de la littérature scientifique ainsi que des observations spécifiques aux EFS franciliennes, basées sur des campagnes de mesures qui seront réanalysées dans le cadre de cette thèse. Parmi les paramètres identifiés, figurent :

- La ligne de RER
- La présence ou non de travaux externes à la gare
- Le nombre de voyageurs par jour (et/ou passage réel des voyageurs)
- La typologie de la gare (mixte ou souterraine)
- L'occurrence de travaux nocturnes sur les rails (ou/et le passage de trains diesel)
- L'occurrence d'un pic de pollution durant la campagne
- La température de la gare
- Les niveaux de CO_2 dans la gare (indiquant le volume de voyageurs)

AIRPARIF (2022a) a écarté certains facteurs d'influence faute de données, à savoir : la vitesse d'entrée en gare des trains, le type de matériel roulant et ses caractéristiques, la caractéristique des tunnels en amont/aval du quai, le rayon de courbure de la gare et les caractéristiques des rails et caténaires.

1.3 Solutions pour améliorer la qualité de l'air dans les EFS

Plusieurs solutions techniques ont été proposées pour réduire les concentrations de polluants, notamment en limitant leur production à la source ou en optimisant les systèmes de gestion de l'air. Cette section examine les approches les plus courantes et prometteuses pour améliorer la qualité de l'air dans les EFS.

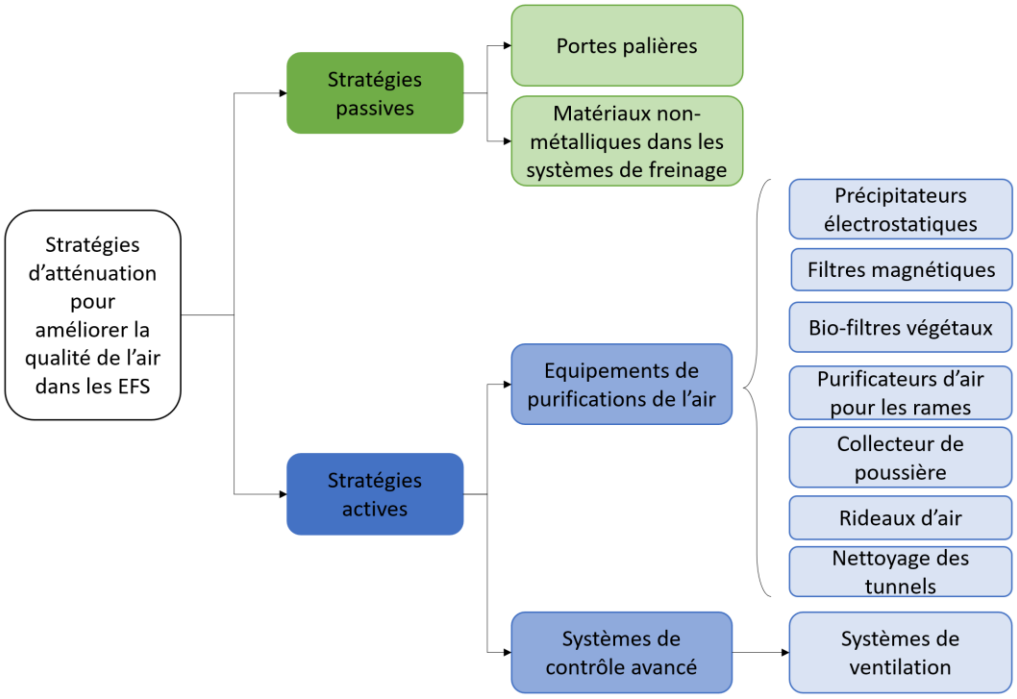


Figure 1.16: Différentes solutions pour améliorer la qualité de l'air dans les EFS

D'un point de vue énergétique et de gestion des EFS, les solutions pour améliorer la qualité de l'air peuvent être classées en deux grandes catégories (Figure 1.16) : les stratégies passives, telles que l'installation de portes palières et les stratégies actives, qui nécessitent de l'énergie pour répondre aux exigences de qualité de l'air dans les EFS. Parmi les stratégies passives, on peut également inclure l'utilisation de matériaux non métalliques dans les systèmes de freinage, permettant de réduire la production de particules fines à la source. Les stratégies actives, quant à elles, se subdivisent en deux types : les équipements de purification de l'air (rideaux d'air, filtres magnétiques, véhicules de lavage de tunnels, précipitateurs électrostatiques, purificateurs pour les rames, collecteurs de poussière et biofiltres végétaux) et les systèmes de contrôle avancé, tels que le système de ventilation automatique.

1.3.1 Stratégies passives

Comme montré dans la Figure 1.16, les deux principales stratégies passives sont les portes palières et les matériaux non-métalliques dans les systèmes de freinage. Les portes palières sont des portes vitrées installées sur les quais qui s'ouvrent en synchronisation avec les portes des trains. Elles créent une barrière physique entre les quais et les tunnels, empêchant ainsi la dispersion des polluants générés par les trains vers les espaces réservés aux passagers. Actuellement, les portes palières disponibles sur le marché comprennent les portes palières intégrales (PSDs, pour Platform Screen Doors), les portes palières à mi-hauteur (HPBDs, pour Half-height Platform Bailout Doors), les portes palières à pleine hauteur (FPBDs, pour Full-height Platform Bailout Doors) et les portes palières à ventilation ajustable (AVPD, pour Adjustable Vent Platform Doors) (Figure 1.17).



Figure 1.17 : Portes palières en EFS : (a) PSDs, (b) FPBDs, (c) HPBDs et (d) AVPDs (Yang et al., 2023)

Plusieurs études (Lee et al., 2009; Kim et al., 2012; Son et al., 2013, 2014b; Han et al., 2015) ont montré que les PSDs contribuent à une amélioration significative de la qualité de l'air sur les quais en empêchant les particules en suspension de pénétrer dans l'enceinte de la station. Selon les études menées par Kim et al. (2012) dans une EFS de Séoul, les concentrations moyennes en PM_{10} et $PM_{2.5}$ ont diminué de 16 % et 12 % après l'installation des PSDs, respectivement. En terme de réduction spatiale, Han et al. (2015) ont montré que la diminution des intrusions de particules issues du trafic ferroviaire est observée aussi bien sur les quais, avec une réduction d'environ 38 %, que dans d'autres espaces des stations, où la réduction atteint environ 34 %. Cependant, cette solution présente des inconvénients. En effet, Son et al. (2014) ont montré qu'après l'installation des PSDs, ils ont constaté une augmentation des concentrations de PM_{10} de 29,9 % à l'intérieur des trains, car l'air du tunnel, entravé par les portes palières, contribue à dégrader la qualité de l'air dans les rames. Comme un voyageur passe généralement plus de temps à l'intérieur du train qu'au sein d'une gare, cela peut se traduire par une hausse de son exposition globale au cours de son trajet (Durand, 2021). Cui et al. (2016) ont constaté que les échantillons collectés dans les stations de métro de Beijing équipées de PSDs étaient plus magnétiques et plus fins que ceux sans portes palières, illustrant que les portes palières n'ont pas réussi à empêcher la migration des particules fines contenant du fer du tunnel vers le quai.

Les FPBDs et HPBDs (Figure 1.17b et c) ont des effets similaires sur les PM. L'effet piston, généré par le passage des trains, peut entraîner la diffusion des particules provenant des tunnels vers les quais. Une étude menée par (Ren et al., 2022) à Tianjin, en Chine, dans deux stations équipées respectivement de PSDs et de HPBDs a montré que les PSDs empêchent une grande partie des particules des tunnels de pénétrer sur les quais.

Les AVPDs (Figure 1.17d) sont une nouvelle génération de portes similaires aux FPBDs, mais équipées de bouches d'aération réglables situées au-dessus ou en dessous des portes. Ces bouches sont fermées en période de chauffage pour éviter l'entrée d'air chaud des tunnels, et ouvertes en dehors de cette période pour ventiler les stations en exploitant l'effet piston des trains. Pan et al. (2019) ont étudié l'impact de ces bouches dans une station de métro à Pékin et ont constaté une réduction de 32 % des concentrations de particules fines lorsque les bouches étaient ouvertes et que la qualité de l'air extérieur

était bonne. À ce jour, peu d'études se sont penchées sur les trois derniers types de portes palières (HPBDs, FPBDs, AVPDs | Yang et al., 2023).

Les systèmes de freinage des trains quant à eux sont l'une des principales sources de particules fines dans les EFS (Section 1.1). Les matériaux métalliques utilisés dans les freins génèrent des particules métalliques lors de la friction avec les roues, contribuant ainsi à la pollution dans les stations. Une solution consiste à remplacer les composants métalliques par des matériaux non-métalliques, tels que des composites ou des céramiques, qui produisent beaucoup moins de particules lors du freinage (Salma et al., 2007; Querol et al., 2012a). Ces nouveaux matériaux réduisent l'usure des freins et, par conséquent, diminuent la quantité de poussières émises, tout en améliorant la durabilité des composants du système. Şahin et al. (2012) ont montré dans leur étude que la dispersion de fer est significativement plus faible lorsque des roues en caoutchouc (Figure 1.18), plutôt que des roues en métal, sont utilisées. Dans cette perspective, Álvarez et al. (2012) ont quantifié les concentrations de métaux issues des roues en caoutchouc dans une station de métro de Mexico. Selon leur étude, les concentrations de fer observées dans cette station sont bien plus faibles que celles rapportés dans d'autres stations, avec 5,85 µg/m^3 contre 94,2 µg/m^3 à Tokyo (Furuya et al., 2001), entre 33 et 102 µg/m^3 à Buenos Aires (Murruni et al., 2009), et entre 37 et 55 µg/m^3 à Rome (Ripanucci et al., 2006) pour les PM_{10}. Pour les $PM_{2.5}$, les concentrations étaient de 2,67 µg/m^3 contre 24,65 µg/m^3 à Helsinki (Aarnio et al., 2005).



Figure 1.18 : Roues en caoutchouc utilisées dans le métro parisien (https://www.theautopian.com/heres-why-some-trains-have-rubber-tires/)

1.3.2 Stratégies actives

1.3.2.1 Les systèmes de ventilation

Les systèmes de ventilation constituent l'une des solutions les plus couramment utilisées dans les EFS. On les retrouve notamment dans de nombreuses grandes villes (Paris, Londres, New York, Tokyo, Séoul) disposant de transports ferroviaires souterrains. Les systèmes de ventilation sont essentiels pour maintenir une bonne qualité de l'air dans les EFS en assurant un renouvellement efficace de l'air et en évacuant les particules présentes. L'installation de systèmes de ventilation modernes, équipés de filtres à particules haute efficacité (HEPA, pour High Efficiency Particulate Air), peut réduire considérablement les concentrations de PM dans l'air ambiant. De plus, l'optimisation des flux d'air à travers les stations et les tunnels, en combinant ventilation mécanique et naturelle, peut améliorer la dispersion des polluants.

Dans une étude menée par Moreno et al. (2014), il a été démontré que les stations de métro de Barcelone, dotées d'accès aux quais aux deux extrémités et bien ventilées grâce à des systèmes de

ventilation, présentaient la meilleure qualité de l'air parmi les enceintes ferroviaires souterraines (EFS) étudiées, notamment les stations Badalona et Paral-lel (Figure 1.11). De même, Kim et al. (2016) ont observé une amélioration de la qualité de l'air dans une EFS de Séoul grâce à l'installation de ventilateurs aux points de ventilation naturelle situés dans un tunnel équipé de portes palières. Le système de ventilation est centralisé et contrôlé par des modules de capteurs et de traitement des données, notamment le PM_{10}, le CO_2, la température et l'humidité ; les ventilateurs se mettaient en marche automatiquement dès que les concentrations dépassaient les seuils d'exposition autorisés, fixés à 150 $\mu g/m^3$ pour les PM_{10} selon la réglementation sur la qualité de l'air intérieur mise en œuvre par le ministère de l'Environnement de Corée.

Récemment, l'étude d'Izadi et al. (2022) a montré que le système UPE (extraction sous le quai, ou under-platform exhaust en anglais) peut réduire efficacement la concentration de particules dans les EFS, en diminuant ces concentrations de plus de 50 %. Cette étude a également démontré que le système UPE est un dispositif efficace pour contrôler la concentration de particules, car il capture une grande partie des émissions avant leur diffusion dans la station. Contrairement au système OPE (extraction au-dessus du quai, ou over-platform exhaust en anglais), qui repose sur une extraction au-dessus du quai, le système UPE limite davantage la propagation des particules dans l'environnement souterrain (Figure 1.19).

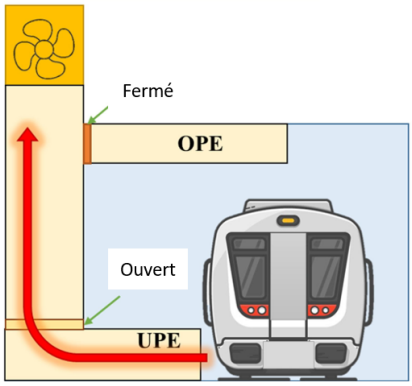


Figure 1.19 : Fonctionnement du système UPE (adapté de Izadi et al. 2022)

Selon Wen et al. (2020), une EFS idéale dotée d'une ventilation et d'un environnement spatial optimaux se compose des éléments suivants (Figure 1.20) : des entrées et conduits d'air situés à l'écart des sources de pollution et d'humidité, un atrium permettant de réduire la vitesse du vent de piston et d'évacuer les polluants, des conduits de ventilation situés au sommet de l'atrium pour l'évacuation des polluants et de la fumée, des conduits inclinés pour moduler le vent de piston et la pression, et des entrées spécifiques adaptées aux conditions climatiques. L'entrée A, coudée, est conçue pour éviter la pénétration du vent froid dans les régions froides, tandis que l'entrée B, de configuration droite, exploite le vent de piston pour favoriser l'évacuation de la chaleur dans les climats tropicaux.

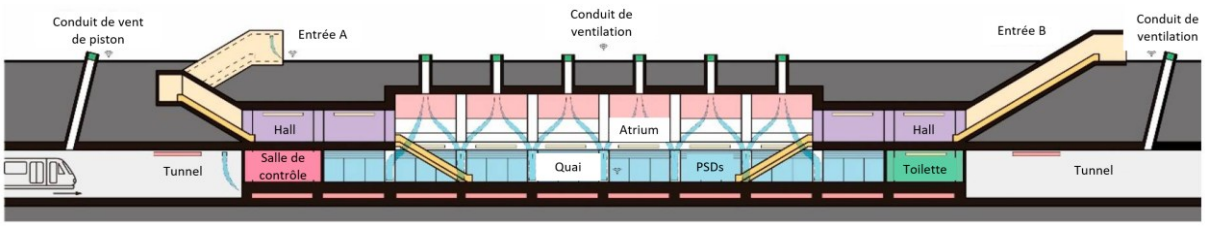


Figure 1.20 : Système de ventilation dans une EFS (adapté de Wen et al. 2020)

Bien que ces systèmes aient des effets positifs notables, leur installation sur un réseau ferroviaire étendu implique une consommation énergétique importante, ce qui représente un inconvénient majeur pour les infrastructures ferroviaires.

1.3.2.2 Equipements de purification de l'air

En complément des systèmes de ventilation, les dispositifs de purification de l'air représentent également une stratégie active et largement adoptée pour améliorer la qualité de l'air dans les EFS. Ces dispositifs capturent les particules grâce à des filtres, réduisant ainsi leur concentration dans l'air ambiant. Ils peuvent être intégrés aux systèmes de ventilation ou installés de manière autonome dans les zones particulièrement polluées, comme les quais ou les tunnels. Dans la littérature, les technologies les plus couramment étudiées pour améliorer la qualité de l'air dans les EFS incluent les précipitateurs électrostatiques, les filtres magnétiques, les bio-filtres végétaux et les collecteurs de poussière.

Les filtres électrostatiques sont particulièrement utilisés pour capter les particules dans les environnements souterrains. Le principe de fonctionnement des filtres électrostatiques est illustré à la Figure 1.21. L'air, initialement neutre (en vert), traverse un champ de fils à très haute tension, symbolisés par les barres verticales vertes. Ces fils, chargés négativement, ionisent les particules en leur transférant une charge négative. Les particules ainsi chargées (en bleu) sont ensuite piégées par des plaques collectrices chargées positivement.

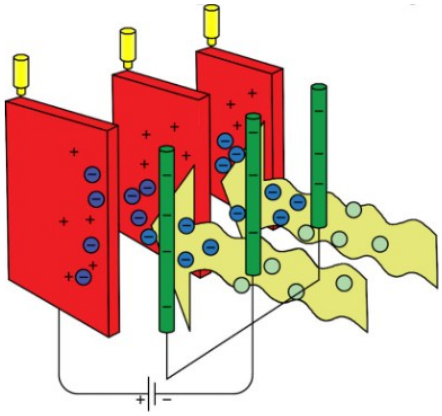


Figure 1.21 : Modèle de précipitateur électrostatique. (https://www.iqsdirectory.com/articles/air-pollution-control/electrostatic-precipitators.html)

Dans une étude menée par Tokarek & Bernis (2006), ils ont évalué l'adaptabilité de la précipitation électrostatique au contexte des EFS et ont exploré la possibilité d'équiper une station entière avec cette technologie. L'évaluation de l'efficacité des précipitateurs a été réalisée en mesurant le nombre de particules avant et après filtration. Leurs résultats ont montré qu'après un an d'utilisation, l'efficacité des filtres électrostatiques diminuait de 15 %, soulignant ainsi la nécessité d'un entretien régulier pour maintenir leur performance. À l'aide d'une simulation basée sur la mécanique des fluides numérique (CFD), ils ont conclu que l'installation de 20 filtres électrostatiques le long du quai d'une EFS parisienne pourrait réduire significativement les concentrations de particules, passant de 230 µg/m³ à 135 µg/m³ en moyenne journalière. Dans l'étude de Pretot et al. (2022), portant sur l'évaluation de l'efficacité des systèmes de filtration dans une EFS du réseau SNCF, les auteurs ont observé une réduction au niveau du quai d'environ 25 % des PM_{10} et de 30 % des $PM_{2.5}$ en comparant deux périodes avec le système en marche à une période sans filtration.

Étant donné que le fer est le composé chimique le plus présent dans les particules (PM) des environnements souterrains, l'application de filtres magnétiques permet également d'éliminer les particules magnétiques présentes dans l'air des EFS (Son et al., 2014a, 2019; Font et al., 2019). Son et al. (2019) ont testé deux types de filtres magnétiques hybrides, à savoir les filtres à cascade magnétique et les filtres aimant-aimant, afin de réduire les concentrations de PM_{10} sur les quais. Les résultats de leurs essais sur le terrain ont montré que l'efficacité de filtration des PM_{10} des filtres à cascade magnétique (40,5 %) était nettement supérieure à celle des filtres aimant-aimant (27 %). Par ailleurs, ils ont observé qu'avec un filtre hybride, l'efficacité de filtration diminue à mesure que la taille des

particules diminue. De plus, leur étude a mis en évidence une réduction de l'efficacité des deux filtres lorsque les concentrations de PM_{10} et la vitesse de l'air augmentent. Au-delà de 8 m/s pour le filtre à cascade magnétique et de 2,4 m/s pour le filtre aimant-aimant, l'efficacité de filtration diminue significativement, suggérant ainsi certaines limitations dans l'application de ces technologies en milieu souterrain.

Dans le même contexte, Kim et al. (2021) ont montré que les bio-filtres végétaux peuvent réduire le taux de fer contenu dans les particules, avec une réduction de 1 % en comparant l'air entrant et l'air sortant. Toutefois, compte tenu de la faible efficacité de filtration observée, il est peu probable que cette réduction ait un impact significatif sur l'exposition des usagers et, par conséquent, sur leur santé.

Le choix du type de filtre dépend de la taille et de la composition des particules à traiter, mais dans tous les cas, un entretien régulier des filtres est nécessaire pour maintenir leur efficacité sur le long terme.

Concernant les collecteurs de poussière, Sim et al. (2018) ont installé des collecteurs de poussière à persiennes sous les trains de la ligne 5 du métro de Séoul et ont mené des tests sur le terrain afin d'évaluer leur efficacité dans la réduction des PM_{10} dans les tunnels souterrains. Les résultats ont montré que plus la vitesse du train est élevée, plus le débit d'air traversant les collecteurs augmente, ce qui entraîne une meilleure capture des particules de plus petite taille. Leur étude a révélé que la taille de coupure des particules diminuait à mesure que la vitesse du train augmentait, passant de 9,7 µm à 5 km/h à 4,9 µm à 65 km/h. Ces résultats confirment que l'efficacité de collecte des particules augmente avec la vitesse du train, atteignant son maximum à 65 km/h, qui correspond généralement à la vitesse la plus élevée des trains de la ligne 5 du métro de Séoul.

Dans les EFS, des rideaux d'air sont également utilisés comme solution pour réduire les particules au niveau des sorties de la gare. Chen et al. (2021), utilisant la CFD et un réseau de neurones artificiels, ont montré que l'efficacité d'interception des particules de différents diamètres (1 µm, 2,5 µm et 10 µm) est d'environ 40 % par rapport à l'absence de rideaux d'air aux sorties de l'EFS.

Outre les systèmes de filtration, des équipements de purification de l'air tels que les véhicules de nettoyage des tunnels sont également utilisés. Cependant, leur déploiement en milieu souterrain est complexe en raison de l'incompatibilité entre le nettoyage à l'eau et la présence d'équipements électriques sensibles dans les tunnels. Cette contrainte limite leur efficacité et nécessite des adaptations spécifiques pour éviter tout risque lié à l'humidité et garantir la sécurité des infrastructures ferroviaires. Chen et al. (2017) ont déployé des véhicules de nettoyage à haute pression pour nettoyer les rails et les parois des tunnels dans le système de métro de Taipei, avec un point de mesure placé sur le quai d'une EFS en amont. Les concentrations de PM_{10} et $PM_{2.5}$ ont montré une baisse progressive, des tunnels vers les quais, puis vers les sorties de la station. À court terme, dès le deuxième jour après le nettoyage, une réduction de 81,4 % pour les PM_{10} et de 78 % pour les $PM_{2.5}$ a été observée. À plus long terme, les niveaux de PM_{10} ont diminué de 45,9 % après 3,5 mois et ceux de $PM_{2.5}$ de 71,3 % après deux mois, comparés aux valeurs mesurées la veille du nettoyage. La baisse continue des PM_{10}, observée durant cette étude, pourrait être liée au dépôt des particules sur les parois poreuses du tunnel, ainsi qu'à l'effet filtrant du système de climatisation dans l'EFS.

1.4 Conclusion du chapitre

Ce chapitre a approfondi l'analyse des caractéristiques de la qualité de l'air dans les EFS, mettant en exergue la complexité des particules en suspension. Les sources intérieures, telles que les opérations ferroviaires et les activités humaines, ont été identifiées comme les principales contributrices aux niveaux de particules, tandis que les sources extérieures ajoutent une couche supplémentaire de complexité en influençant également les concentrations.

La variabilité des concentrations de particules a été examinée sous les angles spatial et temporel, et cette analyse a permis de mettre en lumière les principaux facteurs d'influence. Ces facteurs comprennent non seulement la fréquence des trains et l'effet piston, mais aussi la profondeur des stations et la saisonnalité, soulignant l'importance d'une conception et d'une gestion adaptées des infrastructures et des systèmes de ventilation.

Pour améliorer la qualité de l'air dans les EFS, plusieurs stratégies prometteuses ont été discutées. L'utilisation de matériaux non métalliques dans les systèmes de freinage et l'installation de portes palières sont des mesures efficaces, tandis que l'optimisation des systèmes de ventilation peut réduire de manière significative les concentrations de particules, même si leur mise en place peut s'avérer complexe et coûteuse. Ces défis justifient l'exploration de solutions alternatives plus simples à déployer.

En conclusion, assurer la sécurité et la santé des usagers des EFS exige une approche globale qui allie compréhension approfondie des sources de pollution, évaluation rigoureuse des risques et mise en œuvre de solutions technologiques adaptées. Les recherches futures devraient continuer à innover et à évaluer des stratégies efficaces pour la gestion de la qualité de l'air dans ces milieux urbains complexes.

CHAPITRE 2. DISPOSITIFS DE MESURE ET SITES D'ÉTUDE

Le Réseau Express Régional d'Île-de-France, ou RER, est un réseau de transport en commun indispensable qui relie Paris à sa banlieue. Depuis sa création en 1977, il est exploité conjointement par la SNCF et la RATP (Régie autonome des transports parisiens). Le réseau se compose de cinq lignes principales : les lignes A et B, gérées en collaboration par les deux opérateurs, et les lignes C, D et E, dont la gestion est entièrement assurée par la SNCF. Sur les 249 gares que compte le réseau, 157 relèvent de la SNCF, tandis que 92 sont gérées par la RATP. Avec 600 kilomètres de voies ferrées et une fréquentation quotidienne d'environ 2,7 millions de passagers, dont près d'un million rien que pour la ligne A, le RER joue un rôle central dans la mobilité francilienne.

La SNCF, consciente des défis posés par la qualité de l'air dans les EFS, a lancé plusieurs campagnes de mesure depuis 2016 en collaboration avec AIRPARIF et AEF. Ces campagnes couvrent 24 gares franciliennes souterraines ou mixtes[4] de la SNCF, visant à documenter les niveaux de particules et à faciliter la mise en œuvre de plans d'amélioration et la priorisation des interventions. Des mesures en continu sont effectuées dans trois gares stratégiques : Avenue Foch, Sevran Beaudottes et Magenta, et ce, jusqu'à présent. Ces gares, choisies pour des mesures de longue durée, ont été sélectionnées en fonction de l'ancienneté de leurs infrastructures, du type de système de ventilation et des niveaux de concentration en polluants. Sur ces 24 gares, AIRPARIF a instrumenté 17 gares, tandis que AEF a pris en charge 7 autres. Des expérimentations ont également été menées pour tester différents systèmes de filtration de l'air dans les gares d'Avenue Foch et de Sevran-Beaudottes.

Ce chapitre explore les instruments et les méthodologies employés pour la collecte des données relatives à la qualité de l'air dans ces EFS, ainsi que les dispositifs de filtration de l'air expérimentés. Nous aborderons également les différents sites de mesure, en commençant par ceux ayant fait l'objet de mesures de courte durée, puis en nous penchant sur ceux concernés par des mesures de longue durée, et enfin en terminant par les stations extérieures de surveillance de la qualité de l'air.

2.1 Instrumentation

Dans cette section, nous détaillerons les équipements utilisés pour la surveillance de la qualité de l'air dans les EFS, y compris les systèmes de filtration que la SNCF a mis à l'essai.

2.1.1 Instruments de mesures de la qualité de l'air en EFS

Pour évaluer l'exposition des usagers et des travailleurs dans les EFS, le guide de l'INERIS (2023) recommande la surveillance de plusieurs types de polluants, distinguant les polluants prioritaires des polluants complémentaires. Les polluants prioritaires incluent les particules en suspension (PM_{10}, $PM_{2.5}$), les métaux présents dans les PM_{10} (Fe, Ba, Cu, Zn, Sb, Mn, Ni, Pb, As, Cd et Cr), le CO_2 (indicateur de confinement en lien avec le renouvellement de l'air), ainsi que des paramètres de confort tels que la température et l'humidité. Les polluants complémentaires regroupent les composés carbonés (carbone élémentaire et organique), la silice, l'amiante, les HAP (hydrocarbures aromatiques polycycliques), les BTEX (Benzène, Toluène, Ethylbenzène, Xylènes) et les oxydes d'azote (NOx).

Concernant les techniques et appareils de mesure, plusieurs méthodes adaptées aux spécificités des polluants sont mises en œuvre, en respectant les normes françaises et européennes. Les mesures en

[4] Par gares mixtes, on entend des gares dont une partie des quais est en contact à l'air extérieur.

continu des particules PM_{10} et $PM_{2.5}$ sont assurées par des analyseurs automatiques, notamment le TEOM (Tapered Element Oscillating Microbalance), avec ou sans le module FDMS[5] (Filter Dynamic Measurement System). Ces dispositifs opèrent en continu, effectuant des mesures toutes les 15 minutes conformément à la norme NF EN 16450, ce qui permet de capturer les variations temporelles fines des concentrations de particules.

En outre, les paramètres environnementaux, tels que le CO_2, la température et l'humidité relative, sont mesurés en continu à l'aide d'un analyseur de CO_2 par corrélation infrarouge et de capteurs analogiques pour la température et l'humidité relative. Ces dispositifs assurent un suivi en temps réel, avec un pas de temps de 15 minutes, offrant ainsi une surveillance précise des conditions de confinement et de confort.

Conformément au protocole de l'INERIS (2023), les mesures sont effectuées de préférence au milieu des quais sélectionnés afin d'assurer une représentativité des concentrations sur l'ensemble du quai. Les appareils de mesure sont installés à une hauteur comprise entre 0,8 et 2,0 m au-dessus du sol du quai. Ces instruments sont regroupés dans une baie de mesure dédiée, illustrée dans la Figure 2.1.



Figure 2.1 : Baie de mesure de la qualité de l'air installée dans la gare de Musée d'Orsay lors de la campagne de mesure du 10 avril au 26 avril 2018 (AIRPARIF, 2018).

2.1.1.1 Analyseur automatique de particules

Le TEOM est un instrument conçu pour la mesure continue de la concentration massique des particules en suspension dans l'air ambiant. Son principe de fonctionnement repose sur l'oscillation d'un élément conique auquel est fixé un filtre destiné à capturer les particules. Lorsqu'un flux d'air traverse ce filtre, les particules s'y déposent, augmentant ainsi progressivement la masse de l'élément oscillant. La variation de fréquence de cette oscillation est directement proportionnelle à la masse des particules accumulées. Comme dans tout système oscillant, une augmentation de masse entraîne une diminution de la fréquence d'oscillation. La Figure 2.2 présente un schéma détaillant le fonctionnement de cet instrument.

[5] FDMS (Filter Dynamic Measurement System) est un module complémentaire au TEOM permettant de corriger l'évaporation des fractions volatiles des particules.

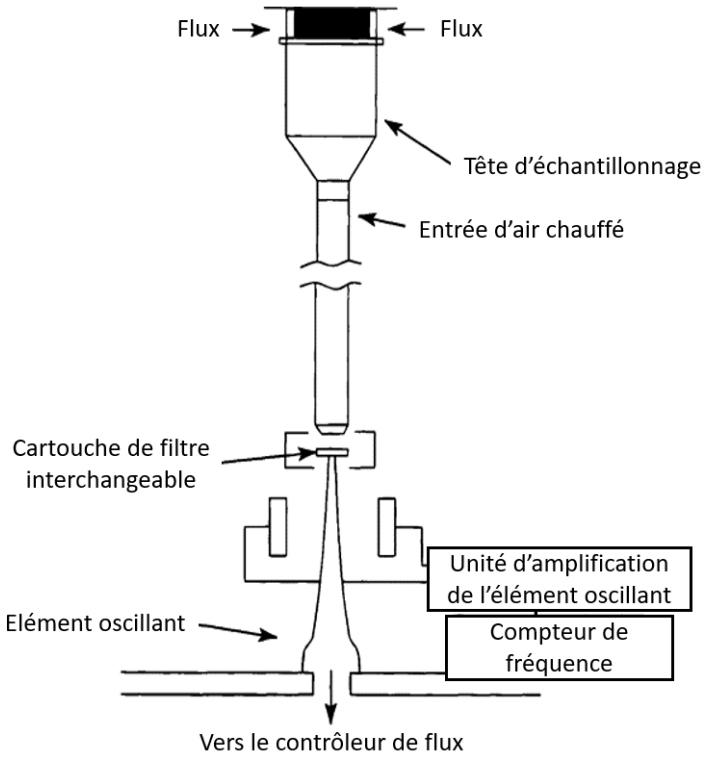


Figure 2.2 : Principe du TEOM (adapté de Su et al. 2020)

Pour les mesures réalisées par AEF, le modèle TEOM 1405-D de Thermo Scientific est utilisé (Figure 2.3a). Cet analyseur dichotomique, opérant avec un pas de temps quart horaire (15 minutes), permet de mesurer simultanément deux fractions de particules en suspension : les particules fines ($PM_{2.5}$) et les particules grossières ($PM_{2.5-10}$). Les particules grossières sont définies comme les particules dont le diamètre aérodynamique est compris entre 2,5 et 10 µm. Son principe est le suivant : ce TEOM 1405-D est équipé d'une tête de prélèvement PM_{10} conforme aux normes de l'US EPA[6], qui capture l'ensemble des particules de diamètre inférieur à 10 µm. Ces particules sont ensuite séparées en deux catégories grâce à un impacteur virtuel, un dispositif conçu pour séparer les particules en fonction de leur taille aérodynamique. Ce système repose sur un principe de division du flux d'air en deux trajets distincts. Une première fraction de l'air est déviée à un angle de 90°. Seules les particules les plus fines, dont le diamètre est inférieur à 2,5 µm, possèdent une faible inertie leur permettant de suivre cette déviation et d'être collectées sur un filtre dédié. En parallèle, une seconde fraction d'air suit une trajectoire en ligne droite. Les particules de diamètre supérieur à 2,5 µm, en raison de leur inertie plus importante, ne parviennent pas à suivre la déviation du flux principal et continuent leur trajectoire rectiligne avant d'être collectées sur un second filtre. Toutefois, une partie des particules fines est également piégée dans le filtre dédié aux particules grossières. En effet, environ 10 % des particules $PM_{2.5}$ sont capturées avec la fraction $PM_{2.5-10}$, tandis que les 90 % restants sont correctement collectés sur le filtre spécifique aux particules fines. La mesure de ces fractions est assurée par deux microbalances coniques (TEOM), maintenues à une température constante de 30 °C afin d'éviter toute condensation d'eau sur le filtre de collection (Misra et al., 2001; Green et al., 2009). Les concentrations en PM_{10} sont donc déduites à partir des données collectées pour chaque fraction (particules fines et particules grossières).

Thermo Scientific recommande des maintenances régulières pour garantir la précision des mesures du TEOM 1405-D. Les filtres pour $PM_{2.5}$ et $PM_{2.5-10}$ doivent être remplacés tous les 30 jours ou

[6] US EPA : United States Environmental Protection Agency, agence fédérale américaine créée en 1970, en charge de la protection de l'environnement et de la santé publique.

lorsque leur charge atteint 100 %. L'entrée d'échantillonnage et l'impacteur virtuel doivent être nettoyés à la même fréquence. Enfin, la pompe d'échantillonnage doit être révisée tous les 18 mois.

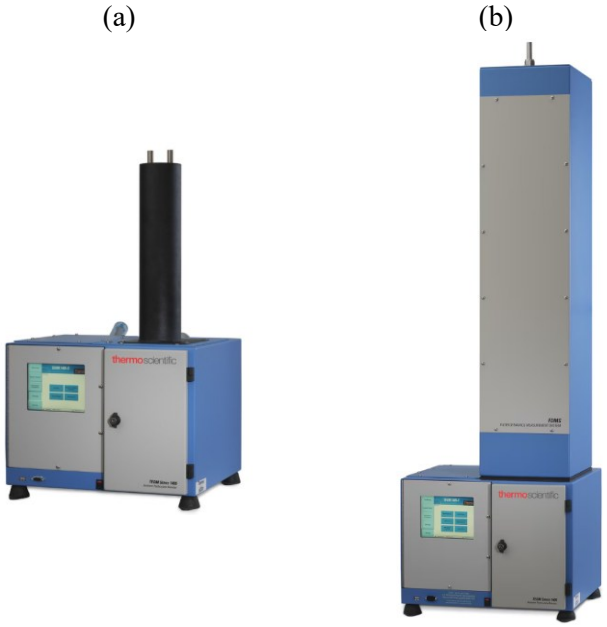


Figure 2.3 : Analyseurs automatiques (a) TEOM 1405-D et (b) TEOM 1405-F (https://www.thermofisher.com/)

Pour les mesures réalisées par AIRPARIF, le modèle TEOM FDMS 1405-F de la société TEI (Thermo Environmental Instruments) est utilisé (Figure 2.3b). Ce dispositif, équipé de la technologie FDMS (Filter Dynamic Measurement System), fournit des mesures en continu avec un pas de temps quart horaire, intégrant automatiquement les fractions volatiles et semi-volatiles des particules. La technologie FDMS assure une correction en temps réel des pertes de masse dues à la volatilisation des particules collectées sur le filtre TEOM. Pour cela, l'échantillon d'air est d'abord soumis à un processus de séchage en amont de la microbalance, réalisé à l'aide d'une membrane déshydratante de type Nafion. Ce séchage permet d'éliminer l'eau liée aux particules avant leur dépôt sur le filtre. Ensuite, le FDMS compense les pertes de masse liées à l'évaporation des fractions volatiles et semi-volatiles des particules en arrêtant temporairement la collecte des échantillons à intervalles réguliers. Pendant cette phase d'arrêt, toute variation de masse du filtre est enregistrée et attribuée à la volatilisation des particules instables. Cette correction permet d'obtenir une estimation plus précise et fiable de la concentration massique des particules. Pour assurer la mesure de la fraction PM_{10} et $PM_{2.5}$ simultanément, 2 TEOM (maintenus à 30°C) ont été utilisés. L'un des analyseurs est doté d'une tête de prélèvement PM_{10} conforme aux normes de l'US EPA, tandis que l'autre est équipé d'un Sharp Cut Cyclone pour les $PM_{2.5}$.

Pour garantir des mesures fiables avec le TEOM 1405-F, Thermo Fisher Scientific recommande plusieurs opérations de maintenance régulières. Le sécheur FDMS doit être remplacé ou rénové une fois par an, ou si nécessaire. Le filtre TEOM doit être remplacé tous les 30 jours ou lorsque la charge atteint 100 %. L'entrée d'échantillonnage doit être nettoyée à chaque remplacement du filtre. Enfin, la pompe d'échantillonnage doit être révisée tous les 18 mois, ou selon les besoins.

Il convient de souligner que les modèles utilisés par AEF et AIRPARIF respectent les exigences de la norme NF EN 12341[7], garantissant ainsi la conformité et la qualité des mesures réalisées. Les spécifications techniques des modèles TEOM 1405-D et TEOM FDMS 1405-F sont présentées dans le Tableau 2.1.

[7] NF EN 12341 « Air ambiant – Méthode normalisée de mesurage gravimétrique pour la détermination de la concentration massique MP10 ou MP2,5 de matière particulaire en suspension », août 2023.

Tableau 2.1 : Spécifications techniques des analyseurs TEOM (https://www.thermofisher.com/)

Caractéristique	Spécifications
Débit	12 l/min
Plage de mesure	0 à 1 000 000 µg/m³
Température de fonctionnement	8 à 25 °C
Précision	± 2 µg/m³ (moyenne sur une heure), ± 1 µg/m³ (moyenne sur 24 heures)
Résolution	0,1 µg/m³

Lors d'une étude menée par AIRPARIF (2018) dans le cadre d'une campagne de mesure du 10 au 26 avril 2018 à la gare de Musée d'Orsay, une comparaison entre ces deux modèles de TEOM a été réalisée. Bien qu'une corrélation supérieure à 0,95 ait été observée entre les deux appareils, les résultats ont mis en évidence une sous-estimation de 10 à 15 % des concentrations en particules (PM_{10} et $PM_{2.5}$) mesurées par le TEOM 1405D par rapport à celles enregistrées par le TEOM FDMS 1405F. AIRPARIF a conclu que ces écarts pourraient s'expliquer à la fois par des différences technologiques entre les instruments et par d'éventuels dysfonctionnements techniques.

2.1.1.2 Analyseur de CO_2 par corrélation infrarouge

Pour la mesure des concentrations en CO_2, un analyseur de CO_2 modèle 410i de Thermo Scientific (Figure 2.4a) par corrélation infrarouge est également introduit dans la baie de mesure en plus des analyseurs automatiques de particule respectant la norme NF EN ISO 16000-26[8], suivant les recommandations de l'INERIS (2023). C'est un instrument couramment utilisé pour la mesure précise et continue des concentrations de dioxyde de carbone dans l'air ambiant. De même qu'avec l'analyseur automatique de particules, les mesures sont faites au pas de temps quart horaire. Le Tableau 2.2 résume les spécifications techniques principales de l'analyseur de CO_2.

Tableau 2.2 : Spécifications techniques de l'analyseur de CO_2 (https://www.thermofisher.com/)

Caractéristique	Spécifications
Débit	1 l/min
Plage de mesure	0 à 10 000 ppm
Température de fonctionnement	5 à 45 °C
Précision	± 1 % de la mesure
Dérive du zéro	Niveau haut : ± 40 ppm ; Standard : ± 1 ppm[9]

Son fonctionnement repose sur la technologie infrarouge non dispersive, qui mesure l'absorption sélective de la lumière infrarouge par les molécules de CO_2 à une longueur d'onde de 4,26 µm. Le faisceau infrarouge est d'abord modulé par un moteur de modulateur, qui alterne son passage à travers une roue de filtre à gaz contenant alternativement du CO_2 et du N_2. Cette roue agit comme un filtre dynamique, permettant d'isoler l'absorption spécifique du CO_2. Après cette étape, le faisceau passe à travers un filtre passe-bande qui sélectionne uniquement la longueur d'onde d'intérêt (4,26 µm), correspondant à l'absorption maximale du CO_2. La radiation filtrée traverse ensuite une cellule de mesure contenant l'échantillon d'air. À l'intérieur de la cellule, un système de miroirs réfléchissants permet d'allonger le chemin optique pour augmenter la sensibilité de la mesure. Le faisceau sort ensuite de la cellule pour atteindre un détecteur infrarouge, où l'intensité résiduelle est mesurée. Le signal capté par le détecteur est amplifié par un préamplificateur avant d'être traité par un système électronique. Ce système applique des corrections et convertit les signaux bruts en résultats linéarisés. Les données ainsi

[8] NF EN ISO 16000-26 « Air intérieur – Partie 26 : Stratégie d'échantillonnage du dioxyde de carbone (CO_2) », octobre 2012.

[9] Le niveau standard est utilisé pour des faibles concentrations, tandis que le niveau haut est adapté à des concentrations plus élevées.

obtenues sont affichées en temps réel sur l'appareil et sont accessibles via des sorties analogiques et numériques (Figure 2.4b).

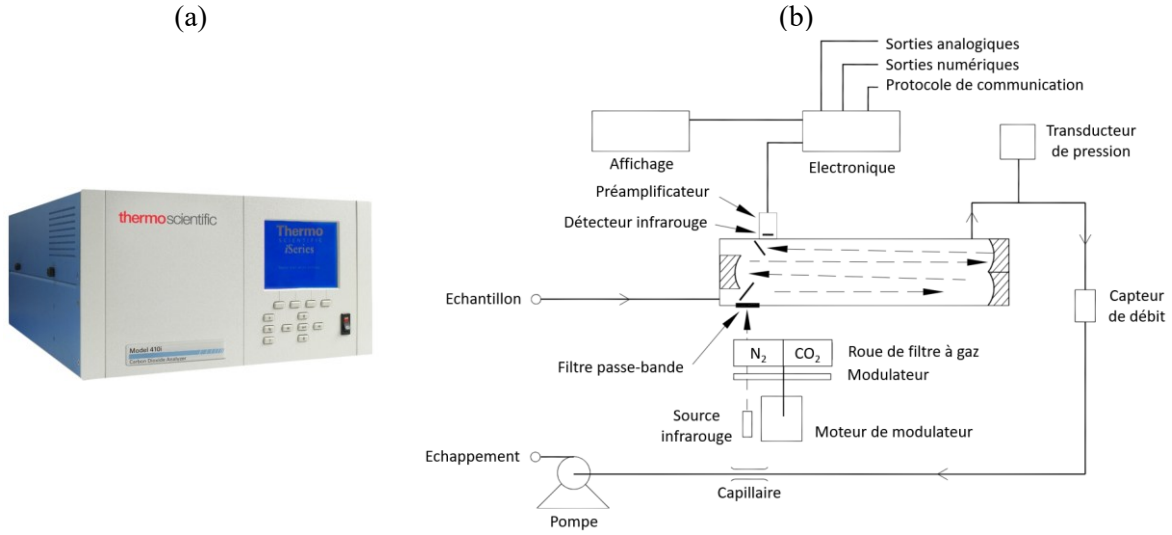


Figure 2.4 : (a) Analyseur de CO_2 modèle 410i, (b) Schéma fonctionnel illustrant son principe de fonctionnement (adapté de https://www.thermofisher.com/)

2.1.1.3 Capteur de température et d'humidité relative

Pour ce qui est des paramètres de confort (Température, Humidité relative), le capteur analogique TRA-102D (0 – 10 Volt) de chez ATAL (https://www.atal.nl/) est également installé dans la baie de mesure avec les analyseurs automatiques de particules et de CO_2 (Figure 2.5).



Figure 2.5 : Capteur de température et d'humidité relative (https://www.atal.nl/)

Ce transmetteur avec sortie analogique 0 – 10 Volt est conçu pour mesurer la température de l'air ambiant (en °C ou °F) et l'humidité relative de l'air ambiant. Les valeurs mesurées et calculées sont affichées sur un écran LCD à deux lignes, avec la température sur la première ligne et l'humidité relative sur la seconde. L'appareil est configuré via un PC à l'aide du câble de communication ATS-AC17, permettant d'assigner les valeurs aux sorties et de calibrer le dispositif.

Le Tableau 2.3 résume les spécifications techniques principales de l'analyseur de température et d'humidité relative.

Tableau 2.3 : Spécifications techniques du capteur analogique de température et d'humidité relative

Caractéristique	Spécifications
Température de fonctionnement de l'électronique	-30 à +80 °C
Compensation en température pour la mesure d'humidité relative	-30 à +125 °C
Précision de la mesure de température	±0,4 °C de -30 à +100 °C ; ±0,4 % au-dessus de +100 °C
Plage d'humidité relative	0 à 100 % RH
Précision de l'humidité relative	±2,5 % RH de 5 à 95 % à 23 °C

2.1.1.4 Contrôle qualité des données

Afin d'assurer que les données collectées atteignent un niveau de qualité optimal, des opérations régulières de vérification, maintenance et calibration des instruments de mesure sont réalisées par AIRPARIF et AEF. Ces deux organismes jouent un rôle complémentaire dans la garantie de la fiabilité et de la représentativité des mesures, en s'appuyant sur des méthodologies adaptées à leurs contextes spécifiques.

Concernant AIRPARIF, un processus structuré de validation est mis en œuvre pour assurer que les données respectent des critères stricts de qualité. Ce processus repose sur des opérations régulières, notamment la vérification technique, la maintenance préventive des instruments et leur étalonnage périodique. La validation des données est réalisée en deux étapes distinctes et complémentaires. Une validation technique est effectuée quotidiennement par un technicien sur place pour détecter rapidement les anomalies ou dysfonctionnements des équipements de mesure, comme des capteurs défectueux. Une validation environnementale est ensuite conduite chaque semaine afin de vérifier que les conditions extérieures ou des événements ponctuels, tels qu'une coupure d'électricité ou une perturbation locale, n'ont pas affecté la représentativité des mesures. Les données jugées non conformes sont exclues de la base de données. Par ailleurs, les données collectées ne sont considérées comme exploitables que si elles répondent à un critère minimal de validité. Par exemple, pour qu'une moyenne horaire soit calculée, au moins 3 mesures enregistrées sur des périodes de 15 minutes doivent être valides, qu'elles soient consécutives ou non (AIRPARIF, 2022b).

De son côté, AEF applique un protocole reposant sur les recommandations techniques des constructeurs et sur l'analyse des niveaux de concentration en particules dans les gares souterraines. La validation des données s'appuie sur le traitement des données brutes, permettant d'identifier les anomalies telles que le bruit électronique, le colmatage des filtres, les dépassements de seuils pour des paramètres comme le débit, la température ou l'hygrométrie, ainsi que les interruptions prolongées supérieures à quinze minutes. Ces interruptions, généralement dues à des coupures électriques, peuvent être volontaires, dans le cadre de maintenances programmées, ou involontaires, résultant de défaillances techniques (AEF, 2021c; b).

Des étapes de prétraitement des données ont été réalisées par nos soins pour compléter les vérifications déjà effectuées par AIRPARIF et AEF. Tout d'abord, une correction de l'horodatage a été effectuée. Les données, initialement enregistrées en temps universel coordonné (UTC), ont été converties en heure locale (heure de Paris, UTC+1 en hiver et UTC+2 en été). Cette conversion, réalisée à l'aide de scripts Python, a pris en compte les changements d'heure saisonniers pour aligner les données avec les périodes de circulation et les variations horaires spécifiques à la région parisienne, garantissant une cohérence temporelle et une interprétation adaptée au contexte local. Par ailleurs, un nettoyage des données non représentatives a été mené pour résoudre un problème identifié dans les concentrations en particules. Certaines valeurs de $PM_{2.5}$ dépassaient celles des PM_{10}, ce qui est anormal, car les particules $PM_{2.5}$, plus fines, doivent logiquement avoir des concentrations inférieures ou égales à celles des PM_{10}.

Ces anomalies, probablement dues à des erreurs de mesure ou à des défaillances instrumentales, ont été corrigées en supprimant les entrées concernées.

2.1.2 Instruments de filtration des particules

Dans le cadre des efforts visant à améliorer la qualité de l'air en milieu ferroviaire souterrain, trois technologies de filtration de l'air ont été testées dans des conditions réelles d'exploitation par la SNCF. Ces systèmes, spécifiquement conçus pour réduire les concentrations de particules PM_{10} et $PM_{2.5}$, ont fait l'objet d'une analyse approfondie dans ce travail de thèse. Les technologies étudiées comprennent la technologie par ionisation électrostatique positive, la technologie par voie humide, et la filtration mécanique.

2.1.2.1 Par ionisation

La première technologie, mise en œuvre par Air Liquide, repose sur un principe d'ionisation électrostatique positive pour capturer les particules en suspension dans l'air. La Figure 2.6 illustre le fonctionnement de ce procédé.

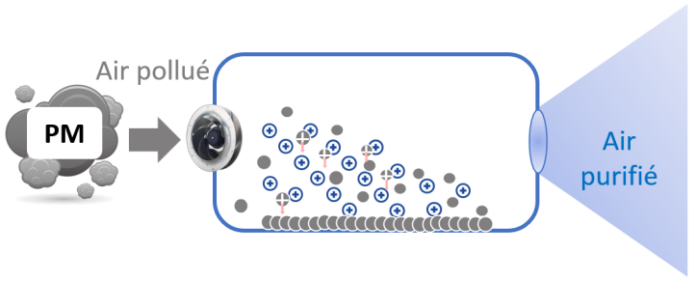


Figure 2.6 : Principe de fonctionnement de la technologie d'ionisation électrostatique positive

L'air pollué, aspiré par un ventilateur, est dirigé dans un compartiment contenant deux plaques parallèles : une cathode et une anode. Une forte différence de potentiel entre ces plaques génère un champ électrostatique qui produit des ions positifs. Ces ions se lient aux particules présentes dans l'air, leur conférant une charge positive. Les particules ainsi chargées sont ensuite attirées et capturées par des plaques collectrices neutres. Une fois les particules éliminées, l'air purifié est réinjecté dans l'environnement, contribuant à améliorer la qualité de l'air pour les usagers et le personnel des infrastructures ferroviaires. Cette technologie se révèle particulièrement efficace pour réduire les concentrations de particules PM_{10} et $PM_{2.5}$. Son efficacité varie cependant en fonction de la taille des particules : plus elles sont grosses, plus le procédé est performant (Tokarek & Bernis, 2006; Zheng et al., 2019).

Cependant, elle présente certains inconvénients. Un risque de production d'ozone peut survenir en raison du champ électrostatique intense (Afshari et al., 2020), tant en tension positive qu'en tension négative. Une étude menée par Boelter & Davidson (1997) a démontré qu'un épurateur CVC en tension négative peut multiplier la production d'ozone par six. Bien que les concentrations d'ozone générées soient généralement faibles, elles doivent être surveillées étroitement, car l'ozone est un polluant secondaire qui peut être nocif pour la santé. A titre d'exemple, un filtre à charbon actif positionné en aval du filtre électrostatique permet de réduire les émissions d'ozone de 6 à 39 % (Poppendieck et al., 2014; Geffre, 2024). Par ailleurs, le fonctionnement optimal de cette technologie nécessite une maintenance régulière, notamment le nettoyage des plaques collectrices, afin de garantir une performance stable et durable dans le temps.

2.1.2.2 Par voie humide

La seconde technologie, par voie humide, développée par Starklab, repose sur un procédé innovant de lavage des particules en suspension dans l'air. La Figure 2.7 illustre le principe de fonctionnement de cette technologie.

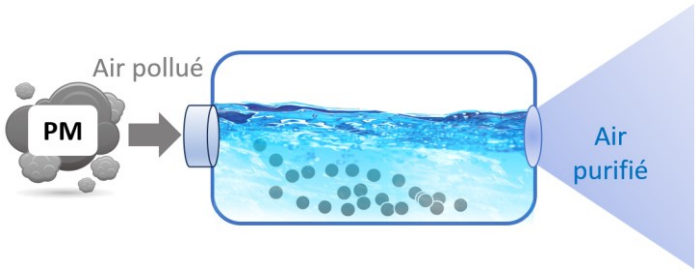


Figure 2.7 : Principe de fonctionnement de la technologie par voie humide

L'air pollué est aspiré par dépression, puis dirigé vers un système de lavage où son passage à travers l'eau génère un bouillonnement. Ce processus facilite le transfert des particules et des gaz polluants vers l'eau. Pour renforcer l'efficacité de cette filtration, du chlorure de calcium est ajouté à l'eau. Cet additif joue un double rôle : il capture les molécules organiques et limite la prolifération de bactéries et de virus, contribuant ainsi à maintenir un environnement plus sain et sécurisé (AIRLAB, 2022).

Ce type de système est conçu pour piéger les particules, en particulier les plus grosses, comme les PM_{10}, plus efficacement que les particules plus fines telles que les $PM_{2.5}$ (Jumlongkul, 2022). Comme la technologie par ionisation électrostatique, cette méthode nécessite une maintenance régulière. Cela inclut notamment le remplacement ou le traitement de l'eau usée afin de garantir des performances optimales et d'éviter tout risque lié à l'accumulation de polluants dans le système.

2.1.2.3 Par filtre mécanique

Enfin, la troisième technologie étudiée repose sur une filtration mécanique, développée par Mann+Hummel. Ce système utilise des filtres mécaniques conventionnels conçus pour piéger efficacement les particules en suspension dans l'air (PM_{10} et $PM_{2.5}$). Bien que sa simplicité et son efficacité soient des atouts, cette technologie nécessite une surveillance et un remplacement régulier des filtres pour garantir une performance continue.

Après avoir examiné les outils dédiés à la surveillance de la qualité de l'air ainsi que les technologies de filtration, nous nous tournons maintenant vers les différents sites où ces mesures et expérimentations ont été réalisées. Cette section présente une exploration approfondie des lieux spécifiques de collecte des données, incluant les gares avec des mesures de courte durée, les gares de référence et les stations de mesure de l'air extérieur.

2.2 Les EFS instrumentées

Parmi les 17 gares instrumentées par AIRPARIF, seules 15 disposaient de mesures exploitables et ont donc été retenues pour cette thèse, dont une gare faisant l'objet de mesures de longue durée, Avenue Foch. Concernant les 7 gares instrumentées par AEF, seulement deux d'entre elles, Sevran-Beaudottes et Magenta, ont été exploitées dans cette étude. Les données des autres gares instrumentées par AEF n'ont pas pu être obtenues. Au total, 17 gares ont été étudiées. Sur ces 17 gares, les gares d'Avenue Foch, de Sevran-Beaudottes et de Magenta ont fait l'objet de mesures de particules en continu sur plusieurs années (5, 4 et 5 années, respectivement).

2.2.1 Gares avec mesures de courte durée

Les campagnes de mesures de courte durée, menées dans 14 gares franciliennes par AIRPARIF, ont joué un rôle clé pour élargir l'analyse de la qualité de l'air à une diversité d'environnements ferroviaires. Bien que limitées à 2 à 3 semaines par site, ces campagnes ont permis de collecter des données essentielles sur les concentrations de PM_{10} et $PM_{2.5}$, offrant ainsi une perspective élargie sur les niveaux de pollution dans le réseau RER. La Figure 2.8 illustre la localisation des 14 gares instrumentées pour ces mesures ponctuelles en Île-de-France. Ces gares couvrent une diversité de configurations architecturales et fonctionnelles : certaines sont souterraines, d'autres mixtes (partiellement ouvertes), offrant ainsi une représentation variée des infrastructures ferroviaires.

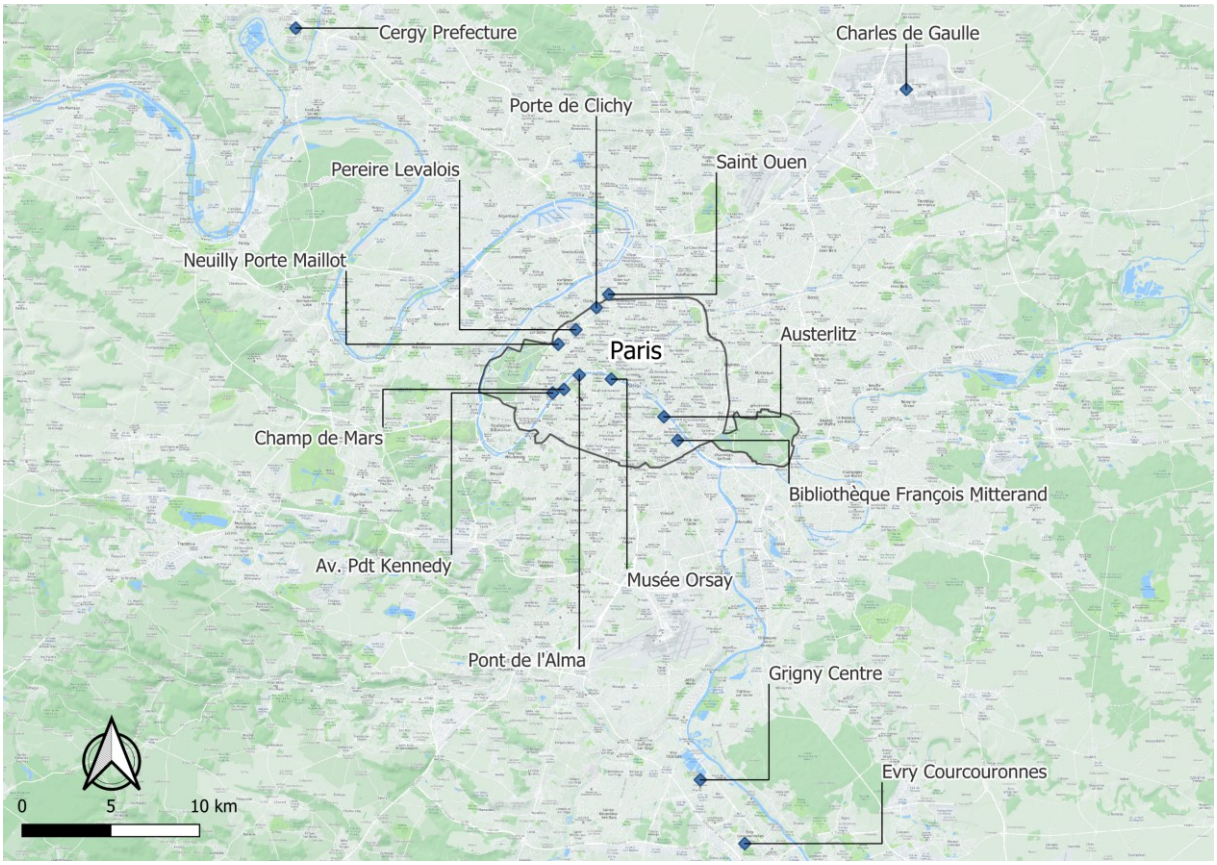


Figure 2.8 : Localisation des gares avec mesures de courte durée en Île-de-France

Le Tableau 2.4 fournit des informations détaillées sur ces gares ayant fait l'objet de campagnes de mesures courtes. Les données incluent le nom de la gare, la ligne de RER desservie, la période de mesure, la profondeur, le type de revêtement des voies, la typologie de la gare, le nombre en moyenne de voyageurs par jour et le nombre moyen de trains.

Compte tenu des facteurs présentés dans le Chapitre 1, certains d'entre eux sont repris dans ce tableau. Pour faciliter la suite de notre analyse sur les facteurs influençant les concentrations en particules, nous avons classé ces facteurs en fonction de leur impact sur la qualité de l'air, selon trois critères illustrés par un code couleur :

- Critère 0 (police normale) : Facteur atténuant – contribue à une réduction des concentrations en particules.
- Critère 1 (*italique*) : Facteur intermédiaire – impact modéré sur les concentrations en particules.
- Critère 2 (**gras**) : Facteur dégradant – entraîne une augmentation des concentrations en particules.

Tableau 2.4 : Caractéristiques des gares instrumentées pour les campagnes de mesures de courte durée.

Nom de la gare	Ligne de RER	Période de mesure	Système de ventilation **	Niveau de profondeur***	Type de revêtement des voies	Type de gare	Nombre de voyageurs par jour (en 2016)	Nombre de trains par jour (JO, SA, DI)
Aéroport Charles de Gaulle 1	B	22/05/2017-11/06/2017	*N*	**-1**	**Ballast**	**Souterraine (2 quais, 4 voies)**	8 830	*(345,* 310*, 309)*
Austerlitz	C	05/09/2016-25/09/2016	*N*	**-2**	**Ballast**	**Souterraine (2 quais, 4 voies)**	**27 536**	**(514, 456, 392)**
Avenue du Président Kennedy	C	02/10/2017-22/10/2017	*N*	0	**Ballast**	Mixte (2 quais, 2 voies)	3 325	(169, 143, 141)
Bibliothèque François Mitterrand	C	17/09/2018-07/10/2018	*N*	*-1*	**Ballast**	**Souterraine (3 quais, 8 voies)**	**47 890**	**(506, 475, 475)**
Cergy Préfecture	A/L*	09/01/2017-29/01/2017	M	*-1*	**Ballast**	*Souterraine (1 quais, 2 voies)*	**20 500**	*(235, 104, 104)*
Champ de Mars	C	18/06/2018-08/07/2018	*N*	-1 (~quai de Seine)	**Ballast**	Mixte (2 quais, 2 voies)	**21 880**	**(432, 352, 317)**
Evry Courcouronnes	D	04/12/2017-24/12/2017	*N*	-1 (extrémité donnant sur l'extérieur)	Béton	Mixte (2 quais, 2 voies)	*17 750*	(195, 149, 125)
Grigny Centre	D	19/06/2017-09/07/2017	*N*	-1 (extrémité donnant sur l'extérieur)	**Ballast**	Mixte (2 quais, 2 voies)	9 940	*(200, 139, 120)*
Musée d'Orsay	C	26/03/2018-15/04/2018	*N*	-1 (extrémité donnant sur l'extérieur)	**Ballast**	Mixte (2 quais, 4 voies)	9 800	**(500, 449, 434)**
Neuilly Porte Maillot	C	18/04/2017-09/05/2017	*N*	*-1*	**Ballast**	*Souterraine (1 quais, 2 voies)*	*11 924*	*(201, 180, 180)*
Pereire Levallois	C	14/05/2018-10/06/2018	*N*	*-1*	**Ballast**	**Souterraine (2 quais, 3 voies)**	9 719	(169, 143, 143)
Pont de l'Alma	C	07/06/2021-28/06/2021	*N*	*-1*	**Ballast**	*Souterraine (2 quais, 2 voies)*	1 011	**319 (JO, SA, DI confondus)**
Porte de Clichy	C	04/09/2017-24/09/2017	*N*	**-3**	Béton	*Souterraine (2 quais, 2 voies)*	8 133	(177, 143, 143)
Saint Ouen	C	20/02/2017-12/03/2017	*N*	**-2**	Béton	Mixte (1 quais, 2 voies)	7 459	(168, 145, 143)

* Ligne L : Ligne du réseau Transilien desservant principalement les banlieues de Paris.
** N : ventilation naturelle ; M : ventilation mécanique
*** Les niveaux indiquent la position de la gare par rapport à la rue : -1 correspond à un niveau juste sous la rue, -2 à un étage plus profond, tandis que 0 est au niveau de la rue.

Il convient de noter que toutes ces gares sont ventilées de manière exclusivement naturelle, sans recours à un système de ventilation mécanique, à l'exception de la gare de Cergy Préfecture, équipée d'un dispositif comprenant 3 extracteurs, 2 centrales de traitement de l'air et 4 extracteurs de désenfumage. Concernant la fréquentation des voyageurs, le comptage a été réalisé en fonction du passage de chaque usager aux bornes de validation des titres de transport.

2.2.2 Gares avec mesures de longue durée

Les gares ayant fait l'objet de mesures de longue durée jouent un rôle central dans cette étude. Elles sont équipées de dispositifs de mesure en continu, fournissant un historique de plus de cinq ans de données. Ces données permettent une analyse approfondie des variations de concentrations de PM sur le long terme. Pour chaque gare de référence, une station de surveillance de la qualité de l'air extérieur d'AIRPARIF a également été utilisée comme site de comparaison pour les mesures de la qualité de l'air ambiant. Le choix de chaque station repose sur sa proximité avec l'EFS concerné, sa localisation en fonction de l'environnement immédiat pour garantir des conditions comparables, et sa nature de station permanente, assurant une mesure continue chaque heure et permettant ainsi un suivi en temps réel de la qualité de l'air.

2.2.2.1 Gare d'Avenue Foch

Mise en service en 1854, la gare d'Avenue Foch est située sur la ligne RER C, desservant plusieurs axes majeurs tels que Pontoise, Versailles-Château, Saint-Quentin-en-Yvelines, Saint-Martin d'Étampes, Dourdan-la-Forêt et Massy-Palaiseau. Depuis avril 2018, elle est équipée d'instruments de mesure permettant le suivi en continu des concentrations de particules en suspension. Cette gare souterraine de configuration simple, sans correspondance, est située au 85 avenue Foch dans le 16e arrondissement de Paris. Le quai central se trouve à une faible profondeur de 5 m sous terre (Figure 2.9) et bénéficie d'une ventilation naturelle, sans recours à un système de ventilation mécanique. La gare ne dispose pas de portes palières (Figure 2.9b).

(a) (b)

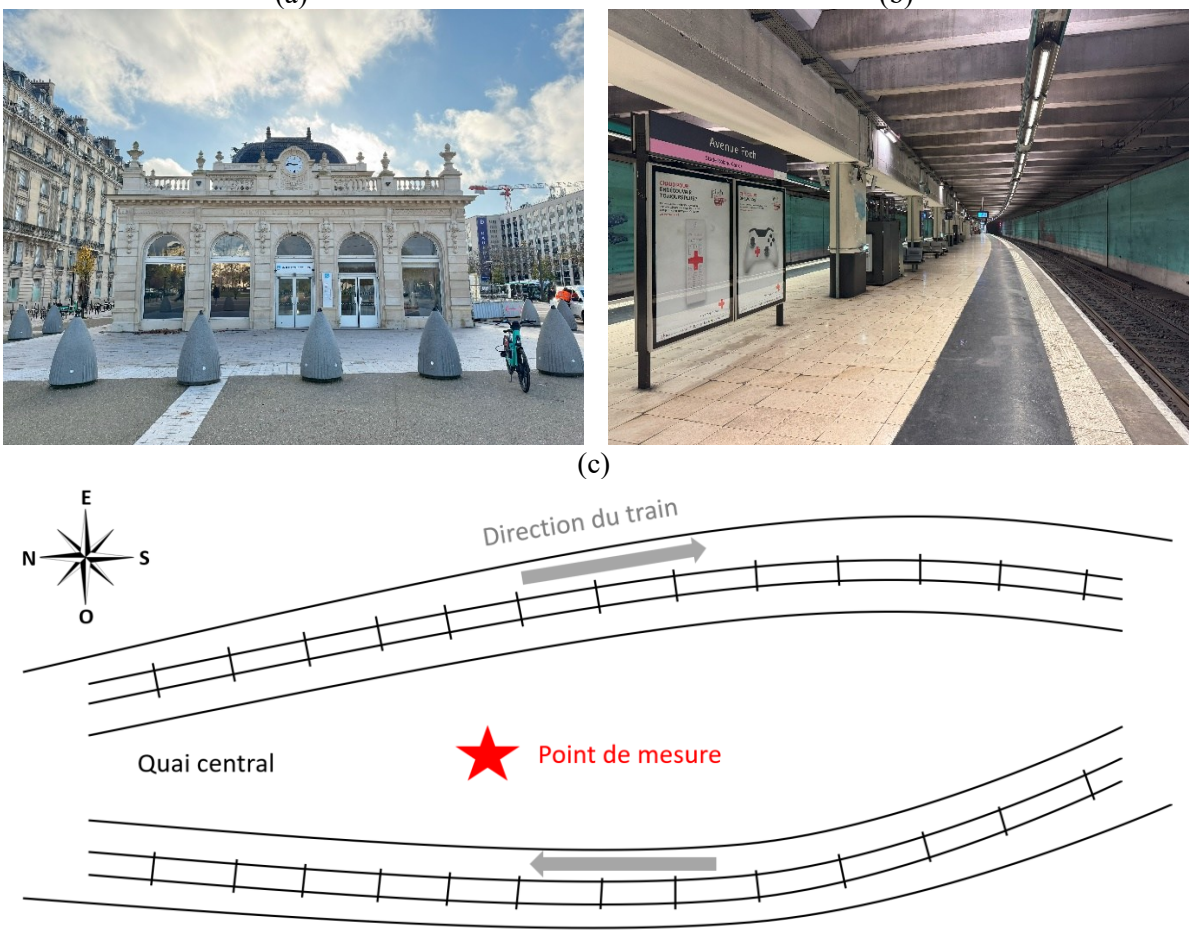


Figure 2.9 : Illustration de la gare d'Avenue Foch : (a) vue extérieure, (b) quai, et (c) plan simplifié de la gare

Avec un flux moyen de 2 301 voyageurs par jour, la gare d'Avenue Foch est l'une des moins fréquentées du réseau RER en Île-de-France (données SNCF, 2021). Le trafic ferroviaire y est régulier, avec 169 trains en circulation les jours de semaine et 143 trains par jour le weekend, empruntant les deux voies encadrant un quai central.

Le Tableau 2.5 présente les principales caractéristiques de la gare d'Avenue Foch, en mettant en avant ses spécificités architecturales, telles que sa profondeur et son revêtement, ainsi que son système de ventilation. Il fournit également des informations détaillées sur le trafic ferroviaire quotidien, la fréquentation moyenne des voyageurs et la période de mesure associée. Les mises en forme de police (normale, italique, gras) utilisées dans ce tableau reprennent les critères définis dans le Tableau 2.4.

Tableau 2.5 : Caractéristiques principales de la gare d'Avenue Foch

Caractéristiques	Détails
Ligne de RER	C
Date d'ouverture	**1854**
Profondeur	*~ 5 m*
Type de revêtement des voies	**Ballast**
Type de gare	*Souterraine (1 quai, 2 voies)*
Système de ventilation	*Ventilation naturelle*
Fréquentation par jour	2 301
Traffic ferroviaire par jour (JO, WE)	169 trains JO, 143 trains WE
Période de mesure	11/04/2018 – en cours

Dans le cadre de notre étude, nous avons retenu la période allant du 11 avril 2018 au 31 décembre 2022. D'après les données fournies par AIRPARIF, les paramètres analysés dans cette gare incluent les concentrations en PM_{10}, $PM_{2.5}$, ainsi que le CO_2, la température et l'humidité relative mesurés sur le quai avec un pas de temps quart horaire. Les statistiques descriptives de ces paramètres sont présentées dans le Tableau 2.6, qui en synthétise les principaux résultats.

Tableau 2.6 : Statistiques descriptives des données collectées à Avenue Foch durant la période étudiée

Variable	Points de données	Unité	Plage (min ; max)	Plage (25è ; 75è)	Moyenne	Ecart-type
PM_{10}	141 428	$\mu g/m^3$	(0,1 ; 4 492,9)	(60,1 ; 214,8)	146,6	102,2
$PM_{2.5}$	138 741	$\mu g/m^3$	(0,1 ; 2 684,2)	(29 ; 84,3)	60,2	42,2
CO_2	143 817	ppm	(351 ; 1 456)	(436,3 ; 526)	494,9	83,1
RH	142 767	%	(26,6 ; 90,4)	(51 ; 66)	58,5	10,3
T	142 691	°C	(5,6 ; 33,3)	(16,9 ; 23,9)	20,4	4,5

Durant la période considérée (11 avril 2018 - 31 décembre 2022), plusieurs événements ont impacté significativement les dynamiques des concentrations en particules dans la gare d'Avenue Foch. Ces perturbations incluent :

- **Les phases de dépollution des particules :**

Dans le cadre des initiatives de la SNCF visant à améliorer la qualité de l'air dans les EFS d'Île-de-France, deux expérimentations de dépollution des particules ont été menées. Ces initiatives ont permis de tester différents systèmes de filtration dans des conditions réelles d'exploitation, comme détaillé dans la section 2.1.2.

  - 27/05/2019 – 09/09/2019 : Expérimentation Air Liquide (technologie par ionisation positive)

Dans le cadre de cette expérimentation, six modules de filtration ont été installés le long des quais. Ces modules se distinguent par leurs systèmes de récupération des poussières, utilisant différentes

techniques : aspiration, bac à eau, vibrations et pulvérisation d'eau. Le constructeur indique qu'ils offrent un débit nominal de 52 000 m³/h. La Figure 2.10 illustre l'emplacement précis des six modules de filtration installés sur le quai, ainsi que la localisation du TEOM, utilisé pour mesurer les concentrations en particules avant, pendant et après filtration. Ces six modules n'ont pas tous fonctionné simultanément tout au long de la période d'expérimentation mais on ne dispose pas d'informations supplémentaires.

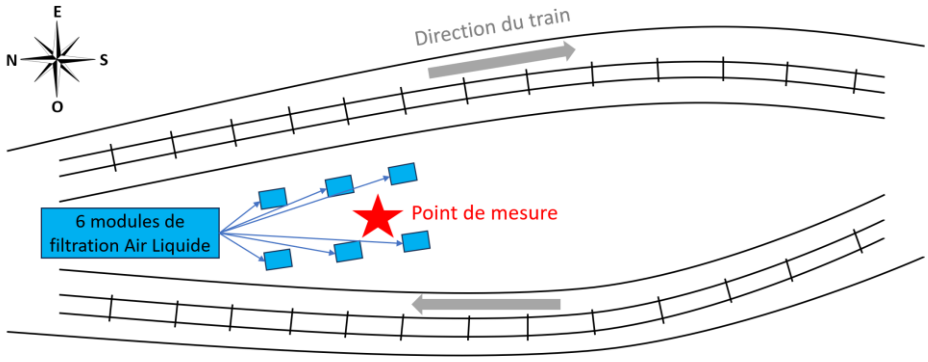


Figure 2.10 : Disposition des modules de filtration Air Liquide et localisation du TEOM

- 20/01/2020 – 15/03/2020 : Expérimentation Starklab (technologie par voie humide)

Dans le cadre de l'expérimentation menée par Starklab, six modules de filtration ont été installés autour de la baie de mesure. La Figure 2.11 illustre leur agencement sur le quai. À l'instar des tests réalisés par Air Liquide, ces modules ne fonctionnaient pas tous simultanément tout au long de la période d'expérimentation, et on n'en sait pas plus. Toutefois, selon le constructeur, leur débit nominal atteint 15 000 m³/h lorsqu'ils sont tous en fonctionnement.

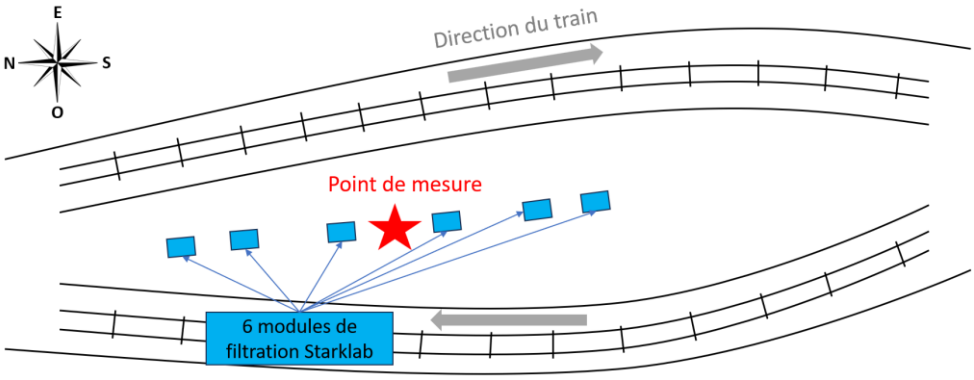


Figure 2.11 : Disposition des modules de filtration Starklab et localisation du TEOM

- **Les confinements liés au COVID-19 :** 17/03/2020 – 10/05/2020 ; 30/10/2020 – 14/12/2020 ; 03/04/2021 – 02/05/2021

Les confinements imposés par le gouvernement ont profondément modifié les habitudes de déplacement et entraîné une diminution significative de la demande en transport de passagers.

- **Les travaux d'été CASTOR :** 07/07/2021 – 01/09/2021 ; 15/07/2022 – 20/08/2022

Des programmes de maintenance récurrents ont été menés par la SNCF pendant les mois d'été afin de renforcer les sections centrales et souterraines de la ligne RER C, qui, en raison de leur ancienneté et de leur sensibilité aux variations des niveaux de la Seine, nécessitent une attention régulière. Ces travaux, réalisés durant les étés 2021 et 2022 dans l'EFS spécifiée, ont impacté les flux de circulation et entraîné la fermeture temporaire de la station.

- **Les grèves** : 03/04/2018 – 28/06/2018 (grèves de deux jours tous les quatre jours) ; 05/12/2019 – 26/12/2019 ; 01/01/2020 – 12/01/2020

Les grèves provoquent d'importantes perturbations du réseau de transport national. Cela peut entraîner l'annulation ou le retard des services ferroviaires et une réduction du nombre de trains circulant sur la ligne RER C.

2.2.2.2 Gare de Sevran Beaudottes

La gare de Sevran-Beaudottes, située sur la ligne B du RER (desservant Aéroport Charles de Gaulle 2, Mitry-Claye, et Saint-Rémy-lès-Chevreuse), se trouve au cœur de la commune de Sevran, dans le département de la Seine-Saint-Denis, en Île-de-France. Elle est stratégiquement localisée à proximité immédiate du centre commercial Beau Sevran et de plusieurs infrastructures publiques.

Cette gare mixte combine une partie souterraine, où se trouvent deux quais et deux voies situés à une profondeur de 5 mètres sous terre (Figure 2.12b et Figure 2.12c), et une partie aérienne comprenant le hall, les guichets et la salle des pas perdus (Figure 2.12a). Tout comme la gare d'Avenue Foch, elle n'est pas équipée d'un système de ventilation mécanique ; la circulation de l'air repose exclusivement sur une ventilation naturelle, facilitée par les mouvements des trains et les ouvertures structurelles.

(a) (b)



(c)

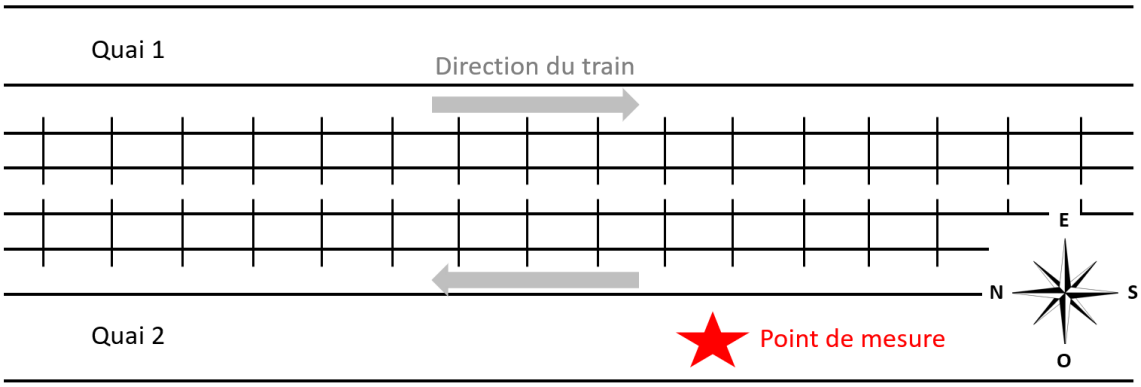


Figure 2.12 : Illustration de la gare de Sevran Beaudottes : (a) vue extérieure, (b) quai et (c) plan simplifié de la gare

En termes de trafic ferroviaire, la gare compte 229 trains par jour en semaine et 154 trains par jour le weekend, selon les données réelles de 2022 (AEF, 2022c).

Concernant la fréquentation, cette gare accueille en moyenne 19 862 voyageurs par jour les jours ouvrés, soit un total annuel d'environ 7,2 millions de passagers enregistrés en 2022 (AEF, 2022c).

Le Tableau 2.7 résume les principales caractéristiques de la gare de Sevran-Beaudottes, en mettant en évidence ses spécificités architecturales, son système de ventilation, ainsi que les données liées au trafic ferroviaire et à la fréquentation des voyageurs. Les mises en forme de police utilisées dans ce tableau reprennent également les critères définis dans le Tableau 2.4.

Tableau 2.7 : Caractéristiques principales de la gare de Sevran-Beaudottes

Caractéristiques	Détails
Ligne de RER	B
Date d'ouverture	*1976*
Profondeur	*~ 5 m*
Type de revêtement des voies	Béton
Type de gare	*Souterraine (2 quais, 2 voies)*
Système de ventilation	*Ventilation naturelle*
Fréquentation par jour	*19 862 les jours ouvrés*
Traffic ferroviaire par jour (JO, WE)	*229 trains JO, 154 trains WE*
Période de mesure	01/01/2018 – en cours

Concernant la période considérée, l'étude initiale pour la gare de Sevran-Beaudottes couvrait une période allant de juin 2018 à décembre 2022, afin de garantir une cohérence avec les analyses menées sur la gare d'Avenue Foch. Cependant, un changement d'instrumentation survenu le 22 décembre 2021 a profondément modifié les caractéristiques des données collectées. La Figure 2.13 met en évidence une augmentation significative du niveau de bruit dans les concentrations mesurées après cette date, suggérant des modifications dans les protocoles ou dans la précision des mesures.

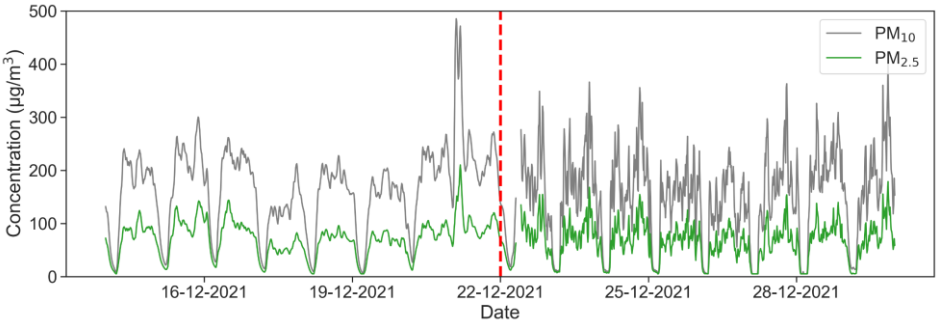


Figure 2.13 : Concentrations journalières de PM_{10} et $PM_{2.5}$ à Sevran-Beaudottes, mettant en évidence l'impact du changement d'instrumentation le 22 décembre 2021.

Afin d'assurer une représentativité optimale des résultats et de minimiser les biais potentiellement introduits par ces variations, la période d'analyse a été limité du 6 juin 2018 au 21 décembre 2021. Selon les données fournies par AEF, les paramètres étudiés dans cette gare incluent les concentrations en PM_{10}, $PM_{2.5}$, ainsi que la température, mesurés sur le quai avec un pas de temps quart horaire. Le Tableau 2.8 résume les statistiques descriptives des paramètres collectés durant cette période d'étude. Les données de CO_2 et d'humidité relative ne sont pas disponibles dans les ressources fournies pour cette gare.

Tableau 2.8 : Statistiques descriptives des données collectées à Sevran-Beaudottes durant la période étudiée

Variable	Points de données	Unité	Plage (min ; max)	Plage (25è ; 75è)	Moyenne	Ecart-type
PM_{10}	108 480	µg/m³	(5 ; 1069,7)	(96,8 ; 294,9)	200,9	125,9
$PM_{2.5}$	108 480	µg/m³	(5 ; 882,8)	(41,3 ; 123,5)	84,8	52,4
T	108 480	°C	(4,5 ; 35,6)	(14,8 ; 24,3)	19,4	5,4

Durant la période d'étude, cette gare a été impactée par plusieurs événements majeurs ayant perturbé les dynamiques des concentrations en particules. Ces périodes incluent :

- **La phase de dépollution des particules :**

La gare a également fait l'objet d'une expérimentation visant à tester une technologie de filtration des particules développée par Mann+Hummel. Dans ce cadre, des colonnes de filtration ont été installées sur les quais entre novembre 2020 et juin 2021, afin d'évaluer leur efficacité dans un environnement ferroviaire souterrain.

- **Les confinements liés au COVID-19 :** 17/03/2020 – 10/05/2020 ; 30/10/2020 – 14/12/2020 ; 03/04/2021 – 02/05/2021
- **Les grèves :** 07/06/2018 – 28/06/2018 (grèves de deux jours tous les quatre jours) ; 05/12/2019 – 26/12/2019 ; 01/01/2020 – 12/01/2020

Notons qu'aucuns travaux de type CASTOR n'ont eu lieu durant cette période.

2.2.2.3 Gare de Magenta

La gare de Magenta, mise en service en juillet 1999, est une station relativement complexe sur le plan architectural et occupe un rôle stratégique dans le réseau ferroviaire parisien. Elle se situe entre la Gare du Nord, qui permet des correspondances avec les lignes 4 et 5 du métro ainsi que les RER B et D, et la Gare de l'Est. Bien que relativement récente, elle n'est pas un terminus de ligne. Elle figure toutefois parmi les dernières gares de la ligne E. Parmi les trois gares de référence du réseau SNCF, elle est la seule dotée d'un système de ventilation mécanique, assurant une ventilation de deux types : désenfumage et confort.

La gare de Magenta est située à 30,1 m sous le niveau de la rue du Faubourg Saint-Denis. Elle dispose de deux quais principaux : le premier quai est subdivisé en Quai 1 et Quai 2, tandis que le second se divise en Quai 3 et Quai 4. Au total, la gare comprend quatre voies, toutes situées sur un sol en béton (Figure 2.14c).

L'organisation des voies repose sur une disposition en trois tunnels distincts. Le tunnel central, situé entre les quais 2 et 3, abrite deux voies principales. Les deux tunnels latéraux, chacun équipé d'une seule voie, sont situés respectivement face au Quai 1 et face au Quai 4 (Figure 2.14c).

En termes d'exploitation, les trois tunnels sont utilisés de manière différenciée en fonction des horaires. Aux heures de pointe, les quatre voies sont en service, optimisant ainsi la capacité de la gare pour gérer le flux important de passagers. En revanche, aux heures creuses, seuls les tunnels latéraux restent actifs pour permettre une circulation minimale, tandis que les trains en terminus stationnent dans le tunnel central.

Chaque jour ouvré, la gare de Magenta voit circuler en moyenne 440 trains. En termes de fréquentation, elle enregistre 90 537 voyageurs par jour ouvré selon les données de 2022 (AEF, 2022b), se classant ainsi au sixième rang des neuf gares souterraines les plus fréquentées du réseau RER.



Figure 2.14 : Illustration de la gare de Magenta : (a) vue extérieure, (b) quai, et (c) plan simplifié de la gare

Le Tableau 2.9 résume les principales caractéristiques de la gare de Magenta, en mettant en évidence ses spécificités architecturales, son système de ventilation, ainsi que les données liées au trafic ferroviaire et à la fréquentation des voyageurs. Comme pour les deux autres gares, les mises en forme de police utilisées dans ce tableau reprennent les critères définis dans le Tableau 2.4.

Tableau 2.9 : Caractéristiques principales de la gare de Magenta

Caractéristiques	Détails
Ligne de RER	E
Date d'ouverture	1999
Profondeur	**30,1 m**
Type de revêtement des voies	Béton
Type de gare	**Souterraine (2 quais, 4 voies)**
Système de ventilation	Ventilation mécanique
Fréquentation par jour	**90 537 les jours ouvrés**
Traffic ferroviaire par jour	**440 les jours ouvrés**
Période de mesure	20/02/2017 – en cours

Pour cette gare la période d'analyse dans ce travail va de 20 février 2017 à 21 décembre 2021. En effet, comme pour Sevran Beaudottes, un changement d'instrumentation est également survenu le 22

décembre 2021, modifiant les caractéristiques des mesures. La Figure 2.15 illustre cette transition, mettant en évidence une augmentation significative des fluctuations dans les concentrations mesurées après cette date.

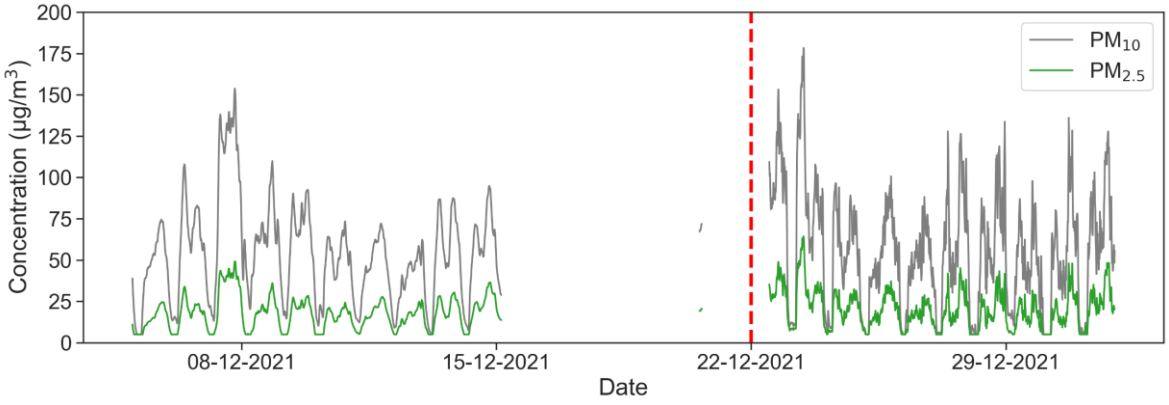


Figure 2.15 : Concentrations journalières de PM_{10} et $PM_{2.5}$ à Magenta, mettant en évidence l'impact du changement d'instrumentation le 22 décembre 2021.

Les concentrations en PM_{10}, $PM_{2.5}$ et la température ont été mesurées sur le quai avec un pas de temps quart horaire, conformément aux données fournies par AEF. Le Tableau 2.10 synthétise les principales statistiques descriptives issues de ces mesures. Comme pour la gare de Sevran Beaudottes, les données de CO_2 et d'humidité relative ne sont pas disponibles.

Tableau 2.10 : Statistiques descriptives des données collectées à Magenta durant la période étudiée

Variable	Points de données	Unité	Plage (min ; max)	Plage (25è ; 75è)	Moyenne	Ecart-type
PM_{10}	135 399	µg/m³	(1,7 ; 1901,2)	(29,5 ; 78,9)	58,7	45,9
$PM_{2.5}$	135 399	µg/m³	(0,1 ; 285,4)	(10,8 ; 28,2)	21,0	14,2
T	135 399	°C	(5,6 ; 33,3)	(16,9 ; 23,9)	20,4	4,5

Enfin, il n'y pas eu d'expérimentation de dépollution dans cette gare. Ainsi les périodes notables incluent :

- **Les confinements liés au COVID-19 :** 17/03/2020 – 10/05/2020 ; 30/10/2020 – 14/12/2020 ; 03/04/2021 – 02/05/2021
- **Les travaux d'été CASTOR** : 12/07/2019 – 26/08/2019 ; 10/07/2022 – 30/08/2022
- **Les grèves :** 03/05/2018 – 28/06/2018 (grèves de deux jours tous les quatre jours) ; 05/12/2019 – 26/12/2019 ; 01/01/2020 – 12/01/2020

2.2.3 Station de mesures de l'air extérieur

À des fins de comparaison avec les données collectées dans les EFS, des stations de surveillance de la qualité de l'air extérieur, opérées par AIRPARIF, ont été sélectionnées. Ces stations mesurent les concentrations en particules (PM_{10} et $PM_{2.5}$) ainsi que d'autres paramètres environnementaux, au pas de temps horaire, dans des environnements extérieurs proches des gares étudiées, en utilisant des équipements comparables à ceux déployés dans les EFS.

Les trois stations retenues, chacune associée à une gare de référence, ont été choisies en raison de leur proximité géographique et/ou de la similitude de leurs environnements respectifs avec les gares étudiées (Figure 2.16).

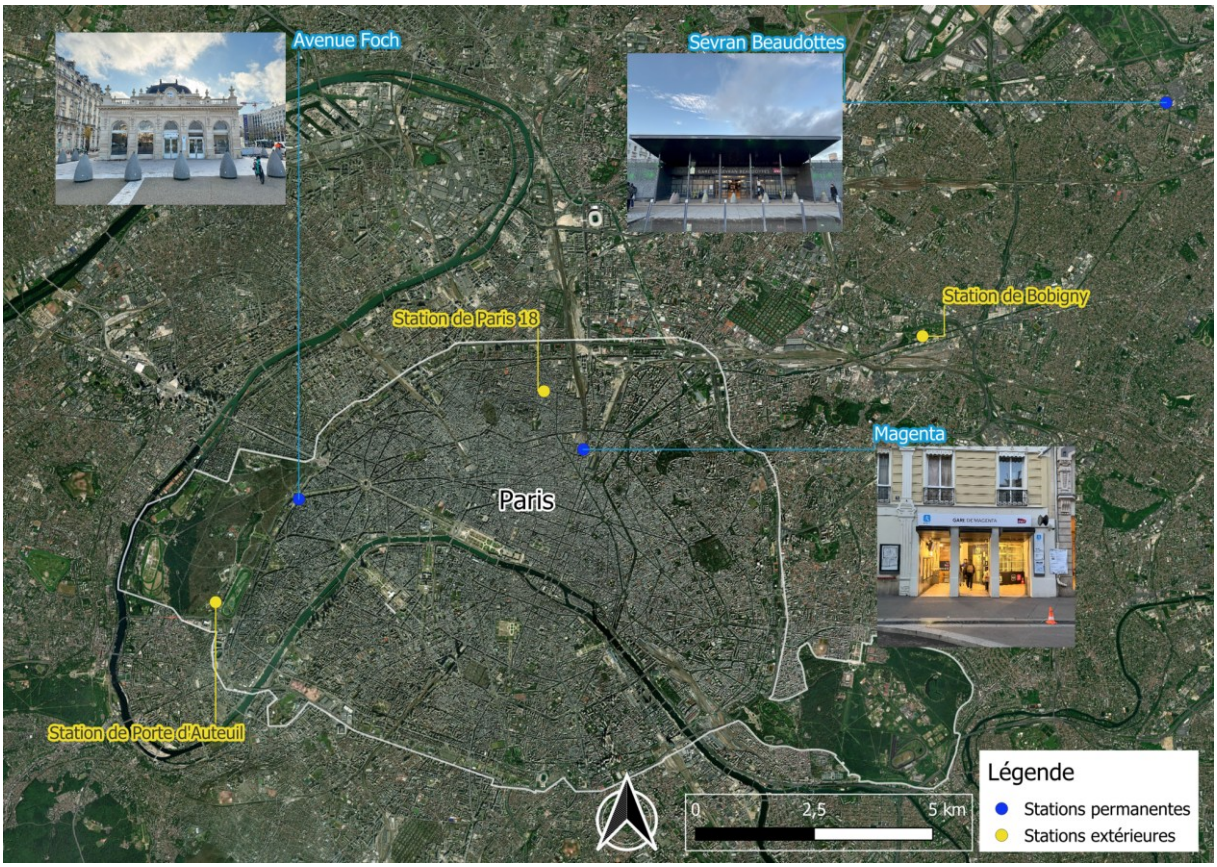


Figure 2.16 : Localisation des trois gares de référence (Avenue Foch, Sevran Beaudottes, Magenta) et des stations de mesures de l'air extérieur correspondantes (Porte d'Auteuil, Bobigny, Paris 18è)

- **Station de la Porte d'Auteuil**

La station de surveillance de la qualité de l'air de fond urbain située au Boulevard Périphérique de la Porte d'Auteuil (48°51'04.5"N, 2°15'07.6"E) a été choisie pour sa proximité avec la gare d'Avenue Foch, située dans le même arrondissement. Cette station reflète les caractéristiques d'un environnement urbain résidentiel adjacent au Bois de Boulogne et aux périphéries, offrant ainsi une base de comparaison pertinente pour les niveaux de pollution mesurés à la gare d'Avenue Foch. Les mesures effectuées à cette station portent sur PM_{10}, $PM_{2.5}$, NO_2, NO et NOx au cours de la période étudiée.

- **Station de Bobigny**

La station de surveillance de la qualité de l'air de fond urbain de Bobigny, située près du Parc Départemental de la Bergère (48°54'09.9"N, 2°27'16.6"E), est utilisée comme référence pour la gare de Sevran Beaudottes. Bien qu'elle soit légèrement éloignée de cette gare, ses données reflètent les caractéristiques environnementales d'une zone périurbaine typique, permettant une comparaison cohérente avec les niveaux de pollution mesurés dans l'EFS de Sevran Beaudottes. Durant la période considérée, cette station a permis de surveiller les concentrations de PM_{10}, $PM_{2.5}$, NO_2, NO et NOx.

- **Station de Paris 18ème**

La station de surveillance de la qualité de l'air de fond urbain située dans le 18è arrondissement de Paris (48°53'30.6"N 2°20'45.3"E), a été sélectionnée comme point de référence pour la gare de Magenta. Cette station reflète un environnement urbain dense et animé, correspondant ainsi aux conditions environnementales de la zone entourant la gare de Magenta. Les données collectées à cette station incluent les concentrations de PM_{10}, NO_2, NO, NOx et O_3 sur l'ensemble de la période étudiée. En revanche, les données relatives au $PM_{2.5}$ ne sont disponibles qu'à partir de 2022.

2.3 Conclusion du chapitre

Ce chapitre a présenté les méthodes et outils employés pour la collecte des données sur la qualité de l'air en EFS.

Tout d'abord, les instruments de mesure ont été détaillés, couvrant à la fois les analyseurs automatiques de particules, les analyseurs de CO_2 par corrélation infrarouge, ainsi que les capteurs de température et d'humidité relative. Ces équipements, essentiels pour caractériser l'environnement souterrain, ont été complétés par la description des instruments de filtration des particules testés en conditions réelles.

Ensuite, une attention particulière a été portée aux sites de mesure considérés dans cette thèse, distinguant les campagnes de mesures de courte durée menées dans 14 gares du réseau SNCF et les gares sélectionnées pour une surveillance continue des concentrations en particules. Parmi elles, les gares d'Avenue Foch, de Sevran-Beaudottes et de Magenta ont été équipées d'analyseurs automatiques pour mesurer les concentrations en particules (PM_{10} et $PM_{2.5}$), ainsi que, dans certains cas, les concentrations en CO_2, la température et l'humidité relative. Toutefois, pour les gares de Sevran-Beaudottes et Magenta, seules les données relatives aux PM_{10}, $PM_{2.5}$ et à la température ont été incluses dans la base de données disponible. Les baies de mesure ont été positionnées au centre des quais dans chaque gare afin d'assurer une représentativité optimale des niveaux de pollution. Les mesures, effectuées avec un pas de temps de 15 minutes, permettent une analyse fine des variations temporelles de la qualité de l'air. Le Tableau 2.11 résume les dispositifs de mesure utilisés dans chaque gare ainsi que les paramètres surveillés.

Tableau 2.11 : Synthèse des dispositifs de mesure et des paramètres mesurés dans les gares de référence

Gare	Modèle TEOM	Pas de temps	PM_{10}	$PM_{2.5}$	CO_2	Température	Humidité relative
Avenue Foch	FDMS 1405-F	15 minutes	✓	✓	✓	✓	✓
Sevran Beaudottes	1405-D		✓	✓	-	✓	-
Magenta	1405-D		✓	✓	-	✓	-

Ces mesures ont été complétées par les données issues des stations de surveillance de l'air extérieur, permettant une comparaison entre la qualité de l'air intérieur et extérieur (Tableau 2.12).

Tableau 2.12 : Synthèse des paramètres mesurés dans les stations de surveillance de l'air extérieur

Station	Pas de temps	PM_{10}	$PM_{2.5}$	NO_2	NO	NOx	O_3
Porte d'Auteuil	1h	✓	✓	✓	✓	✓	-
Bobigny		✓	✓	✓	✓	✓	-
Paris 18ème		✓	-	✓	✓	✓	✓

Enfin, les expérimentations menées dans les gares d'Avenue Foch et de Sevran Beaudottes ont été introduites, mettant en avant les tests de trois technologies de filtration des particules : l'ionisation électrostatique, la filtration par voie humide et la filtration classique. Ces expérimentations constituent une base essentielle pour évaluer l'efficacité des dispositifs testés et guider les stratégies d'amélioration de la qualité de l'air dans les EFS.

Les données collectées et les expérimentations décrites dans ce chapitre constituent ainsi les fondations méthodologiques sur lesquelles s'appuieront les analyses approfondies des chapitres suivants.

CHAPITRE 3. MÉTHODOLOGIES D'ANALYSE DES DONNÉES

L'analyse des concentrations de particules dans les EFS repose sur trois approches méthodologiques principales, développées ou utilisées pour répondre aux enjeux définis précédemment, notamment pour l'analyse des mesures déjà obtenues par la SNCF et ses partenaires, l'établissement de profils type, la détermination des facteurs influençant les concentrations en particules, et l'évaluation de l'efficacité des systèmes de filtration en milieu ferroviaire souterrain.

La première méthode vise à déterminer de manière rigoureuse, à partir de mesures sur des périodes plus ou moins longues, les profils journaliers types des concentrations de particules dans les EFS, séparément pour les jours ouvrés et les jours de weekend. Elle conduit également à un coefficient d'amplitude journalier (Daily Amplitude Coefficient – DAC), représentant l'amplitude de ce profil par rapport au profil moyen. Ce coefficient permet de « traduire » ou « transformer » l'ensemble des mesures journalières en un seul nombre représentatif de la journée, facilitant ainsi leur confrontation à d'autres types de mesures. Il convient de souligner que cette approche est mise en œuvre pour la première fois, marquant une avancée importante dans l'analyse des concentrations de particules en milieu ferroviaire souterrain.

La deuxième méthode utilise les modèles additifs généralisés (GAM, Generalized Additive Models en anglais), pour analyser ces mêmes mesures et modéliser les relations entre les concentrations de particules et divers facteurs explicatifs, tels que des variables environnementales et opérationnelles. Cette méthode permet d'étudier les facteurs influençant les concentrations de particules dans les EFS.

A la différence des deux autres méthodes, la troisième vise à prédire sur des temps courts (quelques jours) l'évolution des concentrations en particules dans l'EFS. Elle recourt à un modèle de réseau de neurones LSTM (Long Short-Term Memory), conçu pour exploiter les dépendances temporelles dans les séries de données et prédire les concentrations de particules à court terme.

La gare d'Avenue Foch a été choisie comme cas d'étude pour le développement de ces méthodes en raison de son jeu de données exceptionnel couvrant cinq années de mesures au pas de temps quart horaire. Initialement développées pour les PM_{10}, ces méthodes ont ensuite été appliquées aux $PM_{2.5}$ et étendues à d'autres gares. L'ensemble des analyses réalisées dans ce travail de thèse, du traitement des données à la modélisation, a été conduit à l'aide de Python.

3.1 Prétraitement et nettoyage des données

Bien que des traitements préliminaires aient déjà été réalisés, notamment les vérifications de la qualité des données effectuées par Airparif, la correction de l'horodatage et l'élimination des données non représentatives (comme les cas anormaux où $PM_{2.5} > PM_{10}$), des étapes complémentaires ont été nécessaires pour répondre aux spécificités des données utilisées dans cette étude.

- **Identification des périodes notables et nettoyage manuel**

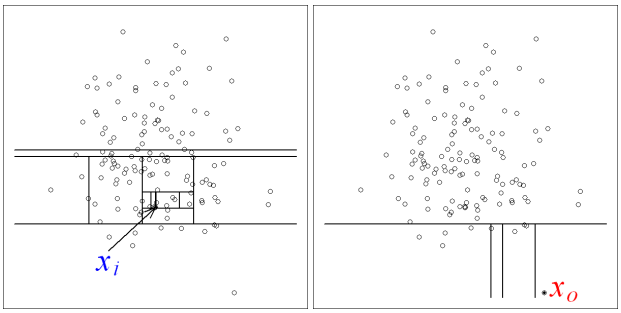
Certaines périodes ont été identifiées comme non représentatives des conditions normales de fonctionnement de la gare en raison d'événements ou d'interventions spécifiques. Pour garantir la validité des analyses, seules les données correspondant à des périodes où le fonctionnement de la gare est jugé « normal » ont été retenues. Ces périodes anormales incluent : les expérimentations de dépollution, les confinements liés à la COVID-19, les travaux estivaux Castor dans certaines EFS et les grèves.

Ces exclusions ont concerné des jours complets et ont été réalisées manuellement, en s'appuyant sur les historiques et les rapports annuels fournis par AIRPARIF ou AEF, les propriétaires des bases de données.

- **Traitement des anomalies**

Outre les périodes notables, des anomalies (ou outliers) ont également été identifiées comme susceptibles d'altérer les profils types. Plusieurs approches statistiques ont été testées pour détecter ces anomalies, telles que les statistiques descriptives (histogramme, boxplot, plage interquartile), les tests statistiques (score Z), les méthodes basées sur la proximité (K-Nearest Neighbor, Local Outlier Factor) et les approches basées sur les arbres (Isolation Forest). Parmi ces méthodes, l'algorithme Isolation Forest s'est révélé être la solution la plus satisfaisante pour notre jeu de données. Les autres méthodes se sont révélées peu discriminantes, excluant des points qui ne sont pas des outliers tout en en laissant passer d'autres. Grâce à sa capacité d'ajustement des hyperparamètres, tels que le nombre d'arbres ($n_estimators$) et le taux de contamination ($contamination$), il permet de capturer efficacement les outliers tout en limitant les faux positifs, offrant ainsi une robustesse supérieure pour notre jeu de données. Implémenté en Python via la bibliothèque $Scikit-learn$ (module $sklearn.ensemble.IsolationForest$), cet algorithme combine performance et flexibilité, permettant une intégration aisée dans notre chaîne de traitement des données.

Cet algorithme, développé par Liu et al. (2008), permet d'identifier les anomalies dans un jeu de données, en particulier les points qui diffèrent de manière significative des autres. Il utilise une méthode appelée forêt d'arbres d'isolement, ou $iTrees$. Une $iTree$ est un arbre de décision qui partitionne les données de manière répétée pour isoler les groupes de points. À chaque partition, les données sont séparées en deux groupes, ce qui permet de localiser plus facilement les anomalies. Les points anormaux sont isolés plus rapidement, car ils nécessitent moins de partitions que les points normaux. Cela se traduit par des chemins plus courts, car chaque partition rapproche davantage un point anormal de son isolement.



(a) Isolating x_i (b) Isolating x_o

Figure 3.1 : Partitionnement des points dans une distribution gaussienne pour l'isolement des anomalies (Liu et al., 2008)

La Figure 3.1, telle qu'illustrée par Liu et al. (2008), montre comment les points sont séparés dans une distribution gaussienne composée de 135 points de données. Un point normal, noté x_i, nécessite plusieurs partitions pour être isolé, tandis qu'un point anormal, noté x_0, en nécessite beaucoup moins. Dans le cas (a), le point x_i est isolé après 12 partitions, tandis que le point anormal x_0 (cas b) est isolé après seulement 4 partitions.

Quantitativement, pour déterminer si une instance est considérée comme une anomalie, un score d'anomalie est calculé. Le score d'anomalie s pour une instance x et une taille d'échantillon n est défini par la formule suivante :

$$s(x,n) = 2^{-\left(\frac{E(h(x))}{c(n)}\right)} \quad (3.1)$$

où $h(x)$ représente la longueur du chemin d'un point x, c'est-à-dire le nombre d'arêtes (ou divisions) que le point x traverse dans un $iTree$, une structure arborescente utilisée pour isoler les points de données. Le parcours commence à partir du nœud racine de l'arbre et se poursuit jusqu'au nœud externe où le point x est isolé, indiquant ainsi le nombre de partitions nécessaires pour isoler ce point. $E(h(x))$ désigne la moyenne des longueurs de chemin calculée sur un ensemble d'arbres d'isolement, indiquant ainsi la distance moyenne parcourue pour isoler un point x dans plusieurs arbres. De plus, $c(n)$ fait référence à la longueur moyenne du chemin d'une recherche infructueuse dans un arbre binaire de recherche, c'est-à-dire la distance parcourue lorsqu'un élément recherché n'est pas trouvé dans l'arbre.

Le score s peut être interprété comme suit :

- Les instances avec un score d'anomalie (s) supérieur à 0,5 sont identifiées comme des anomalies.
- Les instances présentant une valeur (s) inférieure ou égale à 0,5 sont catégorisées comme des instances normales.

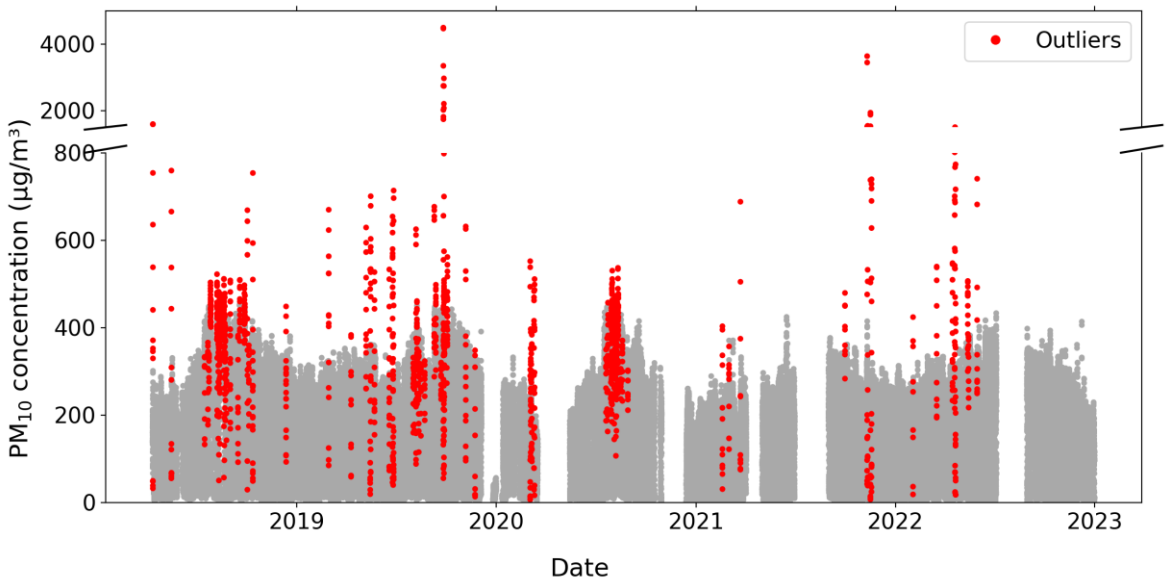


Figure 3.2 : Anomalies détectées à Avenue Foch

Dans cette étude, nous avons utilisé un jeu de données avec des caractéristiques spécifiques : une taille de $n = 141\,428$ points de données, une dimensionnalité de $d = 1$, et un taux de contamination de 0,004. Ce taux de contamination a été choisi après avoir effectué plusieurs tests avec différentes valeurs $(0{,}1\ ;\ 0{,}01\ ;\ 0{,}001\ ;\ 0{,}004)$. Il a été défini en fonction du volume de nos données, car il reflète la proportion de points de données susceptibles d'être des anomalies, soit 0,4 %. La Figure 3.2 montre les anomalies détectées à Avenue Foch, représentées par des points rouges, sur la concentration de PM_{10} durant la période considérée.

Le Tableau 3.1 fournit un aperçu détaillé du processus de nettoyage des données pour les gares d'Avenue Foch, Sevran Beaudottes et Magenta.

À Avenue Foch, la majorité des suppressions de données, représentant environ 21 % de la base initiale, sont dues aux opérations spécifiques de la station. Seulement 5 % des suppressions concernent les données brutes (dont 4 % de données invalides et 1 % d'anomalies). Au total, 26 % de la base de données originale ont été éliminés, aboutissant à un ensemble final de 105 339 points de données pour les PM_{10} et 102 774 points pour les $PM_{2.5}$.

Tableau 3.1 : Résumé du nettoyage des données et proportion de points supprimés aux gares Avenue Foch, Sevran Beaudottes et Magenta

Etapes de nettoyage de données	Proportion des points de données retirés de la base de données initiale		
	Avenue Foch	**Sevran Beaudottes**	**Magenta**
Expérimentations de dépollution des particules	9 %	17 %	-
COVID-19	7 %	4 %	5 %
Travaux d'été	1 %	-	-
Grèves	4 %	2 %	2 %
Données invalides	4 %	0 %	0 %
Anomalies (Outliers)	1 %	1 %	2 %
Total	26 %	24 %	9 %

Pour Sevran Beaudottes, une situation similaire à celle d'Avenue Foch a été observée. Environ 23 % des données ont été éliminées en raison des opérations spécifiques à la gare, tandis que seulement 1 % des données ont été supprimées en tant qu'outliers. Au total, 24 % des données initiales ont été retirées pour les deux classes de particules (PM_{10} et $PM_{2.5}$), résultant en une base finale de 83 376 points pour les PM_{10} et 83 250 points pour les $PM_{2.5}$.

À Magenta, les perturbations majeures, telles que les confinements liés à la pandémie de COVID-19 et les grèves, ont significativement impacté les données. Ces événements ont conduit à la suppression de 7 % des données initiales (5 % attribuables à la COVID-19 et 2 % aux grèves). Par ailleurs, 2 % des points ont été identifiés comme outliers. Finalement, seulement 9 % des données initiales ont été exclues lors du processus de nettoyage, un taux inférieur à celui observé dans les autres gares. Cela a permis de conserver un total de 121 921 points de données pour les PM_{10} et 122 188 points pour les $PM_{2.5}$. Cela s'explique principalement par l'absence de période de dépollution des particules à Magenta, ainsi que par le nettoyage préalable des données liées aux travaux CASTOR effectué par AEF.

Il est à noter qu'aucune donnée invalide ($PM_{2.5} > PM_{10}$) n'a été détectée pour Sevran Beaudottes et Magenta. Cette absence s'explique par les différences technologiques entre les instruments utilisés par AEF (TEOM 1405-D) et AIRPARIF (TEOM FDMS 1405-F), comme détaillé au Chapitre 2.

- **Rééchantillonnage temporel des données**

Contrairement à la méthode DAC, les modèles GAM et LSTM ont nécessité un prétraitement supplémentaire, notamment un rééchantillonnage temporel des données. Les données initialement collectées à un pas de temps quart horaire (96 points par jour) ont été agrégées à un pas horaire (24 points par jour).

Pour le modèle GAM, ce rééchantillonnage était nécessaire pour aligner la fréquence des données de quai avec celle d'autres bases de données, telles que celles provenant des stations de surveillance de l'air ambiant, qui sont disponibles à un pas horaire. En revanche, pour le modèle LSTM, le rééchantillonnage a été effectué afin de simplifier le jeu de données tout en conservant ses caractéristiques essentielles. Cette étape a permis de réduire la complexité computationnelle tout en préservant les informations nécessaires pour les prévisions à court terme.

3.2 Méthode basée sur le Coefficient d'Amplitude Journalier (DAC)

D'après la littérature (Birenzvige et al., 2003; Raut et al., 2009; Querol et al., 2012a; Walther et al., 2017), deux profils journaliers des concentrations en particules dans les EFS ont été identifiés : l'un pour les jours ouvrés et l'autre pour les weekends, ces deux profils étant distincts. Toutefois, la difficulté réside dans la manière de déterminer rigoureusement et précisément ces profils. Cette problématique

faisait partie de la demande initiale de la SNCF, qui souhaitait disposer d'un profil journalier de référence. Pour y répondre, nous avons développé une méthodologie appelée méthode du Coefficient d'Amplitude Journalier (DAC). Cette approche, fondée sur des ajustements linéaires, a d'abord été élaborée à la gare d'Avenue Foch avant d'être appliquée à l'ensemble des gares du réseau. Bien que l'accent ait été mis sur les fractions PM_{10}, des analyses similaires ont également été réalisées pour les $PM_{2.5}$.

L'objectif principal de cette méthode est de construire un profil type journalier représentatif des concentrations en particules, en dépassant un simple moyennage. Elle vise à identifier et éliminer les profils atypiques susceptibles de biaiser les résultats tout en capturant fidèlement les variations caractéristiques des concentrations journalières. La méthodologie et les résultats obtenus dans la gare d'Avenue Foch sont détaillés dans Rakotonirinjanahary et al. (2024) (Annexe 1).

Après les étapes de prétraitement et de nettoyage des données décrites plus haut, un profil moyen initial est d'abord calculé. Chaque profil individuel journalier est ensuite ajusté linéairement au profil moyen journalier.

Ainsi, à chaque ajustement, un DAC, correspondant à l'amplitude de l'ajustement, ainsi qu'un coefficient de détermination R^2 sont calculés. Les profils non représentatifs, identifiés par des valeurs R^2 inférieures à un seuil défini en fonction de la base de données propre à chaque gare, sont ensuite écartés. Enfin, un profil moyen final est recalculé à partir de la base de données épurée, offrant ainsi un profil type fiable et pertinent pour les analyses subséquentes. Cette méthodologie est résumée dans la Figure 3.3 qui illustre les étapes clés de l'approche.

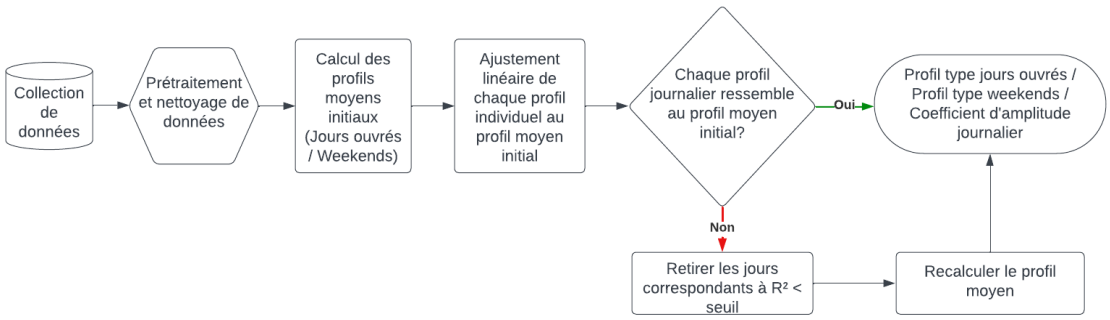


Figure 3.3 : Diagramme pour l'établissement de profil type journalier des concentrations en particules (adapté de Rakotonirinjanahary et al. 2024)

Inspirée des travaux de Raut et al. (2009) et de Walther and Bogdan (2017), ainsi que des analyses exploratoires menées dans cette étude, la méthode distingue les jours ouvrés des weekends afin de mieux représenter les variations de concentration. Cette approche produit deux profils types journaliers distincts — l'un pour les jours ouvrés et l'autre pour les weekends — tout en générant une métrique appelée Coefficient d'Amplitude Journalier (DAC).

3.2.1 Calcul de profils moyens initiaux (jours ouvrés et weekends séparés)

Les profils moyens initiaux sont calculés en distinguant deux catégories principales de jours : les jours ouvrés et les weekends. Cette distinction est essentielle pour capturer les variations caractéristiques des concentrations de PM_{10} selon le type de jour. Les jours ouvrés incluent tous les jours de la semaine, du lundi au vendredi. En revanche, les weekends regroupent les samedis, les dimanches et les jours fériés. Les jours fériés ont été indexés manuellement dans la base de données. Ces profils journaliers moyens (jours ouvrés ou weekend) sont calculés à partir de l'ensemble des données agrégées (jours ouvrés ou weekend).

Les résultats de ce calcul sont illustrés à la Figure 3.4, qui met en évidence les différences marquées entre les profils des jours ouvrés et ceux des weekends à la gare d'Avenue Foch. Ces profils moyens initiaux constituent une étape cruciale pour l'ajustement linéaire de chaque profil individuel.

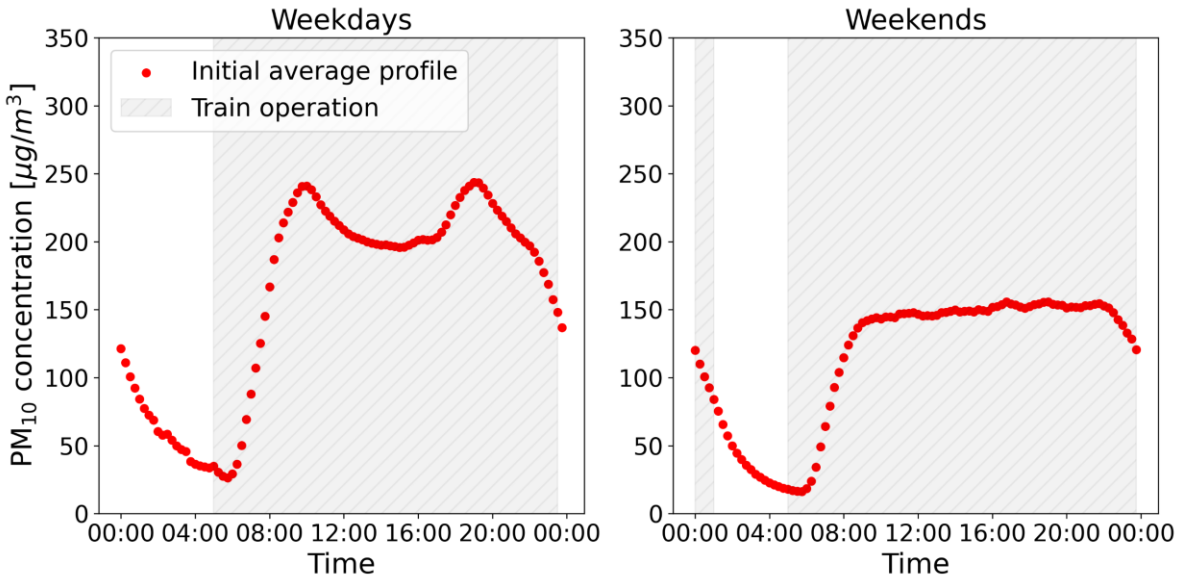


Figure 3.4 : Profils moyens journaliers initiaux de la concentration de PM_{10} pour les jours ouvrés (à gauche) et les weekends (à droite) à Avenue Foch

3.2.2 Ajustement linéaire et coefficient d'amplitude journalier (DAC)

Afin d'identifier les profils journaliers individuels qui s'écartent du profil journalier moyen, une approche basée sur un modèle de régression linéaire a été adoptée. Cette méthode permet d'établir une relation entre le profil de concentrations en PM d'un jour individuel, représentées par y (variable dépendante), et le profil journalier moyen de PM, représenté par x (variable indépendante).

Dans cette étude, nous supposons une relation linéaire entre ces deux variables, exprimée par l'équation : $y = DAC.x$. L'ordonnée à l'origine est considérée comme nulle. La pente obtenue, désignée ci-après sous le nom de Coefficient d'Amplitude Journalier (DAC), est calculée sur 96 points (4*24 = 96) de données pour chaque profil individuel. Ce DAC quantifie l'amplitude d'un profil individuel par rapport au profil journalier moyen. Lorsque le DAC est supérieur à 1, cela indique que la concentration de PM durant le jour d'intérêt est supérieure à celle du profil moyen. Inversement, un DAC inférieur à 1 signifie que la concentration de PM durant cette journée est inférieure au profil moyen. Par ailleurs, dans le cadre de cette méthode, il est supposé que toutes les concentrations journalières de PM commencent à 0 µg/m³, ce qui justifie l'absence d'un terme d'intercept dans l'équation. En effet, bien que les concentrations ne descendent pas réellement à 0, les valeurs très faibles proches de 5 µg/m³, en ordonnée à l'origine, sont considérées comme négligeables par rapport à l'amplitude de la journée. De plus, durant notre analyse, il est nécessaire de fixer l'ordonnée à l'origine, car sans cela, nous risquerions d'avoir trop de points élevés et bas, et pas assez de points pendant la montée et la descente des concentrations (Figure 3.5b).

Pour une illustration de cette analyse, la Figure 3.5 présente la distribution des données (Figure 3.5a) et les résultats de la régression (Figure 3.5b) pour le profil journalier de PM_{10} du 10 novembre 2022, ainsi que le profil moyen de PM_{10} sur la période considérée. Pour cette journée spécifique, le DAC était égal à 1,33, indiquant que les concentrations de PM_{10} étaient en moyenne 33% plus élevées ce jour-là par rapport au profil moyen.

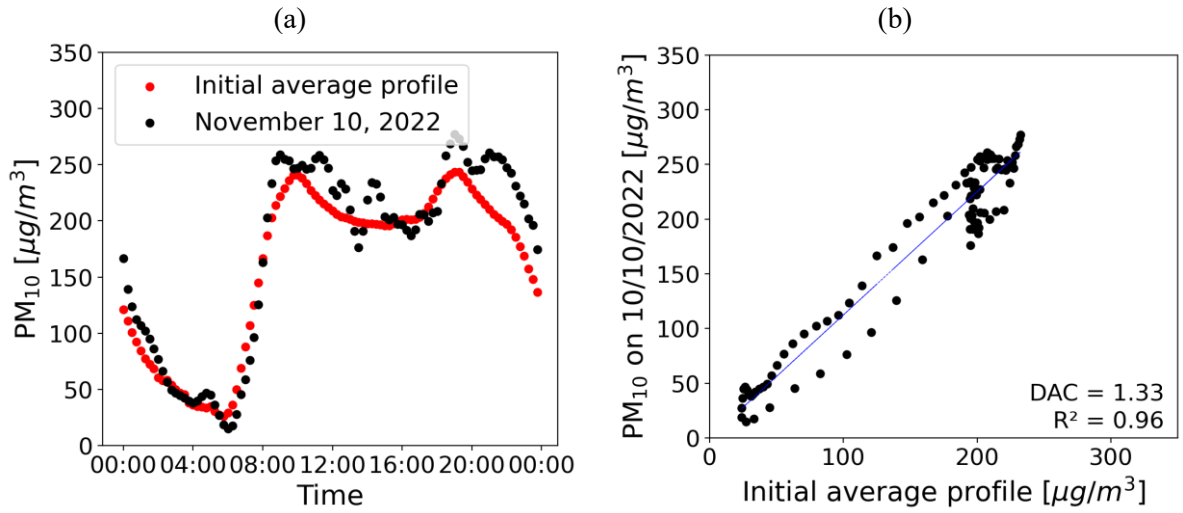


Figure 3.5 : (a) Comparaison des profils : profil moyen initial (ligne rouge pointillée) et profil du 10 novembre 2022 (ligne noire pointillée), (b) Modèle de régression linéaire, basé sur chaque mesure de PM du profil journalier, pour le profil moyen de PM_{10} et le profil de PM_{10} de ce jour-là (Rakotonirinjanahary et al., 2024)

3.2.3 Elimination des profils atypiques

Après avoir ajusté linéairement chaque profil individuel avec le profil moyen initial, on peut associer à chaque profil individuel un R^2. Comme détaillé dans notre premier article (Article 1, en annexe), cette métrique a été employée pour filtrer et exclure les profils journaliers déviant significativement du profil moyen initial.

Pour la gare d'Avenue Foch, un seuil de $R^2 = 0{,}67$ a été appliqué pour l'exclusion des profils. Ce seuil a été déterminé après une analyse approfondie des distributions des valeurs de R^2, comme illustré par les histogrammes dans les Figure 3.6 (a), (b), (c) et (d). Ce seuil peut varier d'une gare à l'autre. Ce choix permet d'optimiser l'équilibre entre l'inclusion de données représentatives et l'exclusion de variations atypiques qui pourraient biaiser les résultats de l'étude. Les profils en dessous de ce seuil ont montré une concordance insuffisante avec le profil moyen.

L'application de ce critère peut être illustrée par deux exemples concrets. Le profil journalier du 10 novembre 2022 (Figure 3.5a) présente un coefficient de détermination R^2 de 0,96. Ce profil montre une excellente corrélation avec le profil moyen, même si des écarts visuels sont observables, notamment au niveau des deux pics de la journée. Cependant, ces variations restent alignées avec les tendances générales, justifiant la conservation de ce jour dans l'analyse. En revanche, le profil journalier du 2 novembre 2022 (Figure 3.7) affiche un coefficient R^2 de 0,4. Des concentrations très faibles autour de la mi-journée (entre 12 h 00 et 16 h 00), ce qui diffère fortement du profil moyen où les concentrations restent plus élevées. Des pics prononcés le matin et le soir ne suivent pas les amplitudes observées dans le profil moyen. Par conséquent, ce profil ne répond pas au seuil fixé et serait exclu de l'analyse finale afin d'éviter de biaiser les résultats avec des données atypiques.

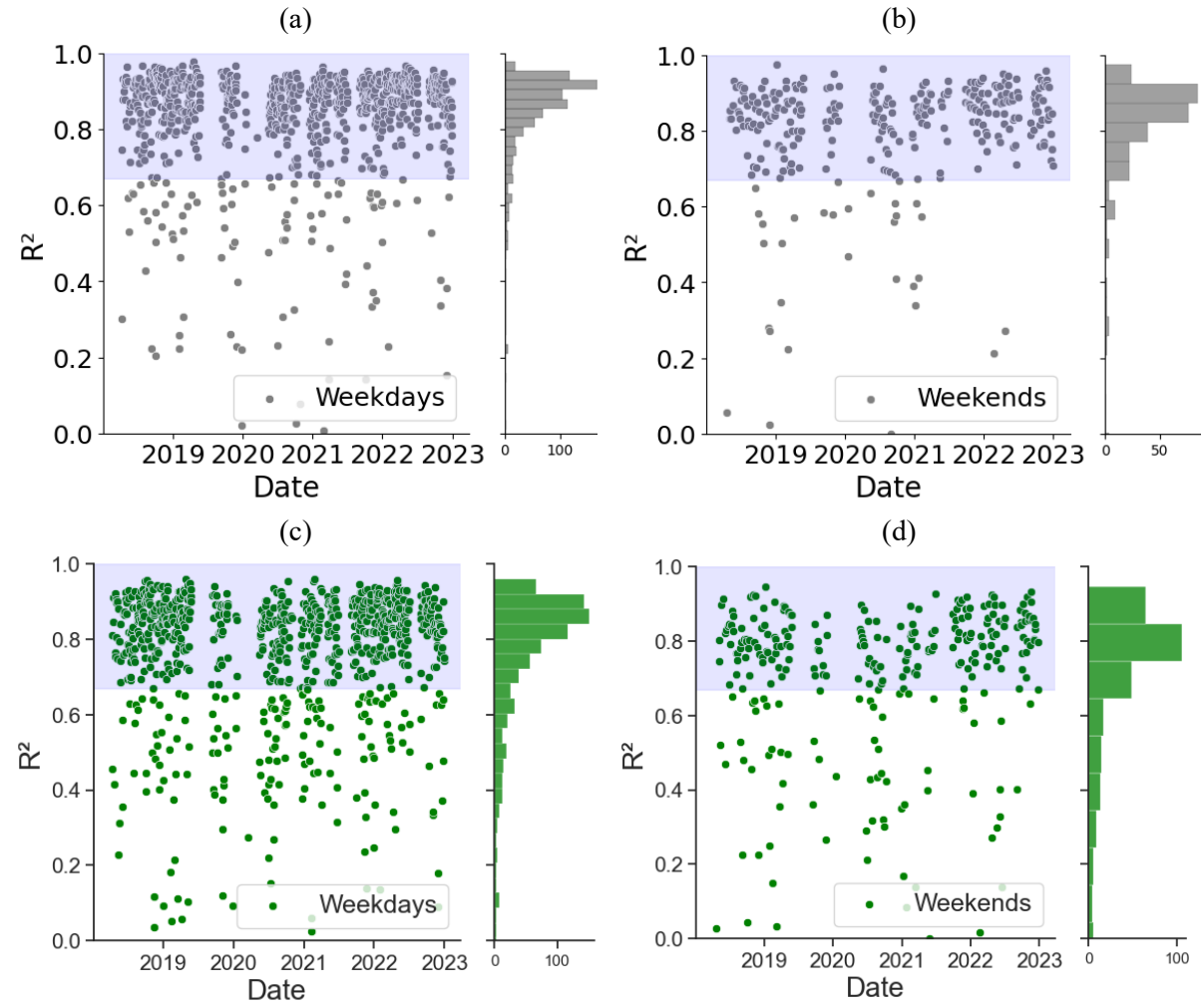


Figure 3.6 : Variation temporelle du coefficient de détermination (R^2) pour les PM_{10} (en gris) et les $PM_{2.5}$ (en vert) à la gare d'Avenue Foch : (a) jours ouvrés pour PM_{10}, (b) weekends pour PM_{10}, (c) jours ouvrés pour $PM_{2.5}$, et (d) weekends pour $PM_{2.5}$

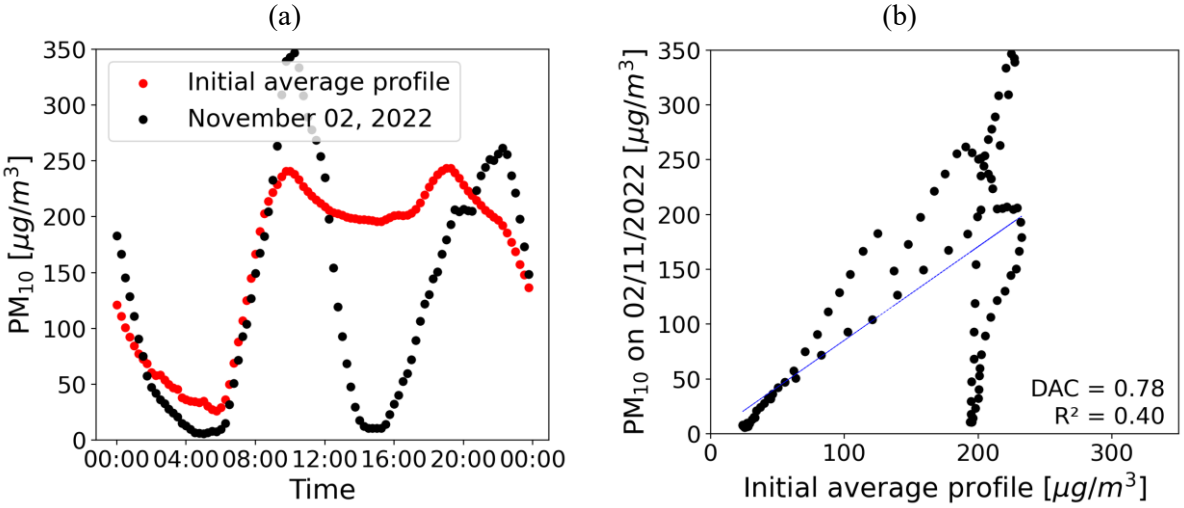


Figure 3.7 : (a) Comparaison des profils : profil moyen initial (ligne rouge pointillée) et profil du 2 novembre 2022 (ligne noire pointillée), (b) Modèle de régression linéaire, basé sur chaque mesure de PM du profil journalier, pour le profil moyen de PM_{10} et le profil de PM_{10} de ce jour-là

L'application de ce seuil de R^2 a abouti à la rétention de 88 % (72 576 points de données) et 80 % (25 632 points de données) des profils pour les jours de semaine et les weekends respectivement pour les PM_{10}. Pour les $PM_{2.5}$, ceci a abouti à la rétention de 85 % (63 264 points de données) et 78 % (20 256 points de données) des profils pour les jours de semaine et les weekends respectivement. Ce processus d'exclusion a permis de raffiner notre base de données, assurant que les profils retenus reflètent véritablement un comportement typique des concentrations de PM_{10} et $PM_{2.5}$.

Ce seuil peut varier d'une gare à l'autre. Pour la gare d'Avenue Foch, c'est cette valeur qui s'est avérée la plus appropriée.

Pour la gare de Sevran-Beaudottes, nous avons également appliqué un seuil de $R^2 = 0{,}67$ pour filtrer les profils non représentatifs, comme illustré dans la Figure A. 2 (Annexe 3) pour les deux classes de particules. Ce seuil, déjà validé pour les données de la gare d'Avenue Foch, s'est à nouveau révélé parfaitement adapté aux caractéristiques du jeu de données de la gare de Sevran-Beaudottes, confirmant ainsi sa pertinence et son applicabilité dans un autre environnement ferroviaire. À l'issue de ce processus, 88 % (56 928 points de données) et 83 % (21 120 points de données) des profils ont été conservés pour les jours ouvrés (Figure A. 2a) et les weekends (Figure A. 2b) respectivement pour les PM_{10}. En ce qui concerne les $PM_{2.5}$, 88 % (56 352 points de données) et 82 % (20 832 points de données) des profils ont été retenus pour les jours ouvrés (Figure A. 2c) et les weekends (Figure A. 2d) respectivement.

Pour la gare de Magenta, une adaptation spécifique a été nécessaire. Contrairement aux seuils de $R^2 = 0{,}67$ retenus pour les gares d'Avenue Foch et de Sevran-Beaudottes, un seuil de $R^2 = 0{,}6$ a été adopté ici. Ce choix s'appuie sur l'analyse des distributions présentées dans les histogrammes situés à droite de chaque sous-figure de la Figure A. 3 (Annexe 3), qui mettent en évidence une concentration significative de profils au-dessus de ce seuil. Cette adaptation reflète les spécificités des données collectées à Magenta et leur distribution particulière.

Cette étape méthodologique a permis de conserver 81,9 % (79 008 points de données) des profils pour les jours ouvrés (Figure A. 3a) et 71,7 % (26 208 points de données) pour les weekends (Figure A. 3b) en ce qui concerne les PM_{10}. Pour les $PM_{2.5}$, les pourcentages de profils retenus sont légèrement inférieurs, avec 69,6 % (66 9012 points de données) pour les jours ouvrés (Figure A. 3c) et 62,2 % (22 464 points de données) pour les weekends (Figure A. 3d). Ces taux de rétention, bien que variables selon la classe de particules et le type de jour, garantissent une base de données robuste et représentative pour le calcul des profils types finaux.

3.2.4 Application aux gares avec des mesures courte durée

Le Tableau 3.2 récapitule les périodes de mesure pour les différentes gares ayant fait l'objet de campagnes de courte durée, ainsi que le nombre de profils retenus après l'application de notre méthodologie.

On observe une grande variabilité du nombre de profils retenus selon les gares et les périodes de mesure. Par exemple, à la gare de Grigny Centre, seulement 7 profils sur 15 ont été retenus pour les jours ouvrés et 3/6 les weekends. Le profil journalier type obtenu, illustré dans la Figure 3.8, montre une tendance représentative des données disponibles, mais il soulève une question importante : quelle est la justesse de ce profil, compte tenu du faible nombre de profils retenus et de la durée limitée de la campagne de mesure ?

Tableau 3.2 : Périodes de mesure et nombre de profils journaliers retenus pour les jours ouvrés et les weekends dans les gares avec des mesures de courte durée.

Nom de la gare	Période de mesure	Profils retenus	
		Jours ouvrés	Weekends
Austerlitz	05/09/2016 – 25/09/2016	15/15	5/6
Avenue Président Kennedy	02/10/2017 – 22/10/2017	10/15	0/6
Bibliothèque François Mitterrand	17/09/2018 – 07/10/2018	6/15	3/6
Cergy Préfecture	09/01/2017 – 29/01/2017	15/15	3/6
Charles de Gaulle 1	22/05/2017 – 11/06/2017	12/14	5/6
Champ de Mars	18/06/2018 – 08/07/2018	13/15	5/6
Evry Courcouronnes	04/12/2017 – 24/12/2017	14/15	5/6
Grigny Centre	19/06/2017 – 09/07/2017	7/15	3/6
Musée Orsay	26/03/2018 – 15/04/2018	11/15	1/6
Neuilly Porte Maillot	18/04/2017 – 08/05/2017	15/15	4/6
Pereire Levallois	14/05/2018 – 10/06/2018	11/14	4/4
Pont de l'Alma	12/06/2021 – 24/06/2021	7/8	3/4
Porte de Clichy	04/09/2017 – 22/09/2017	14/15	4/4
Saint Ouen	20/02/2017 – 12/03/2017	10/15	1/6

Cette interrogation est cruciale pour optimiser les efforts de mesure tout en garantissant des résultats robustes et représentatifs. D'un point de vue opérationnel, il s'agit de déterminer comment réaliser des mesures sur la durée la plus courte possible, afin de limiter les coûts, tout en obtenant des résultats suffisamment fiables pour la SNCF.

Pour répondre à cette question, une approche par fenêtre glissante a été mise en œuvre. Cette méthode permet d'évaluer progressivement la justesse des profils types en fonction de différentes durées d'échantillonnage.

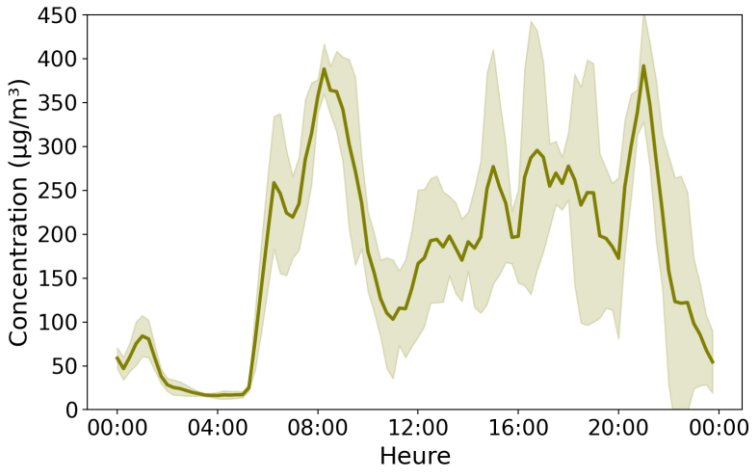


Figure 3.8 : Profil journalier type obtenu à la gare de Grigny Centre durant les jours ouvré. La bande colorée correspond à 1σ.

La procédure consiste à appliquer la méthodologie sur des sous-fenêtres glissantes couvrant des durées croissantes, comme illustré dans la Figure 3.9. Le processus débute avec une fenêtre glissante d'une semaine, puis s'étend progressivement jusqu'à inclure l'ensemble du jeu de données (n semaines). Pour chaque sous-fenêtre ($w = 1,2,3,\dots,n$), un profil type journalier est recalculé en utilisant uniquement les données contenues dans cette fenêtre. La sous-fenêtre est ensuite décalée le long de l'axe temporel, permettant ainsi de capturer toutes les combinaisons possibles de données.

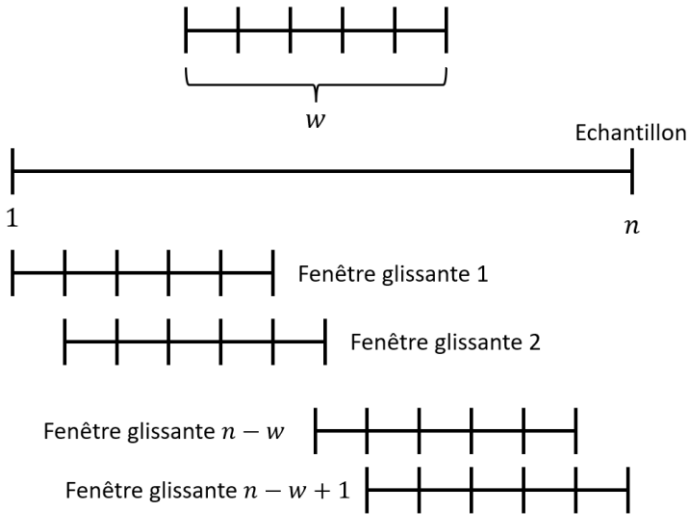


Figure 3.9 : Illustration du principe de la fenêtre glissante pour l'échantillonnage progressif des données (adapté de Polanco-Martínez 2019)

Chaque profil type journalier obtenu à partir d'une fenêtre glissante est comparé au profil de référence global, établi sur la totalité de la période d'étude. Cette comparaison est réalisée à l'aide de l'erreur absolue moyenne (MAE), un indicateur quantifiant la différence entre les profils calculés sur des sous-périodes et le profil global. Par définition, la MAE est définie comme suit :

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{3.2}$$

où N représente le nombre total d'observations, y_i est la valeur réelle de la variable à l'instant i, $\hat{y}_i$ est la valeur estimée ou prédite à l'instant i.
Dans notre cas, $N = 96$ correspond au nombre total de points correspondant aux données quart horaires dans une journée.

Pour chaque durée d'échantillonnage, une MAE moyenne est calculée, accompagnée de son écart type, permettant ainsi d'évaluer la précision des profils recalculés en fonction de la durée de la fenêtre d'échantillonnage. Une MAE faible indique une forte correspondance entre le profil recalculé et le profil de référence.

3.2.5 Evaluation de la réduction des PM pour les systèmes de filtration

Pour quantifier l'efficacité des systèmes de filtration, nous avons calculé le ratio des DAC sur une période suffisamment longue lorsque le système de filtration était en place (DAC_{Y_filt}) par rapport à une période de même durée sans filtration (DAC_{Y_nofilt}) :

$$Ratio = \frac{DAC_{Y_filt}}{DAC_{Y_nofilt}} \tag{3.3}$$

Un ratio inférieur à 1 indique que le niveau de concentration des particules (PM) pendant la période de l'expérimentation est plus faible que celui de la période sans expérimentation.

À partir de ce ratio, la réduction en pourcentage des concentrations de PM, représentant la performance du système de filtration, peut être dérivée comme suit :

$$Réduction\ (\%) = \left(1 - \frac{\overline{Ratio_{in_EXP}}}{\overline{Ratio_{out_EXP}}}\right) \times 100 \tag{3.4}$$

où $\overline{Ratio_{in_EXP}}$ est la moyenne des ratios DAC pendant la période d'expérimentation, $\overline{Ratio_{out_EXP}}$ est la moyenne des ratios DAC en dehors de la période d'expérimentation.

Pour évaluer la variabilité de ces réductions, nous avons également calculé l'écart-type ($\sigma_{réduction}$) des ratios DAC. Cela permet d'estimer la cohérence de la performance du système de filtration au cours de la période expérimentale. L'écart-type est calculé comme suit :

$$\sigma_{réduction}(\%) = Réduction(\%) \times \sqrt{\left(\frac{\sigma_{in_EXP}}{\overline{Ratio_{in_EXP}}}\right)^2 + \left(\frac{\sigma_{out_EXP}}{\overline{Ratio_{out_EXP}}}\right)^2} \quad (3.5)$$

où σ_{in_EXP} est l'écart-type des ratios DAC pendant la période d'expérimentation, σ_{out_EXP} est l'écart-type des ratios DAC en dehors de la période d'expérimentation.

3.3 Modèle additif généralisé (GAM)

Pour répondre à la problématique des facteurs influençant les concentrations en particules et évaluer l'efficacité des systèmes de filtration, nous avons adopté le modèle additif généralisé (GAM), introduit par Hastie and Tibshirani (1986).

Le GAM est une extension des modèles linéaires et des modèles linéaires généralisés. Il est conçu pour modéliser les relations entre une variable cible (ici, les concentrations en particules) et plusieurs paramètres explicatifs. Ces derniers incluent des facteurs environnementaux, opérationnels ou externes, qui peuvent interagir de manière complexe avec la variable cible. Ces paramètres explicatifs peuvent être représentés à travers des variables catégorielles ou continues. Les variables catégorielles incluent des facteurs qualitatifs ayant des niveaux distincts. Ces variables influencent directement les concentrations de particules, mais leur effet ne peut pas être modélisé comme une relation continue. Les variables continues, en revanche, décrivent des phénomènes mesurés sur une échelle numérique continue, tels que des facteurs environnementaux ou temporels. Ces variables peuvent avoir des relations complexes avec les concentrations de particules, ce qui justifie l'utilisation de fonctions lissées dans le GAM.

Le modèle GAM s'appuie sur une forme générale qui combine des relations linéaires et non linéaires entre les prédicteurs et la variable cible. Elle est exprimée comme suit :

$$g\big(\mathbb{E}(y_i)\big) = \beta_0 + \sum_{j=1}^{m} \alpha_j C_j + \sum_{k=1}^{p} S_k(x_{ik}) + \varepsilon_i \quad (3.6)$$

Où i représente l'indice, allant de 1 à n, correspondant au nombre total d'observations. De même, j varie de 1 à m, où m correspond au nombre de variables catégorielles incluses dans le modèle, tandis que k varie de 1 à p, p étant le nombre de variables continues. La valeur attendue de la variable réponse, notée $\mathbb{E}(y_i)$, est reliée aux prédicteurs par une fonction de liaison $g(\cdot)$.

Le paramètre β_0 désigne l'ordonnée à l'origine, représentant la valeur de référence de la variable réponse lorsque toutes les variables explicatives sont à leur niveau de référence. Les coefficients α_j sont associés aux variables catégorielles C_j, tandis que les fonctions lisses $S_k(x_{ik})$ capturent les relations non linéaires entre les prédicteurs continus x_{ik} et la variable réponse. Enfin, le terme ε_i représente les résidus, c'est-à-dire les écarts entre les valeurs observées et celles prédites par le modèle.

En général, les fonctions de liaison $g(\cdot)$ utilisées dans les modèles GAM dépendent des caractéristiques de la variable réponse et des distributions sous-jacentes. Parmi les fonctions de liaison les plus courantes, on retrouve :

- Fonction identité ($g(z) = z$) : utilisée lorsque la variable réponse peut être modélisée directement sur son échelle d'origine.

- Fonction logarithme ($g(z) = \log(z)$) : souvent utilisée avec des distributions Gamma ou de Poisson pour garantir des prédictions strictement positives.
- Fonction logit $\left(g(z) = \log\left(\frac{z}{1-z}\right)\right)$: adaptée pour modéliser des probabilités dans des contextes où la variable réponse est binaire.
- Fonction racine carrée ($g(z) = \sqrt{z}$) : utilisée pour stabiliser la variance lorsque la variable réponse est modélisée avec une distribution de Poisson.
- Fonction inverse $\left(g(z) = \frac{1}{z}\right)$: appliquée dans certains modèles avec des distributions Gamma, pour des variables avec des relations inverses entre prédicteurs et variable réponse.

Dans notre cas, étant donné notre objectif de capturer avec précision l'influence des facteurs environnementaux sur les concentrations de particules, nous avons opté pour la fonction de liaison identité ($g(z) = z$) dans notre modèle GAM. Ce choix est particulièrement adapté, car il permet de conserver les résultats du modèle sur l'échelle originale de la variable réponse, facilitant ainsi une interprétation directe des effets des prédicteurs, tels que les expérimentations de filtration, sur les niveaux de PM.

3.3.1 Critères de sélection

Les variables explicatives ont été sélectionnées en fonction de leur disponibilité, de leur pertinence environnementale, de leur influence potentielle sur les concentrations de particules et de l'absence de multi-colinéarité excessive. Pour évaluer la colinéarité entre les variables, nous avons calculé le facteur d'inflation de la variance (VIF pour Variance Inflation Factor). Seules les variables avec un VIF acceptable, indiquant une colinéarité modérée, ont été retenues dans le modèle. Cette étape garantit que chaque variable explicative apporte une information unique et significative dans la prédiction des concentrations de particules, sans redondance avec les autres variables incluses.

Le VIF est défini par la relation :

$$VIF = \frac{1}{1 - R^2} \tag{3.7}$$

Selon Zuur et al. (2009), une valeur seuil de 5, voire 3, peut être utilisée pour identifier et exclure les variables colinéaires. Dans cette étude, nous avons adopté un seuil de 3 pour garantir une stabilité optimale du modèle GAM, conformément aux recommandations de Zuur et al. (2009) et Belušić et al. (2015).

Une approche itérative a été mise en œuvre pour gérer la colinéarité : une variable était retirée à chaque itération, le VIF était recalculé, et ce processus est répété jusqu'à ce que toutes les variables explicatives restantes respectent le seuil de VIF fixé.

En appliquant ce processus, seules les variables explicatives présentant un VIF faible, reflétant une colinéarité modérée, ont été conservées dans le modèle. Cette étape garantit que chaque variable explicative apporte une information unique et significative dans la prédiction des concentrations de particules, sans redondance avec les autres variables incluses.

Le Tableau 3.3 illustre le processus d'analyse de colinéarité réalisé pour identifier et retirer les variables explicatives présentant une forte colinéarité à Avenue Foch.

Les itérations successives ont conduit à l'élimination des paramètres fortement colinéaires, comme $PM_{2.5}$ (ou PM_{10} lorsque $PM_{2.5}$ est la variable réponse), $PM_{2.5_out}$ (ou PM_{10_out} lorsque $PM_{2.5}$ est la variable réponse), NO_{out} et NOx_{out}. À chaque itération, les variables jugées redondantes ont été retirées.

Tableau 3.3 : VIF des variables explicatives du GAM à Avenue Foch, pour la variable cible PM_{10}

Nombre d'itérations	1	2	3	4	5
PM_{10}	12,9	12,9	1,9	1,8	1,8
$PM_{2.5}$	12,5	12,5	Retiré	Retiré	Retiré
CO_2	1,5	1,5	1,5	1,5	1,5
RH	1,7	1,7	1,7	1,7	1,7
T	2,0	2,0	1,9	1,7	1,6
PM_{10_out}	3,8	3,8	3,8	3,8	1,9
$PM_{2.5_out}$	2,9	2,9	2,9	2,9	Retiré
NO_{2_out}	4,2	4,2	4,1	2,2	2,1
NO_{out}	∞	4,9	4,7	Retiré	Retiré
NOx_{out}	∞	Retiré	Retiré	Retiré	Retiré
Time	1,6	1,6	1,6	1,5	1,5
Week	1,2	1,2	1,2	1,2	1,2
DOW	1,1	1,1	1,1	1,1	1,1

Tableau 3.4 : VIF des variables explicatives du GAM à Sevran-Beaudottes, pour la variable cible PM_{10}

Nombre d'itérations	1	2	3	4
PM_{10}	33,8	33,7	1,7	1,7
$PM_{2.5}$	34,1	34,1	Retiré	Retiré
T	1,5	1,5	1,5	1,5
PM_{10_out}	4,3	4,3	4,3	2,1
$PM_{2.5_out}$	4,1	4,1	4,1	Retiré
NO_{2_out}	1848,1	1,8	1,8	1,8
NO_{out}	8767,3	1,9	1,9	1,9
NOx_{out}	14977,5	Retiré	Retiré	Retiré
Time	1,4	1,4	1,4	1,4
Week	1,2	1,2	1,2	1,2
DOW	1	1	1	1

Dans le cas de la gare de Sevran-Beaudottes (Tableau 3.4), le processus a permis d'éliminer les variables $PM_{2.5}$ et $PM_{2.5_out}$ lorsque PM_{10} était la variable cible, et inversement, PM_{10} et PM_{10_out} ont été retirés lorsque $PM_{2.5}$ était la variable cible. À l'issue de ce processus, les variables retenues incluent PM_{10} (ou $PM_{2.5}$ selon la cible), T, PM_{10_out} (ou $PM_{2.5_out}$ selon la cible), NO_{2_out}, NO_{out}, Time, Week, et DOW (Day of week pour jour de la semaine).

Le Tableau 3.5 présente les résultats du VIF pour les variables explicatives prises en compte dans le modèle GAM pour la gare de Magenta. Ainsi, les variables explicatives retenues pour cette gare incluent PM_{10}, T, PM_{10_out}, NO_{2_out} , NO_{out}, O_{3_out}, Time, Week et DOW

Tableau 3.5 : VIF des variables explicatives du GAM à Magenta, pour la variable cible PM_{10}

Nombre d'itérations	1	2	3
PM_{10}	12,9	12,9	1,8
$PM_{2,5}$	14,2	14,1	Retiré
T	1,5	1,5	1,5
PM_{10_out}	2,0	2,0	1,8
NO_{2_out}	2445,1	2,8	2,8
NO_{out}	7370,1	1,8	1,8
NOx_{out}	14655,6	Retiré	Retiré
O_{3_out}	2,5	2,5	2,5
Time	1,4	1,4	1,4
Week	1,3	1,3	1,3
DOW	1	1	1

3.3.2 Identification et traitement des variables catégorielles

Dans le modèle final, ces variables sont représentées par des coefficients linéaires associés à chaque catégorie. Par exemple :

- **Jour de la semaine (DOW)** : Cette variable catégorielle permet de différencier l'impact des jours ouvrés (lundi à vendredi) et des weekends (samedi et dimanche) sur les concentrations de particules. Chaque jour a son propre coefficient pour capturer les spécificités opérationnelles et environnementales liées au trafic ou à l'activité humaine.
- **Périodes expérimentales (Exp)** : Ces variables catégorielles indiquent la présence ou l'absence des systèmes de filtration testés lors des périodes expérimentales. Les coefficients associés à ces variables permettent de quantifier l'impact direct de chaque système sur les concentrations de particules.

Les variables catégorielles ne sont pas lissées dans le GAM, car elles ne possèdent pas de continuité entre les niveaux. Elles sont intégrées linéairement sous forme de termes indicateurs (ou « dummy variables »).

3.3.3 Identification et traitement des variables continues

- **Facteurs environnementaux :** Les variables telles que la température (T), l'humidité relative (RH) et les concentrations en dioxyde de carbone (CO_2) ont été introduites dans le modèle avec des termes lissés. Ces fonctions lissées (représentées par $s(\cdot)$) permettent de capturer les relations non linéaires entre ces paramètres et les concentrations de particules.
- **Temps ($Time$)** : Une approche spécifique a été adoptée pour cette variable, car les profils temporels ne sont pas identiques selon le type de jour, principalement en raison des différences dans les fréquences de train entre les jours ouvrés et les weekends. Plutôt que de considérer le temps comme une simple échelle continue, nous avons stratifié cette variable par type de jour ($Daytype$). Concrètement, cela signifie que le modèle distingue les effets temporels selon qu'il s'agisse d'un jour ouvré ou d'un weekend, en utilisant une fonction lissée pondérée par la variable catégorielle Daytype ($s(Time, by = Daytype)$). Cette approche permet ainsi de mieux capturer les variations horaires tout en tenant compte des différences de fonctionnement entre les jours ouvrés et les weekends.

- **Saisonnalité hebdomadaire ($Week$)** : Pour capturer les tendances hebdomadaires, une fonction lissée a été appliquée à la variable $Week$. Cette variable continue représente le numéro de la semaine dans l'année et permet de modéliser les variations saisonnières lentes au cours de l'année.

3.3.4 Équation finale

Le modèle final pour Avenue Foch se présente comme suit :

$$PM = \beta_0 + \alpha_{DOW}DOW + \alpha_{Exp1}Exp1 + \alpha_{Exp2}Exp2 + s(CO_2) + s(T) + s(RH) + s(PM_{10_out}) + s(NO_{2_out}) + s(Time, by = Daytype) + s(Week) + \varepsilon \quad (3.8)$$

Ce modèle combine des termes linéaires pour les variables catégorielles $(DOW, Exp1, Exp2)$ et des termes lisses $(s(\cdot))$ pour les variables continues $(CO_2, T, RH, PM_{out}, NO_{2_out}, Time, Week)$. Chaque composante a pour but de capturer les différentes influences sur les concentrations de particules dans la gare d'Avenue Foch.

Pour la gare de Sevran-Beaudottes, le modèle, destiné à analyser les concentrations de particules, prend la forme suivante :

$$PM = \beta_0 + \alpha_{DOW}DOW + \alpha_{Exp}Exp + s(T) + s(PM_{out}) + s(NO_{2_out}) + s(NO_{out}) + s(Time, by = Daytype) + s(Week) + \varepsilon \quad (3.9)$$

Ce modèle intègre à la fois des effets paramétriques, liés aux jours de la semaine (DOW) et à la période d'expérimentation (Exp), ainsi que des termes non paramétriques représentés par des fonctions lisses (s) pour la température (T), les concentrations extérieures de polluants $(PM_{out}, NO_{2_out}, NO_{out})$, l'heure de la journée (Time, modulé par le type de jour, $Daytype$), et les variations hebdomadaires (Week).

Enfin, le modèle GAM pour la gare de Magenta est formulé comme suit :

$$PM = \beta_0 + \alpha_{DOW}DOW + s(T) + s(PM_{out}) + s(NO_{2_out}) + s(NO_{out}) + s(O_{3_out}) + s(Time, by = Daytype) + s(Week) + \varepsilon \quad (3.10)$$

Ce modèle inclut à la fois des effets paramétriques associés aux jours de la semaine (DOW), ainsi que des termes non-paramétriques représentés par des fonctions lisses (s) pour la température (T), les concentrations extérieures de polluants $(PM_{out}, NO_{2_{out}}, NO_{out}, O_{3_out})$, l'heure de la journée ($Time$ modulé par le type de jour $Daytype$) et les variations hebdomadaires $(Week)$.

L'utilisation du modèle additif généralisé (GAM) permet de produire plusieurs types de sorties essentielles pour interpréter et évaluer les relations entre les concentrations de particules (variable cible) et les variables explicatives. Ces sorties offrent une vision globale des effets des prédicteurs, qu'ils soient linéaires ou non linéaires, tout en permettant une analyse détaillée de l'ajustement du modèle et de la pertinence des relations modélisées.

Le tableau des coefficients paramétriques résume les contributions des variables catégorielles et des termes linéaires inclus dans le modèle. Chaque coefficient est accompagné de sa valeur estimée, de son écart-type, de sa valeur statistique (t-value), et du niveau de significativité (p-value). Ces paramètres sont cruciaux pour évaluer l'effet des variables catégorielles et permettent de quantifier l'impact de chaque catégorie par rapport à une référence (β_0). Par exemple, un coefficient positif (α) pour un jour

spécifique pourrait indiquer une augmentation des concentrations en particules par rapport à un jour de référence.

Le GAM propose également une évaluation statistique de la significativité des termes lissés, souvent rapportée sous forme de tableau. Ce tableau inclut :

- Le degré de liberté effectif (Effective Degrees of Freedom, EDF), qui reflète la complexité du terme lissé. Un EDF proche de 1 indique une relation proche d'une droite, tandis qu'un EDF élevé suggère une relation fortement non linéaire.
- Le degré de liberté de référence (Reference Degrees of Freedom, Ref.df), qui indique la complexité maximale que le modèle pourrait utiliser pour représenter la relation. En d'autres termes, il définit la limite supérieure pour l'EDF.
- La statistique F, qui évalue si un terme lissé apporte une contribution significative au modèle. Elle est calculée comme le rapport entre la variance expliquée par le terme lissé et la variance non expliquée (résiduelle).
- Le niveau de significativité (p-value), permettant de déterminer si un terme lissé contribue de manière significative à l'explication de la variable cible.

Ces 4 paramètres permettent de vérifier la pertinence des relations non linéaires modélisées et d'identifier les prédicteurs ayant un impact significatif sur les concentrations de particules.

Les graphiques de dépendance partielle sont essentiels pour visualiser les relations non linéaires entre les prédicteurs continus et la variable cible dans un modèle GAM. Ils représentent la contribution estimée d'un prédicteur à la variable cible, en maintenant les autres variables constantes. Sur ces graphiques, l'axe des abscisses correspond aux valeurs du prédicteur analysé, tandis que l'axe des ordonnées indique l'effet partiel estimé (souvent centré autour de zéro) sur la variable cible. Ainsi, ces courbes permettent de visualiser comment les concentrations de particules varient en fonction des variables explicatives, en isolant l'influence propre de chacune d'elles. Ils offrent une interprétation visuelle claire des relations complexes capturées par les termes lissés (*smooth terms*), facilitant la compréhension des effets non linéaires. Pour les variables non continues, les effets sont représentés sous forme de points ou de segments discrets. Chaque niveau de la variable est alors associé à un effet estimé sur la variable cible, permettant de comparer directement l'influence relative des différentes modalités.

3.3.5 Evaluation de la réduction des PM à l'aide du modèle GAM

Pour quantifier l'efficacité du système de filtration, la réduction en pourcentage des concentrations de PM peut être calculée comme suit :

$$Réduction\ (\%) = \left(\frac{\alpha_{Exp}}{\beta_0}\right) \times 100 \tag{3.11}$$

où $Réduction\ (\%)$ représente le pourcentage de réduction des concentrations de PM dû à l'effet du système de filtration, α_{Exp} est le coefficient associé à l'expérimentation, indiquant la variation des concentrations de PM lorsque le système de filtration est activé, β_0 est l'ordonnée à l'origine, correspondant à la concentration moyenne de PM en l'absence de filtration (condition de référence).

Ainsi, $\frac{\alpha_{Exp}}{\beta_0}$ représente la réduction fractionnaire de la concentration de pollution causée par le système de filtration par rapport à la concentration de référence.

L'écart-type ($\sigma_{réduction}$) de cette réduction est calculé comme suit :

$$\sigma_{réduction}(\%) = Réduction(\%) \times \sqrt{\left(\frac{\sigma_{Exp}}{\alpha_{Exp}}\right)^2 + \left(\frac{\sigma_{\beta_0}}{\beta_0}\right)^2} \quad (3.12)$$

où $\sigma_{réduction}(\%)$ représente l'écart-type de la réduction en pourcentage, fournissant une mesure de l'incertitude associée à l'estimation de la réduction, σ_{Exp} est l'écart-type du coefficient α_{Exp}, reflétant la variabilité des changements de concentration de PM liés au système de filtration, σ_{β_0} est l'écart-type de l'ordonnée à l'origine, indiquant la variabilité de la concentration de PM de référence.

3.4 Réseau de neurones récurrents à mémoire à long et court terme (LSTM)

Pour répondre aux besoins de prévision des concentrations en particules et à l'évaluation de l'efficacité des systèmes de filtration, nous avons adopté un modèle de réseau de neurones récurrents à mémoire à long et à court terme (LSTM) introduit par Hochreiter and Schmidhuber (1997), particulièrement adapté pour capturer les dépendances temporelles comme la qualité de l'air. Le choix de cette méthode a été motivé par PMForecast (Rahmani et al., 2024), un outil basé sur LSTM, développé par Maryam Rahmani durant sa thèse sous la direction de Suzanne Crumeyrolle pour prédire la pollution atmosphérique en particules extérieures et par ses résultats très prometteurs dans l'air extérieur. Dans notre cas, l'utilisation de cette approche vise principalement à évaluer les performances des systèmes de filtration dans les EFS. Cela est réalisé en comparant les concentrations de particules prédites par le modèle avec les valeurs réelles mesurées au cours des périodes d'expérimentation. Cette méthodologie permet de quantifier l'impact des systèmes de filtration sur les concentrations de particules, en tenant compte des variations temporelles et des conditions spécifiques observées dans les EFS.

3.4.1 Architecture et fonctionnement interne d'une cellule LSTM

La Figure 3.10 illustre la structure d'une cellule LSTM (Long Short-Term Memory), utilisée pour modéliser des dépendances à long terme dans des séries temporelles. Les LSTM disposent de mécanismes de contrôle (portes) permettant de conserver ou d'oublier des informations selon leur pertinence grâce à trois portes principales : la porte d'oubli (f_t), la porte d'entrée (i_t), et la porte de sortie (o_t).

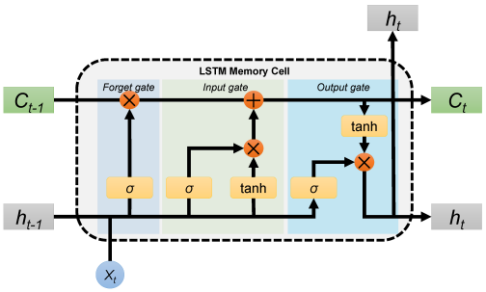


Figure 3.10 : Architecture interne d'une cellule LSTM (Fan et al., 2020)

- **Entrées et états :**
  - x_t : entrée à l'instant t
  - h_{t-1} : état caché du pas de temps précédent.
  - C_{t-1} : état mémoire précédent.
- **Porte d'oubli (f_t)** : détermine quelles informations issues de la mémoire précédente (C_{t-1}) doivent être conservées ou oubliées.

- **Porte d'entrée (i_t) et état candidat ($\widetilde{C}_t$)** : permettent d'ajouter de nouvelles informations pertinentes à la mémoire.
- **Mise à jour de la mémoire (C_t)** : combine l'information retenue par la porte d'oubli et les nouvelles données ajoutées par la porte d'entrée.
- **Porte de sortie (o_t)** : régule quelles informations de la mémoire mise à jour (C_t) seront utilisées pour produire l'état caché actuel (h_t).
- **Sorties :**
  - h_t: état caché, utilisé pour les prédictions ou transmis à la cellule suivante.
  - C_t : mémoire mise à jour, transmise au pas de temps suivant.

Ces mécanismes permettent aux LSTM de mémoriser les informations pertinentes sur de longues séquences tout en ignorant celles qui sont inutiles, rendant ce modèle particulièrement adapté aux séries temporelles et aux processus séquentiels complexes (Li et al., 2017; Tran et al., 2023).

3.4.2 Entraînement et validation du modèle

Pour entraîner et valider le modèle LSTM, nous avons utilisé les données disponibles avant le début de chaque expérimentation, couvrant une période allant de quelques mois à plusieurs années, en fonction du début de l'expérimentation concernée.

La gestion des valeurs manquantes a représenté un défi majeur en raison de la présence de longues périodes sans données. En effet, les modèles LSTM ne peuvent pas traiter directement les valeurs manquantes. Les méthodes classiques, telles que l'interpolation linéaire ou le moyennage glissant, se sont révélées inadaptées pour reconstruire de manière fiable les données manquantes sur de longues périodes. Par conséquent, la solution retenue a été de supprimer les pas de temps contenant des valeurs manquantes. Cette approche garantit que seules des données complètes sont utilisées pour l'entraînement du modèle.

Les données ainsi nettoyées et prétraitées ont été soigneusement sélectionnées afin de refléter les conditions de base, sans l'influence d'événements notables. Cela permet de créer un modèle capable de prédire les concentrations de particules dans des conditions de référence, qui pourront ensuite être comparées aux données mesurées durant les périodes expérimentales.

Le jeu de données a été divisé en deux ensembles selon une répartition classique : 80 % des données ont été utilisées pour l'entraînement, tandis que les 20 % restants ont été réservés pour tester et valider les performances du modèle. Cette division permet de s'assurer que le modèle est évalué sur des données qu'il n'a pas vues lors de son apprentissage, garantissant ainsi une validation robuste de sa capacité de généralisation.

Enfin, le modèle LSTM a été configuré pour prédire les concentrations de particules avec un horizon de prévision d'une heure, en se basant sur les données des 48 heures précédentes. Ce choix permet de capturer les dépendances temporelles pertinentes tout en étant adapté aux objectifs de prévision. L'utilisation d'une fenêtre temporelle glissante de deux jours fournit un historique suffisant pour détecter les tendances tout en limitant la complexité computationnelle.

3.4.3 Hyperparamètres et optimisation du modèle

Pour maximiser les performances du modèle, plusieurs hyperparamètres ont été optimisés à l'aide d'une recherche empirique. Les configurations finales sont les suivantes :

- Taux d'apprentissage (learning rate) : Défini à 64 après avoir testé les valeurs $[16, 32, 64, 128]$. Ce paramètre détermine la taille des pas effectués par l'algorithme d'optimisation lors de la mise à jour des poids.

- Batch size : Défini à 64 après avoir testé les valeurs $[16, 32, 64, 128]$. Ce paramètre représente le nombre d'échantillons traités simultanément lors de chaque itération.
- Nombre d'époques (epochs) : Fixé à 200 après avoir évalué des valeurs entre $[50, 100, 200]$. Les époques correspondent au nombre de passages complets effectués sur le jeu de données d'entraînement.

Pour éviter le surapprentissage (overfitting) et améliorer la généralisation du modèle, une stratégie d'arrêt anticipé (Early Stopping) a été mise en place, comme recommandé dans Rahmani et al. (2024). Cette technique interrompt l'entraînement dès que les performances cessent de s'améliorer ou commencent à se dégrader, empêchant ainsi un entraînement inutile qui pourrait nuire à la capacité du modèle à généraliser.

Une description détaillée de la méthodologie ainsi que des résultats de performance du modèle sont présentées dans l'article 2. Cette publication offre une analyse approfondie des performances de l'algorithme et de son adéquation pour la prévision de la qualité de l'air en milieu souterrain.

3.4.4 Métriques d'évaluation des prédictions

Pour évaluer la qualité des prédictions réalisées par le modèle LSTM, plusieurs métriques d'évaluation ont été utilisées. Ces métriques permettent de quantifier l'écart entre les valeurs prédites par le modèle et les valeurs observées, tout en offrant différentes perspectives sur la précision et la performance du modèle. Les métriques retenues dans cette étude sont les suivantes :

- **L'erreur quadratique moyenne (MSE – Mean Squared Error)** mesure l'écart moyen entre les valeurs prédites ($\hat{y}_i$) et les valeurs réelles (y_i), en élevant ces écarts au carré pour pénaliser les grandes erreurs :

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{3.13}$$

- **L'erreur absolue moyenne (MAE – Mean Absolute Error)** représente l'écart absolu moyen entre les prédictions et les observations, ce qui en fait une mesure plus robuste face aux grandes erreurs. Elle mesure directement l'écart moyen entre les valeurs prédites et réelles, en restant sur la même échelle que les données (Equation 3.2).
- **La racine de l'erreur quadratique moyenne (RMSE - Root Mean Squared Error)**, obtenue en prenant la racine carrée de la MSE, amplifie l'impact des grandes erreurs. Cela permet d'avoir une vue plus sensible des écarts importants. La RMSE est particulièrement utile lorsqu'il est crucial de minimiser les grandes erreurs dans les prédictions :

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{3.14}$$

- **Le coefficient de détermination R^2** mesure la proportion de la variance des valeurs réelles (y_i) expliquée par les prédictions ($\hat{y}_i$). Il est défini comme suit :

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{3.15}$$

où $\bar{y}$ est la moyenne des valeurs réelles. Un R^2 proche de 1 indique un excellent ajustement du modèle.

- Enfin, **l'erreur absolue moyenne pondérée (WMAPE - Weighted Mean Absolute Percentage Error)** pondère les erreurs absolues par rapport à la somme des valeurs réelles. Exprimée en pourcentage, cette métrique offre une perspective relative de la précision des prédictions :

$$WMAPE = \frac{\sum_{i=1}^{N}|y_i - \hat{y}_i|}{\sum_{i=1}^{N}|y_i|} \times 100 \quad (3.16)$$

3.4.5 Précision des prévisions des concentrations de PM sur différents horizons temporels

Afin d'évaluer la performance du modèle de prédiction des concentrations de PM, nous avons évalué la précision de la prédiction des niveaux de PM pour les ensembles d'entraînement et de test, sur des horizons de prévision allant de 1 à 72 heures (Figure 3.11). La Figure 3.11a et la Figure 3.11b montrent les scores moyens de R^2 et RMSE sur différentes fenêtres de prévision [1, 6, 12, 24, 48, 72 heures]. Pour assurer une prévision précise, chaque fois qu'une prédiction était effectuée, la valeur prédite était réinjectée dans le jeu de données historique. Ce processus consistait à supprimer le point de données le plus ancien et à ajouter la nouvelle prédiction comme la donnée la plus récente. Par exemple, lors de la prévision pour les 12 heures suivantes, le modèle génère des prédictions horaires de manière itérative, qui sont ensuite moyennées pour calculer le R^2 global pour cette période. Une diminution de la précision est observée à mesure que la fenêtre de prévision s'allonge, avec un score R^2 tombant sous 0,5 au-delà de 72 heures, ce qui indique les limitations du modèle pour prédire les niveaux de PM sur des périodes supérieures à 3 jours.

Ainsi, pour l'évaluation des résultats issus de cette méthode, l'accent a été mis sur les 3 premiers jours de prévision.

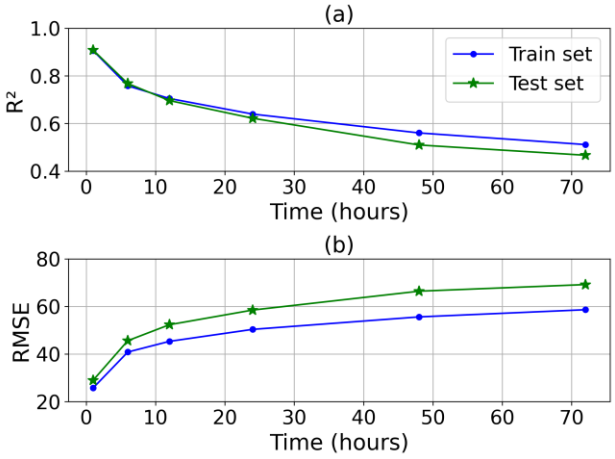


Figure 3.11 : Évaluation de la performance du modèle LSTM : (a) Précision évaluée par R², (b) RMSE (µg/m³). Les lignes bleues indiquent la performance sur les ensembles d'entraînement, tandis que les lignes vertes représentent la performance sur les ensembles de test (Article 2, en annexe)

3.4.6 Evaluation de la réduction des PM à l'aide du modèle LSTM

Pour quantifier l'efficacité du système de filtration, la réduction en pourcentage des concentrations de PM peut être calculée comme suit :

$$Réduction(\%) = \frac{1}{N}\sum_{t=1}^{N}\left(1 - \frac{PM_{LSTM,t}}{PM_{GT,t}}\right) \times 100 \quad (3.17)$$

où $Réduction(\%)$ représente la réduction moyenne en pourcentage, correspondant à la différence moyenne entre les valeurs de PM prédites par le modèle LSTM et les valeurs réelles mesurées (Ground Truth), N est le nombre total d'intervalles de temps, $PM_{LSTM,t}$ est la concentration de PM prédite par le modèle LSTM à l'instant t, et $PM_{GT,t}$ est la concentration de PM mesurée à l'instant t.

L'écart-type ($\sigma_{Réduction}$) de cette réduction est calculé comme suit :

$$\sigma_{Réduction}(\%) = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(\left(1 - \frac{PM_{LSTM,t}}{PM_{GT,t}}\right) \times 100 - Réduction(\%)\right)^2} \quad (3.18)$$

3.5 Conclusion du chapitre

Ce chapitre a présenté les trois approches méthodologiques principales développées ou utilisées dans cette étude pour analyser les concentrations en particules. Chaque méthode apporte une perspective unique et complémentaire à la problématique.

La méthode basée sur le Coefficient d'Amplitude Journalier (DAC) permet de suivre l'évolution des concentrations de particules sur l'ensemble des périodes expérimentales, en fournissant des profils journaliers typiques. Elle quantifie les variations intra-journalières grâce à une métrique simple et efficace. Le DAC offre ainsi une approche claire et rapide pour visualiser les performances des systèmes de filtration, en comparant différentes périodes (avec et sans système de filtration), tout en capturant les tendances saisonnières et temporelles dans la réduction des PM. Cependant, son efficacité repose sur la disponibilité de données continues et non perturbées. Les périodes anormales, telles que les pandémies ou les grèves, peuvent perturber les résultats. Malgré ces limitations, le DAC reste un outil précieux pour les évaluations à long terme, à condition d'avoir suffisamment de données fiables.

En contrastant avec le DAC, le Modèle Additif Généralisé (GAM) offre une approche plus flexible et statistiquement robuste. Contrairement au DAC, qui se concentre sur les variations globales des concentrations, le GAM permet d'isoler les effets des variables explicatives. Le GAM modélise des relations non linéaires et capture les interactions complexes entre les PM et divers prédicteurs, offrant une interprétation détaillée des effets marginaux de chaque variable. Cependant, le modèle GAM présente aussi des défis, comme l'évaluation des tendances de réduction sur toute la période d'expérimentation et la disponibilité des variables explicatives nécessaires, ce qui peut compliquer son application dans certains contextes.

Le modèle LSTM, quant à lui, représente une approche avancée d'apprentissage automatique, particulièrement efficace pour prédire à court terme les évènements de forte pollution dans les EFS, ce qui le rend précieux pour les applications en temps réel. Toutefois, au fur et à mesure que l'horizon de prévision s'allonge, les incertitudes augmentent, ce qui rend les prévisions à long terme moins fiables. En outre, bien que le LSTM offre des performances prédictives puissantes, il est également coûteux en termes de calcul et nécessite de grandes quantités de données. De plus, son manque d'interprétabilité par rapport au GAM complique la compréhension précise de la manière dont les variables influencent les résultats.

Ensemble, ces trois approches permettent une évaluation complète et rigoureuse des systèmes de filtration. Tandis que le DAC et le GAM fournissent une vue d'ensemble robuste sur toute la période d'expérimentation, le LSTM offre une évaluation détaillée et dynamique à court terme. Cette complémentarité méthodologique constitue une base solide pour répondre aux enjeux définis dans cette thèse, alliant précision, robustesse et adaptabilité aux différents contextes opérationnels.

CHAPITRE 4. ANALYSE DES CONCENTRATIONS EN PARTICULES DANS LES EFS

Après avoir exploré les dynamiques des concentrations en particules dans les EFS, présenté notre base de données riche et détaillée, et expliqué les méthodologies adaptées pour analyser ces données dans les chapitres précédents, nous nous attachons maintenant à appliquer cette méthodologie afin d'obtenir des perspectives plus approfondies sur les variations des concentrations en particules. Dans ce chapitre, nous utiliserons ces méthodes pour examiner la variabilité et les tendances à long terme des concentrations en particules observées dans les gares de référence. Nous caractériserons ensuite les profils journaliers types et testerons la robustesse et la transférabilité de notre méthode sur des stations équipées de systèmes de mesure pendant des périodes courtes. Par la suite, nous proposerons une classification des gares basée sur leurs caractéristiques spécifiques et identifierons les principaux facteurs influençant les concentrations en particules. Enfin, nous évaluerons l'incertitude des profils obtenus en fonction de la période de mesure pour assurer la fiabilité de nos analyses.

4.1 Variabilité et tendances à long terme des concentrations en particules dans les gares de références

Cette section se concentre sur l'analyse de la variabilité et des tendances à long terme des concentrations en particules (PM_{10} et $PM_{2.5}$) dans les trois gares de référence : Avenue Foch, Sevran Beaudottes et Magenta. L'objectif est d'identifier les évolutions temporelles de la pollution particulaire dans ces gares, afin de mieux comprendre les dynamiques de particules au sein des EFS.

La Figure 4.1 illustre l'évolution à long terme des concentrations quart-horaires de PM_{10} mesurées dans les trois gares de référence : (a) Avenue Foch entre 2018 et 2022, (b) Sevran Beaudottes entre 2018 et 2021, et (c) Magenta entre 2017 et 2021. Ces graphiques présentent à la fois les concentrations originales de PM_{10}, représentées par des points gris, ainsi que leurs DAC, indiqués en bleu. De manière similaire, la Figure 4.2 représente l'évolution des $PM_{2.5}$ pour les mêmes périodes, avec les concentrations originales en points verts et leurs DAC en bleu.

Les concentrations en particules (PM_{10} et $PM_{2.5}$) dans les trois gares de référence révèlent des différences significatives, tant en termes d'amplitudes que de tendances saisonnières. Pour les trois gares, les concentrations en particules fines dépassent largement les recommandations de l'Organisation mondiale de la santé (WHO, 2021), qui fixe les seuils journaliers à 15 µg/m³ pour les $PM_{2.5}$ et 45 µg/m³ pour les PM_{10}. Ces observations confirment les conclusions de la revue menée par Passi et al. (2021).

À Avenue Foch, les concentrations quart-horaires en PM_{10} peuvent atteindre 450 µg/m³ (Figure 4.1a), tandis que celles en $PM_{2.5}$ atteignent jusqu'à 175 µg/m³ (Figure 4.2a). La gare de Sevran Beaudottes enregistre les niveaux les plus élevés, avec des concentrations en PM_{10} atteignant 500 µg/m³ (Figure 4.1b) et en $PM_{2.5}$ s'élevant à 200 µg/m³ (Figure 4.2b). À l'inverse, la gare de Magenta présente les concentrations les plus faibles parmi les trois gares étudiées, avec des niveaux maximaux de 150 µg/m³ pour les PM_{10} (Figure 4.1c) et de 50 µg/m³ pour les $PM_{2.5}$ (Figure 4.2c).

Ces valeurs particulièrement basses observées dans cette dernière sont assez étonnantes étant donné qu'elle est aussi l'une des gares les plus profondes du réseau SNCF. Néanmoins, plusieurs critères caractéristiques de cette gare peuvent expliquer ces observations. En effet, tout d'abord la présence d'un système d'extraction d'air performant, de systèmes de désenfumage et d'infrastructures modernes, liées à sa mise en service relativement récente, en 1999, qui contribue à améliorer la qualité de l'air sur les quais. De plus, la structure spacieuse de la gare de Magenta (deux quais, quatre voies et une hauteur de

plafond importante) favorise une meilleure dispersion et dilution des particules, limitant ainsi leur accumulation dans l'environnement souterrain (Moreno et al., 2014).

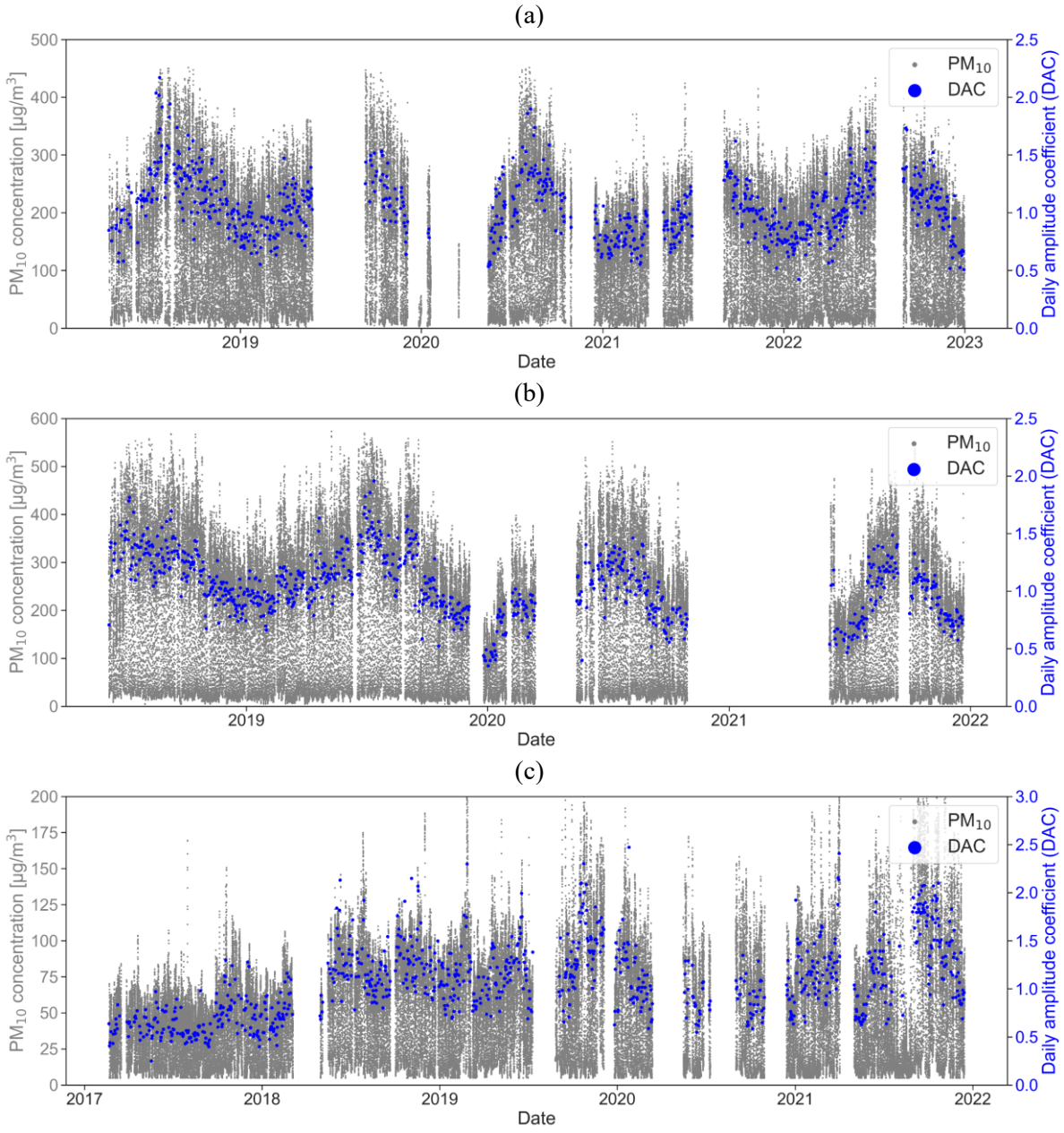


Figure 4.1 : Évolution temporelle du DAC (points bleus) et des concentrations originales de PM_{10} (points gris) mesurées à la gare de (a) Avenue Foch (2018-2022), (b) Sevran Beaudottes (2018-2021) et (c) Magenta (2017-2021)

En revanche, les concentrations un peu plus élevées à Sevran Beaudottes pourraient s'expliquer par l'absence d'accès direct à l'air extérieur, contrairement à Avenue Foch, où l'entrée de la gare débouche directement sur le quai. Cette connexion avec l'extérieur favorise une ventilation naturelle, bien que ces deux gares soient relativement peu profondes (niveau -1).

Ces observations rejoignent les conclusions de Moreno et al. (2014) et Alves et al. (2017). En effet, selon Moreno et al. (2014), les concentrations en PM (PM_{10} et $PM_{2.5}$) dans les systèmes ferroviaires à Barcelone sont fortement influencées par la présence ou l'absence de ventilation mécanique, ainsi que la conception des EFS.

Alves et al. (2017) montrent que les variations des concentrations de PM (PM_{10} et $PM_{2.5}$) dans les EFS sont principalement associées à des différences dans la conception des gares et des tunnels, l'âge des infrastructures, la fréquence des trains, les systèmes de ventilation et de climatisation, la densité des usagers, la géométrie des rails et la qualité de l'air extérieur.

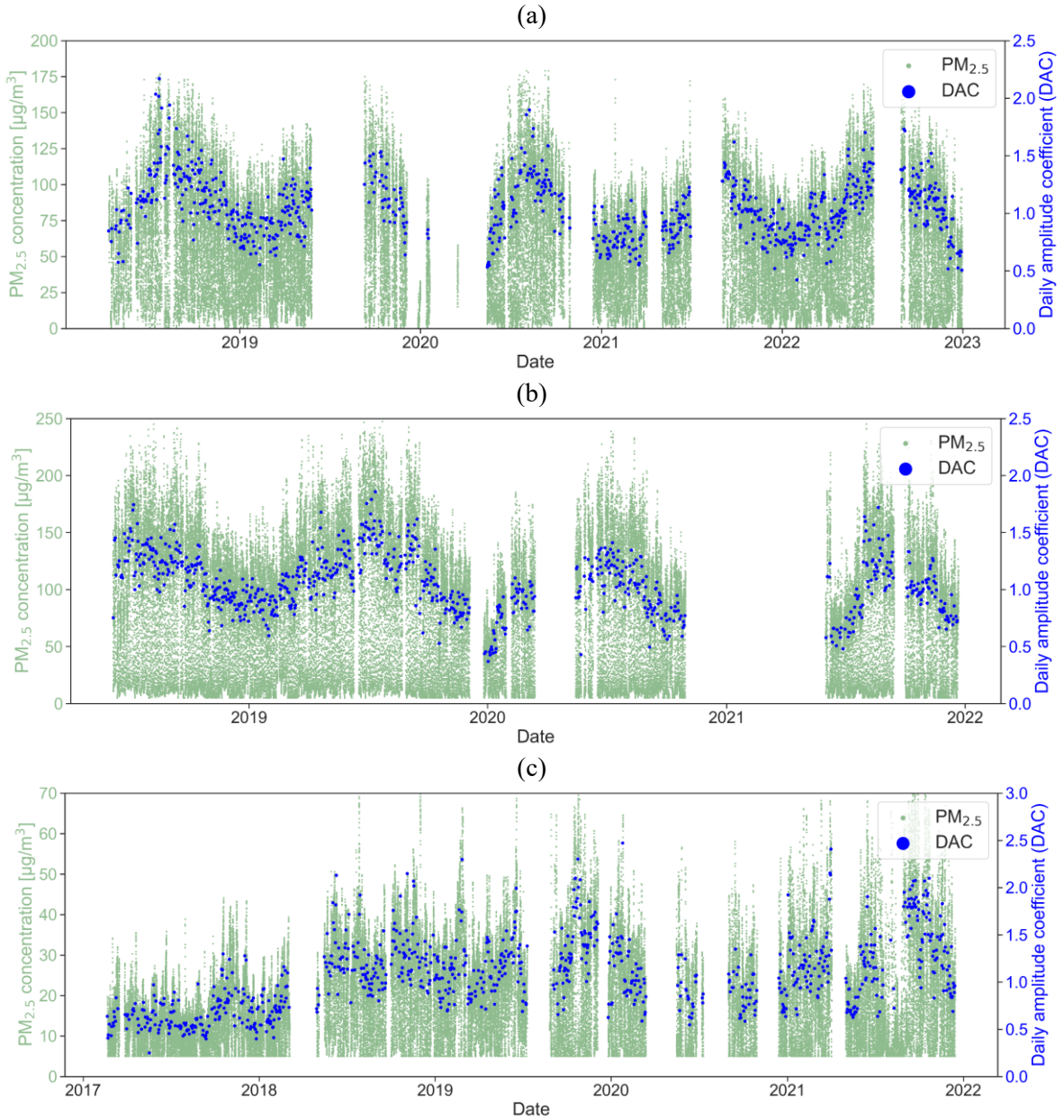


Figure 4.2 : Évolution temporelle du DAC (points bleus) et des concentrations originales de $PM_{2.5}$ (points verts) mesurées à la gare de (a) Avenue Foch (2018-2022), (b) Sevran Beaudottes (2018-2021) et (c) Magenta (2017-2021)

En ce qui concerne les tendances saisonnières, des comportements distincts se dégagent pour chaque gare. À Avenue Foch (Figure 4.1a et Figure 4.2a), les concentrations en particules sont nettement plus élevées durant l'été (juillet, août, septembre) que durant l'hiver (décembre à février). Les niveaux maximaux de PM_{10} passent de 250 µg/m³ en hiver à 450 µg/m³ en été, tandis que les $PM_{2.5}$ varient de 100 µg/m³ en hiver à 150 µg/m³ en été.

Un schéma similaire est observé à Sevran Beaudottes (Figure 4.1b et Figure 4.2b), où les concentrations en PM_{10} augmentent également en été (juillet, août, septembre), atteignant 500 µg/m³ contre 300 µg/m³ en hiver (décembre à février). Les concentrations en $PM_{2.5}$ suivent cette même tendance, passant de 150 µg/m³ en hiver à 200 µg/m³ en été.

En revanche, à Magenta (Figure 4.1c et Figure 4.2c), cette dynamique saisonnière n'est pas observée. Les concentrations maximales sont atteintes en automne (septembre, octobre, novembre), tandis que le reste de l'année, les niveaux restent relativement stables. Contrairement à Avenue Foch et Sevran Beaudottes, la gare de Magenta présente une évolution annuelle différente, marquée par une augmentation progressive des concentrations en PM_{10} et $PM_{2.5}$ tout au long de la période de mesure. Selon AEF (2021c), cette tendance pourrait s'expliquer par le démarrage des travaux EOLE (Est-Ouest Liaison Express) sur la ligne E à partir de février 2018, travaux qui se déroulent chaque été. Ces travaux visent à étendre la ligne vers l'ouest de Paris. En plus de l'impact des travaux, une autre explication possible de cette tendance pourrait être liée à la détérioration de l'efficacité du système de ventilation de la gare. Comme tout système mécanique, les systèmes de ventilation peuvent subir une usure au fil du temps. Sans maintenance adéquate, leur efficacité peut diminuer, ce qui entraînerait une augmentation des concentrations en particules. Toutefois, les informations concernant les périodes de maintenance et les intervalles de fonctionnement de ce système installé à la gare de Magenta ne sont pas disponibles.

Il convient de noter que toutes ces observations sont en opposition avec les tendances rapportées dans la littérature (Martins et al., 2015; Van Ryswyk et al., 2017; Luo et al., 2021a; Gokul et al., 2023). À notre connaissance actuelle, aucune étude n'a décrit les tendances observées dans les EFS parisiennes.

La majorité des travaux existants portent sur les réseaux de métro, et non sur les trains, ce qui pourrait expliquer certaines différences. Cette divergence peut être attribuée à des disparités structurelles majeures entre les gares du réseau de trains et les gares du réseau de métro, notamment la présence ou l'absence de systèmes de climatisation, ainsi que l'utilisation d'un pantographe caténaire pour l'alimentation électrique des trains qui est une source de particules supplémentaire en raison des frottements et des arcs électriques. Par ailleurs, les concentrations élevées observées durant la période estivale pourraient être partiellement expliquées par la forte fréquentation touristique à Paris favorisant la remise en suspension des particules déjà présentes sur les quais. En été, la température sur les quais, similaire à celle de l'extérieur, favorise une usure accrue des freins et des rails, ce qui explique les concentrations plus élevées de particules observées. Ce phénomène est cohérent avec les travaux de Fruhwirt et al. (2024), qui ont montré qu'une augmentation exponentielle des émissions de particules se produit lorsque la température de surface critique des systèmes de freinage est dépassée. En hiver, les températures plus basses limitent cette élévation thermique, réduisant ainsi l'usure des composants et les émissions de particules associées, ce qui correspond à nos observations. Cette tendance est corroborée par Walia et al. (2019), qui ont montré que l'énergie de freinage des blocs en fonte est plus faible en hiver qu'en été, réduisant ainsi la génération de chaleur et l'usure des composants.

Ces tendances saisonnières se reflètent également dans les variations temporelles du DAC (Figure 4.1 et Figure 4.2). Comme prévu, cette métrique représente l'amplitude journalière des concentrations, mettant en évidence les fluctuations journalières caractéristiques de chaque site. A l'échelle mensuelle, la Figure 4.3 illustre l'évolution mensuelle du DAC moyen des PM_{10} (en gris) et des $PM_{2.5}$ (en vert) dans les trois gares étudiées : Avenue Foch (a, b), Sevran Beaudottes (c, d) et Magenta (e, f). Ces évolutions sont également comparées aux DAC moyens des PM_{10} et $PM_{2.5}$ (en violet) dans leurs stations extérieures respectives : Porte d'Auteuil (a, b), Bobigny (c, d) et Paris 18ème (e, f). En plus de se démarquer des observations rapportées pour d'autres EFS dans le monde, nos trois EFS présentent une variabilité saisonnière distincte de celle de l'air extérieur, où les niveaux de particules sont généralement plus élevés en hiver et plus faibles en été (Figure 4.3).

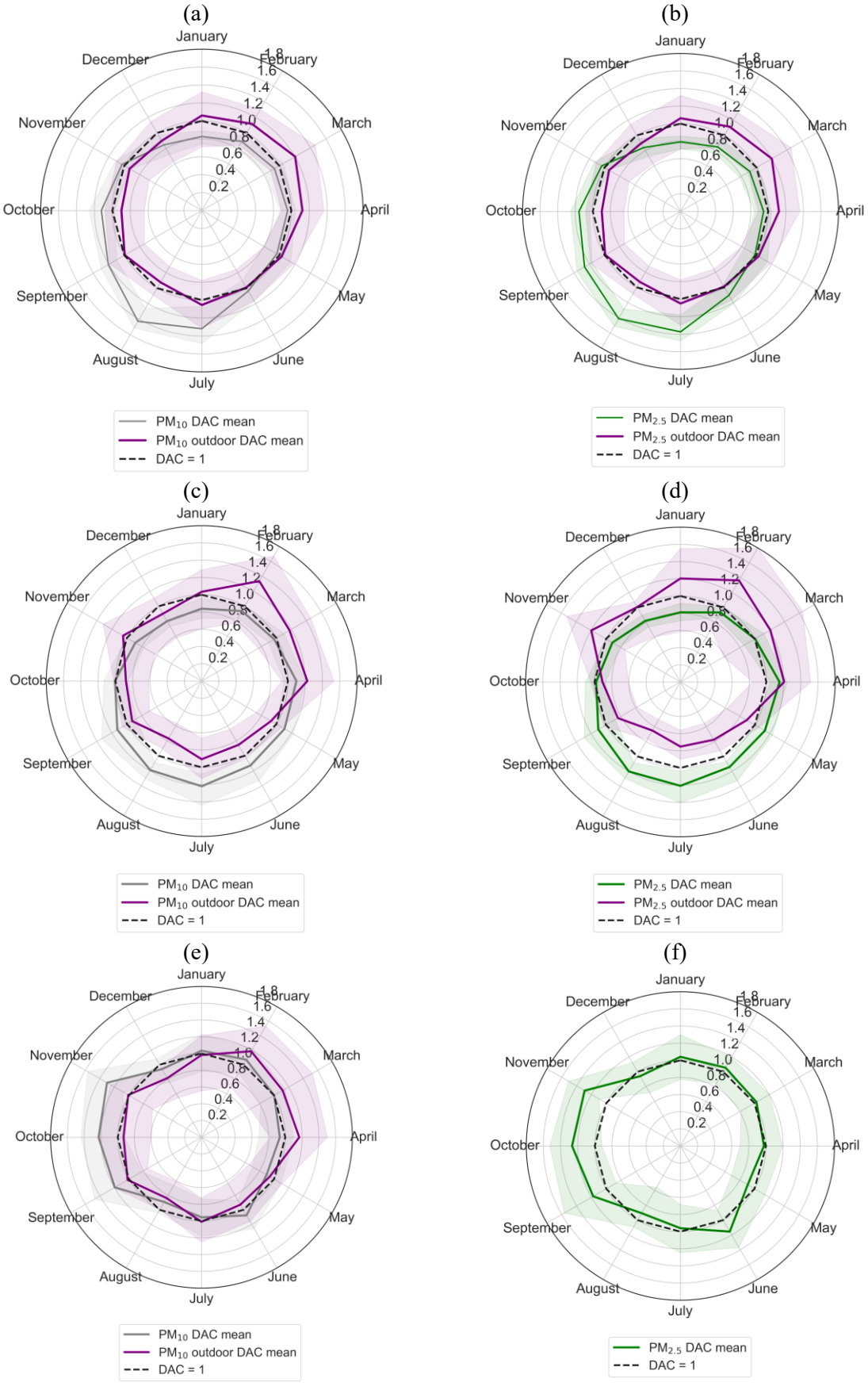


Figure 4.3 : Évolution mensuelle du DAC moyen des PM_{10} (gris) et $PM_{2.5}$ (vert) dans les gares d'Avenue Foch (a, b), Sevran Beaudottes (c, d) et Magenta (e, f), comparée à celle des PM_{10} et $PM_{2.5}$ (violet) de leurs stations extérieures respectives : Porte d'Auteuil (a, b), Bobigny (c, d) et Paris 18ème (e, f)

Cette variabilité saisonnière observée dans l'air extérieur s'explique en grande partie par la dynamique de la couche limite atmosphérique (CLA). En hiver, la CLA est basse et souvent associée à des phénomènes d'inversion thermique, qui piègent les polluants près du sol et limitent leur dispersion. Ce phénomène conduit à des concentrations plus élevées en particules dans l'air extérieur. En été, la CLA devient plus épaisse et plus turbulente, facilitant une meilleure dispersion des polluants.

Par conséquent, les concentrations en particules diminuent de manière significative pendant cette période. À l'intérieur des EFS, l'influence de la dynamique atmosphérique est inexistante, car les niveaux de particules y sont principalement déterminés par des facteurs internes.

Du point de vue opérationnel, à la gare d'Avenue Foch, une stabilité relative autour d'un DAC égal à 1 est observée au cours des mois de printemps (mars, avril, mai), ce qui en fait une période idéale pour mener des campagnes de mesure et établir des profils de référence. Pour Sevran Beaudottes, cette période de stabilité se situe principalement durant les mois de février et mars, tandis que pour Magenta, les mois propices s'étendent de janvier à avril. Cependant, cette variabilité temporelle entre les sites soulève une difficulté opérationnelle majeure puisqu'il n'est pas possible de déterminer à l'avance une période propice pour mener des campagnes de mesure. Cette absence de prévisibilité complique la planification des études et nécessite une approche adaptative, en tenant compte des spécificités de chaque site.

Le rapport $PM_{2.5}/PM_{10}$ fournit des informations complémentaires sur la dynamique et les sources des particules fines dans les EFS. La Figure 4.4 présente l'évolution temporelle de ce ratio à la gare d'Avenue Foch sur la période 2018 – 2022, ainsi que sa distribution selon différentes échelles temporelles : (b) par année, (c) par mois, (d) par jour de la semaine et (e) par période de la journée. Pour les échelles annuelle, mensuelle et journalière, la distribution a été calculée uniquement en journée, c'est-à-dire de 5h à minuit. Des analyses similaires sont illustrées dans la Figure 4.5 pour la gare de Sevran Beaudottes et dans la Figure 4.6 pour celle de Magenta.

En moyenne, ce ratio est mesuré à 0,44 à la gare d'Avenue Foch (Figure 4.4a) et à Sevran Beaudottes (Figure 4.5a), ce qui indique que les particules $PM_{2.5}$ représentent environ 44 % des PM_{10}. À Magenta, il est légèrement inférieur, avec une valeur moyenne de 0,39 (Figure 4.6a). Ces résultats confirment la prédominance des PM_{10} dans les trois gares de référence. La prépondérance des PM_{10} n'est pas surprenante et s'explique par leur forte production liée aux processus d'usure mécanique dans les environnements ferroviaires souterrains (Colombi et al., 2013; Olofsson et al., 2013; Moreno et al., 2014, 2015b; b; Martins et al., 2015; Wang et al., 2016; He et al., 2018; Lee et al., 2018; Durand, 2021) et de remise en suspension via l'effet piston (Moreno et al. 2014, 2015b; Martins et al. 2015; Wang et al. 2016; He et al. 2018). Comparées aux valeurs rapportées dans la littérature pour d'autres métropoles, les ratios $PM_{2.5}/PM_{10}$ observés dans ces gares sont légèrement plus faibles. Par exemple, ils varient généralement entre 0,6 et 0,8 à Beijing, Chine (Wang et al., 2022), autour de 0,5 à Téhéran, Iran (Parchami et al., 2022), entre 0,5 et 0,7 à Séoul, Corée du Sud (Kwon et al., 2015), et atteignent environ 0,7 à Chennai, Inde (Passi et al. 2021b). Les proportions les plus basses dans les EFS parisiennes suggèrent une spécificité propre à chaque environnement et soulignent l'importance d'adapter les stratégies de gestion de la qualité de l'air en fonction des caractéristiques propres à chaque gare souterraine, notamment en optimisant les systèmes de ventilation et en réduisant les émissions à la source.

Toutefois, des fluctuations du ratio sont observées dans les trois gares au cours du temps (Figure 4.4a, Figure 4.5a, Figure 4.6a). Nous allons analyser ces variations à différentes échelles temporelles.

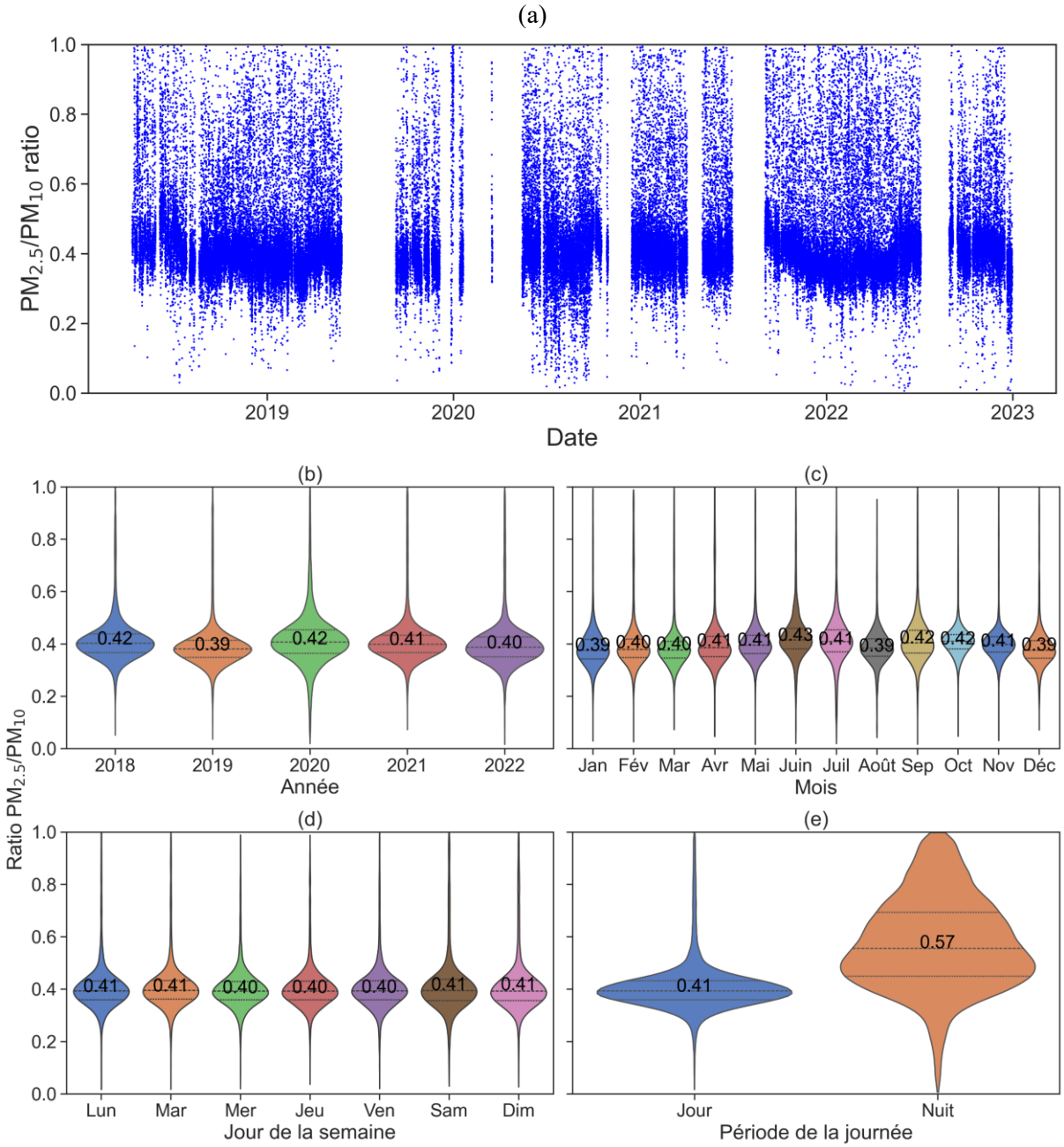


Figure 4.4 : (a) Évolution temporelle du ratio $PM_{2.5}/PM_{10}$ à la gare d'Avenue Foch et sa distribution selon différentes échelles temporelles à Avenue Foch : (b) par année, (c) par mois, (d) par jour de la semaine et (e) par période de la journée

À l'échelle annuelle, les tendances diffèrent selon les gares. À Avenue Foch, le ratio reste globalement stable autour de 0,41 (Figure 4.4b). À Sevran Beaudottes, le ratio reste relativement stable entre 2018 et 2020, avant de connaître une augmentation plus marquée en 2021, atteignant 0,46 (Figure 4.5b). Cette hausse pourrait être directement liée aux travaux de creusement et à la réalisation des paliers souterrains pour la nouvelle ligne 16 du Grand Paris Express, dont les travaux ont commencé début 2021. Le diagramme en violon observée dans les distributions de 2021 s'explique par la dispersion des $PM_{2.5}$. Ces particules, plus légères, restent en suspension dans l'air plus longtemps, augmentant ainsi leur concentration relative par rapport aux PM_{10}. À Magenta, le ratio montre également une augmentation initiale, avant de se stabiliser à partir de 2018 autour de 0,37. Cette tendance à la hausse dans les gares de Sevran Beaudottes et Magenta peut s'expliquer par plusieurs facteurs. Tout d'abord, le vieillissement des infrastructures ferroviaires joue un rôle important. Comme l'ont montré Martins et al. (2016), avec le temps, la surface des rails devient plus rugueuse, ce qui augmente les micro-fractures

et la fragmentation des matériaux, favorisant ainsi la formation de $PM_{2.5}$. De plus, à Magenta, les travaux à proximité liés au projet EOLE (comme suggéré par AEF) pourraient également contribuer à cette augmentation, notamment en amplifiant l'usure des infrastructures et la production de $PM_{2.5}$.

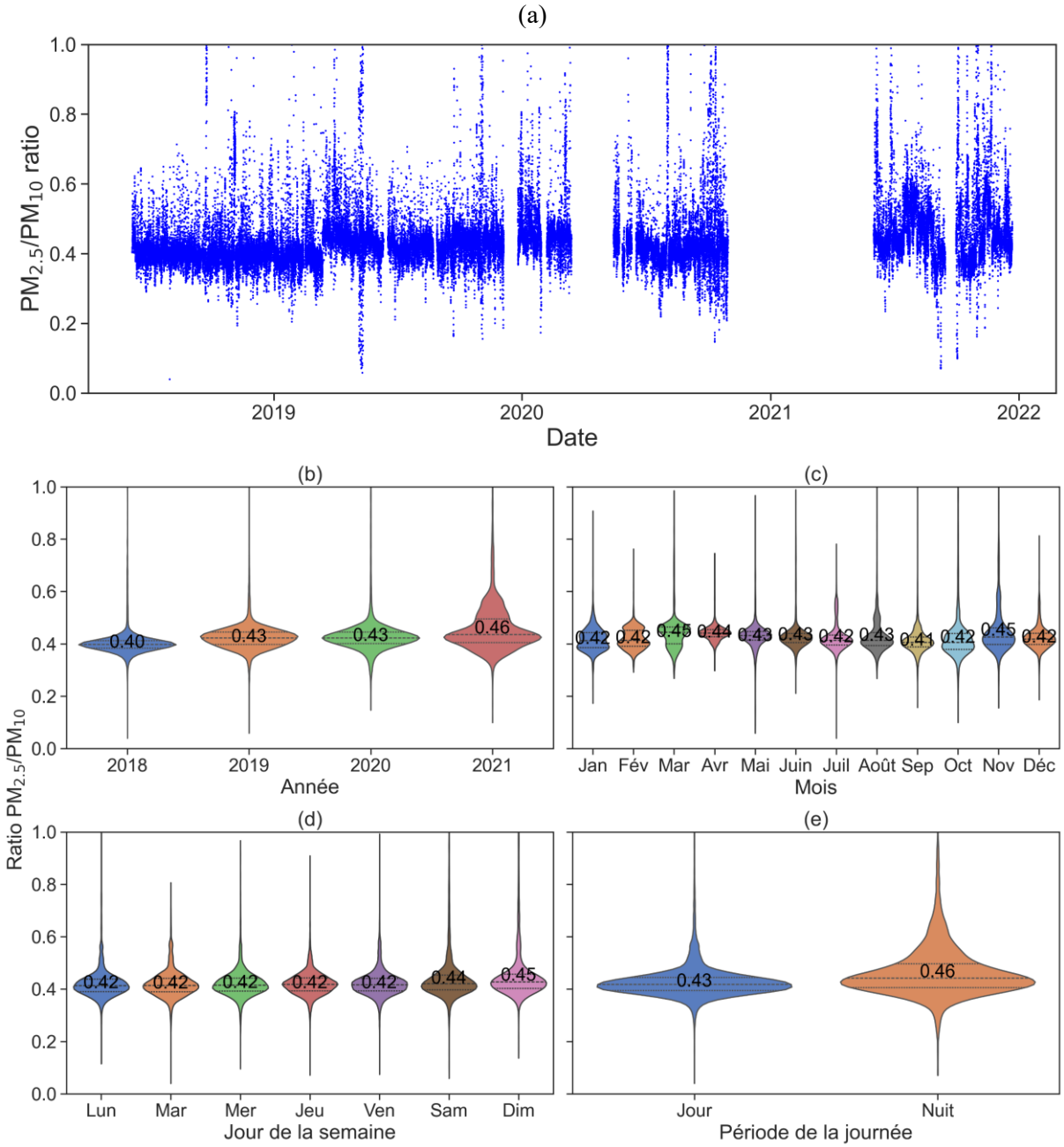


Figure 4.5 : (a) Évolution temporelle du ratio $PM_{2.5}/PM_{10}$ à la gare de Sevran Beaudottes et sa distribution selon différentes échelles temporelles à Sevran Beaudottes : (b) par année, (c) par mois, (d) par jour de la semaine et (e) par période de la journée

À l'échelle mensuelle, le ratio $PM_{2.5}/PM_{10}$ montre une relative stabilité dans les trois gares étudiées, bien que certaines variations saisonnières soient observées. Pour Avenue Foch, le ratio est globalement constant, avec un minimum en août (0,39) et un maximum en juin (0,43). À Sevran Beaudottes, le ratio reste également relativement stable, mais un pic est observé en novembre, où il atteint 0,45, tandis que le minimum est enregistré en septembre (0,41). À Magenta, le ratio montre une plus grande variabilité. Il atteint un minimum en avril (0,34) et un maximum en hiver, durant les mois de décembre, janvier et février, autour de 0,37.

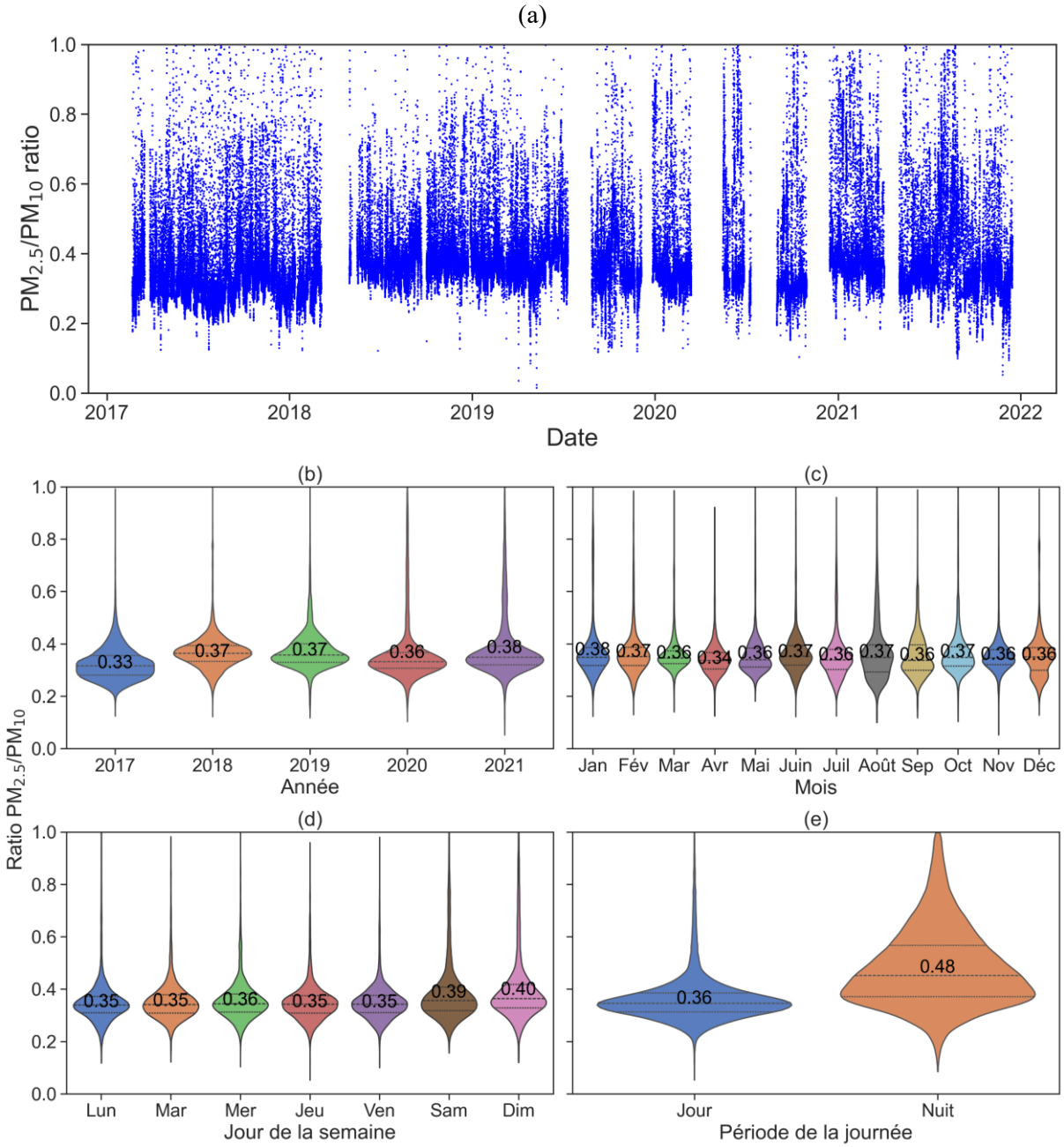


Figure 4.6 : (a) Évolution temporelle du ratio $PM_{2.5}/PM_{10}$ à la gare de Magenta et sa distribution selon différentes échelles temporelles à Magenta : (b) par année, (c) par mois, (d) par jour de la semaine et (e) par période de la journée

Selon **les jours de la semaine**, ce ratio se maintient en moyenne à 0,44 pour la gare d'Avenue Foch, 0,43 pour Sevran Beaudottes et 0,38 pour Magenta les jours ouvrés. Cependant, le weekend, on observe une légère augmentation de ce ratio dans les trois gares. Cette tendance pourrait s'expliquer par la réduction de l'activité ferroviaire durant le weekend. En effet, une fréquence réduite des trains diminue l'usure des freins et des rails, qui sont des sources majeures de PM_{10}. Par conséquent, la proportion relative de $PM_{2.5}$ augmente, ce qui se traduit par une augmentation du ratio $PM_{2.5}/PM_{10}$ (Aarnio et al., 2005; Martins et al., 2016).

Le rapport $PM_{2.5}/PM_{10}$ varie significativement **en fonction de la période de la journée**. Il est généralement plus faible le jour et augmente la nuit (Figure 4.4e, Figure 4.5e, Figure 4.6e). À la gare d'Avenue Foch, ce rapport passe de 0,41 le jour à 0,57 la nuit. À Sevran Beaudottes, l'augmentation est moins marquée, avec un ratio évoluant de 0,43 à 0,46. Enfin, à Magenta, le rapport augmente de 0,36 le jour à 0,48 la nuit. Ces observations s'expliquent principalement par l'arrêt des activités ferroviaires à

minuit, qui réduit la génération de PM_{10}, associé au dépôt plus lent des $PM_{2.5}$ par rapport aux PM_{10} (Lai & Nazaroff, 2000; Zhao & Wu, 2006; Walther et al., 2017).

Pour illustrer ce dernier, nous avons étudié le temps de dépôt des particules dans chaque gare durant les périodes d'inactivité des trains. En ajustant les courbes de concentration des PM_{10} et $PM_{2.5}$ à une décroissance exponentielle, nous avons pu quantifier cette différence de comportement entre les deux classes de particules (Figure 4.7). A des fins de comparaison avec les résultats obtenus par Walther et al. (2017), nous avons également examiné les temps de dépôt relatif à la fraction $PM_{2.5\text{-}10}$, illustré dans la Figure A. 7 (Annexe 4).

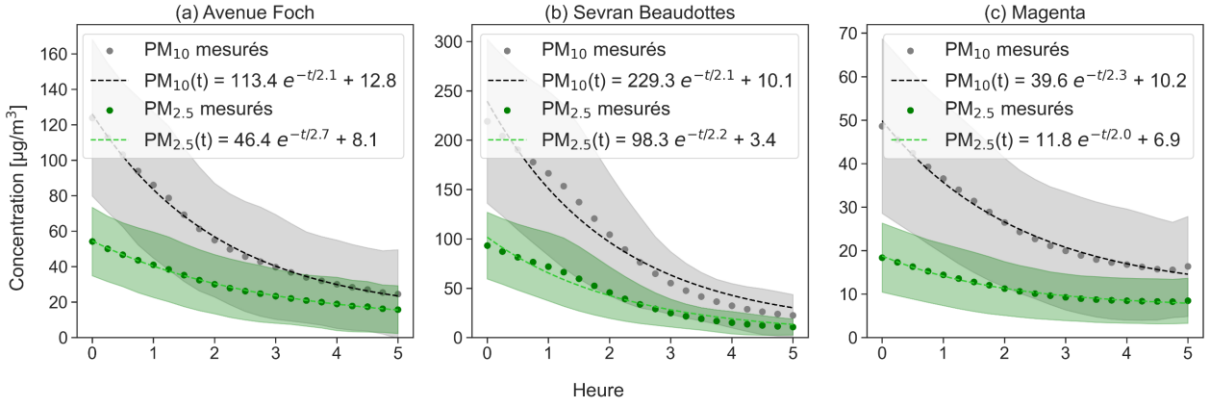


Figure 4.7 : Analyse du temps de dépôt nocturne des concentrations de PM_{10} (en gris) et $PM_{2.5}$ (en vert) dans les gares de (a) Avenue Foch, (b) Sevran Beaudottes et (c) Magenta. Les lignes pointillées représentent les ajustements exponentiels des courbes de concentration.

À Avenue Foch, les temps de dépôt calculés sont de 2,1 h pour les PM_{10}, de 1,9 h pour les $PM_{2.5\text{ - }10}$ et de 2,7 h pour les $PM_{2.5}$. À Sevran Beaudottes, ces durées sont respectivement de 2,1 h, 1,9 h et 2,2 h. En revanche, à Magenta, les temps de dépôt sont de 2,3 h pour les PM_{10}, de 2,4 h pour les $PM_{2.5\text{-}10}$ et de 2 h pour les $PM_{2.5}$. À l'exception de Magenta, nos résultats sont cohérents avec ceux trouvés par Walther et al. (2017), où le temps de dépôt pour les $PM_{2.5}$ varie entre [2,2 h ; 500 h] et pour les $PM_{2.5\text{-}10}$ entre [0,1 h ; 2,2 h]. Selon les études de Lai and Nazaroff (2000) ainsi que de Zhao and Wu (2006), la vitesse de dépôt est plus élevée pour les particules de plus grand diamètre, ce qui confirme la tendance observée dans nos analyses. Ces résultats sont valables dans le domaine des particules comprises entre 0,1 µm et 10 µm (Figure 1.5). Les observations à Magenta semblent indiquer un dépôt plus rapide des $PM_{2.5}$ par rapport au PM_{10}.

Une des différences majeures de ces trois gares est la présence d'un système de ventilation mécanique à Magenta qui pourrait entrainer un dépôt ou une extraction plus efficace des $PM_{2.5}$. Walther et al. (2017) soulignent que la vitesse du vent au niveau du quai a un impact significatif sur le taux de dépôt. Cependant, nous ne disposons pas d'informations détaillées sur les horaires de fonctionnement du système de ventilation installé à Magenta ne nous permettant donc pas de conclure sur ce point.

4.2 Profils journaliers dans les 3 EFS avec des mesures de longue durée

En plus des variations à long terme des concentrations de particules, des variations journalières significatives sont observées pour les trois gares de référence (Figure 4.1). Dans la littérature (Birenzvige et al., 2003; Raut et al., 2009; Walther & Bogdan, 2017), un cycle diurne différent se dessine entre les jours ouvrés et les weekends, directement lié la fréquence des trains. En examinant de plus près une semaine de données à la gare d'Avenue Foch (Figure 4.8), on constate que du lundi au vendredi (zone grise), la concentration quotidienne de PM_{10} pendant la journée suit une tendance similaire, puis diminue durant le weekend (zone verte).

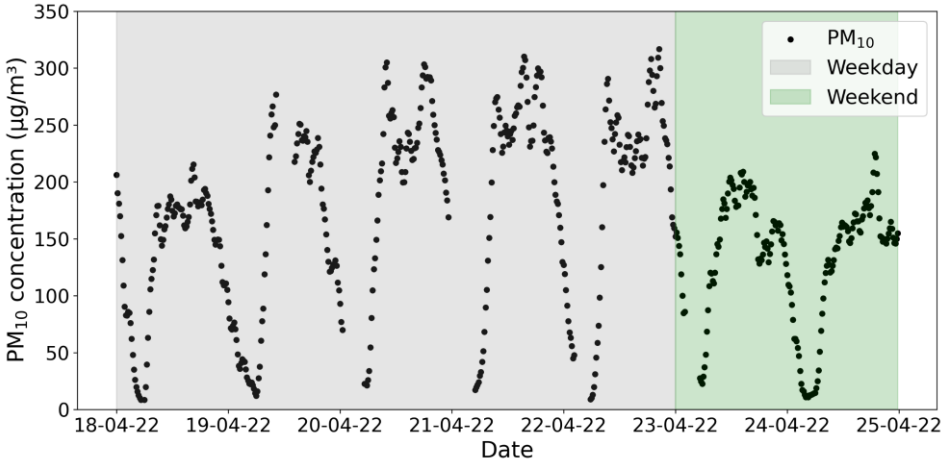


Figure 4.8 : Profil temporel des concentrations en PM_{10} sur une semaine (18/04/2022 – 25/04/2022) à la gare d'Avenue Foch

Dans cette section, nous analysons les profils journaliers en utilisant la méthodologie développée sur Avenue Foch, Sevran Beaudottes et Magenta. Rappelons que cette méthode génère deux profils journaliers type (jours ouvrés et weekends), ainsi qu'un DAC associé à chaque profil journalier individuel. L'objectif est de se concentrer sur l'analyse des profils journaliers types des trois gares avec des mesures longue durée, afin d'étudier les dynamiques des particules à l'échelle de la journée, tout en identifiant les similitudes et les spécificités propres à chaque gare.

Pour commencer, nous vérifions que les profils journaliers des jours ouvrés (lundi à vendredi) sont similaires entre eux, de même que ceux des weekends (samedi et dimanche). Cette analyse est menée pour chacune des trois gares. Les résultats confirment que les profils des jours ouvrés sont comparables, tout comme ceux des weekends, comme le montrent la Figure A. 4 et la Figure A. 5 en Annexe 4. Cette homogénéité nous permet de consolider les données en un profil journalier type pour chaque catégorie (jours ouvrés et weekends).

La Figure 4.9 présente les profils journaliers types des concentrations de PM_{10} mesurées sur les quais des gares de (a) Avenue Foch, (b) Sevran Beaudottes et (c) Magenta, en distinguant les jours ouvrés (à gauche) des weekends (à droite). De même, la Figure 4.10 illustre les profils correspondants pour les $PM_{2.5}$ dans ces trois gares. Ces profils types sont obtenus en calculant la moyenne des profils journaliers pour chaque catégorie : du lundi au vendredi pour les jours ouvrés, et du samedi au dimanche pour les weekends. Les écarts-types associés, représentés sur les figures, sont calculés en appliquant les lois de propagation des incertitudes aux écarts-types individuels des jours ouvrés et des weekends. Pour un ensemble de n jours, l'écart-type combiné $\sigma_{combiné}$ est donné par :

$$\sigma_{combiné} = \frac{\sqrt{\sum_{i=1}^{n} \sigma_i^2}}{n} \tag{4.19}$$

où : σ_i est l'écart-type du jour i. n est le nombre de jours dans la catégorie (5 pour les jours ouvrés, 2 pour les weekends).

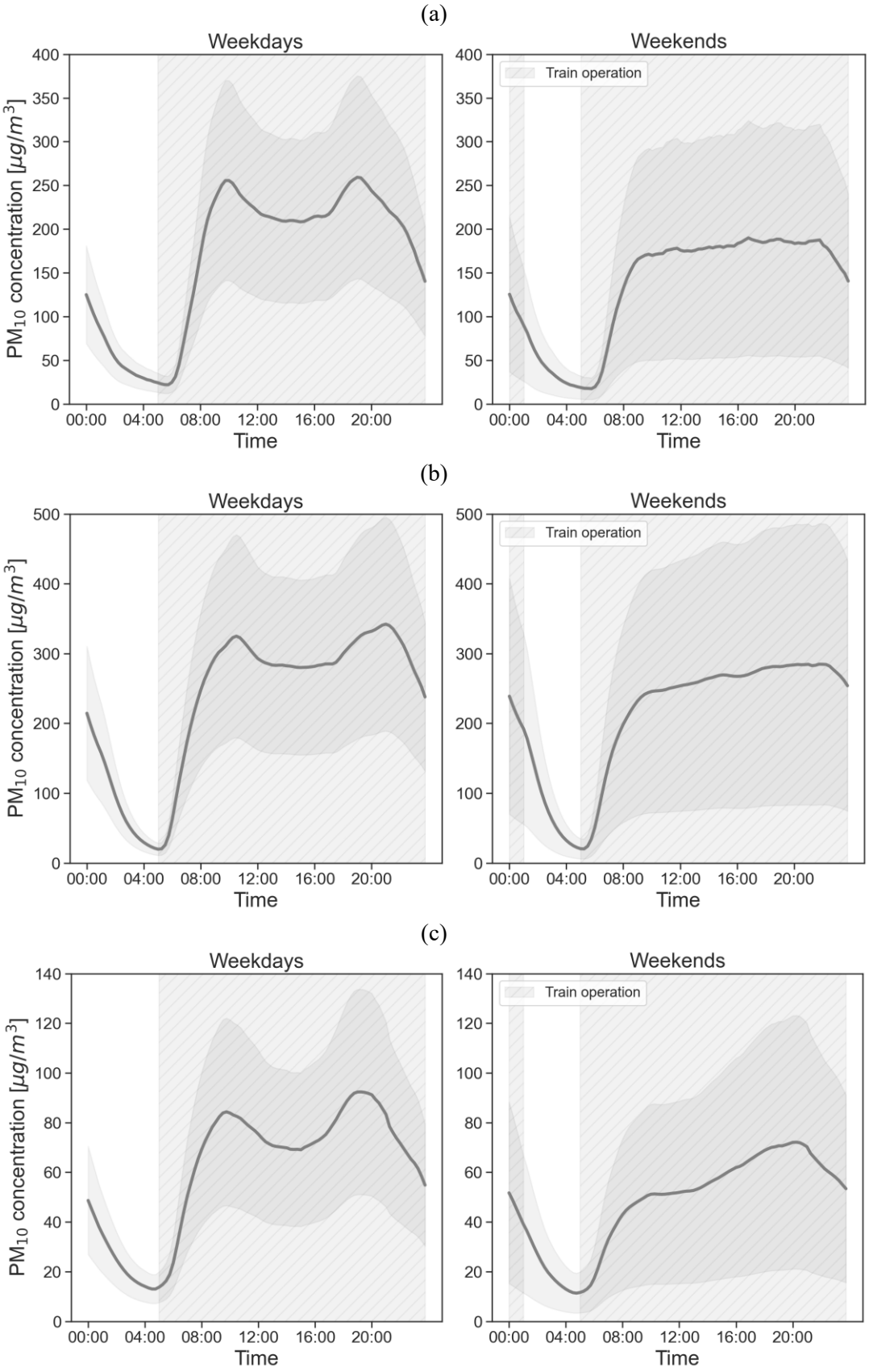


Figure 4.9 : Profils types des PM_{10} pour les jours de semaine (à gauche) et les weekends (à droite) dans les gares de référence : (a) Avenue Foch, (b) Sevran Beaudottes et (c) Magenta. La zone hachurée indique la période d'exploitation des trains. Les zones ombrées représentent un écart type (σ=1)

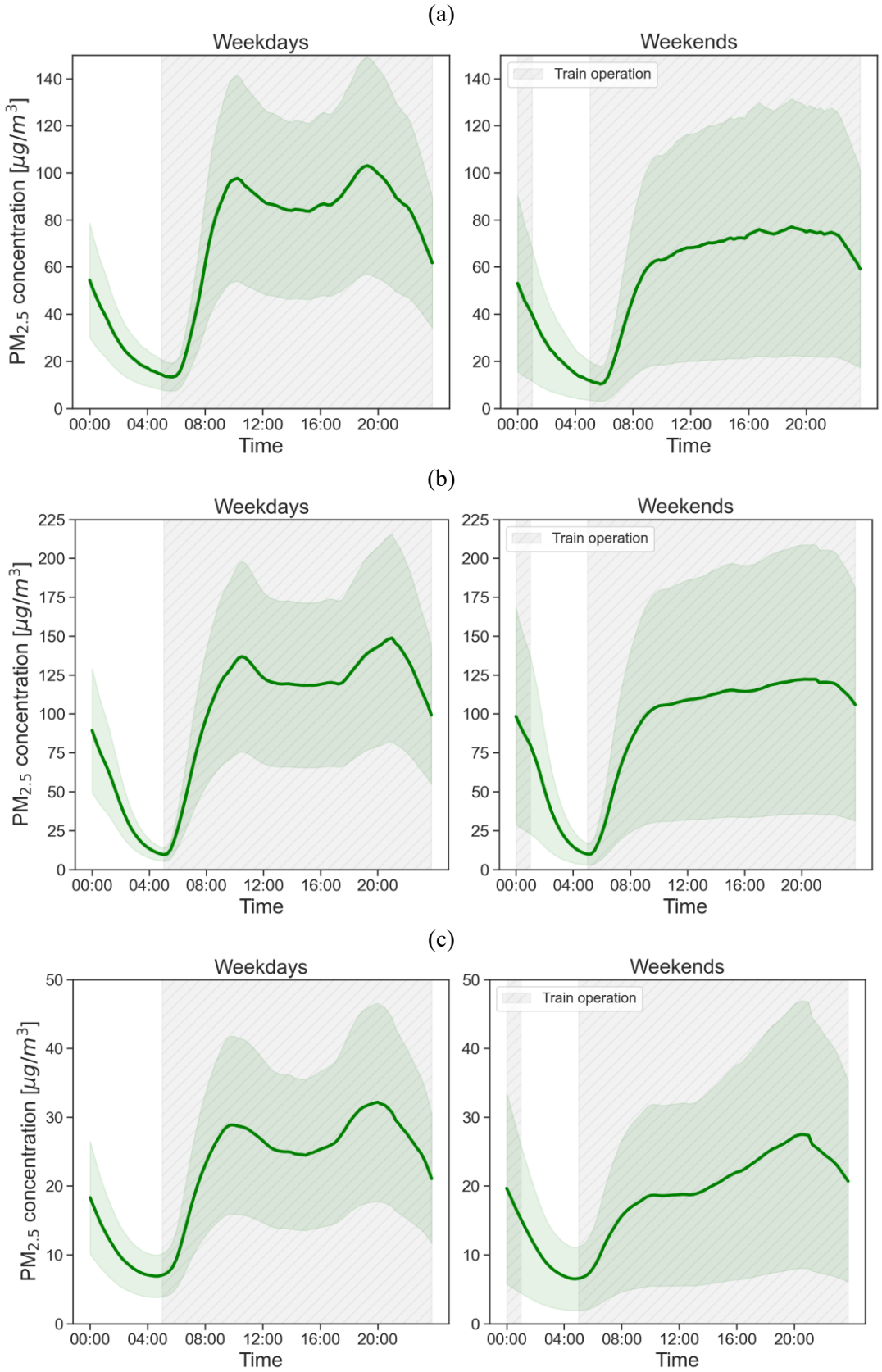


Figure 4.10 : Profils types des $PM_{2.5}$ pour les jours de semaine (à gauche) et les weekends (à droite) dans les gares de référence : (a) Avenue Foch, (b) Sevran Beaudottes et (c) Magenta. La zone hachurée indique la période d'exploitation des trains. Les zones ombrées représentent un écart type (σ=1)

Les trois gares présentent des profils journaliers similaires, caractérisés par un cycle différent entre les jours ouvrés et les weekends (Birenzvige et al., 2003; Raut et al., 2009; Walther & Bogdan,

2017). Les jours ouvrés se distinguent par deux pics de concentration, correspondant aux heures de pointe du matin et du soir, tandis que les weekends montrent des tendances plus stables avec des concentrations globalement plus faibles. Ces observations sont valables pour les deux fractions de particules (PM_{10} et $PM_{2.5}$).

Les différences entre les gares se manifestent principalement par des amplitudes de concentrations variables et des pics survenant à des moments distincts, reflétant les particularités opérationnelles et contextuelles de chaque site.

À la gare d'**Avenue Foch** (Figure 4.9a et Figure 4.10a), les heures de pointe du matin et du soir sont observées vers 10 h et 19h. Les niveaux maximaux atteignent 250 µg/m³ pour les PM_{10} et 100 µg/m³ pour les $PM_{2.5}$. La concentration moyenne journalière, calculée pendant les heures de circulation des trains, s'élève à 188 µg/m³ pour les PM_{10} et 75 µg/m³ pour les $PM_{2.5}$. En revanche, les weekends présentent des concentrations plus faibles, de l'ordre de 157 µg/m³ pour les PM_{10} et 60 µg/m³ pour les $PM_{2.5}$, sans pics marqués. Les niveaux restent relativement constants après 8h (environ 170 µg/m³ pour les PM_{10} et 70 µg/m³ pour les $PM_{2.5}$), avec une légère augmentation. Ces variations s'expliquent par les différences d'exploitation ferroviaire entre les jours ouvrés et les weekends. En semaine, les trains circulent de 05h00 à 23h30, avec une fréquence élevée (environ toutes les 8 minutes) pendant les heures de pointe et toutes les 15 minutes en dehors de ces périodes, ce qui entraîne les deux pics observés. Le weekend, les trains fonctionnent de 05h00 à 01h00 le lendemain, avec une fréquence constante d'environ un train toutes les 15 minutes. On observe également un délai d'une heure entre le début des opérations ferroviaires (05h00) et l'augmentation des concentrations de PM_{10} (à partir de 06h00), aussi bien en semaine que le weekend. Cette observation est en accord avec les travaux de Birenzvige et al. (2003) et Raut et al. (2009), qui ont montré que les concentrations de particules suivent étroitement l'intensité du trafic, mais avec un décalage temporel d'environ une heure. Ce délai s'explique principalement par la dispersion des particules due à la resuspension causée par l'intensité du trafic et l'affluence des passagers.

Les profils journaliers de la gare de **Sevran Beaudottes** (Figure 4.9b et Figure 4.10b) présentent des similitudes avec ceux d'Avenue Foch, mais avec des concentrations moyennes journalières plus élevées durant les jours ouvrés (270 µg/m³ pour les PM_{10} et 115 µg/m³ pour les $PM_{2.5}$) et durant le weekend (240 µg/m³ pour les PM_{10} et 100 µg/m³ pour les $PM_{2.5}$). Les jours ouvrés, deux pics sont observés : un premier vers 10h00 (325 µg/m³ pour les PM_{10} et 140 µg/m³ pour les $PM_{2.5}$) et un second, plus prononcé, vers 21h00 (350 µg/m³ pour les PM_{10} et 150 µg/m³ pour les $PM_{2.5}$). Ce deuxième pic, plus tardif qu'à Avenue Foch, peut s'expliquer par des horaires d'exploitation prolongés en soirée, avec une fréquence horaire atteignant jusqu'à 20 trains par heure (1 train toutes les 3 minutes). Pour les weekends, à l'instar d'Avenue Foch, les concentrations restent relativement constantes (250 µg/m³ pour les PM_{10} et 120 µg/m³ pour les $PM_{2.5}$), mais avec une augmentation plus marquée en journée.

Enfin, **Magenta** (Figure 4.9c et Figure 4.10c) suit également la variation diurne entre jours ouvrés et weekends, avec des concentrations plus faibles que dans les deux autres sites. Les concentrations moyennes journalières s'élèvent à 70 µg/m³ pour les PM_{10} et 25 µg/m³ pour les $PM_{2.5}$ en semaine, et à 50 µg/m³ pour les PM_{10} et 20 µg/m³ pour les $PM_{2.5}$ le weekend. Les jours ouvrés, les profils présentent deux pics : un premier vers 10h (80 µg/m³ pour les PM_{10} et 30 µg/m³ pour les $PM_{2.5}$) et un second, plus prononcé, en soirée vers 19h (90 µg/m³ pour les PM_{10} et 33 µg/m³ pour les $PM_{2.5}$). En weekend, les concentrations augmentent progressivement pour atteindre un maximum vers 20h (60 µg/m³ pour les PM_{10} et 28 µg/m³ pour les $PM_{2.5}$). Une particularité de cette gare est la différence de concentrations entre le samedi et le dimanche (Figure A. 4c et Figure A. 5c, Annexe 4), malgré une fréquence de trains identique, entre autres due à une fréquentation différente ces deux jours. Contrairement aux deux autres gares étudiées, la gare de Magenta se distingue par son rôle de gare de connexion, notamment grâce à sa liaison directe avec la gare du Nord. Cette particularité influence probablement les dynamiques de fréquentation et, par conséquent, les profils de concentration des particules observés.

- **Comparaison avec les stations de mesure de l'air extérieur**

Pour enrichir notre analyse, il est pertinent de comparer ces profils types de concentrations en particules avec ceux de l'air extérieur, en utilisant la même méthode que celle développée pour les EFS. Cette comparaison est également l'occasion de démontrer l'applicabilité de notre méthode aux mesures extérieures. Nous avons vérifié que, tout comme avec les EFS, les profils du lundi au vendredi et ceux du samedi au dimanche sont équivalents. Ces résultats sont comparables et illustrés dans la Figure A. 8 et la Figure A. 9 en Annexe 5.

La Figure 4.11 présente les profils journaliers types des concentrations en PM_{10} relevées dans les stations de mesure extérieures associées : (a) Porte d'Auteuil (pour Avenue Foch), (b) Bobigny (pour Sevran Beaudottes), et (c) Paris 18ème (pour Magenta). Ces profils distinguent les jours ouvrés (à gauche) des weekends (à droite), permettant ainsi de mettre en évidence les dynamiques spécifiques de chaque site. Les profils journaliers des concentrations en $PM_{2.5}$, qui complètent cette analyse, sont illustrés dans la Figure A. 10 (Annexe 5) à l'exception de la station de Paris 18ème qui n'a pas de $PM_{2.5}$. Les écarts types pour ces profils sont également calculés sur la base de l'équation 4.19.

La station de Porte d'Auteuil enregistre les concentrations les plus élevées, avec une moyenne de 37 µg/m³ les jours ouvrés et de 29 µg/m³ les weekends. Cette situation s'explique par sa proximité immédiate avec des axes routiers majeurs.

Pendant les jours ouvrés, les profils journaliers de PM_{10} montrent une augmentation marquée le matin, entre 6h00 et 10h00, atteignant un pic vers midi. Les valeurs maximales varient selon les stations : entre 35 et 40 µg/m³ pour Porte d'Auteuil, entre 20 et 25 µg/m³ pour Bobigny, et entre 24 et 28 µg/m³ pour Paris 18ème. Cette hausse matinale est principalement attribuable à l'intensification des activités humaines et du trafic routier en début de journée, tandis que le pic de midi est lié aux heures de pointe associées aux déplacements professionnels et scolaires (Onat & Stakeeva, 2013; Krall et al., 2021). Les concentrations diminuent progressivement au cours de l'après-midi, en raison de l'évolution de la couche limite atmosphérique (CLA). La CLA, plus haute en fin de matinée et en début d'après-midi, favorise la dilution des particules, ce qui explique cette baisse (Yuval et al., 2020). Une légère remontée des concentrations est toutefois observée en fin d'après-midi, vers 16h00, coïncidant avec le trafic de retour. Cette tendance globale sur l'ensemble de la journée a été récemment corroborée lors des Jeux Olympiques de Paris en 2024, où des variations similaires ont été observées (Bougault et al., 2024).

Le weekend, les profils montrent des concentrations globalement plus faibles et plus stables, oscillant autour de 30 µg/m³ pour Porte d'Auteuil et 20 µg/m³ pour Bobigny et Paris 18ème tout au long de la journée. Cette réduction reflète une diminution significative du trafic routier et des activités anthropiques, caractéristique des jours non ouvrés (Motallebi et al., 2003).

Les profils de PM extérieurs diffèrent donc fortement en termes d'allure et d'amplitude par rapport aux profils mesurés en EFS. La Figure 4.12 illustre la comparaison des concentrations moyennes de PM_{10} et $PM_{2.5}$ mesurées sur les quais des trois gares de référence et leurs stations extérieures associées : (a) Avenue Foch – Porte d'Auteuil, (b) Sevran Beaudottes – Bobigny, et (c) Magenta – Paris 18ème. Les résultats sont présentés pour les jours ouvrés (à gauche) et les weekends (à droite), couvrant la période d'exploitation des trains (de 05h00 à 00h00 en semaine et de 05h00 à 01h00 le weekend).

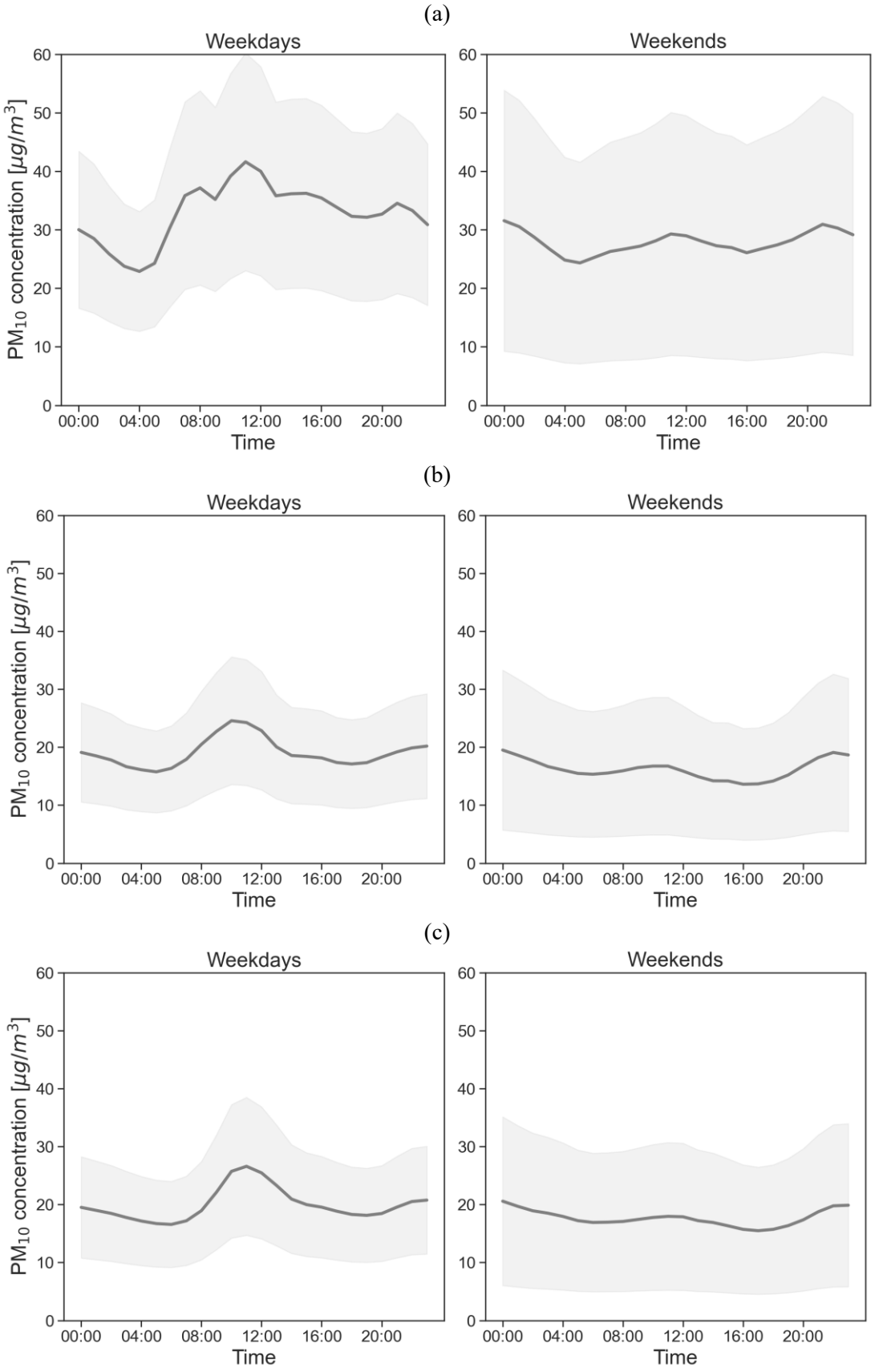


Figure 4.11 : Profils types des PM_{10} pour les jours de semaine (à gauche) et les weekends (à droite) dans les stations de mesure extérieur : (a) Porte d'Auteuil, (b) Bobigny et (c) Paris 18ème. Les zones ombrées représentent un écart type (σ=1)

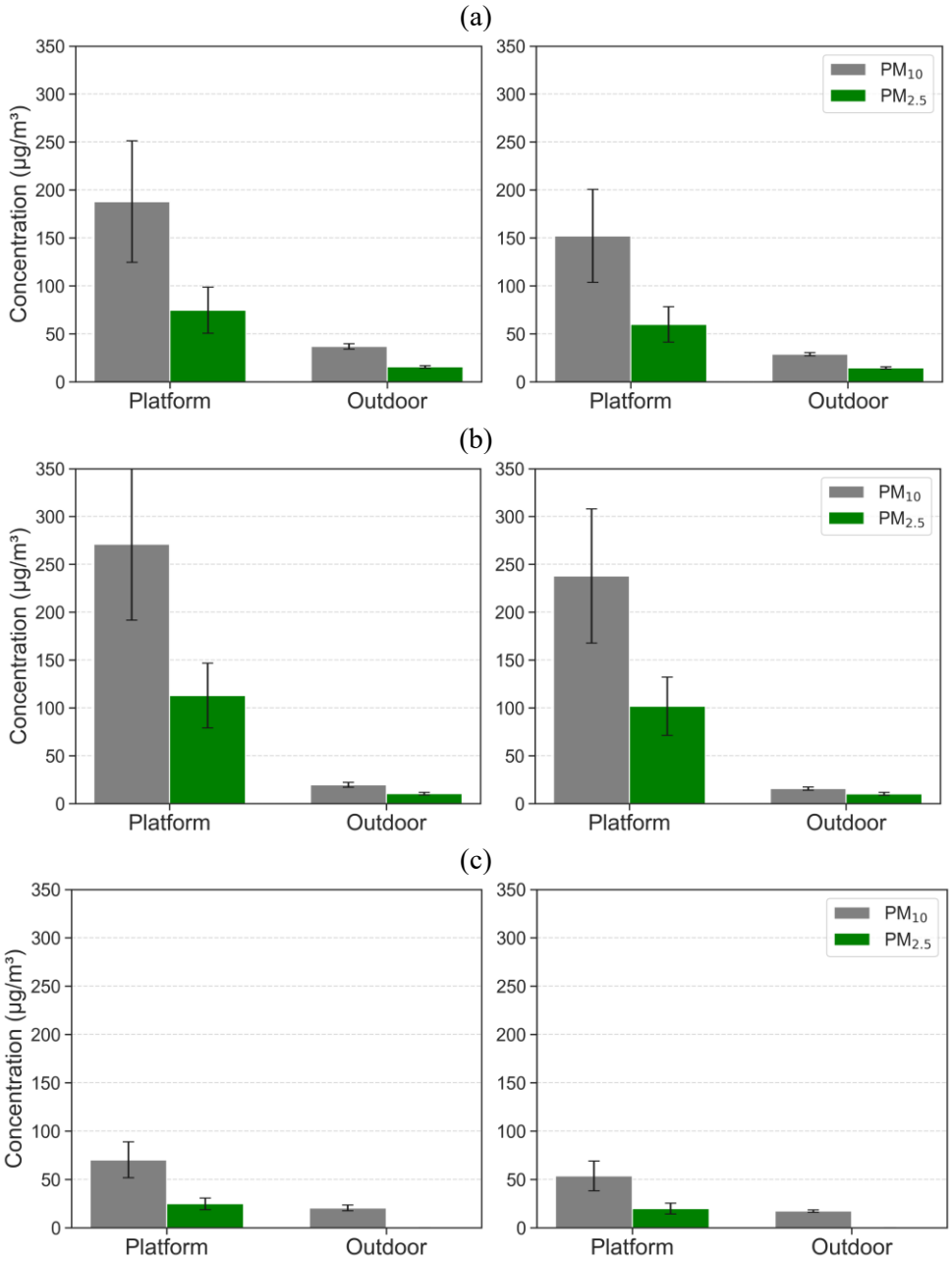


Figure 4.12 : Comparaison des concentrations moyennes de PM_{10} et $PM_{2.5}$ entre les gares de référence et leurs stations extérieures associées : (a) Avenue Foch et Porte d'Auteuil, (b) Sevran Beaudottes et Bobigny, et (c) Magenta et Paris 18ème, pour les jours ouvrés (à gauche) et les weekends (à droite).

Pour les jours ouvrés, les concentrations mesurées sur les quais sont significativement plus élevées que celles enregistrées en extérieur. Cette tendance se maintient également pendant les weekends, bien qu'une diminution générale des concentrations soit observée, tant sur les quais qu'en extérieur. En termes d'ordre de grandeur, les concentrations sur les quais sont environ 5 fois plus élevées pour les PM_{10} et 4 fois plus élevées pour les $PM_{2.5}$ à la gare d'Avenue Foch, aussi bien les jours ouvrés que les weekends. À Sevran Beaudottes, les concentrations sont environ 15 fois plus élevées pour les PM_{10} et 10 fois plus élevées pour les $PM_{2.5}$, tant en semaine que le weekend. Enfin, à Magenta, les concentrations sont environ 3 fois plus élevées pour les PM_{10}, aussi bien les jours ouvrés que les weekends. Compte tenu des résultats rapportés dans la littérature, les trois gares de référence étudiées présentent des ratios comparables aux valeurs typiques observées dans les environnements ferroviaires urbains. Par exemple, une étude réalisée à Sydney a mis en évidence un facteur 3 entre les concentrations de PM_{10} mesurées dans une gare souterraine et celles de l'air extérieur (Mohsen et al., 2018). De même,

à Naples (Cartenì et al., 2015), un ratio de 5 a été observé, soulignant une accumulation importante des particules en milieu souterrain. À Philadelphie, en Pennsylvanie, Karim et al. (2024) ont quantifié des ratios de 5,7 pour les $PM_{2.5}$ et de 7,6 pour les PM_{10}, montrant une différence encore plus marquée entre l'environnement ferroviaire et l'air extérieur. À Paris, les travaux de Raut et al. (2009) ont démontré que durant la journée, les concentrations de PM_{10} et $PM_{2.5}$ dans les gares souterraines sont environ 5 à 30 fois supérieures à celles mesurées à l'extérieur soulignant des concentrations exagérément hautes dans les EFS parisiennes.

4.3 Profils journaliers dans les gares avec des mesures de courte durée

La Figure 4.13 présente les profils journaliers types de PM_{10} ainsi que leurs écarts-types pour toutes les stations dotées de mesures de courte durée, en distinguant les jours ouvrés et les weekends. Les profils journaliers types de $PM_{2.5}$ sont quant à eux illustrés dans la Figure A. 6 (Annexe 4). Les profils montrent une variabilité notable ainsi que des divergences significatives dans les niveaux de concentration entre les stations, illustrant les disparités structurelles, opérationnelles et les particularités des environnements locaux des EFS. Notons que ces campagnes de mesure ont été réalisées à différents moments de l'année, intégrant ainsi un effet saisonnier qui peut influencer les différences de concentrations observées.

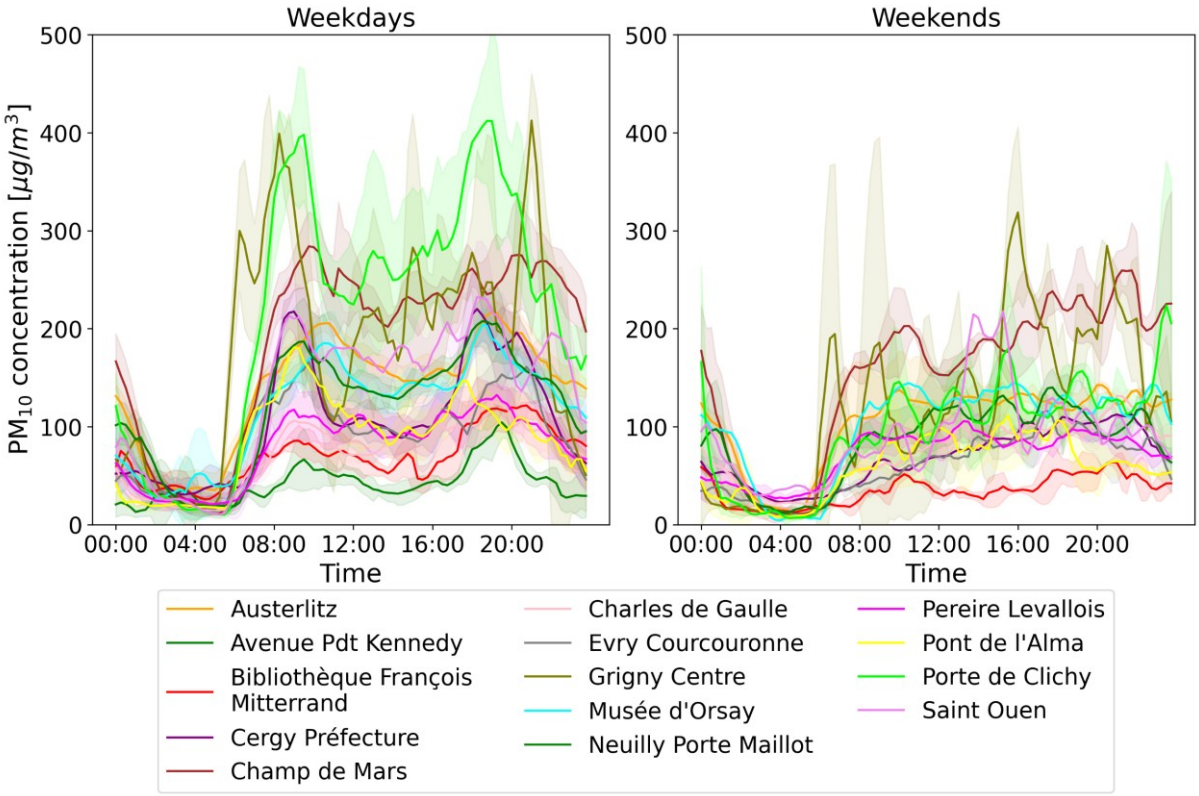


Figure 4.13 : Profils journaliers et écarts-types des concentrations en PM_{10} pour les jours ouvrés et les weekends dans les stations ayant fait l'objet de mesures de courte durée (campagnes de 2 à 3 semaines)

Les tendances journalières observées confirment les profils caractéristiques des trois gares de référence, aussi bien pour les jours ouvrés que pour les weekends. Malgré un certain niveau de bruit dans ces profils, ils demeurent informatifs et cohérents avec les observations antérieures des profils journaliers. Ce bruit peut s'expliquer par le fait que les mesures ont été réalisées sur de courtes périodes et également par le nombre de profils retenus selon notre méthodologie, décrite dans le Tableau 3.2 du Chapitre 3. Parmi les gares étudiées, Austerlitz, Charles de Gaulle 1 et Neuilly Porte Maillot se distinguent par un nombre élevé de profils journaliers retenus : 15/15, 12/14 et 15/15 pour les jours

ouvrés, et 5/6, 5/6 et 4/6 pour les weekends, respectivement. Ces résultats témoignent d'une grande uniformité et de profils relativement lisses. En revanche, des stations comme Avenue Président Kennedy ou Bibliothèque François Mitterrand montrent un nombre nettement inférieur de profils retenus, révélant ainsi une variabilité significative des profils journaliers d'un jour à l'autre (Tableau 3.2). Cette variabilité marque une limitation de notre méthode, étant donné que des séries temporelles courtes compliquent la détermination d'un profil journalier distinctif.

L'examen des concentrations moyennes pour chaque gare constitue un indicateur clé permettant d'identifier les facteurs influençant les niveaux de particules. Toutefois, les campagnes de mesure dans les gares avec des mesures de courte durée ayant été réalisées à différentes périodes de l'année (Tableau 3.2), une normalisation des concentrations a été nécessaire pour établir des références fiables pour ces gares à mesures limitées. Sur la base des comportements saisonniers similaires observés à Avenue Foch et Sevran-Beaudottes (concentrations élevées en été et faibles en hiver, à l'exception de Magenta), nous avons émis l'hypothèse que toutes les gares étudiées suivaient une dynamique comparable. La gare de Magenta a été exclue de cette analyse, car elle est la seule des trois, équipée de ventilation mécanique. Ainsi, la normalisation des profils de concentration pour chaque gare avec des mesures de courte durée a été réalisée en se référant aux valeurs du DAC mensuel calculé pour Avenue Foch sur la période 2018-2022 (Tableau 4.1). L'objectif de cette normalisation est de corriger les concentrations mesurées en fonction du mois de la campagne, afin de garantir une meilleure comparabilité entre les différentes périodes de mesure. Cette approche se traduit par la formule suivante :

$$PM_{normalisé} = \frac{PM_{mesuré}}{DAC} \tag{4.20}$$

Par exemple, dans le cas d'Evry Courcouronnes, où les campagnes de mesure ont été réalisées en décembre, la valeur médiane du DAC restitué dans la gare d'Avenue Foch pour ce mois est de 0,87. Pour les campagnes s'étendant sur deux mois, comme à Neuilly Porte Maillot (mesures effectuées entre avril et mai), nous avons utilisé la moyenne des DAC estimés dans la gare d'Avenue Foch correspondant à ces deux mois, soit 0,99 (Tableau 4.1).

La Figure 4.14 présente une comparaison directe des concentrations normalisées de PM_{10} pour chaque gare, en distinguant les jours ouvrés (barres bleues) des weekends (barres vertes). Les symboles en forme de vague au-dessus de certaines barres indiquent la proximité de ces stations à la Seine, tandis que les icônes de ventilateurs montrent la présence de systèmes de ventilation mécanique dans les gares. Les symboles représentant des flèches opposées désignent les gares mixtes, illustrant un contact avec l'air extérieur. Ces concentrations sont obtenues en calculant la moyenne des concentrations sur la période d'exploitation du train. Les gares sont classées par ordre décroissant des concentrations moyennes observées les jours ouvrés. Le faible nombre de profils retenus pour les weekends ne permet pas d'établir un classement fiable des concentrations (Tableau 3.2).

Les concentrations de PM_{10} les plus élevées sont observées à la gare de Sevran-Beaudottes, où les niveaux moyens atteignent environ 271 µg/m³ les jours ouvrés et 238 µg/m³ les weekends. À l'opposé, les concentrations les plus faibles sont enregistrées à Avenue Président Kennedy, avec une moyenne d'environ 40 µg/m³ les jours ouvrés. On peut noter que Magenta, malgré sa profondeur, fait partie des gares les moins polluées en particules du réseau. Il convient de souligner que les gares étudiées dans ce travail de thèse présentent des niveaux moyens de pollution aux particules dont la moitié affiche des niveaux faibles, comme en témoigne la cartographie récemment établie par Airparif (https://www.airparif.fr/stations-de-metro-et-de-rer).

Tableau 4.1 : DAC mensuel à la gare d'Avenue Foch (Rakotonirinjanahary et al., 2024)

DAC Médiane (25th -75th percentile)	2018	2019	2020	2021	2022	2018-2022
Janvier	-	0,88 (0,84-0,99)	-	0,72 (0,69-0,79)	0,84 (0,80-0,90)	0,83 (0,74-0,89)
Février	-	0,96 (0,87-1,08)	-	0,81 (0,72-0,89)	0,94 (0,87-0,98)	0,89 (0,80-0,96)
Mars	-	0,95 (0,87-1,06)	-	0,82 (0,74-0,93)	1,02 (0,94-1,15)	0,95 (0,82-1,06)
Avril	-	1,00 (0,94-1,09)	-	-	0,98 (0,90-1,05)	0,99 (0,93-1,08)
Mai	-	1,13 (0,90-1,18)	0,69 (0,62-0,81)	0,87 (0,76-0,91)	1,14 (1,07-1,26)	0,99 (0,81-1,16)
Juin	0,95 (0,86-0,99)	-	0,96 (0,89-1,13)	0,98 (0,84-1,04)	1,17 (1,09-1,30)	1,01 (0,90-1,17)
Juillet	1,31 (1,14-1,61)	-	1,23 (1,00-1,48)	-	-	1,27 (1,07-1,51)
Août	1,43 (1,32-1,55)	-	1,35 (1,26-1,50)	-	-	1,38 (1,31-1,54)
Septembre	1,29 (1,18-1,40)	-	1,16 (1,06-1,26)	1,10 (1,05-1,15)	1,15 (1,09-1,21)	1,16 (1,09-1,26)
Octobre	1,20 (1,11-1,29)	1,27 (1,20-1,43)	-	0,99 (0,96-1,06)	1,12 (1,03-1,21)	1,14 (1,02-1,27)
Novembre	1,11 (0,96-1,23)	1,10 (1,04-1,19)	-	0,96 (0,86-1,01)	0,97 (0,91-1,03)	1,02 (0,92-1,11)
Décembre	0,98 (0,97-1,06)	-	-	0,88 (0,81-0,95)	0,72 (0,66-0,78)	0,87 (0,74-0,98)

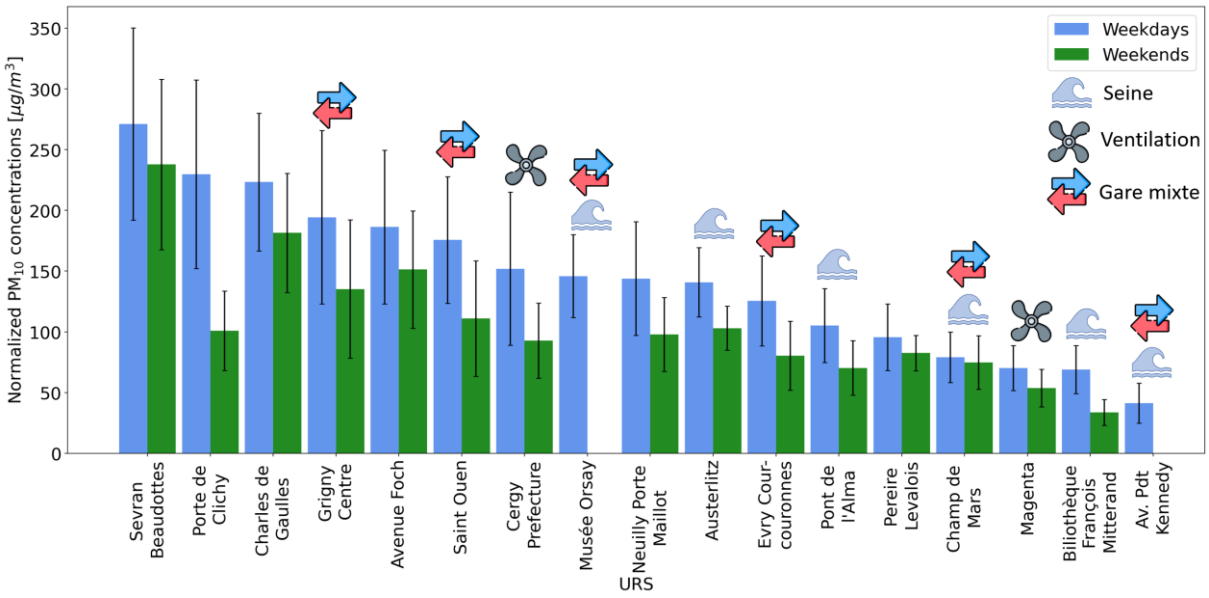


Figure 4.14 : Concentrations normalisées en PM_{10} (jours ouvrés et weekends) pour les différentes EFS

De manière générale, il est clairement établi que les concentrations en particules sont significativement plus élevées les jours ouvrés que les weekends. Toutefois, ces différences sont plus ou moins marquées en fonction des gares. Par exemple, à Grigny Centre, les concentrations atteignent

213 µg/m^3 les jours ouvrés contre 163 µg/m^3 les weekends, tandis qu'à Porte de Clichy, elles s'élèvent à environ 230 µg/m^3 en semaine et chutent à 101 µg/m^3 le weekend. Pour cette dernière gare, presque tous les profils journaliers ont été retenus après élimination des profils atypiques (14/15 pour les jours ouvrés et 4/4 pour les weekends).

Les gares mixtes, comme Avenue du président Kennedy et Champ de Mars, affichent des concentrations nettement plus faibles, en moyenne inférieures à 100 µg/m³. Ces résultats confirment que leur conception partiellement ouverte facilite une ventilation naturelle accrue, permettant une meilleure dispersion des particules.

Par ailleurs, les gares situées en grande majorité à proximité de la Seine, telles que Champ de Mars, Avenue du Président Kennedy, Pont de l'Alma, Musée d'Orsay, Austerlitz et Bibliothèque François-Mitterrand, présentent également des concentrations plus faibles. Des études menées à New York ont montré que la proximité d'une rivière pouvait significativement réduire les concentrations de particules, grâce à l'effet de dispersion des masses d'air au-dessus des cours d'eau (Luglio et al., 2023). Cette dynamique explique en partie les niveaux de pollution relativement bas observés dans les gares bordant la Seine (Figure 4.14). L'effet de cette ventilation naturelle semble même grandement limiter l'impact des facteurs d'influence tels que la nature du revêtement des voies. En effet, les gares de Bibliothèque François Mitterrand et Avenue du Président Kennedy, bien qu'équipées de ballast (Moreno et al., 2017) facteurs aggravant pour les concentrations de PM, affichent des concentrations plus faibles.

Par ailleurs, les gares comme Sevran Beaudottes et Charles de Gaulle, les deux seules gares de cette étude situées sur la ligne B, présentent les concentrations les plus élevées. Les trains de la ligne B sont parmi les plus anciens du réseau SNCF (modèles MI 84 et MI 79). Selon une étude récente (Tu & Olofsson, 2021), en raison de leur technologie vieillissante, ces trains sont susceptibles d'émettre davantage de particules.

La gare de Magenta, quant à elle, présente des particularités notables qui la distinguent des autres EFS. Bien qu'elle soit l'une des gares les plus profondes du réseau parisien, elle figure parmi celles affichant les concentrations les plus faibles en particules. Cette observation diffère de ce qui a été rapporté par Colombi et al. (2013), Kamani et al. (2014), et Van Ryswyk et al. (2017), qui associent généralement des concentrations élevées à une plus grande profondeur dans les enceintes souterraines. Cette performance remarquable s'explique principalement par la présence d'un système de ventilation mécanique sophistiqué, qui assure un renouvellement constant et efficace de l'air, même à des profondeurs importantes. En outre, la gare bénéficie de systèmes de désenfumage avancés et d'infrastructures modernes, liées à sa mise en service relativement récente en 1999. Ces caractéristiques techniques expliquent les niveaux moyens observés de 70 µg/m³ en semaine et 53 µg/m³ le weekend, malgré son caractère entièrement souterrain.

4.4 Incertitude des profils en fonction de la période de mesure

L'une des questions soulevées lors de cette étude, particulièrement en relation avec les campagnes de mesure de quelques semaines, concerne l'incertitude associée aux profils en fonction de la période de mesure. Pour explorer cette question, nous avons analysé l'évolution des profils types selon la durée d'échantillonnage des données. Cette étude a été principalement conduite à la gare d'Avenue Foch, qui bénéficie d'une base de données particulièrement complète.

La Figure 4.15 présente l'évolution de la moyenne glissante du MAE en fonction du nombre de semaines d'échantillonnage (courbe en bleu), ainsi que l'évolution de l'écart-type (courbe en olive) pour les profils journaliers moyens des concentrations en particules. Les barres d'erreur associées à la courbe bleue représentent une incertitude correspondant à un écart-type (σ=1), mettant en évidence la variabilité des données au fil des semaines.

La moyenne glissante du MAE montre une diminution progressive à mesure que la fenêtre d'échantillonnage augmente. Cela indique que plus la fenêtre d'observations est large, plus le profil journalier de référence correspondant gagne en fiabilité. Avec une semaine de données, le MAE est environ de 15 µg/m³. Cependant, dès que la période d'observation s'étend à 10 semaines, le MAE chute significativement pour atteindre environ 10 µg/m³. Au-delà de 25 semaines, la diminution du MAE devient moins marquée, se stabilisant autour de valeurs proches de 5 µg/m³, montrant une convergence des profils vers un état stable.

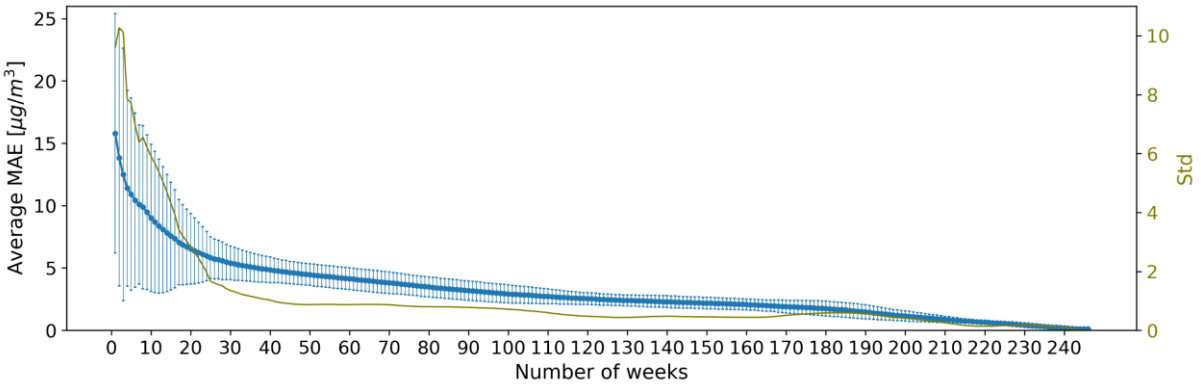


Figure 4.15 : Moyenne glissante de l'erreur absolue moyenne (points en bleu, avec des barres d'erreur correspondant à un écart-type) accompagnée de l'évolution glissante de l'écart-type (ligne continue en olive).

L'évolution de l'écart-type suit une tendance similaire, diminuant de manière significative au fil des semaines. Cette diminution de l'écart-type reflète une réduction de la variabilité des profils moyens, renforçant la robustesse et la précision des profils typiques obtenus avec une période d'échantillonnage prolongée. Selon ces observations, pour une semaine de données, l'incertitude est de 8,5% ± 5,1%, ce qui indique qu'une seule semaine de mesures suffit pour obtenir un profil journalier de référence avec une incertitude raisonnable. Toutefois, pour réduire davantage cette incertitude et garantir une précision accrue des profils, une période d'échantillonnage de 25 semaines, soit environ 6 mois, est recommandée.

Afin de valider ces conclusions, nous avons sélectionné des périodes aléatoires dans notre base de données, correspondant à des fenêtres d'échantillonnage croissantes : une semaine (01/04/2022 – 07/04/2022), deux semaines (01/04/2022 – 14/04/2022), et ainsi de suite jusqu'à six semaines (01/04/2022 – 12/05/2022). La Figure 4.16 illustre l'évolution des profils pour les jours ouvrés et les weekends, en partant d'une semaine de données et en incrémentant la fenêtre d'échantillonnage par tranches hebdomadaires jusqu'à couvrir l'intégralité des cinq années de données.

Pour les jours ouvrés (Figure 4.16a), les profils obtenus dès la première semaine (courbe bleue) montrent déjà des schémas horaires similaires à ceux du profil sur la période complète (courbe noire). Cependant, une variabilité plus importante est observée, notamment aux heures de pointe, en raison de l'insuffisance de données pour stabiliser le profil. À mesure que la période d'échantillonnage augmente (par exemple, 4 semaines ou 5 semaines), les profils deviennent de plus en plus proches du profil global, avec une réduction significative de la variabilité.

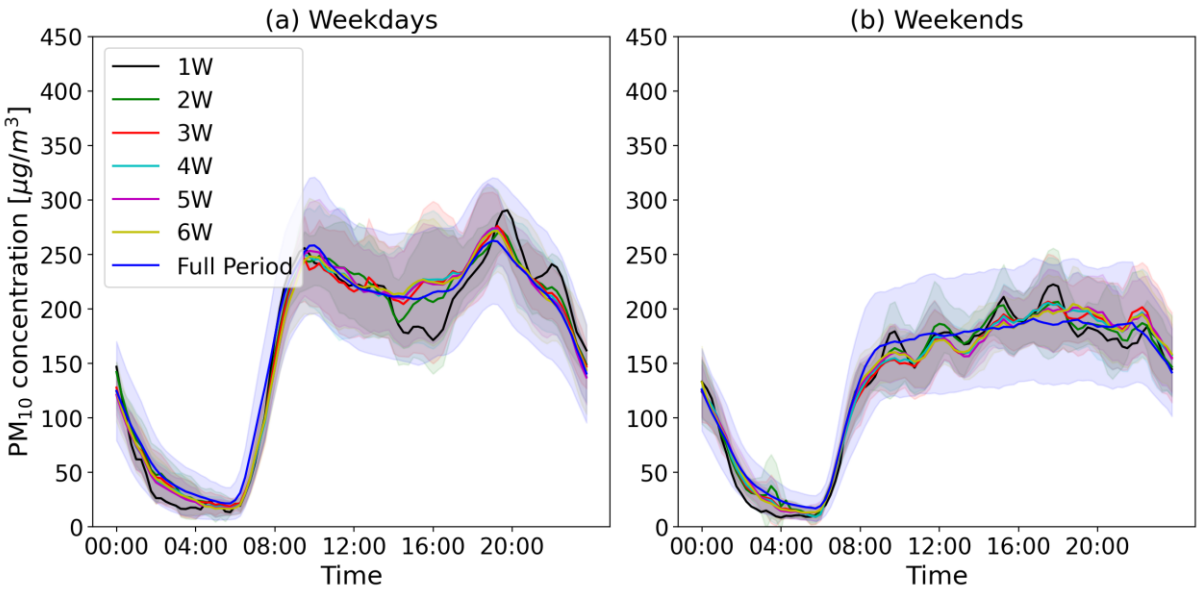


Figure 4.16 : Variation des profils types pour (a) les jours ouvrés et (b) les weekends au fil du temps. Les zones ombrées représentent un écart-type (σ=1).

Pour les weekends (Figure 4.16b), un comportement similaire est observé. Les profils obtenus avec une semaine de données montrent des variations plus marquées, notamment pendant les heures d'activité ferroviaire. À partir de la quatrième semaine, les profils commencent à converger vers celui de la période complète, confirmant que l'augmentation de la durée de l'échantillonnage réduit la variabilité et améliore la précision des profils typiques.

Comme le montrent la Figure 4.16 (a) et la Figure 4.16 (b), même avec une période de mesure d'une semaine (courbe bleue), il est possible d'obtenir un profil fiable, chaque point de mesure quart-horaire se rapprochant du profil établi sur cinq ans (courbe noire) avec une marge d'erreur inférieure à 10 %.

En partant de ces conclusions, bien que 6 semaines de mesure soit la période que nous recommandons pour les campagnes de mesure, nous pouvons conclure que les campagnes de mesures de 2 à 3 semaines déjà réalisées par la SNCF sont suffisantes pour l'analyse des profils correspondants en prenant en compte l'incertitude associée au profil.

4.5 Conclusion du chapitre

Ce chapitre a permis de mettre en application la méthodologie développée pour l'analyse à long terme et à court terme des concentrations de particules dans les EFS. Il a également démontré la transférabilité de cette méthode, tant pour les campagnes de longue ou de courte durée.

L'analyse à long terme des concentrations en particules a révélé une tendance saisonnière unique dans les EFS parisiennes, distincte de celle observée dans les EFS d'autres pays. Contrairement à la plupart des environnements, les concentrations de particules sont plus élevées en été (juillet, août, septembre) et plus faibles en hiver (décembre, janvier, février).

L'étude des ratios entre les $PM_{2.5}$ et les PM_{10} dans les gares de référence a montré des valeurs plus faibles (0,39 à 0,44) que celles rapportées dans la littérature, indiquant une prédominance des PM_{10}. À l'échelle annuelle, les tendances de ce ratio varient selon les gares. La gare d'Avenue Foch présente un ratio globalement stable tout au long de la période d'étude, tandis qu'à Sevran Beaudottes et à Magenta, une augmentation progressive de ce ratio a été observée. À l'échelle mensuelle, le ratio reste relativement stable pour les trois gares étudiées. En revanche, à l'échelle hebdomadaire, des ratios plus

faibles sont observés en semaine par rapport aux weekends, ce qui peut être principalement attribué à la réduction de l'activité ferroviaire durant les weekends. Ce ratio présente également des variations entre le jour (plus faible) et la nuit (plus élevé), ces dernières étant partiellement dues aux différences de vitesse de dépôt entre les deux classes de particules.

Concernant les profils journaliers des concentrations en particules, une tendance diurne marquée a été constatée à l'échelle quotidienne, caractérisée par deux pics prononcés lors des heures de pointe durant les jours ouvrés. En revanche, les concentrations durant les weekends sont généralement plus faibles et nettement constantes. Un décalage d'environ une heure est observé entre les pics de concentration et le début de l'activité ferroviaire, ce retard pouvant être attribué au temps nécessaire pour la remise en suspension des particules. Les amplitudes des profils varient d'une gare à l'autre, reflétant les particularités architecturales et les opérations ferroviaires de chaque site. Nous avons également observé que les profils identifiés sont principalement influencés par les facteurs internes aux EFS. D'importantes différences ont aussi été détectées entre les concentrations intérieures et extérieures, avec des niveaux intérieurs étant de 3 à 15 fois plus élevés que ceux à l'extérieur des EFS.

Dans les gares disposant de mesures de courte durée, les profils journaliers présentent également une tendance diurne, mais avec un signal plus bruité en raison de la durée limitée des mesures. Lors de l'analyse de ces gares, la méthodologie que nous avons développée s'est avérée très utile pour comparer les gares entre elles, malgré des campagnes réalisées à des périodes différentes. La nouvelle métrique introduite dans cette étude, le DAC, a permis de normaliser les concentrations mesurées en fonction du mois d'échantillonnage, offrant ainsi une base de comparaison plus robuste entre les gares, en se référant au DAC mensuel déterminé dans la gare d'Avenue Foch. Cette approche nous a permis de classer les gares et de réaliser une analyse comparative pour identifier les facteurs potentiels influençant les concentrations en particules. Nos résultats ont confirmé que les activités ferroviaires sont les principaux contributeurs aux pics de particules, avec une variabilité significative liée à la structure et à l'exploitation des différentes stations. Ces résultats montrent également que les concentrations observées ne sont pas le fruit d'un seul facteur, mais résultent de l'interaction de plusieurs facteurs contribuant aux variations identifiées.

Enfin, l'analyse de l'incertitude des profils en fonction de la durée des mesures a révélé qu'une semaine de données suffit à produire un profil type avec une incertitude inférieure à 10 %. Cependant, une période d'échantillonnage prolongée améliore significativement la précision, et une durée de six mois est recommandée pour obtenir un profil fiable et représentatif.

CHAPITRE 5. ANALYSE DES FACTEURS D'INFLUENCE SUR LES CONCENTRATIONS EN PARTICULES

Ce dernier chapitre aborde deux aspects cruciaux de notre étude : l'identification des facteurs influençant les concentrations en particules dans les enceintes ferroviaires souterraines (EFS) et l'évaluation de l'efficacité des systèmes de filtration.

Nous débuterons par une analyse détaillée des facteurs d'influence sur les concentrations en particules au sein des trois gares de référence, utilisant pour cela le modèle GAM. Cette analyse sera complétée par une évaluation de l'efficacité des systèmes de filtration installés à Avenue Foch et Sevran Beaudottes en utilisant la même méthodologie.

Enfin, le chapitre se conclura par l'évaluation de l'efficacité de ces systèmes de filtration en utilisant d'autres méthodes, telles que le DAC et le LSTM, permettant une comparaison des performances entre différentes approches.

5.1 Analyse des facteurs d'influence avec le modèle GAM

Cette section vise à évaluer de manière approfondie les facteurs qui modulent les concentrations en particules dans les trois gares de référence. Grâce à l'utilisation du modèle GAM, nous examinons les dynamiques spécifiques à chaque site, en identifiant les déterminants clés des variations observées. Chaque gare fait l'objet d'une analyse individualisée, mettant en lumière les particularités structurelles, opérationnelles et environnementales qui lui sont propres. Cette démarche permet de hiérarchiser les facteurs d'influence les plus significatifs et d'en comprendre l'impact respectif.

5.1.1 Cas d'Avenue Foch

Rappelons que les variables incluses dans le modèle GAM pour cette gare sont les suivantes : les jours de la semaine (DOW), les deux expérimentations ($Exp1, Exp2$) en tant que variables catégorielles, le CO_2, la température (T), l'humidité (RH), l'heure de la journée ($Time$), le numéro de la semaine ($Week$), et les paramètres extérieurs tels que les concentrations en PM, NO_2 et NO en tant que variables continues (voir équation 3.8).

Le Tableau 5.1 et le Tableau 5.2 résument respectivement les résultats des coefficients paramétriques et la significativité approximative des termes lisses pour le PM_{10}. La significativité des résultats est indiquée par des symboles d'étoiles, qui interprètent les valeurs de p de la manière suivante :

- *** : très significatif ($p < 0.001$)
- ** : significatif ($p < 0.01$)
- * : marginalement significatif ($p < 0.05$)
- Pas d'étoile : non significatif ($p \geq 0.05$)

La Figure 5.1 illustre l'effet de ces facteurs d'influence sur les concentrations de PM_{10} telles que modélisées par le modèle GAM. L'axe des ordonnées représente les valeurs de la fonction de lissage (par exemple, s(T) montre la tendance des concentrations de PM_{10} en fonction des variations de la température), tandis que le degré de liberté effectif (EDF) indique la complexité de la courbe. L'axe des abscisses représente les valeurs mesurées des facteurs d'influence. La courbe noire solide indique la

tendance des concentrations de PM_{10} avec la variation des facteurs d'influence, tandis que les lignes pointillées représentent l'intervalle de confiance à 95 % des concentrations de PM_{10}.

Tableau 5.1 : Résultats des termes paramétriques (linéaires) du modèle GAM pour les concentrations de PM_{10} à Avenue Foch

Term	Estimate	Std.Error	t value	Pr(>|t|)
β_0	156,8	0,8	204,9	$< 2e^{-16}$ ***
DOW2	6,8	1,1	6,4	$2e^{-10}$ ***
DOW3	5,8	1,1	5,3	$9,8e^{-8}$ ***
DOW4	4,2	1,1	3,9	$9,8e^{-5}$ ***
DOW5	0	1,1	0	1,0
DOW6	-32,4	1,1	-30,2	$< 2e^{-16}$ ***
DOW7	-28,9	1,1	-26,3	$< 2e^{-16}$ ***
Exp1-On	-20,5	1,3	-15,5	$< 2e^{-16}$ ***
Exp2-On	1,3	2	0,7	0,5

Tableau 5.2 : Résultats des termes lisses (non linéaires) du modèle GAM pour les concentrations de PM_{10} à Avenue Foch

Smooth Term	Edf	Ref.df	F	p-value
s(CO_2)	8	8,7	163,2	$< 2e^{-16}$ ***
s(RH)	7,2	8,2	23,7	$< 2e^{-16}$ ***
s(T)	8	8,7	99,8	$< 2e^{-16}$ ***
s(PM_{10_out})	1	1,1	43,9	$< 2e^{-16}$ ***
s(NO_{2_out})	4,2	5,2	3,8	0,002 **
s(Time):Weekdays	9	9	3331,3	$< 2e^{-16}$ ***
s(Time):Weekends	8,9	9	880,1	$< 2e^{-16}$ ***
s(Week)	4	4	367,2	$< 2e^{-16}$ ***

D'après le Tableau 5.1, nous observons les effets de divers facteurs catégoriels, y compris les deux expérimentations de filtration. $DOW1$ (qui représente le lundi), $Exp1-Off$ et $Exp2-Off$ ne sont pas inclus dans le tableau car ils servent de catégories de référence dans notre modèle. Cela signifie que les effets de tous les autres jours de la semaine (DOW) sont mesurés par rapport au lundi, et que les effets de la présence de l'expérimentation 1 ou de l'expérimentation 2 sont comparés à ceux observés en leur absence. L'expérimentation 1 montre une réduction significative des concentrations de PM_{10}, avec une estimation de coefficient de -20,5 µg/m^3 et une erreur standard de 1,3 µg/m^3 (Figure 5.1). Cela se traduit par une réduction de 13,1 % (Equation 3.11) des concentrations de PM_{10} pendant l'expérience 1, avec une marge d'erreur de ± 0,8 % (Equation 3.12). Ainsi, la réduction des concentrations de PM_{10} pendant l'expérience 1 est d'environ 13,1 % ± 0,8 %, mettant en évidence l'impact considérable et statistiquement significatif de la présence du système de filtration. Comme illustré dans la Figure 5.1, l'expérimentation 2 entraîne une augmentation moins marquée, qui n'était pas statistiquement significative (p = 0,5, Tableau 5.1), avec une estimation de coefficient positive de 1,3 µg/m^3 et une erreur standard de 2,0 µg/m^3. Cela conduit à une augmentation approximative de 1,3 % ± 2 % des concentrations de PM_{10} pendant l'expérience 2. L'expérimentation 1 montre une significativité statistique forte ($p < 2e^{-16}$), tandis que l'expérience 2 n'était pas statistiquement significative. L'expérimentation 1 a eu un effet plus substantiel sur les concentrations de PM_{10} par rapport à l'expérimentation 2, avec une réduction de 13,1 % ± 0,8 %.

Concernant les jours de la semaine (DOW), le GAM indique des différences dans les concentrations de PM_{10} selon les jours (Figure 5.1). Les jours de semaine ($DOW2$ à $DOW4$) présentent des niveaux de PM_{10} plus élevés que les weekends ($DOW6$ et $DOW7$), comme le montrent les coefficients paramétriques dans le Tableau 5.1, en raison des variations de l'activité humaine et des horaires opérationnels. Plus précisément, les coefficients pour $DOW2$ (6,8), $DOW3$ (5,8) et $DOW4$ (4,2)

sont positifs et significativement différents de zéro ($p < 2e^{-16}$), tandis que $DOW6$ et $DOW7$ montrent des valeurs négatives et significatives (-32,4 et -28,9, respectivement), indiquant une diminution des niveaux de PM_{10} par rapport aux jours de semaine. Notons que le coefficient pour $DOW5$ est de 0, avec une p = 1, indiquant que la concentration de PM ce jour-là est statistiquement indiscernable de celle de DOW1 (Figure 5.1).

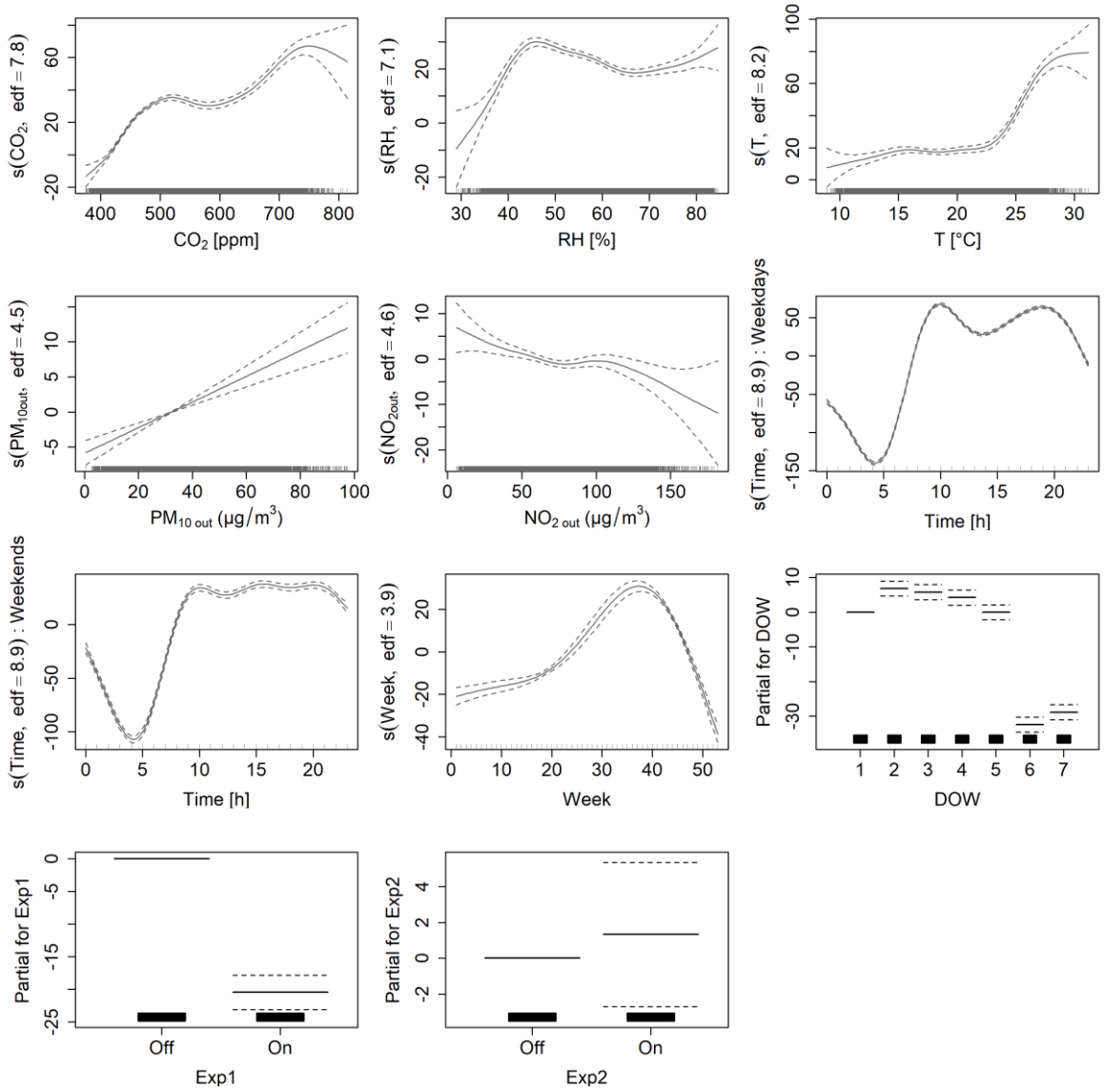


Figure 5.1 : Effet des facteurs d'influence sur les concentrations de PM_{10} dans le modèle GAM à la gare d'Avenue Foch

Le Tableau 5.2 détaille la significativité des termes lissés liés au CO_2, à l'humidité relative (RH), à la température (T), aux PM_{10} extérieures (PM_{10_out}), au NO_2 extérieur (NO_{2_out}), au temps ($Time$) et à la semaine ($Week$). Tous les termes lissés sont hautement significatifs ($p < 2e{-}16$), renforçant l'importance de ces variables dans l'influence des concentrations de PM_{10}. Cependant, le terme pour NO_2 suggère un niveau de significativité plus faible ($p < 0{,}01$), indiquant un impact moins prononcé sur les niveaux de PM_{10}.

Le terme lissé pour l'heure de la journée, avec une valeur F de 3331,3 pour les jours de semaine et de 880,1 pour les weekends, exerce l'influence la plus forte, indiquant que cette variable est le principal facteur déterminant des concentrations de PM_{10}. En effet, l'heure de la journée joue un rôle critique, avec le terme lissé pour le temps montrant un schéma hautement significatif et non linéaire

(Figure 5.1). La variation temporelle reflète la fréquence des trains (Birenzvige et al., 2003; Raut et al., 2009; Şahin et al., 2012; Querol et al., 2012b; Walther & Bogdan, 2017), avec des pics de concentrations de PM_{10} pendant les heures de pointe en semaine et des niveaux plus stables le weekend.

Le deuxième facteur le plus important est le numéro de la semaine, avec une valeur F de 367,2 (Tableau 5.2). La variation saisonnière représentée par la variable semaine influence significativement les concentrations de PM dans la station. Le GAM capture la variabilité au cours de l'année, qui pourrait refléter des changements à long terme des conditions environnementales ou des schémas d'utilisation des stations souterraines. Notamment, les concentrations de PM atteignent un pic en été et diminuent en hiver (Figure 5.1), comme souligné dans le chapitre 4. Toutefois, les mécanismes à l'origine de cette tendance saisonnière demeurent mal élucidés. Cette incertitude est renforcée par le fait que des tendances inverses sont observées ailleurs, comme à Stockholm, en Suède (Tu & Olofsson, 2024), et à Séoul, en Corée du Sud (Park et al., 2024).

Le troisième facteur le plus important est le CO_2, qui est associé au nombre de personnes présentes dans la station à un moment donné (Yang et al., 2023). Toutefois, dans le contexte d'une gare souterraine, le CO_2 pourrait également être un traceur des influences environnementales externes. Comme illustré dans la Figure 5.1, une relation non linéaire positive significative est observée entre les niveaux de CO_2 et les concentrations de PM_{10}. À mesure que les niveaux de CO_2 augmentent, les concentrations de PM_{10} augmentent également, en particulier à des niveaux élevés de CO_2. Cette corrélation est cohérente avec les travaux de Park and Ha (2008), qui ont démontré que les concentrations de PM_{10} et de $PM_{2.5}$ sont fortement corrélées au CO_2.

En revanche, la température (T), l'humidité relative (RH) et les PM_{10} extérieures, et encore plus le NO_2 extérieur, n'ont pas un impact fort sur les concentrations de PM_{10}. Cependant, la Figure 5.1 montre que T, RH et les PM_{10} extérieures ont chacune une relation non linéaire positive avec les concentrations de PM_{10}, bien que ces effets soient moins prononcés que ceux du CO_2.

Comme observé pour les PM_{10}, les résultats pour les $PM_{2.5}$ dans le Tableau 5.3 montrent une réduction significative des concentrations de particules pendant l'expérimentation 1 (technologie par ionisation). La diminution de 16,4 µg/m^3 correspond à une réduction d'environ 26 % ± 0,8 % (équations 3.11 et 3.12). Cependant, il existe une différence marquée dans l'efficacité de l'expérimentation 2 (technologie par voie humide). Pour les $PM_{2.5}$, l'expérimentation 2 entraîne une réduction de seulement 2,9 µg/m^3, correspondant à une diminution de 4,6 % ± 1,3 % (équations 3.11 et 3.12). Bien que statistiquement significative, cette réduction faible indique à nouveau que cette méthode de filtration est moins efficace par rapport à l'expérimentation 1. Cela contraste avec le résultat non significatif observé pour les PM_{10}. Ces résultats indiquent que si l'expérimentation 1 a été efficace pour les $PM_{2.5}$ et les PM_{10}, l'expérimentation 2 a eu un effet négligeable et statistiquement non significatif sur les concentrations de PM_{10}.

Tableau 5.3 : Résultats des termes paramétriques (linéaires) du modèle GAM pour les concentrations de $PM_{2.5}$ à Avenue Foch

Term	Estimate	Std.Error	t value	Pr(>|t|)
β_0	63,1	0,3	204,1	$< 2e^{-16}$ ***
DOW2	3,1	0,4	7,2	$4,9e^{-13}$ ***
DOW3	2,4	0,4	5,6	$2,3e^{-8}$ ***
DOW4	2	0,4	4,6	$5,1e^{-6}$ ***
DOW5	0,4	0,4	1	0,3
DOW6	-12	0,4	-27,5	$< 2e^{-16}$ ***
DOW7	-10,9	0,4	-24,5	$< 2e^{-16}$ ***
Exp1-On	-16,4	0,5	-30,7	$< 2e^{-16}$ ***
Exp2-On	-2,9	0,8	-3,5	0,0005 ***

Tableau 5.4 : Résultats des termes lisses (non linéaires) du modèle GAM pour les concentrations de $PM_{2.5}$ à Avenue Foch

Smooth Term	Edf	Ref,df	F	p-value
s(CO_2)	7,8	8,6	176,3	< $2e^{-16}$ ***
s(RH)	7,1	8,1	20,5	< $2e^{-16}$ ***
s(T)	8,2	8,8	88,1	< $2e^{-16}$ ***
s(PM_{10_out})	4,5	5,5	32,8	< $2e^{-16}$ ***
s(NO_{2_out})	4,6	5,7	4,2	0,0005 ***
s(Time):Weekdays	9	9	2541,3	< $2e^{-16}$ ***
s(Time):Weekends	8,9	9	647	< $2e^{-16}$ ***
s(Week)	4	4	301,5	< $2e^{-16}$ ***

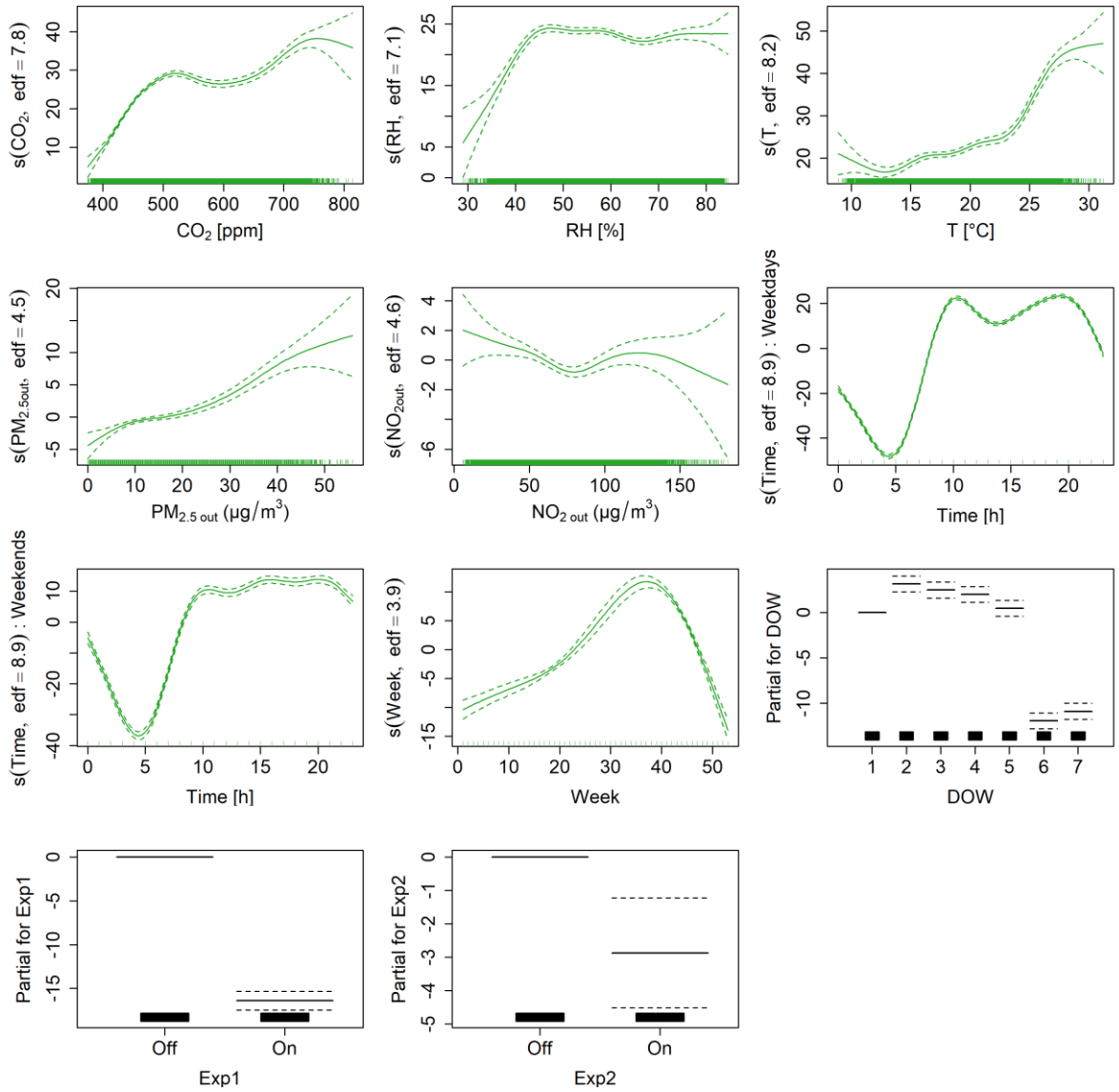


Figure 5.2 : Effet des facteurs d'influence sur les concentrations de $PM_{2.5}$ dans le modèle GAM à la gare d'Avenue Foch

Comme illustré dans le Tableau 5.3, le Tableau 5.4 et la Figure 5.2, le modèle GAM continue de révéler une relation complexe et non linéaire entre les concentrations de $PM_{2.5}$ et les variables influentes (CO_2, RH, T, PM_{10_out}, NO_{2_out}, heure de la journée, numéro de la semaine, jour de la semaine, expérimentation 1 et expérimentation 2). Ces résultats sont cohérents avec ceux obtenus pour les PM_{10} (Figure 5.2).

5.1.2 Cas de Sevran Beaudottes

Pour la gare de Sevran Beaudottes, les paramètres analysés incluent la température au niveau du quai ainsi que les concentrations des polluants extérieurs, qui varient selon la variable cible : les concentrations de PM_{10} extérieures pour les PM_{10} intérieures, et les concentrations de $PM_{2.5}$ extérieures pour les $PM_{2.5}$ intérieures. Les concentrations de NO_2 et de NO sont également incluses, offrant une occasion unique d'explorer les interactions entre les conditions extérieures et les concentrations de particules à l'intérieur de la gare. Les variables temporelles, telles que l'heure de la journée ($Time$), le numéro de la semaine ($Week$) et le jour de la semaine (DOW), enrichissent l'analyse. De plus, les périodes spécifiques liées à l'expérimentation (Exp, filtration classique) menée dans cette gare sont prises en compte, permettant une meilleure compréhension des dynamiques journalières et hebdomadaires, ainsi que des impacts des interventions ciblées durant l'expérimentation (voir équation 3.9).

Les résultats des analyses sont présentés dans le Tableau 5.5 et le Tableau 5.6, qui résument respectivement les coefficients paramétriques et la significativité approximative des termes lisses pour les concentrations de PM_{10}. Par ailleurs, la Figure 5.3 illustre l'effet de ces facteurs d'influence sur les concentrations de PM_{10}, telles que modélisées par le modèle GAM. Le Tableau 5.5 met en évidence les effets significatifs des jours de la semaine (DOW) et des périodes expérimentales (Exp) sur les concentrations de PM_{10}. Des écarts marqués dans les niveaux de PM_{10} sont observés selon les jours de la semaine. Par rapport au lundi, qui sert de jour de référence ($DOW1$), les concentrations diminuent progressivement en approchant du weekend, avec des réductions substantielles le samedi (-41 µg/m³) et le dimanche (-40 µg/m³). Ces diminutions sont attribuables à une baisse de l'activité ferroviaire et une fréquentation réduite durant les weekends. La Figure 5.3 illustre clairement cette variabilité, démontrant les différences entre les jours ouvrés et les weekends, en particulier à travers les termes temporels $s(Time): Weekdays$ et $s(Time): Weekends$. Cependant, il est important de noter que les profils de variation des DOW à Sevran Beaudottes diffèrent de ceux observés à Avenue Foch. À Sevran, les concentrations commencent à décroître dès le lundi, contrairement à Avenue Foch où une augmentation est enregistrée à partir du lundi, atteignant un pic le mardi avant de diminuer durant le weekend. Cette divergence reflète les caractéristiques opérationnelles distinctes des deux gares.

L'expérimentation (Exp) a également un effet significatif sur les concentrations de PM_{10} ($p < 2e^{-16}$, selon le Tableau 5.5). L'installation des systèmes de filtration à la gare de Sevran Beaudottes a entraîné une diminution de 77,3 µg/m³, correspondant à une réduction de 34,4% ± 0,8% (Equations 3.11 et 3.12). Cette observation est conforme aux résultats obtenus par Pretot et al. (2022), qui ont documenté des baisses similaires (environ 30% de réduction) dans les niveaux de PM_{10} à la même station. La réduction significative des PM_{10}, illustrée dans la Figure 5.3, confirme l'efficacité des stratégies de filtration mises en œuvre.

Tableau 5.5 : Résultats des termes paramétriques (linéaires) du modèle GAM pour les concentrations de PM_{10} à Sevran-Beaudottes

Term	Estimate	Std.Error	t value	Pr(>|t|)
(Intercept)	224,8	1,3	174,2	$< 2e^{-16}$ ***
DOW2	-6,5	1,8	-3,6	0,0003 ***
DOW3	-8,1	1,8	-4,5	$6e^{-6}$ ***
DOW4	-7,2	1,8	-4,0	$6e^{-5}$ ***
DOW5	-12,4	1,8	-7,0	$3e^{-12}$ ***
DOW6	-41,1	1,8	-23,2	$< 2e^{-16}$ ***
DOW7	-39,4	1,8	-21,8	$< 2e^{-16}$ ***
ExpOn	-77,3	1,7	-46,2	$< 2e^{-16}$ ***

Tableau 5.6 : Résultats des termes lisses (non linéaires) du modèle GAM pour les concentrations de PM_{10} à Sevran-Beaudottes

Smooth term	edf	Ref.df	F	p-value
s(T)	8,3	8,9	75,9	$< 2e^{-16}$ ***
$s(PM_{10_out})$	6,6	7,6	44,4	$< 2e^{-16}$ ***
$s(NO_{2_out})$	5,1	6,1	5,3	$2e^{-5}$ ***
$s(NO_{out})$	1,4	1,7	0,4	0,5
s(Time): Weekdays	9,0	9,0	2529,5	$< 2e^{-16}$ ***
s(Time): Weekends	8,9	9,0	635,7	$< 2e^{-16}$ ***
s(Week)	4,9	5,0	90,6	$< 2e^{-16}$ ***

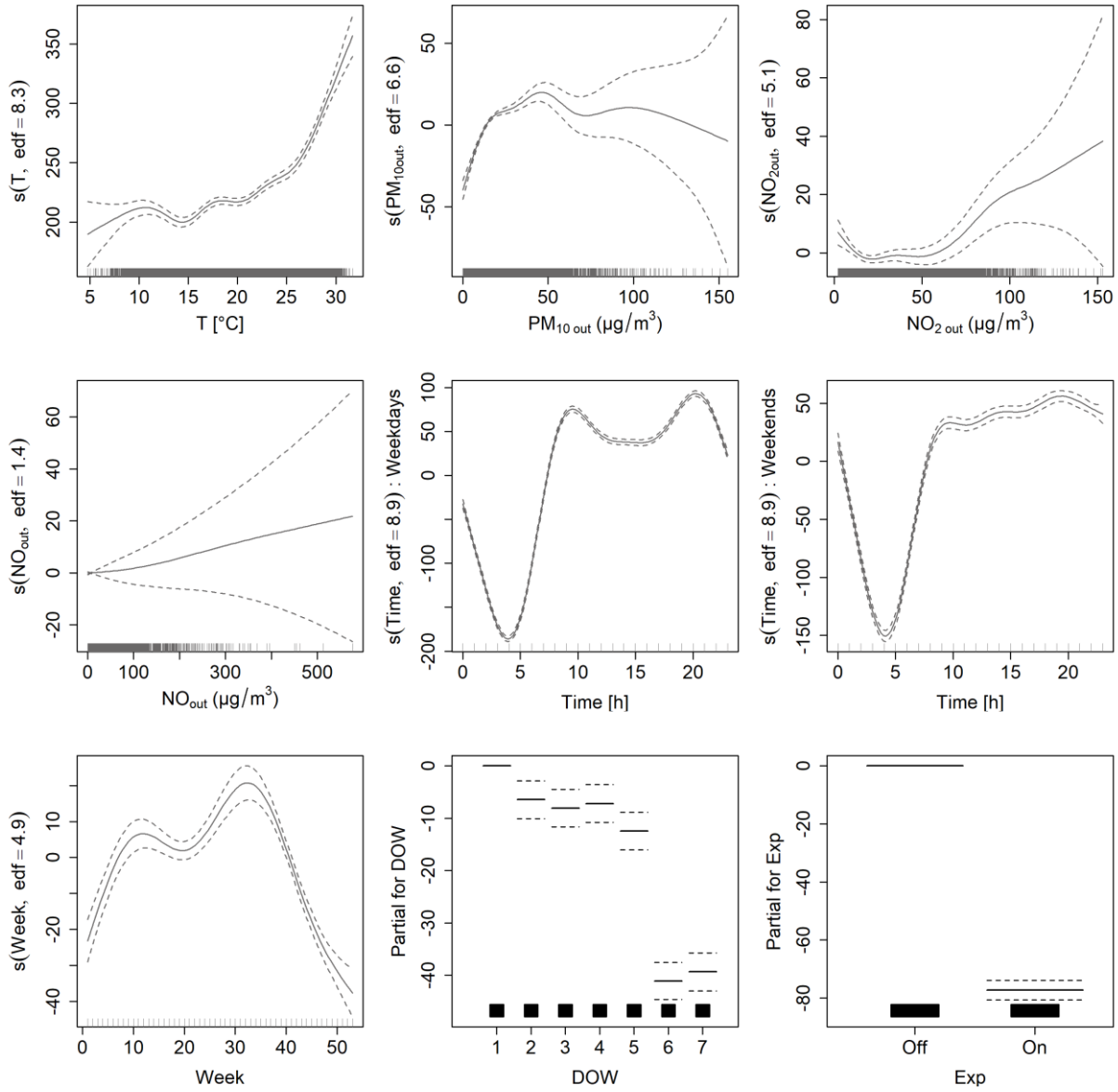


Figure 5.3 : Effet des facteurs d'influence sur les concentrations de PM_{10} dans le modèle GAM à la gare de Sevran Beaudottes

Le Tableau 5.6 examine les relations non linéaires entre les facteurs explicatifs et les concentrations de PM_{10}. A l'exception de NO_{out}, tous les termes du modèle sont statistiquement significatifs, soulignant la pertinence de chaque paramètre. Similairement à la gare d'Avenue Foch, le terme lisse associé au temps ($Time$) affiche des valeurs de F très élevées, respectivement de 2529,5 pour les jours ouvrés et de 635,7 pour les weekends, confirmant son influence prédominante sur les concentrations de particules. Cette influence est particulièrement marquée durant les jours ouvrés. En effet, l'heure de la journée détermine significativement les niveaux de particules, comme le démontrent les termes lisses avec des motifs non linéaires significatifs (Figure 5.3). La variation temporelle reflète les patterns de déplacement quotidiens, étroitement liés à la fréquence des trains (Figure 5.3).

Après le paramètre $Time$, le paramètre $Week$ se révèle être le second facteur le plus influent, avec une valeur de F de 90,6. Le modèle GAM reproduit fidèlement les tendances saisonnières observées par la méthode DAC. Notamment, les concentrations de PM_{10} atteignent un pic en été et diminuent en hiver, comme illustré à la gare d'Avenue Foch. Cependant, il est pertinent de noter qu'à la gare de Sevran Beaudottes, un pic supplémentaire se manifeste au printemps avant d'atteindre son maximum en été. Bien que ce phénomène soit à peine perceptible à la gare d'Avenue Foch, il est clairement évident au printemps à Sevran Beaudottes où nous observons également un pic de concentrations de particules.

Par ailleurs, la température sur les quais des gares (F = 75,9) joue un rôle prépondérant, tout comme les concentrations extérieures de PM_{10} (F = 44,4). Bien que ces dernières affichent des dynamiques très différentes à court et à long terme, comme nous l'avons discuté dans le chapitre précédent, elles restent significativement importantes dans notre modèle. Bien que l'impact du NO_2 sur les concentrations de PM_{10} soit statistiquement significatif, il est relativement faible (F = 5,3), ce qui suggère que ce polluant a une influence très limitée dans les EFS. Le terme $s(NO_{out})$ est non significatif (p = 0,6).

Pour les $PM_{2.5}$, les résultats présentés dans les tableaux Tableau 5.7 et Tableau 5.8 ainsi que dans la Figure 5.4 associée montrent des similarités notables avec ceux relatifs aux PM_{10}. Toutefois, des points spécifiques méritent d'être soulignés.

Tableau 5.7 : Résultats des termes paramétriques (linéaires) du modèle GAM pour les concentrations de $PM_{2.5}$ à Sevran Beaudottes

Term	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	93,9	0,5	175,3	$< 2e^{-16}$ ***
DOW2	-2,2	0,7	-2,9	0,004 **
DOW3	-2,5	0,7	-3,4	0,0007 ***
DOW4	-2,2	0,7	-3,0	0,003 **
DOW5	-4,6	0,7	-6,3	$4e^{-10}$ ***
DOW6	-16,1	0,7	-21,9	$< 2e^{-16}$ ***
DOW7	-14,8	0,7	-19,8	$< 2e^{-16}$ ***
ExpOn	-31,2	0,7	-44,7	$< 2e^{-16}$ ***

Comme observé pour les PM_{10}, les analyses paramétriques révèlent une variation significative des concentrations de $PM_{2.5}$ selon le jour de la semaine. Les coefficients associés aux $DOW2$, $DOW3$, $DOW4$, $DOW5$, $DOW6$ et $DOW7$ sont tous négatifs, indiquant une décroissance marquée pendant les weekends ($DOW6$ et $DOW7$). Ces observations concordent avec les résultats obtenus pour les PM_{10}, confirmant une réduction des concentrations en fin de semaine, associée à une diminution des activités ferroviaires et une fréquentation moindre. Cependant, il est à noter que la diminution observée en milieu de semaine (jours ouvrés) n'était pas détectable avec la méthode DAC (Figure A. 4b et Figure A. 5b, Annexe 4). Les raisons de cette divergence restent à élucider.

L'expérimentation (Exp) montre également un effet significatif sur les $PM_{2.5}$, avec une diminution de 31,2 µg/m³, correspondant à une réduction de 33,2% ± 0,8%. Cette réduction est légèrement inférieure à celle observée pour les PM_{10}.

Tableau 5.8 : Résultats des termes lisses (non linéaires) du modèle GAM pour les concentrations de $PM_{2.5}$ à Sevran Beaudottes

Smooth term	edf	Ref.df	F	p-value
s(T)	8,2	8,8	82,3	$< 2e^{-16}$ ***
s($PM_{2.5_out}$)	5,5	6,5	17,2	$< 2e^{-16}$ ***
s(NO_{2_out})	1,0	1,0	26,6	$9e^{-7}$ ***
s(NO_{out})	1,0	1,0	0,3	0,6
s(Time): Weekdays	9,0	9,0	2632,2	$< 2e^{-16}$ ***
s(Time): Weekends	8,9	9,0	667,3	$< 2e^{-16}$ ***
s(Week)	4,9	5,0	154,7	$< 2e^{-16}$ ***

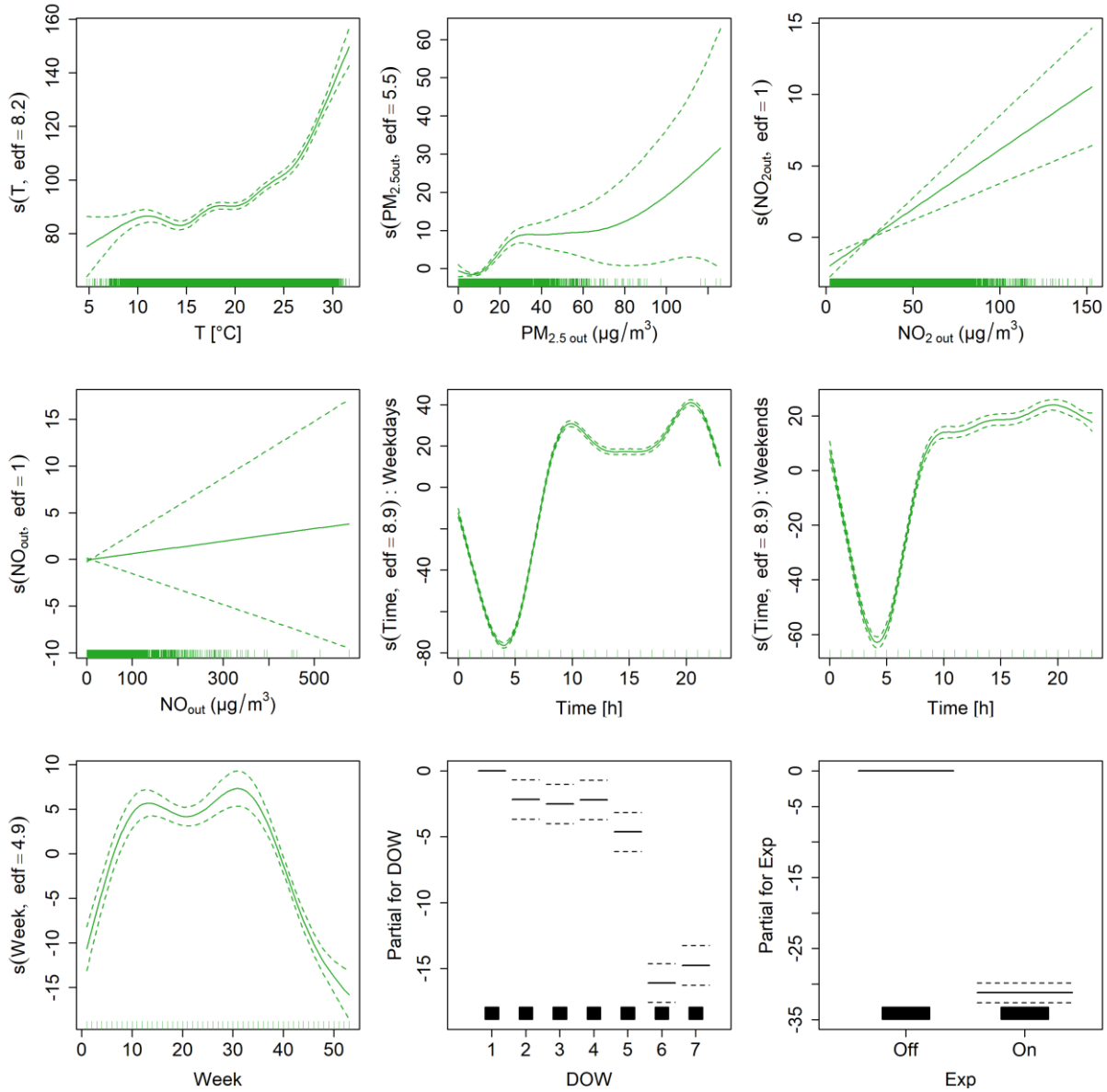


Figure 5.4 : Effet des facteurs d'influence sur les concentrations de $PM_{2.5}$ dans le modèle GAM à la gare de Sevran Beaudottes

Les analyses des termes lisses (Tableau 5.8) révèlent également des tendances similaires à celles observées pour les PM_{10}. L'heure de la journée, représentée par le terme lisse $s(Time)$, demeure le facteur le plus influent, avec des pics notables aux heures de pointe durant les jours ouvrés et une variation moins marquée les weekends. De même, la variable $Week$ indique une tendance saisonnière analogue, avec des concentrations maximales en été et minimales en hiver.

Les facteurs environnementaux, tels que la température $s(T)$ et les $PM_{2.5}$ extérieures, exercent également un impact significatif. La relation entre la température et les concentrations de $PM_{2.5}$ est positive et non linéaire, suggérant que les niveaux de $PM_{2.5}$ augmentent avec l'accroissement des températures. Ce comportement est conforme à celui observé pour les PM_{10}. En revanche, contrairement aux PM_{10}, les $PM_{2.5}$ affichent une relation linéaire avec le NO_{2_out} (edf = 1) et le NO_{out} (edf = 1), bien que l'effet du NO_{out} ne soit pas statistiquement significatif (p=0.57).

En somme, à Sevran-Beaudottes, trois principaux groupes de facteurs émergent comme déterminants pour les concentrations en particules (PM_{10} et $PM_{2.5}$) : les facteurs temporels, incluant les variations journalières, hebdomadaires et le paramètre numéro de la semaine; les facteurs environnementaux, tels que la température au niveau du quai, les concentrations extérieures de PM et de NO_2; et les interventions ciblées, notamment l'installation de systèmes de filtration qui ont prouvé leur efficacité à réduire les niveaux de concentrations en particules.

5.1.3 Cas de Magenta

Après l'analyse quantitative menée sur les gares d'Avenue Foch et de Sevran-Beaudottes, la gare de Magenta offre une nouvelle perspective pour explorer les facteurs influençant les concentrations en particules dans les EFS. Située au cœur de Paris, cette gare moderne et profondément enterrée (33m de profondeur) constitue un cas particulier en raison de ses infrastructures récentes, mises en service en 1999, et de son système de ventilation mécanique avancé. L'étude de cette gare permet d'analyser les dynamiques propres à un environnement souterrain complexe et de tester la capacité du modèle GAM à capturer les interactions complexes entre les paramètres environnementaux, structurels et opérationnels.

Le Tableau 5.9, le Tableau 5.10 ainsi que la Figure 5.5 illustrent les résultats correspondant au modèle GAM appliqué aux concentrations de PM_{10} dans la gare de Magenta. Le Tableau 5.9 présente les résultats des coefficients paramétriques (linéaires) du modèle. Le Tableau 5.10, quant à lui, détaille les résultats des termes lisses (non linéaires) du modèle. La Figure 5.5 offre une visualisation graphique des tendances décrites dans les tableaux.

Tableau 5.9 : Résultats des termes paramétriques (linéaires) du modèle GAM pour les concentrations de PM_{10} à Magenta

Term	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	65,1	0,4	161,7	$< 2e^{-16}$ ***
DOW2	-0,8	0,6	-1,5	0,1
DOW3	-2,7	0,6	-4,8	$2e^{-6}$ ***
DOW4	-2,5	0,6	-4,4	$1e^{-5}$ ***
DOW5	-2,7	0,6	-4,7	$2e^{-6}$ ***
DOW6	-16,5	0,6	-29,2	$< 2e^{-16}$ ***
DOW7	-18,2	0,6	-32,2	$< 2e^{-16}$ ***

Tableau 5.10 : Résultats des termes lisses (non linéaires) du modèle GAM pour les concentrations de PM_{10} à Magenta

Smooth term	edf	Ref.df	F	p-value
s(T)	8,3	8,9	99,8	$<2e^{-16}$ ***
s(PM_{10_out})	6,4	7,4	42,4	$<2e^{-16}$ ***
s(NO_{2_out})	6,2	7,2	29,4	$<2e^{-16}$ ***
s(NO_{out})	4,7	5,8	15,9	$<2e^{-16}$ ***
s(O_{3_out})	3,7	4,6	9,6	$<2e^{-16}$ ***
s(Time): Weekdays	9,0	9,0	1275,2	$<2e^{-16}$ ***
s(Time): Weekends	8,7	9,0	270,8	$<2e^{-16}$ ***
s(Week)	5,0	5,0	340,7	$<2e^{-16}$ ***

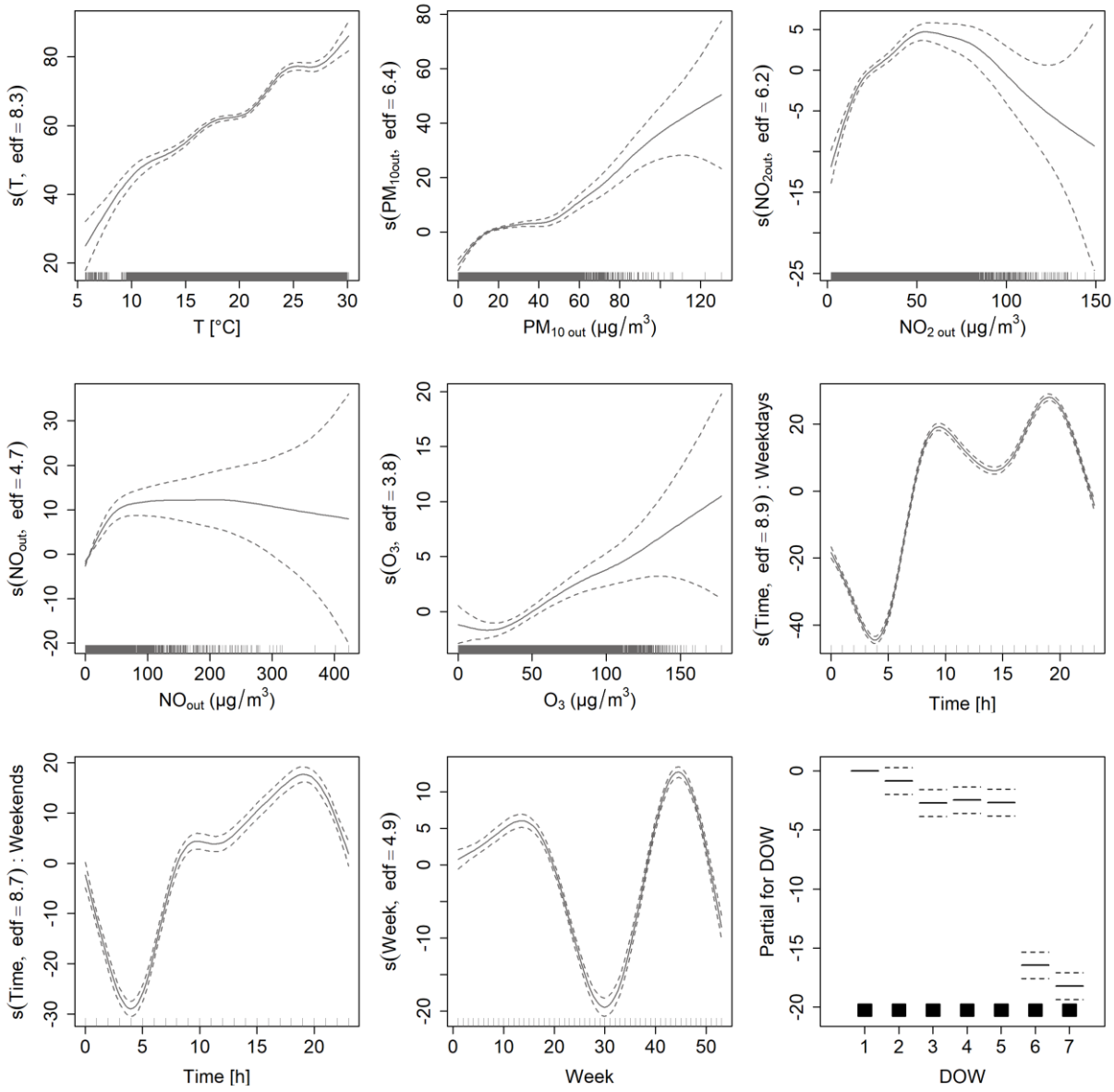


Figure 5.5 : Effet des facteurs d'influence sur les concentrations de PM_{10} dans le GAM à la gare de Magenta

Concernant les jours de la semaine (DOW), des coefficients négatifs sont observés de $DOW2$ à $DOW7$, indiquant des réductions significatives des concentrations de PM_{10} par rapport au lundi ($DOW1$), jour de référence. Ces diminutions sont particulièrement prononcées durant les weekends, avec des réductions estimées à -16,5 µg/m³ pour le samedi ($DOW6$) et -18,2 µg/m³ pour le dimanche

($DOW7$), comme illustré dans la Figure 5.5. Ces résultats suggèrent une fréquentation réduite de la gare pendant ces jours, similaire à celle observée à la gare de Sevran Beaudottes. Notons que le coefficient pour le mardi ($DOW2$) n'est pas statistiquement significatif, indiquant que les concentrations de PM_{10} ce jour-là sont comparables à celles du lundi. Comme observé précédemment à la gare de Sevran Beaudottes, la gare de Magenta présente une tendance similaire où les concentrations de PM_{10} diminuent progressivement au cours des jours ouvrés, en contraste avec Avenue Foch où cette tendance n'est pas évidente.

En ce qui concerne les termes lisses, le Tableau 5.10 révèle que, comme pour les deux autres gares de référence, le paramètre Time (F = 1275,2 pour les jours ouvrés et F = 270,8 pour les weekends) et le numéro de la semaine Week (F = 340,7) sont les facteurs prédominants. Le paramètre Week met en évidence des variations saisonnières significatives, avec des concentrations de PM_{10} généralement plus élevées durant les mois de septembre, octobre et novembre, ce qui confirme les observations issues de la méthode DAC. D'après la Figure 5.5, un creux marqué apparaît en été, principalement en raison des travaux d'été CASTOR qui ont été exclus de la base de données. Un pic de concentration est également observé au printemps pour cette gare.

Parmi les facteurs environnementaux, la température (T) occupe une place notable, confirmant son rôle significatif ($p < 2e^{-16}$) dans les variations de PM_{10}. La Figure 5.5 montre que la température et les concentrations de PM_{10} évoluent de manière synchronisée ; en effet, à des températures plus élevées correspondent des concentrations plus élevées de particules.

Enfin, bien que l'influence des paramètres externes, tels que les concentrations extérieures de PM_{10}, de NO_2, de NO et de O_3, soit relativement faible (F = 42,4 ; 29,4 ; 15,9 ; 9,6, respectivement), tous ces paramètres sont statistiquement significatifs. Ces résultats suggèrent que, malgré la profondeur de cette gare, qui est l'une des plus importantes du réseau SNCF, l'air extérieur contribue à la variation des concentrations de particules au niveau du quai, amenant des polluants tels que les PM_{10}, le NO_2, le NO et l'O_3, principalement via le système de ventilation de la gare. Ces observations corroborent les conclusions d'études antérieures (Fortain, 2008; Lee et al., 2018; Pan et al., 2019; Son et al., 2021).

Concernant les $PM_{2.5}$, le Tableau 5.11 et le Tableau 5.12 présentent respectivement les résultats des coefficients paramétriques et la significativité statistique approximative des termes lisses. La Figure 5.6 illustre l'effet de ces facteurs d'influence sur les concentrations de $PM_{2.5}$. Étant donné l'absence de données disponibles pour les $PM_{2.5}$ extérieurs, notre modèle a utilisé à la place les données des PM_{10} extérieurs, comme indiqué dans le Tableau 5.12.

Tableau 5.11 : Résultats des termes paramétriques (linéaires) du modèle GAM pour les concentrations de $PM_{2.5}$ à Magenta

Term	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	23,2	0,1	168,6	$< 2e^{-16}$ ***
DOW2	-0,1	0,2	-0,5	0,6
DOW3	-0,4	0,2	-2,3	0,02 *
DOW4	-0,9	0,2	-4,9	$1e^{-6}$ ***
DOW5	-0,8	0,2	-4,3	$2e^{-5}$ ***
DOW6	-4,6	0,2	-24,0	$< 2e^{-16}$ ***
DOW7	-4,9	0,2	-25,2	$< 2e^{-16}$ ***

Tableau 5.12 : Résultats des termes lisses (non linéaires) du modèle GAM pour les concentrations de $PM_{2.5}$ à Magenta

Smooth term	edf	Ref.df	F	p-value
s(T)	7,0	8,1	115,4	$<2e^{-16}$ ***
$s(PM_{10_out})$	5,3	6,3	139,6	$<2e^{-16}$ ***
$s(NO_{2_out})$	5,9	6,9	34,8	$<2e^{-16}$ ***
$s(NO_{out})$	3,7	4,7	23,1	$<2e^{-16}$ ***
$s(O_{3_out})$	6,7	7,7	14,7	$<2e^{-16}$ ***
s(Time): Weekdays	8,9	9,0	1029,2	$<2e^{-16}$ ***
s(Time): Weekends	8,6	9,0	247,9	$<2e^{-16}$ ***
s(Week)	5,0	5,0	281,0	$<2e^{-16}$ ***

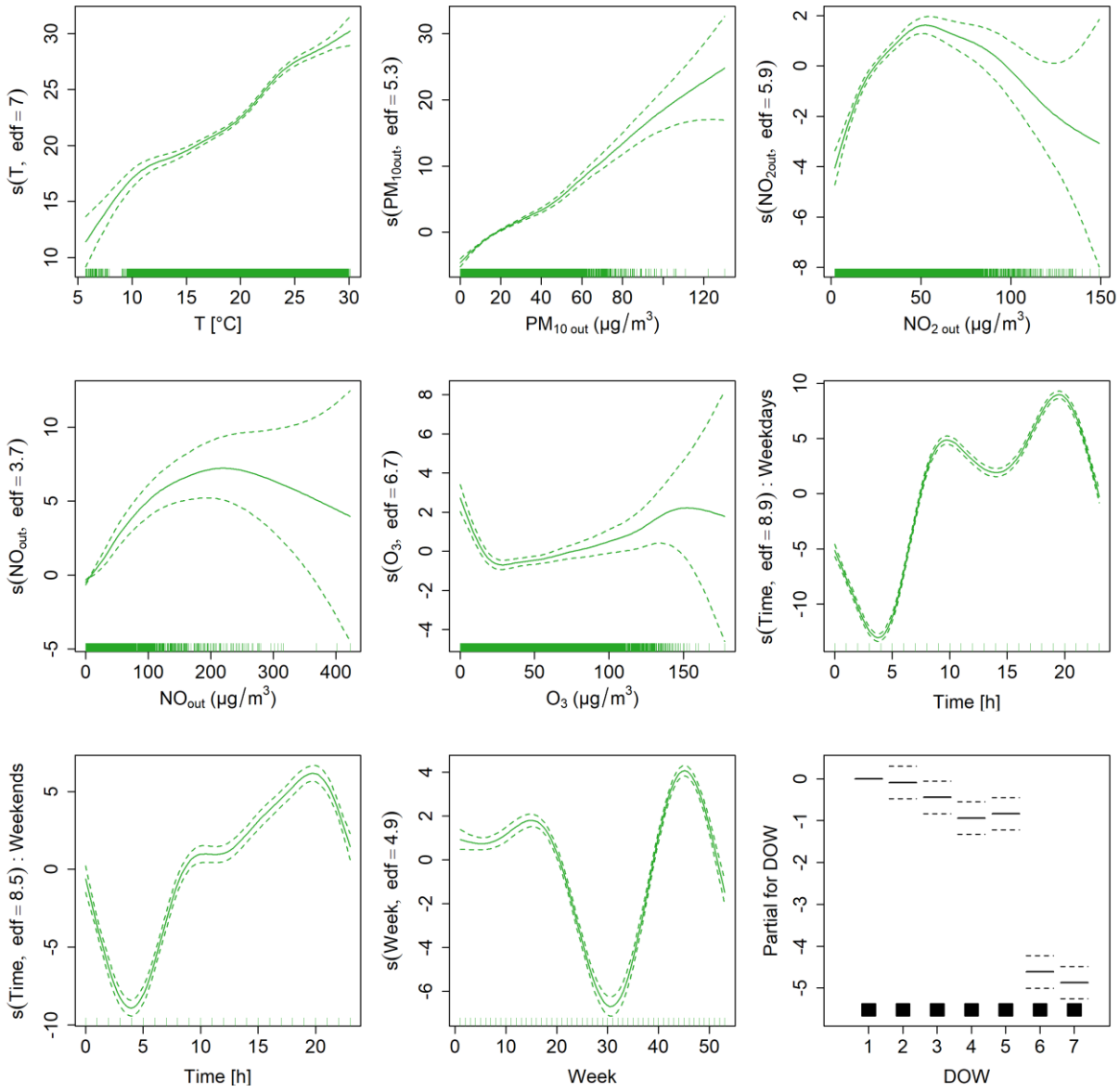


Figure 5.6 : Effet des facteurs d'influence sur les concentrations de $PM_{2.5}$ dans le GAM à la gare de Magenta

Les coefficients paramétriques révèlent des tendances similaires à celles observées pour les PM_{10}, avec une diminution progressive des concentrations de particules du lundi au dimanche. Il est important de souligner que les jours $DOW2$ et $DOW3$ présentent une significativité moindre par rapport aux autres jours de la semaine. Cela est illustré par le fait que $DOW2$ ($p = 0.6$), non significatif, est au même niveau

que $DOW1$, qui sert de référence. De même, pour $DOW3$ (p = 0.02), les niveaux de concentration le mercredi sont plus proches de ceux du lundi, comme le montre la Figure 5.6.

En ce qui concerne les termes lisses, il est essentiel de noter que, selon les tests statistiques présentés dans le Tableau 5.12, tous sont significatifs, soulignant l'importance de chaque paramètre inclus dans le modèle. Comme pour les PM_{10}, les facteurs les plus influents sont l'heure de la journée pour les jours ouvrés et les weekends ainsi que le numéro de la semaine. La Figure 5.6 illustre les tendances observées à court et long terme des concentrations de particules, conformément aux observations faites avec la méthode DAC. Cependant, le facteur des PM_{10} extérieurs semble avoir un impact plus marqué sur les concentrations de $PM_{2.5}$ que la température.

Les résultats obtenus pour la station Magenta rejoignent les conclusions observées à Sevran, mettant en évidence deux groupes principaux de facteurs influençant les concentrations de particules (PM_{10} et $PM_{2.5}$) : les facteurs temporels et les facteurs environnementaux. Si les dynamiques internes à la gare — notamment au niveau du quai — jouent un rôle prépondérant, les apports extérieurs ne sont pas négligeables et participent également à la variabilité des concentrations de particules dans les EFS.

5.2 Evaluation de l'efficacité des systèmes de filtration

Dans cette dernière section, nous nous concentrons sur l'évaluation de l'efficacité des systèmes de filtration installés dans les gares d'Avenue Foch et de Sevran Beaudottes, l'un des objectifs principaux de cette thèse par la SNCF. En complément des résultats prometteurs obtenus avec le modèle GAM, nous avons approfondi notre analyse en appliquant deux méthodes complémentaires : la méthode DAC et le modèle LSTM.

5.2.1 Evaluation des deux technologies de filtration à Avenue Foch

5.2.1.1 Evaluation avec la méthode DAC

Pour l'expérimentation 1 (filtration par ionisation), les données du DAC de l'année 2019 (lorsque l'expérimentation 1 a été menée à la station) ont été comparées à une moyenne des années 2018, 2020, 2021 et 2022. Pour l'expérimentation 2 (filtration par voie humide), réalisée en 2020, les données du DAC ont été comparées avec une moyenne des années 2018, 2019, 2021 et 2022. Cette approche par moyenne permet une comparaison robuste en atténuant les effets des variations annuelles et des perturbations atypiques, assurant ainsi une analyse plus cohérente et complète de l'efficacité du système de filtration.

La Figure 5.7 illustre l'évolution des ratios DAC : le panneau (a) montre la comparaison pour les PM_{10} en 2019, année de l'expérimentation 1, par rapport aux données moyennes de 2018, 2020, 2021 et 2022 ; la bande rouge indique la durée de l'expérimentation 1. Le panneau (b) présente la comparaison pour les PM_{10} en 2020, année de l'expérimentation 2, par rapport aux données moyennes de 2018, 2019, 2021 et 2022 ; la bande bleue représente la période de l'expérimentation 2. Les panneaux (c) et (d) fournissent des comparaisons similaires pour les $PM_{2.5}$. La ligne pointillée bleue représente le ratio DAC moyen des années utilisées comme référence pendant les périodes de filtration, tandis que la ligne pointillée rouge montre le ratio DAC moyen pour les années expérimentales.

Dans l'expérimentation 1, les ratios DAC des PM_{10} (Figure 5.7a) et des $PM_{2.5}$ (Figure 5.7c) montrent une diminution, passant d'environ $1{,}1 \pm 0.2$ à $0{,}8 \pm 0{,}1$ pour les PM_{10}, et de $1{,}1 \pm 0{,}2$ à $0{,}8 \pm 0{,}2$ pour les $PM_{2.5}$. Cela correspond à une réduction de 17,8 % ± 3,2 % pour les PM_{10} et de 25,9 % ± 6,6 % pour les $PM_{2.5}$ (Equations 3.3, 3.4 et 3.5). Ces résultats montrent que le système de filtration installé lors de l'expérimentation 1 a entraîné une réduction significative des concentrations de particules, et s'est révélé plus efficace pour filtrer les $PM_{2.5}$ que les PM_{10}. Pour les deux classes de particules, une diminution du ratio DAC est observée durant le premier mois de l'expérience. Cependant, entre août et

septembre—avant la fin de l'expérience—une augmentation du ratio DAC est notée. Ces observations pourraient être dues à la variabilité annuelle des concentrations de particules, à des interventions techniques sur le système de filtration, à la stabilisation de son efficacité, ou même à un artefact lié au nombre de points de données retenus pour l'analyse. En l'absence d'informations supplémentaires, et dans le cadre de cette méthode globale d'analyse de la performance du système, nous n'aborderons pas spécifiquement cet effet.

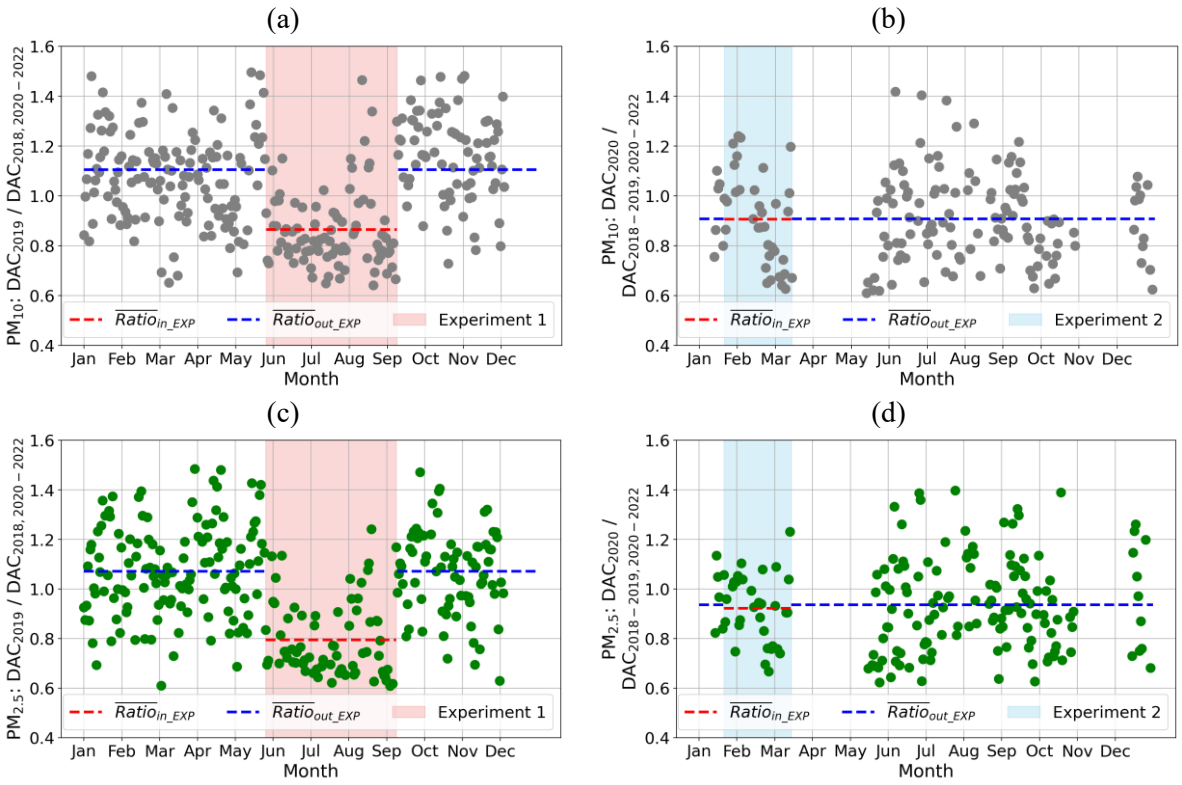


Figure 5.7 : Évolution des ratios DAC pour PM_{10} et $PM_{2.5}$: (a) PM_{10}, expérimentation 1 ; (b) PM_{10}, expérimentation 2 ; (c) $PM_{2.5}$, expérimentation 1 ; et (d) $PM_{2.5}$, expérimentation 2.

En revanche, les Figure 5.7 (b) et (d) ne révèlent aucun changement significatif durant l'expérimentation 2, avec un ratio DAC fluctuant autour de 1. Cela suggère que l'expérimentation 2 a eu peu d'effet observable, entraînant des réductions de seulement 0,4 % ± 0,1 % pour les PM_{10} et 3,7 % ± 1,0 % pour les $PM_{2.5}$. Il est important de noter que l'année 2020 a été particulièrement marquée par des grèves survenues juste avant l'expérience et le début de la pandémie de COVID-19 peu après. Ces événements ont limité les données disponibles pour l'analyse, rendant difficile toute conclusion définitive sur l'efficacité de l'Expérimentation 2.

5.2.1.2 Evaluation avec le modèle LSTM

Rappelons que le LSTM, contrairement à la méthode DAC et au modèle GAM, permet d'analyser uniquement les premiers jours, et non l'ensemble de la période de filtration. L'efficacité de la réduction à l'aide de cette méthode a été calculée en comparant les trois premiers jours de prédictions aux concentrations réelles de PM pendant cette période (fonctionnement des trains). Cependant, nous avons étendu la prévision à cinq jours pour obtenir une compréhension plus large des tendances de concentration de PM sur un cycle complet de jours de semaine. Cela a permis une comparaison plus complète des concentrations de PM prédites et observées, notamment en évaluant les variations dues aux activités en semaine et aux conditions environnementales.

La Figure 5.8 présente les concentrations réelles de PM (ligne bleue) et les concentrations de PM prédites par le modèle LSTM (ligne rouge) pour les PM_{10} : (a) expérimentation 1, (b) expérimentation

2 ; et pour les $PM_{2.5}$: (c) expérimentation 1, (d) expérimentation 2 pendant les cinq premiers jours de chaque expérimentation.

De manière générale, le modèle reproduit efficacement les profils quotidiens des concentrations de PM pour les deux fractions de particules dans les deux configurations, l'expérimentation 1 et l'expérimentation 2. Pour plus de clarté, la comparaison sur cinq jours correspond aux jours ouvrés (lundi à vendredi). Comme illustré dans le Chapitre 4, deux pics de concentrations de PM sont généralement observés pendant les heures de pointe du matin (10h00) et du soir (19h00), en plus d'une augmentation des niveaux de PM au début des opérations des trains et d'une diminution vers la fin. De plus, nous observons une tendance hebdomadaire similaire à celle trouvée avec le modèle GAM, où les concentrations de PM augmentent du lundi au mercredi, puis diminuent du mercredi au vendredi.

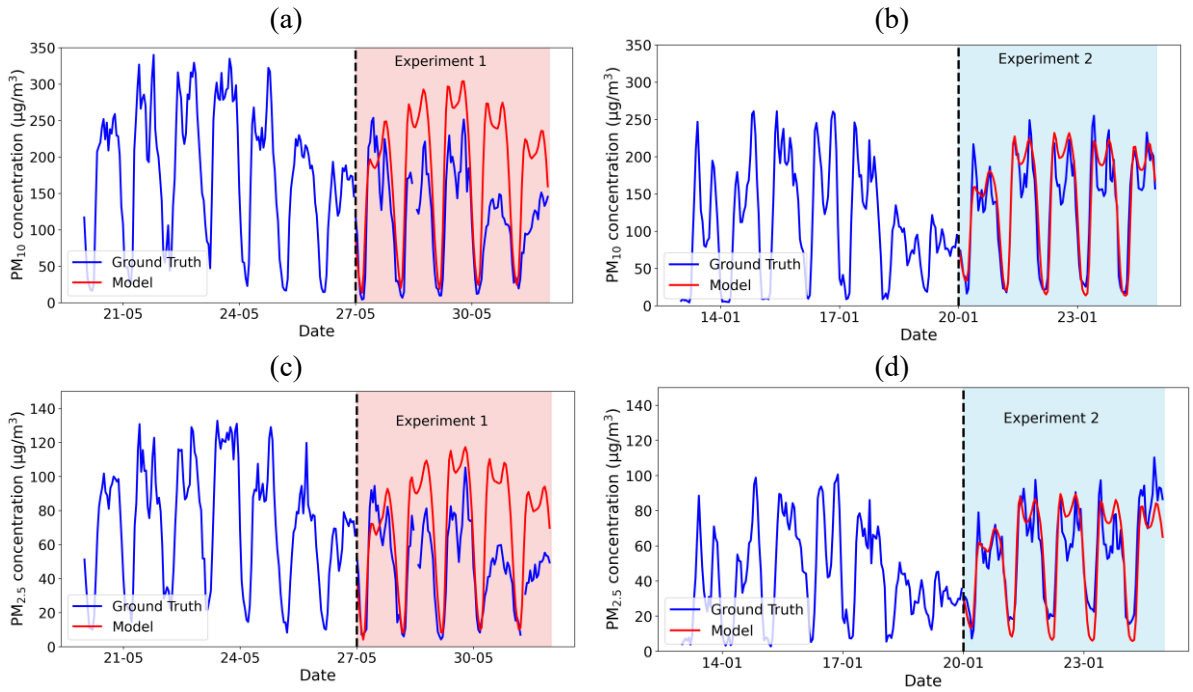


Figure 5.8 : Comparaison des concentrations mesurées de PM (ligne bleue) et prédites par le modèle LSTM (ligne rouge) : (a) PM_{10}, expérimentation 1 ; (b) PM_{10}, expérimentation 2 ; (c) $PM_{2.5}$, expérimentation 1 ; (d) $PM_{2.5}$, expérimentation 2.

Dans l'Expérimentation 1, une réduction significative des concentrations de PM_{10} et $PM_{2.5}$ est observée, notamment à partir du deuxième jour, avec des réductions de 26,8 % ± 6,1 % pour les PM_{10} et de 22,6 % ± 4,8 % pour le $PM_{2.5}$ pendant les trois premiers jours (équations 3.17 et 3.18). Le premier jour, il n'y a pas de différence entre les valeurs mesurées et prédites en raison de la phase d'activation initiale du système de filtration, qui peut nécessiter du temps pour atteindre une efficacité opérationnelle complète après l'installation (Figure 5.8a et Figure 5.8c).

En revanche, l'Expérimentation 2 montre peu de différence entre les mesures et les prédictions (Figure 5.8b), avec une réduction de 5,3 % ± 1,2 % sur les trois premiers jours. Pour les $PM_{2.5}$, la Figure 5.8d révèle des résultats comparables pendant la journée, avec seulement une diminution de 2,7 % ± 0,7 % pendant les trois premiers jours de l'expérience. La nuit, les prédictions du modèle LSTM sont systématiquement inférieures aux valeurs mesurées. Cet effet n'étant pas observé pour les PM_{10} (Figure 5.8b), les raisons de cette divergence ne sont pas encore identifiées.

Les résultats des trois méthodes, résumés dans le Tableau 5.13, montrent que l'expérimentation 1 est plus efficace que l'expérimentation 2 pour les deux classes de particules, les PM_{10} et les $PM_{2.5}$. Par ailleurs, l'expérimentation 1 capture les $PM_{2.5}$ plus efficacement que les PM_{10}.

La méthode DAC et le modèle GAM produisent des résultats comparables, tant pour les PM_{10} que pour les $PM_{2.5}$, dans les deux expérimentations. En comparaison, le modèle LSTM présente des

écarts significatifs : pour les PM_{10} lors de l'expérimentation 1, il indique une réduction d'environ 27 %, contre 15 % pour la méthode DAC et le modèle GAM. De même, pour les $PM_{2.5}$ lors de l'expérimentation 2, il suggère une réduction de 5 %, alors que les deux autres méthodes ne montrent aucune réduction (~0 %). Il est important de noter que le modèle LSTM se limite aux premiers jours de l'expérimentation et ne couvre pas l'ensemble de la période. On peut supposer que l'efficacité des systèmes utilisés évolue dans le temps.

Tableau 5.13 : Taux de réductions de PM_{10} et $PM_{2.5}$ (%) évalués par les trois méthodes

Méthode		DAC	GAM	LSTM
PM_{10}	Exp.1	17,8 ± 3,2	13,1 ± 0,8	26,8 ± 6,1
	Exp.2	0,4 ± 0,1	Non significatif	5,3 ± 1,2
$PM_{2.5}$	Exp.1	25,9 ± 6,6	26 ± 0,8	22,6 ± 4,8
	Exp.2	3,7 ± 1,0	4,6 ± 1,3	2,7 ± 0,7

5.2.2 Évaluation de la technologie de filtration à Sevran Beaudottes

La station Sevran Beaudottes a été choisie comme site clé pour tester l'efficacité d'une technologie de filtration mécanique, mise en œuvre de novembre 2020 à juin 2021. Cette section détaille les méthodes utilisées pour évaluer la réduction des concentrations de particules engendrée par cette technologie, les périodes d'expérimentation, et examine les résultats obtenus. Il est important de noter que, pour cette station, la méthode LSTM n'a pas pu être appliquée. En effet, la semaine précédant le début de l'expérimentation coïncidait avec une période de confinement, rendant impossible la prédiction d'une situation non perturbée. L'utilisation de la méthode LSTM nécessite en effet des conditions stables, non influencées par des événements majeurs, afin de prédire les concentrations en l'absence de perturbations expérimentales ou externes notables.

Pour Sevran Beaudottes, en complément des résultats obtenus par le modèle GAM, cette section présente également les résultats acquis en appliquant la méthode DAC.

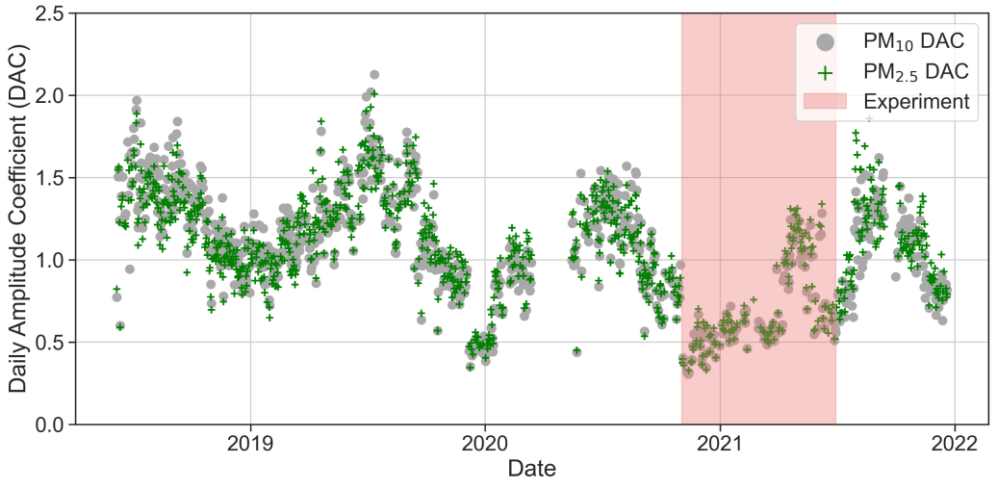


Figure 5.9 : Evolution temporelle des PM_{10} (points gris) et des PM2.5 (puces vertes)

La Figure 5.9 illustre l'évolution temporelle du DAC pour les particules PM_{10} et $PM_{2.5}$. Étant donné que la période d'expérimentation est suffisamment longue, il est également nécessaire d'avoir une période hors expérimentation tout aussi étendue pour obtenir des résultats plus représentatifs. Par conséquent, au lieu de comparer sur une seule année et de couvrir la période de novembre 2020 à juin 2021 pour cette gare, nous avons comparé la période de juin 2020 à décembre 2021 (période incluant l'expérimentation) avec celle de juin 2018 à décembre 2019.

La Figure 5.10 illustre l'évolution des ratios DAC : le panneau (a) compare les concentrations de PM_{10} entre la période 2020-2021, correspondant à l'expérimentation, et la période de référence 2018-2019. Le panneau (b) présente une comparaison similaire pour les $PM_{2.5}$, avec la bande rouge indiquant la durée de l'Expérimentation.

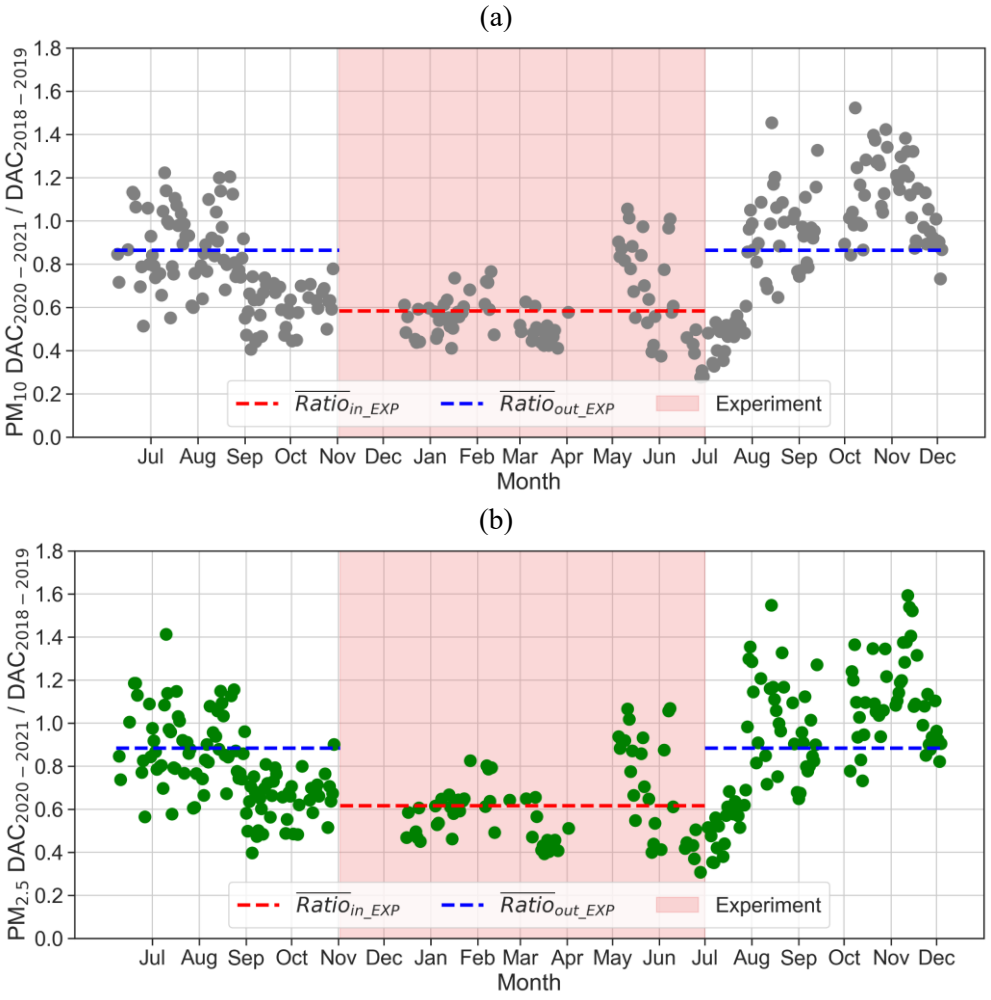


Figure 5.10 : Évolution des ratios DAC pour (a) PM_{10} et (b) $PM_{2.5}$ entre 2018-2019 et 2020-2021

L'analyse montre une diminution notable des ratios DAC pour les deux classes de particules durant les périodes d'expérimentation, mise en évidence par la bande rose. Pour les PM_{10}, le ratio DAC chute de 0,9 ± 0,3 à 0,6 ± 0,2, représentant une réduction de 32,4% ± 14%. Pour les $PM_{2.5}$, la réduction est du même ordre de grandeur, passant également de 0,9 ± 0,3 à 0,6 ± 0,2, ce qui équivaut à une réduction de 30,2% ± 13% (voir équations 3.3, 3.4 et 3.5).

Ces résultats montrent que le système de filtration installé à Sevran Beaudottes a entraîné une baisse significative des concentrations de particules. Il est notable que pour les deux tailles de particules, une augmentation du ratio DAC est observée avant la fin de l'expérimentation. Cette tendance pourrait être due à des interruptions de fonctionnement du système installé, qui n'opérait pas de manière continue dans la gare, et pourrait correspondre à des arrêts temporaires du système. En l'absence d'informations supplémentaires, en raison de la confidentialité entourant les détails du système installé dans cette gare, et dans le cadre de cette analyse globale de la performance du système, cet effet particulier ne sera pas discuté en détail. On observe également une réduction du ratio durant les mois de septembre et octobre 2020, périodes précédant l'expérimentation de filtration pour les deux classes de particules. Cette réduction peut être attribuée à des travaux de maintenance des voies effectués sur la ligne B, qui relie notamment les gares d'Aulnay-sous-Bois et de l'Aéroport Charles de Gaulle 2. Ces travaux, menés du 14 septembre au 11 décembre 2020, ont entraîné des interruptions fréquentes du trafic, y compris à la

gare de Sevran Beaudottes. Par ailleurs, il est intéressant de noter que le ratio reste faible même après l'expérimentation, en particulier durant le mois de juillet 2021. Cette observation peut également être liée à des travaux de maintenance des voies sur le même tronçon, qui ont eu lieu du 1er juillet au 27 août 2021, bien que d'autres facteurs locaux puissent contribuer à cette tendance.

Le Tableau 5.14 présente les réductions des concentrations de particules PM_{10} et $PM_{2.5}$ obtenues grâce aux méthodes DAC et GAM.

Tableau 5.14 : Taux de réductions de PM_{10} et $PM_{2.5}$ (%) évalués par les deux méthodes : DAC et GAM

Méthode	DAC	GAM
PM_{10}	32,4 ± 14	34,4 ± 0,8
$PM_{2.5}$	30,2 ± 13	33,2 ± 0,8

Les résultats obtenus par les deux méthodes sont comparables. Pour les PM_{10}, la méthode DAC a enregistré une réduction de 32,4 % avec une variabilité de ± 14 %, tandis que le modèle GAM a indiqué une réduction de 34,4 % avec une variabilité bien moindre de ± 0,8 %. Pour les $PM_{2.5}$, les réductions observées sont également proches, avec 30,2 % (± 13,1 %) pour la méthode DAC contre 33,2 % (± 0,8 %) pour le modèle GAM. La grande différence d'incertitude entre les deux méthodes peut s'expliquer par leurs approches fondamentalement différentes. La méthode DAC reflète la variabilité brute des données, ce qui entraîne une incertitude élevée. En revanche, le modèle GAM isole l'effet du système de filtration en tenant compte de plusieurs variables explicatives, telles que la température et le jour de la semaine. Cette approche permet de lisser les variations aléatoires et de réduire le bruit dans les données, ce qui se traduit par une incertitude bien plus faible. Cette divergence montre que le modèle GAM est plus robuste que la méthode DAC.

Malgré ces différences, les deux méthodes démontrent une efficacité similaire dans la capture des particules, avec des résultats très proches. Elles indiquent toutes deux que le système installé à Sevran Beaudottes a été aussi efficace pour capturer les PM_{10} que les $PM_{2.5}$. Cela contraste avec les observations à la gare d'Avenue Foch où les particules $PM_{2.5}$ étaient filtrées plus efficacement. Cette disparité souligne des variations potentielles dans l'efficacité de filtration selon le type de système de filtration.

5.3 Conclusion du chapitre

Ce chapitre a fourni une analyse détaillée des facteurs influençant les concentrations de particules dans les trois gares de référence, en se basant sur les données disponibles, et a évalué l'efficacité des systèmes de filtration testés dans les gares d'Avenue Foch et de Sevran Beaudottes.

L'évaluation par le modèle GAM a révélé que, pour les trois gares, le facteur le plus influent est l'heure de la journée, reflétant la fréquence des trains, suivi du numéro de la semaine, qui indique des variations saisonnières. Dans la gare d'Avenue Foch, le CO_2 s'est également avéré être un facteur clé, reflétant principalement le volume de voyageurs présents sur les quais. Les paramètres de confort, tels que la température et l'humidité relative, ont également joué un rôle notable, bien que leur impact soit moindre par rapport aux facteurs mentionnés précédemment. En revanche, les facteurs extérieurs, tels que les concentrations de PM, NO_2, NO et O_3, ont montré une influence très limitée sur les concentrations de particules dans les gares, à l'exception de la gare de Magenta, dotée d'un système de ventilation. Cela suggère que les sources principales de particules proviennent de l'intérieur des gares, et que l'environnement souterrain limite l'infiltration des polluants extérieurs, réduisant ainsi leur impact.

L'évaluation des systèmes de filtration a démontré que les interventions menées à Avenue Foch et à Sevran Beaudottes ont permis de réduire efficacement les concentrations de PM_{10} et de $PM_{2.5}$, bien que les performances varient en fonction de la technologie de filtration utilisée. À Avenue Foch, les

résultats des trois méthodes confirment que l'expérimentation 1, utilisant une technologie d'ionisation positive, a été plus efficace pour réduire les concentrations de PM_{10} et de $PM_{2.5}$ que l'expérimentation 2, basée sur une technologie à eau. Les méthodes DAC et GAM ont indiqué que les $PM_{2.5}$ étaient plus efficacement capturées dans les deux expériences. Cependant, le modèle LSTM a montré une réduction plus marquée des PM_{10} par rapport aux $PM_{2.5}$, mettant en évidence sa sensibilité aux fluctuations à court terme des particules plus grossières. À Sevran Beaudottes, les méthodes DAC et GAM ont montré que le système installé a été aussi efficace pour capturer les PM_{10} que les $PM_{2.5}$. Cette disparité souligne des variations potentielles dans l'efficacité de filtration selon le type de système de filtration. Il est important de noter que, tant à Avenue Foch qu'à Sevran Beaudottes, ces résultats sont localisés au point de mesure et ne reflètent pas nécessairement les conditions sur l'ensemble du quai, étant donné la proximité des systèmes de filtration avec l'emplacement du capteur de particules.

Concernant les trois méthodes d'analyse employées, la méthode DAC est recommandée pour l'analyse des tendances à long terme, en particulier lorsque la cohérence des données et l'absence de perturbations externes sont assurées. Sa capacité à condenser les séries temporelles en indicateurs quotidiens facilement interprétables en fait un outil précieux pour comprendre les tendances générales sur des périodes étendues. La méthode GAM est également adaptée à l'analyse à long terme. Elle offre des résultats statistiquement robustes en prenant en compte de multiples facteurs d'influence, tels que le jour de la semaine et les effets saisonniers. Sa flexibilité pour modéliser des relations complexes et non linéaires la rend particulièrement utile pour isoler l'impact des conditions expérimentales. Cependant, comparée à la méthode DAC, elle nécessite davantage de données et de variables explicatives pour fonctionner de manière optimale. Enfin, le modèle LSTM est idéal pour le monitoring en temps réel et l'évaluation à court terme de la performance des systèmes de filtration. Il permet de détecter rapidement les variations des concentrations de particules, offrant un retour immédiat sur l'efficacité de la filtration. Cependant, il présente certaines limites pour les prédictions à long terme en raison d'une incertitude croissante sur des horizons temporels étendus et il nécessite une année de données préalables pour être efficace.

Ce chapitre a non seulement enrichi notre compréhension des dynamiques de pollution dans les EFS, mais il a également ouvert des perspectives sur la manière d'évaluer efficacement l'efficacité des systèmes de filtration dans ces environnements. Les résultats obtenus ont souligné l'impact des technologies de filtration dans la réduction des concentrations de particules, tout en mettant en lumière les avantages et les limites des différentes méthodes d'analyse (DAC, GAM, LSTM). Ces enseignements constituent une base solide pour formuler des recommandations sur la gestion de la qualité de l'air dans les gares souterraines et pour planifier de futures mesures de filtration adaptées aux spécificités de chaque site.

CONCLUSION GÉNÉRALE ET PERSPECTIVES POUR DE FUTURS TRAVAUX

Ce travail sur la qualité de l'air dans les enceintes ferroviaires souterraines (EFS) met en lumière les nombreux phénomènes qui déterminent l'évolution des concentrations de particules en suspension. Selon la littérature, les opérations ferroviaires, comme la fréquence des trains et la remise en suspension des particules principalement par effet piston constituent les principales sources internes contribuant aux niveaux de pollution. Ces effets sont plus ou moins forts en fonction de la profondeur des stations, la structure architecturale et l'âge de la station. Certaines études (Martins et al., 2015; Van Ryswyk et al., 2017; Luo et al., 2021a; Gokul et al., 2023) basées sur de longues périodes d'observations montrent aussi un effet de la saisonnalité sur les niveaux de pollution, et montrent généralement une augmentation des PM en hiver, probablement en lien avec les niveaux de pollution particulaire observés en extérieur.

Face aux enjeux sanitaires liés à la qualité de l'air dans les EFS, deux grandes catégories de solutions ont été identifiées : les solutions passives et les solutions actives, selon leur consommation énergétique. Les premières, comme l'installation de portes palières ou l'utilisation de matériaux non métalliques dans les systèmes de freinage, visent à réduire la production de particules à la source sans nécessiter d'énergie supplémentaire. Les secondes, telles que les systèmes de ventilation et les équipements de purification de l'air, requièrent une consommation d'énergie pour traiter activement l'air et diminuer les niveaux de pollution. Les observations soulignent l'importance d'une conception et d'une gestion adaptées des infrastructures et des systèmes de ventilation. Elles mettent également en avant la nécessité d'adopter des solutions innovantes, comme l'optimisation des matériaux de freinage ou l'installation de portes palières, pour améliorer durablement la qualité de l'air dans ces environnements confinés.

Ce travail de thèse a exploré en profondeur les dynamiques de pollution particulaire dans les EFS franciliennes du réseau SNCF, en combinant des analyses long et court termes des concentrations en particules (PM_{10} et $PM_{2.5}$), une évaluation des facteurs influençant leur variabilité, ainsi qu'une étude de l'efficacité des systèmes de filtration en se basant sur des campagnes de mesure menées dans 14 gares souterraines du réseau SNCF, sur des périodes allant de quelques semaines à plusieurs années dans les gares d'Avenue Foch, de Sevran-Beaudottes et de Magenta. La base de données ainsi constituée est très riche et permet de documenter les concentrations en particules et leur évolution temporelle, facilitant ainsi la mise en œuvre de plans d'amélioration et la priorisation des interventions. Les paramètres sont mesurés en un point fixe sur les quais de gare, et incluent les concentrations de particules, mesurées à l'aide d'un analyseur automatique de particules TEOM ; les concentrations de CO_2, mesurées par un analyseur à absorption infrarouge ; ainsi que les paramètres de confort thermique (température et humidité). Les périodes de mesure couvrent le fonctionnement normal des EFS, ainsi que trois campagnes de test de systèmes de filtration de particules : l'ionisation positive et la filtration par voie humide, testées à la gare d'Avenue Foch, et la filtration classique, testée à la gare de Sevran Beaudottes.

Pour analyser ces données, trois approches méthodologiques ont été utilisées : la méthode originale DAC (Coefficient d'Amplitude Journalier), développée dans le cadre de cette thèse, le modèle GAM (Modèle Additif Généralisé) et le modèle LSTM (Long Short-Term Memory). Chacune de ces méthodes apporte une perspective unique et permet une évaluation approfondie des dynamiques des niveaux de pollution et des performances des systèmes de filtration. La méthode DAC offre une vision claire des tendances à court et long terme, des profils journaliers des concentrations de particules, et permet des comparaisons entre différentes gares du même réseau. Elle quantifie les variations inter-journalières grâce à une métrique simple et efficace. En comparant différentes périodes, le DAC capture les tendances saisonnières et temporelles de la réduction des particules lorsque l'on compare des périodes avec et sans système de filtration. Cependant, son efficacité repose sur la disponibilité de

données continues et non perturbées. Le modèle GAM, quant à lui, offre une approche plus flexible et statistiquement robuste. Il permet de modéliser des relations non linéaires, capturant ainsi les interactions complexes entre les concentrations de particules et les variables explicatives, tout en isolant les effets de chacune de ces variables. Cette méthode fournit une interprétation détaillée des influences de chaque variable, ce qui en fait un outil puissant pour comprendre et estimer les contributions des facteurs influençant la qualité de l'air. Cependant, le modèle GAM présente aussi des défis, comme l'évaluation des tendances de réduction sur toute la période d'expérimentation et la disponibilité des variables explicatives nécessaires, telles que la fréquence des trains ou la fréquentation des voyageurs. Le modèle LSTM fournit des résultats précieux pour le monitoring en temps réel. En tant qu'approche avancée d'apprentissage automatique, il est particulièrement efficace pour prédire à court terme les événements de forte pollution dans les EFS. Cependant, au fur et à mesure que l'horizon de prévision s'allonge, les incertitudes augmentent, ce qui rend les prévisions à long terme moins fiables.

L'application de la méthode DAC a permis de mettre en évidence des tendances saisonnières et journalières spécifiques dans les concentrations de particules au sein des EFS franciliennes. Les variations à long terme des concentrations de particules (PM_{10} et $PM_{2.5}$) ont révélé des niveaux plus élevés en été et plus faibles en hiver, une tendance qui diffère de celles observées dans la littérature tant pour les concentrations de particules en EFS que les concentrations en air extérieur. Cette divergence peut être attribuée à des disparités structurelles majeures entre les stations étudiées et les stations de métro traitées dans la littérature dans d'autres pays, notamment la présence ou l'absence de systèmes de climatisation, ainsi que l'utilisation d'un pantographe caténaire pour l'alimentation électrique des trains qui est une source de particules supplémentaire en raison des frottements et des arcs électriques. Les concentrations élevées observées durant la période estivale pourraient être partiellement expliquées par la forte fréquentation touristique à Paris favorisant la remise en suspension des particules déjà présentes sur les quais. En été, la température sur les quais, similaire à celle de l'extérieur, favorise une usure accrue des freins et des rails, ce qui explique les concentrations plus élevées de particules observées. En hiver, les températures plus basses de 15°C en moyenne par rapport aux saisons estivales tendent à réduire ainsi l'usure des composants et les émissions de particules associées, ce qui correspond à nos observations.

Nos résultats pour les gares d'Avenue Foch et de Sevran-Beaudottes ont montré que les concentrations en particules ne présentent pas de variation interannuelle significative. En revanche, à la gare de Magenta, les concentrations de PM_{10} et $PM_{2.5}$ suivent une évolution annuelle différente, marquée par une augmentation progressive tout au long de la période de mesure. Cette tendance pourrait s'expliquer par le démarrage des travaux EOLE (Est-Ouest Liaison Express) sur la ligne E à partir de février 2018, travaux qui se déroulent chaque été. En plus de l'impact des travaux, une autre explication possible de cette tendance pourrait être liée à la détérioration de l'efficacité du système de ventilation de la gare. Comme tout système mécanique, les systèmes de ventilation peuvent subir une usure au fil du temps. Sans maintenance adéquate, leur efficacité peut diminuer, ce qui entraînerait une augmentation des concentrations en particules.

Les profils journaliers révèlent des variations liées à la fréquence des trains : deux pics de concentration pendant les jours ouvrés, associés aux heures de pointe, et des concentrations plus faibles et constantes durant les weekends, avec une légère augmentation en journée. Ces profils sont décalés d'une heure par rapport au début du trafic. Ce délai s'explique principalement par la dispersion des particules due à la resuspension causée par l'intensité du trafic et l'affluence des passagers. L'étude du rapport $PM_{2.5}/PM_{10}$ associé aux variations diurnes montre des valeurs plus faibles le jour et plus élevées la nuit. Cette tendance s'explique principalement par l'arrêt des activités ferroviaires à minuit, qui réduit les sources de PM_{10}, associé au dépôt plus lent des $PM_{2.5}$ par rapport aux PM_{10} (Walther et al., 2017).

La méthode DAC nous a permis de classer les gares suivant les niveaux de pollution et de réaliser une analyse comparative pour identifier qualitativement certains des facteurs influençant les concentrations de particules. Les résultats ont mis en évidence l'impact dominant des activités

ferroviaires sur les pics de pollution, ainsi que des différences significatives entre les concentrations intérieures et extérieures. L'analyse approfondie des facteurs influençant les variations des concentrations de particules a été réalisée à l'aide du modèle GAM. Pour les trois gares dotées de mesures de longue durée, le facteur le plus influent est l'heure de la journée, qui correspond à la fréquence des trains, suivi du numéro de la semaine, indiquant des variations saisonnières. À la gare d'Avenue Foch, le CO_2 s'est également révélé être un facteur clé, principalement en lien avec le volume de voyageurs présents sur les quais. Les paramètres de confort, tels que la température et l'humidité relative, ont aussi joué un rôle notable. En revanche, les facteurs extérieurs, tels que les concentrations de PM, NO_2, NO et O_3, ont eu une influence limitée sur les concentrations de particules dans les gares. Cela suggère que les sources de pollution intérieures dominent largement, rendant l'apport extérieur relativement marginal. Cependant, à la gare de Magenta, les facteurs extérieurs se sont révélés plus importants, probablement en raison du système de ventilation mécanique entraînant de l'air extérieur au niveau du quai.

L'évaluation des systèmes de filtration a été réalisée en utilisant les trois méthodes (DAC, GAM et LSTM). Les résultats ont démontré que les expérimentations testées à Avenue Foch et à Sevran Beaudottes ont permis de réduire efficacement les concentrations de PM_{10} et $PM_{2.5}$, bien que les performances varient selon la technologie de filtration utilisée. À Avenue Foch, les résultats des trois méthodes ont confirmé que l'expérimentation 1, utilisant une technologie d'ionisation positive, a été plus efficace pour réduire les concentrations de PM_{10} et de $PM_{2.5}$ que l'expérimentation 2, qui reposait sur une technologie à eau. Les méthodes DAC et GAM, se basant sur l'intégralité de la période de filtration, ont montré que les $PM_{2.5}$ étaient plus efficacement capturées que les PM_{10} par les deux systèmes d'épuration testés. Cependant, le modèle LSTM a révélé une réduction plus marquée des PM_{10} par rapport aux $PM_{2.5}$ sur les trois premiers jours de la période de filtration. Bien que ces résultats puissent sembler contradictoires, ils soulignent la nécessité d'une meilleure compréhension de la dynamique temporelle des systèmes de filtration pour chaque fraction de particules. À Sevran Beaudottes, les méthodes DAC et GAM ont montré que le système de dépollution par filtre installé a été aussi efficace pour capturer les PM_{10} que les $PM_{2.5}$. Cette disparité met en évidence des variations possibles dans l'efficacité des systèmes de filtration selon le type de technologie utilisée. Il est important de noter que, tant à Avenue Foch qu'à Sevran Beaudottes, ces résultats sont localisés au point de mesure et ne reflètent pas nécessairement les conditions sur l'ensemble du quai, en raison de la proximité des systèmes de filtration avec l'emplacement de la baie de mesure.

Ce travail de thèse a non seulement enrichi notre compréhension de la dynamique des niveaux de PM dans les EFS franciliennes, mais il a également ouvert des perspectives pour l'évaluation des systèmes de filtration. Les méthodes développées, alliant robustesse, flexibilité et adaptabilité, constituent une avancée significative pour la gestion de la qualité de l'air dans les environnements ferroviaires souterrains. Les résultats obtenus offrent une base solide pour formuler des recommandations opérationnelles et pour guider les futures recherches dans ce domaine. Enfin, cette étude souligne l'importance d'une approche intégrée, combinant surveillance continue, modélisation avancée et expérimentations sur le terrain, pour relever les défis complexes de la pollution de l'air dans les espaces confinés.

Recommandations issues des résultats

Les résultats obtenus dans cette thèse permettent de formuler quelques recommandations pratiques visant à mieux analyser et suivre la qualité de l'air dans les EFS. Ces recommandations touchent à divers aspects, notamment la collecte des données, les méthodes d'analyse, l'optimisation des mesures et l'amélioration des systèmes de filtration et de ventilation.

Tout d'abord, concernant la collecte et l'accès aux données, il est recommandé de prolonger la durée des campagnes de mesure. Bien que les campagnes de 2 à 3 semaines déjà réalisées par la SNCF aient fourni des résultats significatifs, une durée de six mois serait préférable pour obtenir des profils de

concentration plus précis. L'analyse de l'incertitude des profils obtenus à l'aide de la méthode DAC, en fonction de la durée des mesures, a montré qu'une semaine de données suffit pour obtenir un profil type avec une incertitude inférieure à 10 %. Toutefois, une période d'échantillonnage prolongée permettrait d'améliorer la précision des profils obtenus. Cette approche contribuerait à réduire l'incertitude des analyses et offrirait une meilleure compréhension des variations à long terme des concentrations de particules, un aspect encore mal documenté dans les EFS franciliennes.

Faciliter l'accès à des informations détaillées, telles que le fonctionnement des gares, la configuration des tunnels, le comptage des voyageurs, la maintenance des systèmes de ventilation dans les gares équipées ou les conditions extérieures, est essentiel pour une analyse plus approfondie. La mise à disposition d'une base de données complète sur l'environnement local, incluant des éléments comme la connectivité des tunnels, les travaux à proximité ou les conditions météorologiques, permettrait une meilleure compréhension de l'impact de ces facteurs sur la qualité de l'air dans les EFS. Cela offrirait également la possibilité d'adapter les stratégies en fonction de l'évolution de l'environnement local.

En ce qui concerne les méthodes d'analyse, il est recommandé de combiner les approches DAC, GAM et LSTM pour une évaluation plus complète. Cette combinaison fournit une vision globale et précise des dynamiques de pollution, tout en permettant un monitoring en temps réel à différents horizons temporels. Par ailleurs, tout au long de ce travail de thèse, l'analyse s'est concentrée sur des profils journaliers allant de 00h00 à 23h45. Cependant, la fréquence des trains varie significativement entre les jours ouvrés et les weekends, avec des circulations pouvant se prolonger jusqu'à 01h00 du matin. Pour mieux capturer la dynamique des concentrations de particules en fonction de l'activité ferroviaire, il serait plus judicieux d'étendre la plage horaire d'analyse de 05h00 à 04h45 du lendemain. Cette approche permettrait de mieux refléter les variations liées aux horaires de circulation des trains.

Pour les méthodes de mesure, les concentrations mesurées ont montré qu'on ne peut pas utiliser en l'état les protocoles utilisés en air extérieur. Il est essentiel d'adapter les dispositifs de mesure à ces niveaux élevés de concentration. Cela implique des calibrations régulières des appareils de mesure et une maintenance rigoureuse pour assurer leur précision à long terme. La variabilité temporelle des concentrations de particules a été étudiée en détail, mais la variabilité spatiale dans les EFS reste peu explorée. Initialement, une campagne de mesures utilisant une flotte de micro-capteurs répartis sur l'ensemble du quai était prévue pour 2024. Cependant, en raison de problèmes techniques, cette campagne a été reportée. L'utilisation de capteurs permettrait une couverture complète des concentrations de particules sur l'intégralité du quai, voire dans les tunnels. Une telle configuration serait particulièrement pertinente pour étudier la dynamique des fluides dans les espaces souterrains, en tenant compte des phénomènes de turbulences et de circulation de l'air. Ces facteurs influencent la distribution des particules en suspension, créant des zones de concentration variables. Par ailleurs, l'installation de capteurs extérieurs aux entrées et sorties des gares pour mesurer les concentrations de PM, CO_2 et autres polluants fournirait des informations précieuses pour comprendre l'impact de l'environnement extérieur sur la qualité de l'air dans les EFS.

En ce qui concerne les systèmes de filtration, il est essentiel de collecter des informations sur leur fonctionnement afin d'évaluer leur performance en continu. Cela permettrait d'optimiser leur efficacité et de compléter les travaux réalisés dans cette thèse. Par ailleurs, il serait pertinent d'étudier les propriétés physico-chimiques des particules collectées par ces systèmes ainsi que celles encore présentes dans les EFS. Cette étude permettrait de mieux comprendre la dynamique temporelle de la filtration des particules et d'évaluer l'efficacité des systèmes en fonction des tailles de particules les plus efficacement capturées.

Perspectives pour de futurs travaux

Les résultats de cette étude ouvrent également des perspectives prometteuses pour de futurs travaux de recherche et d'innovation. Ces axes de recherche visent à approfondir les connaissances et à développer de nouvelles solutions pour améliorer la qualité de l'air dans les EFS.

Tout d'abord, il serait pertinent d'analyser l'origine des outliers, des profils atypiques et des événements exceptionnels, tels que les pandémies ou les grèves, qui ont été écartés lors du prétraitement des données. Une telle analyse permettrait de généraliser la méthode développée dans cette thèse en améliorant sa capacité à distinguer les simples outliers des profils atypiques liés à des événements exceptionnels, comme les travaux à proximité de la gare. Cette méthode pourrait être également étendue à l'analyse d'autres polluants ou paramètres, tels que la température et l'humidité. Elle pourrait également être appliquée à d'autres domaines, comme la météorologie, l'économie ou l'analyse des flux de trafic, dès lors que ceux-ci présentent des profils journaliers similaires entre les jours ouvrés (lundi à vendredi) et les weekends (samedi et dimanche). La création d'une bibliothèque Python dédiée pourrait faciliter l'élargissement de son application à ces domaines, en rendant la méthode plus accessible et reproductible pour d'autres chercheurs et professionnels.

La modélisation et la surveillance avancée constituent également des pistes prometteuses. Le modèle LSTM univarié introduit dans cette étude pourrait être amélioré pour renforcer la précision des prévisions à court terme et servir d'outil d'aide à la décision. Par exemple, un modèle LSTM multivarié, comme celui développé par Rahmani et al. (2024) dans leurs travaux sur la qualité de l'air extérieur, pourrait être adapté et appliqué aux EFS. Cette approche permettrait d'intégrer des variables explicatives telles que la fréquence des trains, la température ou la fréquentation des voyageurs, améliorant ainsi les performances prédictives du modèle. Un tel modèle pourrait évaluer les niveaux de pollution pour les heures à venir. Cette application permettrait aux usagers d'adapter leurs modes de transport ou leurs horaires pour éviter une exposition courte à des niveaux de pollution élevés.

Une analyse approfondie de la taille, de la composition chimique et de la morphologie des particules pourrait fournir des informations cruciales pour mieux comprendre la pollution particulaire spécifique aux environnements souterrains. Cette caractérisation permettrait d'identifier les sources principales de particules (par exemple, les métaux issus des systèmes de freinage ou des pantographes caténaires) et d'évaluer leur contribution relative aux concentrations observées. Des campagnes de mesures saisonnières pourraient être menées pour mieux comprendre les différences dans les concentrations et les propriétés des particules entre l'été et l'hiver. Par exemple, l'utilisation de méthodes par impaction permettrait de comparer la distribution des particules entre les deux saisons et d'évaluer si les métaux issus des activités ferroviaires sont plus abondants en été, ainsi que la fraction de particules sur laquelle ils se concentrent. Le ratio $PM_{2.5}/PM_{10}$ observé dans les EFS franciliennes diffère significativement de celui mesuré dans d'autres réseaux de métro à travers le monde. Une étude comparative pourrait être menée pour identifier les facteurs expliquant cette divergence, tels que les différences dans les matériaux utilisés, les systèmes de ventilation, les fréquences de trains ou les pratiques de maintenance.

Une autre perspective prometteuse consisterait à étendre l'application des méthodes utilisées dans ce travail de thèse à d'autres gares, notamment celles gérées par la RATP, qui présentent des configurations architecturales, des niveaux de fréquentation et des caractéristiques techniques distincts. Une analyse comparative sur un large éventail de gares permettrait de mieux comprendre la diversité des résultats obtenus, en identifiant les facteurs influençant les concentrations de particules et en adaptant les méthodes aux spécificités locales. De plus, des études comparatives internationales pourraient être menées pour identifier les facteurs d'influence individuellement et recenser les meilleures pratiques, afin de les adapter aux contextes locaux. Cette approche permettrait non seulement de valider la généralisation des résultats obtenus, mais aussi d'enrichir les connaissances sur la qualité de l'air dans les EFS franciliennes.

Enfin, la collaboration et le partage des connaissances sont essentiels pour accélérer les progrès dans ce domaine. La création de bases de données ouvertes faciliterait les recherches futures et les comparaisons internationales.

BIBLIOGRAPHIE

Aarnio, P., Yli-Tuomi, T., Kousa, A., Mäkelä, T., Hirsikko, A., Hämeri, K., Räisänen, M., Hillamo, R., Koskentalo, T., & Jantunen, M. (2005). The concentrations and composition of and exposure to fine particles (PM2.5) in the Helsinki subway system. *Atmospheric Environment*, *39*(28), 5059-5066. https://doi.org/10.1016/j.atmosenv.2005.05.012

AEF. (2021a). *Mesure en continu de la qualité de l'air en gare de Magenta - 2021* (Rapport interne N° DOC058559-00 / MES022959). AEF. https://eqair.sncf.fr/

AEF. (2021b). *Site de mesures de particules en continu en gare de Magenta - 2020* (Rapport interne N° DOC056443-00 / MES018933). AEF. https://eqair.sncf.fr/

AEF. (2021c). *Site de mesures de particules en continu en gare de Sevran-Beaudottes - 2020* (Rapport interne N° DOC056444-00 / MES019200). AEF. https://eqair.sncf.fr/

AEF. (2022a). *Caractérisation des particules dans l'air des enceintes ferroviaires souterraines* (Note de synthèse N° DOC059277-00 / MES020522). AEF. https://eqair.sncf.fr/

AEF. (2022b). *Mesure en continu de la qualité de l'air en gare de Magenta - 2022* (Rapport interne N° DOC061271-00 / MES022659). AEF. https://eqair.sncf.fr/

AEF. (2022c). *Mesure en continu de la qualité de l'air en gare de Sevran-Beaudottes - 2022* (Rapport interne N° DOC061272-00 / MES022960). AEF. https://eqair.sncf.fr/

Afshari, A., Ekberg, L., Forejt, L., Mo, J., Rahimi, S., Siegel, J., Chen, W., Wargocki, P., Zurami, S., & Zhang, J. (2020). Electrostatic Precipitators as an Indoor Air Cleaner—A Literature Review. *Sustainability*, *12*(21), Article 21. https://doi.org/10.3390/su12218774

Agarwal, D., Trinh, X. T., & Takeuchi, W. (2025). Assessing the Air Quality Impact of Train Operation at Tokyo Metro Shibuya Station from Portable Sensor Data. *Remote Sensing*, *17*(2), Article 2. https://doi.org/10.3390/rs17020235

AIRLAB. (2022). *Expérimentations « Innovons pour l'air de nos stations » synthèse et avis d'AIRLAB* [Rapport]. https://airlab.solutions/projets/innovons-pour-lair-de-nos-stations-92

AIRPARIF. (2017). *Mesures de la Qualité de l'Air Intérieur sur les Quais du RER C en Gare SNCF d'Austerlitz* (p. 41) [Rapport]. https://www.airparif.fr/etudes/2017/mesure-qualite-de-lair-efs-rer-c-sncf-austerlitz-2016

AIRPARIF. (2018). *Traitement de données - Mesure en EFS CIL SNCF* (Note de synthèse NS-040).

AIRPARIF. (2022a). *Recherche des facteurs d'influence des concentrations en polluants mesurés dans les gares du réseau Transilien* (p. 71) [Rapport interne]. AIRPARIF.

AIRPARIF. (2022b). *Mesures de la qualité de l'air intérieur sur les quais des gares souterraines franciliennes de la SNCF* [Rapport final]. AIRPARIF. https://www.airparif.asso.fr/etudes/2021/qualite-de-lair-interieur-quais-gares-sncf-rapport-final-2020

AIRPARIF. (2024). *Cartographie des niveaux de particules sur les quais des gares et stations franciliennes souterraines - 2024* [Résumé]. https://www.airparif.fr/dossiers-fiches-thematiques/2024/cartographie-de-la-qualite-de-lair-dans-les-metros-rer-et

Álvarez, V. M., Figueroa-Lara, J., Romero-Romo, M., Sepúlveda-Sánchez, J., & López-Moreno, T. (2012). Concentrations and properties of airborne particles in the Mexico City subway system. *Atmospheric Environment*, *49*, 284-293. https://doi.org/10.1016/j.atmosenv.2011.11.038

Alves, C. A., Moreno, T., Martins, V., Minguillón, M. C., Querol, X., Mendes, L., Eleftheriadis, K., & Miguel, E. D. (2017). Characterisation of airborne particulate matter in different european subway systems. In H. Yaghoubi (Éd.), *Urban Transport Systems*. InTechOpen. https://doi.org/10.5772/65364

ANSES. (2015). *Pollution chimique de l'air dans les enceintes ferroviaires souterraines et risques pour la santé des travailleurs* (Saisine Nº 2011-SA-0265; Rapport de l'Agence nationale de sécurité sanitaire, de l'alimentation, de l'environnement et du travail, p. 361). ANSES. https://www.anses.fr/fr/system/files/AIR2011sa0265Ra.pdf

ANSES. (2022). *Qualité de l'air ambiant des enceintes ferroviaires souterraines (EFS). Etat des connaissances sur la toxicité des particules et sur les effets sanitaires associés à la pollution de l'air des EFS. Recommandation de concentrations en particules dans l'air des EFS à ne pas dépasser.* (Saisine Nº 2019-SA-0148; Rapport de l'Agence nationale de sécurité sanitaire, de l'alimentation, de l'environnement et du travail, p. 181). ANSES. https://www.anses.fr/fr/system/files/AIR2019SA0148Ra.pdf

Barmparesos, N., Assimakopoulos, V. D., Assimakopoulos, M. N., & Tsairidi, E. (2016). Particulate matter levels and comfort conditions in the trains and platforms of the Athens underground metro. *AIMS Environmental Science*, *3*(2), 199-219. https://doi.org/10.3934/environsci.2016.2.199

Belušić, A., Herceg-Bulić, I., & Klaić, Z. B. (2015). Using a generalized additive model to quantify the influence of local meteorology on air quality in Zagreb. *Geofizika*, *32*(1), Article 1. https://doi.org/10.15233/gfz.2015.32.5

Ben Rayana, T., Hemmendinger, M., Crézé, C., Wild, P., Sauvain, J.-J., Suarez, G., Besançon, S., Méthy, N., Sakthithasan, K., Carillo, G., Debatisse, A., Jouannique, V., Guinhouya, B. C., & Guseva Canu, I. (2022). Analyse exploratoire des mesures de particules ultrafines en temps réel dans des enceintes ferroviaires souterraines de transport public. *Archives des Maladies Professionnelles et de l'Environnement*. https://doi.org/10.1016/j.admp.2022.01.005

Bendl, J., Neukirchen, C., Mudan, A., Padoan, S., Zimmermann, R., & Adam, T. (2023). Personal measurements and sampling of particulate matter in a subway – Identification of hot-spots, spatio-temporal variability and sources of pollutants. *Atmospheric Environment*, 119883. https://doi.org/10.1016/j.atmosenv.2023.119883

Birenzvige, A., Eversole, J., Seaver, M., Francesconi, S., Valdes, E., & Kulaga, H. (2003). Aerosol Characteristics in a Subway Environment. *Aerosol Science and Technology*, *37*(3), 210-220. https://doi.org/10.1080/02786820300941

Boelter, K. J., & Davidson, J. H. (1997). Ozone Generation by Indoor, Electrostatic Air Cleaners. *Aerosol Science and Technology*, *27*(6), 689-708. https://doi.org/10.1080/02786829708965505

Boudia, N., Halley, R., Kennedy, G., Lambert, J., Gareau, L., & Zayed, J. (2006). Manganese concentrations in the air of the Montreal (Canada) subway in relation to surface automobile traffic density. *Science of The Total Environment*, *366*(1), 143-147. https://doi.org/10.1016/j.scitotenv.2005.09.094

Bougault, V., Valorso, R., Sarda-Esteve, R., Baisnee, D., Visez, N., Oliver, G., Bureau, J., Abdoussi, F., Ghersi, V., & Foret, G. (2024). Paris air quality monitoring for the 2024 Olympics and

Paralympics: focus on air pollutants and pollen. *British Journal of Sports Medicine*, *58*(17), 973-982. https://doi.org/10.1136/bjsports-2024-108129

Braniš, M. (2006). The contribution of ambient sources to particulate pollution in spaces and trains of the Prague underground transport system. *Atmospheric Environment*, *40*(2), 348-356. https://doi.org/10.1016/j.atmosenv.2005.09.060

Cartenì, A., & Cascetta, F. (2018). Particulate matter concentrations in a high-quality rubber-tyred metro system: the case study of Turin in Italy. *International Journal of Environmental Science and Technology*, *15*(9), 1921-1930. https://doi.org/10.1007/s13762-017-1566-x

Cartenì, A., Cascetta, F., & Campana, S. (2015). Underground and ground-level particulate matter concentrations in an Italian metro system. *Atmospheric Environment*, *101*, 328-337. https://doi.org/10.1016/j.atmosenv.2014.11.030

Cha, Y., Hedberg, Y., Mei, N., & Olofsson, U. (2016). Airborne Wear Particles Generated from Conductor Rail and Collector Shoe Contact: Influence of Sliding Velocity and Particle Size. *Tribology Letters*, *64*(3), 40. https://doi.org/10.1007/s11249-016-0775-7

Cha, Y., & Olofsson, U. (2018). Effective density of airborne particles in a railway tunnel from field measurements of mobility and aerodynamic size distributions. *Aerosol Science and Technology*, *52*(8), 886-899. https://doi.org/10.1080/02786826.2018.1476750

Chan, L. Y., Lau, W. L., Lee, S. C., & Chan, C. Y. (2002). Commuter exposure to particulate matter in public transportation modes in Hong Kong. *Atmospheric Environment*, *36*(21), 3363-3373. https://doi.org/10.1016/S1352-2310(02)00318-7

Chang, L., Chong, W. T., Wang, X., Pei, F., Zhang, X., Wang, T., Wang, C., & Pan, S. (2021). Recent progress in research on PM2.5 in subways. *Environmental Science: Processes & Impacts*, *23*(5), 642-663. https://doi.org/10.1039/D1EM00002K

Chen, T., Chen, F., Wang, K., Ma, X., Wei, X., Wang, W., Huang, P., Yang, D., Xia, Z., & Zhao, Z. (2021). Acute respiratory response to individual particle exposure (PM1.0, PM2.5 and PM10) in the elderly with and without chronic respiratory diseases. *Environmental Pollution (Barking, Essex: 1987)*, *271*, 116329. https://doi.org/10.1016/j.envpol.2020.116329

Chen, Y.-Y., Lu, C.-Y., Chen, P.-C., Mao, I.-F., & Chen, M.-L. (2017). Analysis of Aerosol Composition and Assessment of Tunnel Washing Performance within a Mass Rapid Transit System in Taiwan. *Aerosol and Air Quality Research*, *17*(6), 1527-1538. https://doi.org/10.4209/aaqr.2017.03.0120

Cheng, Y.-H., Lin, Y.-L., & Liu, C.-C. (2008). Levels of PM10 and PM2.5 in Taipei Rapid Transit System. *Atmospheric Environment*, *42*(31), 7242-7249. https://doi.org/10.1016/j.atmosenv.2008.07.011

Cheng, Y.-H., Liu, Z.-S., & Yan, J.-W. (2012). Comparisons of PM10, PM2.5, Particle Number, and CO2 Levels inside Metro Trains between Traveling in Underground Tunnels and on Elevated Tracks. *Aerosol and Air Quality Research*, *12*(5), 879-891. https://doi.org/10.4209/aaqr.2012.05.0127

Colombi, C., Angius, S., Gianelle, V., & Lazzarini, M. (2013). Particulate matter concentrations, physical characteristics and elemental composition in the Milan underground transport system. *Atmospheric Environment*, *70*, 166-178. https://doi.org/10.1016/j.atmosenv.2013.01.035

Cui, G., Zhou, L., & Dearing, J. (2016). Granulometric and magnetic properties of deposited particles in the Beijing subway and the implications for air quality management. *The Science of the Total Environment*, *568*, 1059-1068. https://doi.org/10.1016/j.scitotenv.2016.06.154

Durand, A. (2021). *Dispersion des particules issues du freinage des trains en stations souterraines.* [Thèse de doctorat, École centrale de Nantes]. https://tel.archives-ouvertes.fr/tel-03526121

Durand, A., Mehel, A., Fokoua, G., Murzyn, F., Puech, S., & Larrarte, F. (2021). Numerical and experimental investigations on brake particle dispersion in the flow generated by a train in an underground station. *Atmospheric Pollution Research*, *12*(10), 101189. https://doi.org/10.1016/j.apr.2021.101189

Fan, H., Jiang, M., Xu, L., Zhu, H., Cheng, J., & Jiang, J. (2020). Comparison of Long Short Term Memory Networks and the Hydrological Model in Runoff Simulation. *Water*, *12*(1), Article 1. https://doi.org/10.3390/w12010175

Figueroa-Lara, J. J., Murcia-González, J. M., García-Martínez, R., Romero-Romo, M., Torres Rodríguez, M., & Mugica-Álvarez, V. (2019). Effect of platform subway depth on the presence of airborne PM2.5, metals, and toxic organic species. *Journal of Hazardous Materials*, *377*, 427-436. https://doi.org/10.1016/j.jhazmat.2019.05.091

Font, O., Moreno, T., Querol, X., Martins, V., Sánchez Rodas, D., de Miguel, E., & Capdevila, M. (2019). Origin and speciation of major and trace PM elements in the Barcelona subway system. *Transportation Research Part D: Transport and Environment*, *72*, 17-35. https://doi.org/10.1016/j.trd.2019.03.007

Fortain, A. (2008). *Caractérisation des particules en gares souterraines* [Thèse de doctorat, La Rochelle]. https://www.theses.fr/2008LAROS231

Fridell, E., Ferm, M., & Ekberg, A. (2010). Emissions of particulate matters from railways – Emission factors and condition monitoring. *Transportation Research Part D: Transport and Environment*, *15*(4), 240-245. https://doi.org/10.1016/j.trd.2010.02.006

Fruhwirt, D., Nöst, T., Leonhardt, P., Leitner, M., Brunnhofer, P., Bode, G., Michael, S., & Rodler, J. (2024). Characterization of emissions from axle-mounted rail disc brakes. *Transportation Research Part D: Transport and Environment*, *130*, 104181. https://doi.org/10.1016/j.trd.2024.104181

Fruhwirt, D., Sturm, P., Bucca, G., Bode, G., Michael, S., & Rodler, J. (2023a). Emissions from railways: Results of tests on a pantograph-catenary test bench. *Transportation Research Part D: Transport and Environment*, *117*, 103667. https://doi.org/10.1016/j.trd.2023.103667

Fruhwirt, D., Sturm, P., Nöst, T., Leonhardt, P., Bode, G., Michael, S., & Rodler, J. (2023b). PM emissions from railways – Results of tests on a wheel-rail test bench. *Transportation Research Part D: Transport and Environment*, *122*, 103858. https://doi.org/10.1016/j.trd.2023.103858

Furuya, K., Kudo, Y., Okinaga, K., Yamuki, M., Takahashi, S., Araki, Y., & Hisamatsu, Y. (2001). *Seasonal variation and their characterization of suspended particulate matter in the air of subway stations*. https://doi.org/10.1081/TMA-100107583

Geffre, E. (2024). *Développement, caractérisation et modélisation d'une fenêtre pariétodynamique filtrante* [Thèse de doctorat, La Rochelle]. https://theses.fr/2024LAROS005

Gokul, P. R., Mathew, A., Bhosale, A., & Nair, A. T. (2023). Spatio-temporal air quality analysis and PM2.5 prediction over Hyderabad City, India using artificial intelligence techniques. *Ecological Informatics*, *76*, 102067. https://doi.org/10.1016/j.ecoinf.2023.102067

Green, D. C., Fuller, G. W., & Baker, T. (2009). Development and validation of the volatile correction model for PM10 – An empirical method for adjusting TEOM measurements for their loss of

volatile particulate matter. *Atmospheric Environment*, *43*(13), 2132-2141. https://doi.org/10.1016/j.atmosenv.2009.01.024

Guo, L., Hu, Y., Hu, Q., Lin, J., Li, C., Chen, J., Li, L., & Fu, H. (2014). Characteristics and chemical compositions of particulate matter collected at the selected metro stations of Shanghai, China. *Science of The Total Environment*, *496*, 443-452. https://doi.org/10.1016/j.scitotenv.2014.07.055

Gustafsson, M., Blomqvist, G., Swietlicki, E., Dahl, A., & Gudmundsson, A. (2012). Inhalable railroad particles at ground level and subterranean stations – Physical and chemical properties and relation to train traffic. *Transportation Research Part D: Transport and Environment*, *17*(3), 277-285. https://doi.org/10.1016/j.trd.2011.12.006

Han, H., Lee, J.-Y., & Jang, K.-J. (2015). Effect of platform screen doors on the indoor air environment of an underground subway station. *Indoor and Built Environment*, *24*(5), 672-681. https://doi.org/10.1177/1420326X14528731

Hastie, T., & Tibshirani, R. (1986). Generalized Additive Models. *Statistical Science*, *1*(3), 297-310.

He, S., Jin, L., Le, T., Zhang, C., Liu, X., & Ming, X. (2018). Commuter health risk and the protective effect of three typical metro environmental control systems in Beijing, China. *Transportation Research Part D: Transport and Environment*, *62*, 633-645. https://doi.org/10.1016/j.trd.2018.04.015

Hernández-Castillo, O., Mugica-Álvarez, V., Castañeda-Briones, M. T., Murcia, J. M., García-Franco, F., & Falcón Briseño, Y. (2014). Aerobiological study in the Mexico City subway system. *Aerobiologia*, *30*(4), 357-367. https://doi.org/10.1007/s10453-014-9334-6

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

Humbal, C., Gautam, S., & Trivedi, U. (2018). A review on recent progress in observations, and health effects of bioaerosols. *Environment International*, *118*, 189-193. https://doi.org/10.1016/j.envint.2018.05.053

INERIS. (2023). *Recommandations pour la réalisation de mesures harmonisées de la qualité de l'air dans les enceintes ferroviaires souterraines* (Guide Ineris-213878-2783593-v1.0). Ineris. https://www.ineris.fr/fr/recommandations-realisation-mesures-harmonisees-qualite-air-enceintes-ferroviaires-souterraines

Izadi, T., Mehrabian, M. A., Sadrizadeh, S., Olofsson, U., & Abouali, O. (2022). The effect of ventilation system with and without under-platform exhaust on the concentration of braking micro-particles inside the subway system. *Tunnelling and Underground Space Technology*, *128*, 104638. https://doi.org/10.1016/j.tust.2022.104638

Johansson, C., & Johansson, P.-Å. (2003). Particulate matter in the underground of Stockholm. *Atmospheric Environment*, *37*(1), 3-9. https://doi.org/10.1016/S1352-2310(02)00833-6

Jumlongkul, A. (2022). Water-based air purifier with ventilation fan system: a novel approach for cleaning indoor/outdoor transitional air during the pandemic. *SN Applied Sciences*, *4*(10), 257. https://doi.org/10.1007/s42452-022-05142-5

Kam, W., Cheung, K., Daher, N., & Sioutas, C. (2011). Particulate matter (PM) concentrations in underground and ground-level rail systems of the Los Angeles Metro. *Atmospheric Environment*, *45*(8), 1506-1516. https://doi.org/10.1016/j.atmosenv.2010.12.049

Kamani, H., Hoseini, M., Seyedsalehi, M., Mahdavi, Y., Jaafari, J., & Safari, G. H. (2014). Concentration and characterization of airborne particles in Tehran's subway system. *Environmental Science and Pollution Research*, *21*(12), 7319-7328. https://doi.org/10.1007/s11356-014-2659-4

Karim, A. S., Malone, M., Bruno, A., Eggler, A. L., Posner, M. A., & Shakya, K. M. (2024). Assessment of air quality in the Philadelphia, Pennsylvania subway. *Journal of Exposure Science & Environmental Epidemiology*, 1-9. https://doi.org/10.1038/s41370-024-00711-9

Kim, B.-W., Jung, H.-J., Song, Y.-C., Lee, M.-J., Kim, H.-K., Kim, J.-C., Sohn, J.-R., & Ro, C.-U. (2010). Characterization of Summertime Aerosol Particles Collected at Subway Stations in Seoul, Korea Using Low-Z Particle Electron Probe X-ray Microanalysis. *Asian Journal of Atmospheric Environment*, *4*. https://doi.org/10.5572/ajae.2010.4.2.097

Kim, G.-S., Son, Y.-S., Lee, J.-H., Kim, I.-W., Kim, J.-C., Oh, J.-T., & Kim, H. (2016). Air Pollution Monitoring and Control System for Subway Stations Using Environmental Sensors. *Journal of Sensors*, *2016*(1), 1865614. https://doi.org/10.1155/2016/1865614

Kim, K. Y., Kim, Y. S., Roh, Y. M., Lee, C. M., & Kim, C. N. (2008). Spatial distribution of particulate matter (PM10 and PM2.5) in Seoul Metropolitan Subway stations. *Journal of Hazardous Materials*, *154*(1), 440-443. https://doi.org/10.1016/j.jhazmat.2007.10.042

Kim, K.-H., Ho, X. D., Ho, Jeon, J.-S., & Kim, J.-C. (2012). A noticeable shift in particulate matter levels after platform screen door installation in a Korean subway station. *Atmospheric Environment*, *49*, 219-223. https://doi.org/10.1016/j.atmosenv.2011.11.058

Kim, T.-H., Choi, B.-H., Kang, M.-S., & Lee, H.-J. (2021). Removal of Iron Oxide from Indoor Air at a Subway Station Using a Vegetation Biofilter: A Case Study of Seoul, Korea. *Atmosphere*, *12*(11), Article 11. https://doi.org/10.3390/atmos12111463

Krall, J. R., Moore, K. D., Joannidis, C., Lee, Y.-C., Pollack, A. Z., McCombs, M., Thornburg, J., & Balachandran, S. (2021). Commuter types identified using clustering and their associations with source-specific PM2.5. *Environmental Research*, *200*, 111419. https://doi.org/10.1016/j.envres.2021.111419

Kwon, S.-B., Jeong, W., Park, D., Kim, K.-T., & Cho, K. H. (2015). A multivariate study for characterizing particulate matter (PM10, PM2.5, and PM1) in Seoul metropolitan subway stations, Korea. *Journal of Hazardous Materials*, *297*, 295-303. https://doi.org/10.1016/j.jhazmat.2015.05.015

Kwon, S.-B., Park, D., Cho, Y., & Park, E.-Y. (2010). Measurement of Natural Ventilation Rate in Seoul Metropolitan Subway Cabin. *Indoor and Built Environment*, *19*(3), 366-374. https://doi.org/10.1177/1420326X10367305

Lai, A. C. K., & Nazaroff, W. W. (2000). Modeling indoor particle deposition from turbulent flow onto smooth surfaces. *Journal of Aerosol Science*, *31*(4), 463-476. https://doi.org/10.1016/S0021-8502(99)00536-4

Lee, J., Byeon, S., & Lee, J. (2009). *The effect of platform screen door (PSD) for fine particles at subway train in Seoul, Korea*.

Lee, Y., Lee, Y.-C., Kim, T., Choi, J. S., & Park, D. (2018). Sources and characteristics of particulate matter in subway tunnels in Seoul, Korea. *International Journal of Environmental Research and Public Health*, *15*(11), Article 11. https://doi.org/10.3390/ijerph15112534

Lei, R., Zhu, F., Cheng, H., Liu, J., Shen, C., Zhang, C., Xu, Y., Xiao, C., Li, X., Zhang, J., Ding, R., & Cao, J. (2019). Short-term effect of PM2.5/O3 on non-accidental and respiratory deaths in

highly polluted area of China. *Atmospheric Pollution Research*, *10*(5), 1412-1419. https://doi.org/10.1016/j.apr.2019.03.013

Li, X., Peng, L., Yao, X., Cui, S., Hu, Y., You, C., & Chi, T. (2017). Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation. *Environmental Pollution*, *231*, 997-1004. https://doi.org/10.1016/j.envpol.2017.08.114

Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation Forest. *2008 Eighth IEEE International Conference on Data Mining*, 413-422. https://doi.org/10.1109/ICDM.2008.17

Luglio, D. G., Huynh, T., Saporito, A., & Gordon, T. (2023). Investigation of a river-tunnel effect on PM2.5 concentrations in New York City subway stations. *Transportation Research Part D: Transport and Environment*, *115*, 103579. https://doi.org/10.1016/j.trd.2022.103579

Luo, H., Han, Y., Cheng, X., Lu, C., & Wu, Y. (2021a). Spatiotemporal Variations in Particulate Matter and Air Quality over China: National, Regional and Urban Scales. *Atmosphere*, *12*(1), Article 1. https://doi.org/10.3390/atmos12010043

Luo, R., Dai, H., Zhang, Y., Wang, P., Zhou, Y., Li, J., Zhou, M., Qiao, L., Ma, Y., Zhu, S., Hu, S., Huang, C., & Shi, H. (2021b). Association of short-term exposure to source-specific PM2.5 with the cardiovascular response during pregnancy in the Shanghai MCPC study. *Science of The Total Environment*, *775*, 145725. https://doi.org/10.1016/j.scitotenv.2021.145725

Ma, C.-J., Matuyama, S., Sera, K., & Kim, S.-D. (2012). Physicochemical Properties of Indoor Particulate Matter Collected on Subway Platforms in Japan. *Asian Journal of Atmospheric Environment*, *6*, 73-82. https://doi.org/10.5572/ajae.2012.6.2.073

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. *International Journal of Environmental Research and Public Health*, *16*(4), Article 4. https://doi.org/10.3390/ijerph16040625

Marcus, H. (2011). Waiting experience at train stations. *Journal of Medical Internet Research - J MED INTERNET RES*.

Martins, V., Moreno, T., Mendes, L., Eleftheriadis, K., Diapouli, E., Alves, C. A., Duarte, M., de Miguel, E., Capdevila, M., Querol, X., & Minguillón, M. C. (2016). Factors controlling air quality in different European subway systems. *Environmental Research*, *146*, 35-46. https://doi.org/10.1016/j.envres.2015.12.007

Martins, V., Moreno, T., Minguillón, M. C., Amato, F., de Miguel, E., Capdevila, M., & Querol, X. (2015). Exposure to airborne particulate matter in the subway system. *Science of The Total Environment*, *511*, 711-722. https://doi.org/10.1016/j.scitotenv.2014.12.013

Midander, K., Elihn, K., Wallén, A., Belova, L., Karlsson, A.-K. B., & Wallinder, I. O. (2012). Characterisation of nano- and micron-sized airborne and collected subway particles, a multi-analytical approach. *Science of The Total Environment*, *427-428*, 390-400. https://doi.org/10.1016/j.scitotenv.2012.04.014

Minguillón, M. C., Reche, C., Martins, V., Amato, F., de Miguel, E., Capdevila, M., Centelles, S., Querol, X., & Moreno, T. (2018). Aerosol sources in subway environments. *Environmental Research*, *167*, 314-328. https://doi.org/10.1016/j.envres.2018.07.034

Misra, C., Geller, M. D., Shah, P., Sioutas, C., & Solomon, P. A. (2001). Development and evaluation of a continuous coarse (PM10-PM2.5) particle monitor. *Journal of the Air & Waste Management Association (1995)*, *51*(9), 1309-1317. https://doi.org/10.1080/10473289.2001.10464360

Mohsen, M., Ahmed, M. B., & Zhou, J. L. (2018). Particulate matter concentrations and heavy metal contamination levels in the railway transport system of Sydney, Australia. *Transportation Research Part D: Transport and Environment*, *62*, 112-124. https://doi.org/10.1016/j.trd.2018.02.015

Moreno, T., Kelly, F. J., Dunster, C., Oliete, A., Martins, V., Reche, C., Minguillón, M. C., Amato, F., Capdevila, M., de Miguel, E., & Querol, X. (2017). Oxidative potential of subway PM2.5. *Atmospheric Environment*, *148*, 230-238. https://doi.org/10.1016/j.atmosenv.2016.10.045

Moreno, T., Martins, V., Querol, X., Jones, T., BéruBé, K., Minguillón, M. C., Amato, F., Capdevila, M., de Miguel, E., Centelles, S., & Gibbons, W. (2015a). A new look at inhalable metalliferous airborne particles on rail subway platforms. *Science of The Total Environment*, *505*, 367-375. https://doi.org/10.1016/j.scitotenv.2014.10.013

Moreno, T., Pérez, N., Reche, C., Martins, V., de Miguel, E., Capdevila, M., Centelles, S., Minguillón, M. C., Amato, F., Alastuey, A., Querol, X., & Gibbons, W. (2014). Subway platform air quality: Assessing the influences of tunnel ventilation, train piston effect and station design. *Atmospheric Environment*, *92*, 461-468. https://doi.org/10.1016/j.atmosenv.2014.04.043

Moreno, T., Reche, C., Rivas, I., Cruz Minguillón, M., Martins, V., Vargas, C., Buonanno, G., Parga, J., Pandolfi, M., Brines, M., Ealo, M., Sofia Fonseca, A., Amato, F., Sosa, G., Capdevila, M., de Miguel, E., Querol, X., & Gibbons, W. (2015b). Urban air quality comparison for bus, tram, subway and pedestrian commutes in Barcelona. *Environmental Research*, *142*, 495-510. https://doi.org/10.1016/j.envres.2015.07.022

Motallebi, N., Tran, H., Croes, B. E., & Larsen, L. C. (2003). Day-of-week patterns of particulate matter and its chemical components at selected sites in California. *Journal of the Air & Waste Management Association (1995)*, *53*(7), 876-888. https://doi.org/10.1080/10473289.2003.10466229

Murruni, L. G., Solanes, V., Debray, M., Kreiner, A. J., Davidson, J., Davidson, M., Vázquez, M., & Ozafrán, M. (2009). Concentrations and elemental composition of particulate matter in the Buenos Aires underground system. *Atmospheric Environment*, *43*(30), 4577-4583. https://doi.org/10.1016/j.atmosenv.2009.06.025

Namgung, H.-G., Kim, J. B., Kim, M.-S., Kim, M., Park, S., Woo, S.-H., Bae, G.-N., Park, D., & Kwon, S.-B. (2017). Size distribution analysis of airborne wear particles released by subway brake system. *Wear*, *372-373*, 169-176. https://doi.org/10.1016/j.wear.2016.12.026

Ngo, S., Brun, G., Gournet, N., Ghozzi, F., Boullet, T., & Carrilho, H. (2023). *Caractérisation et identification des sources de particules dans l'air des enceintes ferroviaires souterraines*. Congrès Français sur les Aérosols. https://doi.org/10.25576/ASFERA-CFA2023-32907

Nyström, A. K. (2012). *Potential mechanisms for acute health effects and lung retention of inhaled particles of different origin*. Karolinska Institutet (Sweden). http://openarchive.ki.se/xmlui/handle/10616/41026

Olofsson, U. (2011). A study of airborne wear particles generated from the train traffic—Block braking simulation in a pin-on-disc machine. *Wear*, *271*(1), 86-91. https://doi.org/10.1016/j.wear.2010.10.016

Olofsson, U., Zhu, Y., Abbasi, S., Lewis, R., & Lewis, S. (2013). Tribology of the wheel–rail contact – aspects of wear, particle emission and adhesion. *Vehicle System Dynamics*, *51*(7), 1091-1120. https://doi.org/10.1080/00423114.2013.800215

Onat, B., & Stakeeva, B. (2013). Personal exposure of commuters in public transport to PM2.5 and fine particle counts. *Atmospheric Pollution Research*, *4*(3), 329-335. https://doi.org/10.5094/APR.2013.037

Pan, S., Du, S., Wang, X., Zhang, X., Xia, L., Liu, J., Pei, F., & Wei, Y. (2019). Analysis and interpretation of the particulate matter (PM10 and PM2.5) concentrations at the subway stations in Beijing, China. *Sustainable Cities and Society*, *45*, 366-377. https://doi.org/10.1016/j.scs.2018.11.020

Parchami, M., Motlagh, M. S., & Pardakhti, A. (2022). Particulate matter concentrations and characterization in urban subway system-case study Tehran, Iran. *Journal of Environmental Health Science and Engineering*, *20*(2), 953-962. https://doi.org/10.1007/s40201-022-00838-7

Park, D., Lee, T., Hwang, D., Jung, W., Lee, Y., Cho, K., Kim, D., & Lee, K. (2014). Identification of the sources of PM10 in a subway tunnel using positive matrix factorization. *Journal of the Air & Waste Management Association*, *64*(12), 1361-1368. https://doi.org/10.1080/10962247.2014.950766

Park, D.-U., & Ha, K.-C. (2008). Characteristics of PM10, PM2.5, CO2 and CO monitored in interiors and platforms of subway train in Seoul, Korea. *Environment International*, *34*(5), 629-634. https://doi.org/10.1016/j.envint.2007.12.007

Park, S.-J., Kim, J., Shim, I.-K., & Lee, J. (2024). Nationwide distribution and factors related to indoor fine particulate matter concentrations in subway stations in Korea: Results of first-year measurement. *Environment International*, 109078. https://doi.org/10.1016/j.envint.2024.109078

Passi, A., Nagendra, S. M. S., & Maiya, M. P. (2021a). 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

Passi, A., Shiva Nagendra, S. M., & Maiya, M. P. (2021b). *Assessment of indoor and outdoor exposure to airborne particles in subway metro stations and their associated health risks*. 85-96. https://doi.org/10.2495/AIR210081

Pini, L., Giordani, J., Gardini, G., Concoreggi, C., Pini, A., Perger, E., Vizzardi, E., Di Bona, D., Cappelli, C., Ciarfaglia, M., & Tantucci, C. (2021). Emergency department admission and hospitalization for COPD exacerbation and particulate matter short-term exposure in Brescia, a highly polluted town in northern Italy. *Respiratory Medicine*, *179*, 106334. https://doi.org/10.1016/j.rmed.2021.106334

Polanco-Martínez, J. (2019). Dynamic relationship analysis between NAFTA stock markets using nonlinear, nonparametric, non-stationary methods. *Nonlinear Dynamics*, *97*, 1-21. https://doi.org/10.1007/s11071-019-04974-y

Poppendieck, D. G., Rim, D., & Persily, A. K. (2014). Ultrafine particle removal and ozone generation by in-duct electrostatic precipitators. *Environmental Science & Technology*, *48*(3), 2067-2074. https://doi.org/10.1021/es404884p

Pretot, P.-E., Schulz, C., Chalet, D., Migaud, J., & Bogdan, M. (2022, avril). Fast method to design air filtration solution at low energy cost in subterranean train stations. *7th World Congress on Civil, Structural, and Environmental Engineering (CSEE'22)*. https://hal.archives-ouvertes.fr/hal-03642567

Qiao, T., Xiu, G., Zheng, Y., Yang, J., & Wang, L. (2015). Characterization of PM and Microclimate in a Shanghai Subway Tunnel, China. *Procedia Engineering*, *102*, 1226-1232. https://doi.org/10.1016/j.proeng.2015.01.250

Querol, X. C., Moreno Pérez, T., Karanasiou, A., Reche Andúgar, C., Alastuey Urós, J. A., Viana Rodríguez, M. del M., Font Piqueras, O., Plana, F., Moreno, N., Miguel, E. de, & Capdevila, M. (2012a). Geochemistry of respirable particles in the Barcelona's subway: how reducing commuting exposure? *Macla: Revista de La Sociedad Española de Mineralogía*, *16*, 151-152.

Querol, X., Moreno, T., Karanasiou, A., Reche, C., Alastuey, A., Viana, M., Font, O., Gil, J., de Miguel, E., & Capdevila, M. (2012b). Variability of levels and composition of PM_{10} and $PM_{2.5}$ in the Barcelona metro system. *Atmospheric Chemistry and Physics*, *12*(11), 5055-5076. https://doi.org/10.5194/acp-12-5055-2012

Rahmani, M., Crumeyrolle, S., Allegri-Martiny, N., Taherkordi, A., & Rouvoy, R. (2024). PMFORECAST: leveraging temporal LSTM to deliver in situ air quality predictions. *Environmental Science and Pollution Research International*. https://doi.org/10.1007/s11356-024-34623-w

Rakotonirinjanahary, V. M., Crumeyrolle, S., Bogdan, M., & Hanoune, B. (2024). A novel method for establishing typical daily profile of PM concentrations in underground railway stations. *Indoor Environments*, *1*(3), 100040. https://doi.org/10.1016/j.indenv.2024.100040

Ramos, M. J., Vasconcelos, A., & Faria, M. (2015). Comparison of Particulate Matter Inhalation for Users of Different Transport Modes in Lisbon. *Transportation Research Procedia*, *10*, 433-442. https://doi.org/10.1016/j.trpro.2015.09.093

Raut, J.-C., Chazette, P., & Fortain, A. (2009). Link between aerosol optical, microphysical and chemical measurements in an underground railway station in Paris. *Atmospheric Environment*, *43*(4), 860-868. https://doi.org/10.1016/j.atmosenv.2008.10.038

Ren, J., He, J., Kong, X., Xu, W., Kang, Y., Yu, Z., & Li, H. (2022). A field study of CO2 and particulate matter characteristics during the transition season in the subway system in Tianjin, China. *Energy and Buildings*, *254*, 111620. https://doi.org/10.1016/j.enbuild.2021.111620

Renzi, M., Stafoggia, M., Michelozzi, P., Davoli, M., Forastiere, F., & Solimini, A. G. (2020). Short-term exposure to PM2.5 and risk of venous thromboembolism: A case-crossover study. *Thrombosis Research*, *190*, 52-57. https://doi.org/10.1016/j.thromres.2020.03.008

Ripanucci, G., Grana, M., Vicentini, L., Magrini, A., & Bergamaschi, A. (2006). Dust in the underground railway tunnels of an Italian town. *Journal of Occupational and Environmental Hygiene*, *3*(1), 16-25. https://doi.org/10.1080/15459620500444004

Russell, H. S., Kappelt, N., Fessa, D., Frederickson, L. B., Bagkis, E., Apostolidis, P., Karatzas, K., Schmidt, J. A., Hertel, O., & Johnson, M. S. (2022). Particulate air pollution in the Copenhagen metro part 2: Low-cost sensors and micro-environment classification. *Environment International*, *170*, 107645. https://doi.org/10.1016/j.envint.2022.107645

Şahin, Ü. A., Onat, B., Stakeeva, B., Ceran, T., & Karim, P. (2012). PM10 concentrations and the size distribution of Cu and Fe-containing particles in Istanbul's subway system. *Transportation Research Part D: Transport and Environment*, *17*(1), 48-53. https://doi.org/10.1016/j.trd.2011.09.003

Salma, I., Pósfai, M., Kovács, K., Kuzmann, E., Homonnay, Z., & Posta, J. (2009). Properties and sources of individual particles and some chemical species in the aerosol of a metropolitan

underground railway station. *Atmospheric Environment*, *43*(22), 3460-3466. https://doi.org/10.1016/j.atmosenv.2009.04.042

Salma, I., Weidinger, T., & Maenhaut, W. (2007). Time-resolved mass concentration, composition and sources of aerosol particles in a metropolitan underground railway station. *Atmospheric Environment*, *41*(37), 8391-8405. https://doi.org/10.1016/j.atmosenv.2007.06.017

Saunders, B. M., Smith, J. D., Smith, T. E. L., Green, D. C., & Barratt, B. (2019). Spatial variability of fine particulate matter pollution (PM2.5) on the London Underground network. *Urban Climate*, *30*, 100535. https://doi.org/10.1016/j.uclim.2019.100535

Sim, J.-B., Woo, S.-H., Yook, S.-J., Kim, J. B., Bae, G.-N., & Oh, S. G. (2018). Baffle dust collector for removing particles from a subway tunnel during the passage of a train. *Journal of Mechanical Science and Technology*, *32*(3), 1415-1421. https://doi.org/10.1007/s12206-018-0245-2

Sitzmann, B., Kendall, M., Watt, J., & Williams, I. (1999). Characterisation of airborne particles in London by computer-controlled scanning electron microscopy. *Science of The Total Environment*, *241*(1), 63-73. https://doi.org/10.1016/S0048-9697(99)00326-5

Son, J., Kim, K., Kwon, S., Park, S.-M., Ha, K., Shin, Y., Ahn, M., Cho, S., Shin, J., Shin, Y., & Lee, G. (2021). Source Quantification of PM10 and PM2.5 Using Iron Tracer Mass Balance in a Seoul Subway Station, South Korea. *Aerosol and Air Quality Research*, *21*(7), 200573. https://doi.org/10.4209/aaqr.200573

Son, Y. S., Dinh, T.-V., Chung, S.-G., Lee, J.-H., & Kim, J.-C. (2014a). Removal of particulate matter emitted from a subway tunnel using magnetic filters. *Environmental Science & Technology*, *48*(5), 2870-2876. https://doi.org/10.1021/es404502x

Son, Y.-S., Jeon, J.-S., Lee, H. J., Ryu, I.-C., & Kim, J.-C. (2014b). Installation of platform screen doors and their impact on indoor air quality: Seoul subway trains. *Journal of the Air & Waste Management Association (1995)*, *64*(9), 1054-1061. https://doi.org/10.1080/10962247.2014.923350

Son, Y.-S., Oh, Y.-H., Choi, I.-Y., Dinh, T.-V., Chung, S.-G., Lee, J., Park, D., & Kim, J.-C. (2019). Development of a magnetic hybrid filter to reduce PM10 in a subway platform. *Journal of Hazardous Materials*, *368*, 197-203. https://doi.org/10.1016/j.jhazmat.2019.01.045

Son, Y.-S., Salama, A., Jeong, H.-S., Jeong, J. Y., Lee, J., Sunwoo, Y., & Kim, J.-C. (2013). The Effect of Platform Screen Doors on PM10 Levels in a Subway Station and a Trial to Reduce PM10 in Tunnels. *Asian Journal of Atmospheric Environment*, *7*, 38-47. https://doi.org/10.5572/ajae.2013.7.1.038

Su, X., Sutarlie, L., & Loh, X. J. (2020). Sensors and Analytical Technologies for Air Quality: Particulate Matters and Bioaerosols. *Chemistry - An Asian Journal*, *15*. https://doi.org/10.1002/asia.202001051

Sun, B., Song, J., Wang, Y., Jiang, J., An, Z., Li, J., Zhang, Y., Wang, G., Li, H., Alexis, N. E., Jaspers, I., & Wu, W. (2021). Associations of short-term PM2.5 exposures with nasal oxidative stress, inflammation and lung function impairment and modification by *GSTT1*-null genotype: A panel study of the retired adults. *Environmental Pollution*, *285*, 117215. https://doi.org/10.1016/j.envpol.2021.117215

Tokarek, S., & Bernis, A. (2006). An exemple of particle concentration reduction in Parisian subway stations by electrostatic precipitation. *Environmental Technology*, *27*(11), 1279-1287. https://doi.org/10.1080/09593332708618746

Tran, H. D., Huang, H.-Y., Yu, J.-Y., & Wang, S.-H. (2023). Forecasting hourly PM2.5 concentration with an optimized LSTM model. *Atmospheric Environment*, *315*, 120161. https://doi.org/10.1016/j.atmosenv.2023.120161

Tu, M., & Olofsson, U. (2021). PM10 in underground stations from different types of trains. *Transportation Research Part D: Transport and Environment*, *95*, 102867. https://doi.org/10.1016/j.trd.2021.102867

Tu, M., & Olofsson, U. (2024). Estimating PM levels on an underground metro platform by exploring a new model-based factor research. *Atmospheric Environment: X*, *22*, 100261. https://doi.org/10.1016/j.aeaoa.2024.100261

Van Ryswyk, K., Anastasopolos, A. T., Evans, G., Sun, L., Sabaliauskas, K., Kulka, R., Wallace, L., & Weichenthal, S. (2017). Metro Commuter Exposures to Particulate Air Pollution and PM2.5-Associated Elements in Three Canadian Cities: The Urban Transportation Exposure Study. *Environmental Science & Technology*, *51*(10), 5713-5720. https://doi.org/10.1021/acs.est.6b05775

Vilcassim, M. J.Ruzmyn, Thurston, G. D., Peltier, R. E., & Gordon, T. (2014). Black Carbon and Particulate Matter (PM2.5) Concentrations in New York City's Subway Stations. *Environmental Science & Technology*, *48*(24), 14738-14745. https://doi.org/10.1021/es504295h

Walia, M. S., Vernersson, T., Lundén, R., Blennow, F., & Meinel, M. (2019). Temperatures and wear at railway tread braking: Field experiments and simulations. *Wear*, *440-441*, 203086. https://doi.org/10.1016/j.wear.2019.203086

Walther, E., & Bogdan, M. (2017). A novel approach for the modelling of air quality dynamics in underground railway stations. *Transportation Research Part D: Transport and Environment*, *56*, 33-42. https://doi.org/10.1016/j.trd.2017.07.014

Walther, E., Bogdan, M., & Cohen, R. (2017). Modelling of airborne particulate matter concentration in underground stations using a two size-class conservation model. *Science of The Total Environment*, *607-608*, 1313-1319. https://doi.org/10.1016/j.scitotenv.2017.07.090

Wang, J., Zhao, L., Zhu, D., Gao, H. O., Xie, Y., Li, H., Xu, X., & Wang, H. (2016). Characteristics of particulate matter (PM) concentrations influenced by piston wind and train door opening in the Shanghai subway system. *Transportation Research Part D: Transport and Environment*, *47*, 77-88. https://doi.org/10.1016/j.trd.2016.05.006

Wang, S., Qin, T., Tu, R., Li, T., Chen, G. I., Green, D. C., Zhang, X., Feng, J., Liu, H., Hu, M., & Fu, Q. (2024). Indoor air quality in subway microenvironments: Pollutant characteristics, adverse health impacts, and population inequity. *Environment International*, *190*, 108873. https://doi.org/10.1016/j.envint.2024.108873

Wang, X., Xia, L., Pei, F., Chang, L., Chong, W. T., Wang, Z., & Pan, S. (2022). Particulate Matter (PM2.5 and PM10) Concentration of Subway Transfer Stations in Beijing, China. *Sustainability*, *14*(3), Article 3. https://doi.org/10.3390/su14031552

Wen, Y., Leng, J., Shen, X., Han, G., Sun, L., & Yu, F. (2020). Environmental and Health Effects of Ventilation in Subway Stations: A Literature Review. *International Journal of Environmental Research and Public Health*, *17*(3), 1084. https://doi.org/10.3390/ijerph17031084

WHO. (2021). *WHO Global Air Quality Guidelines: Particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide: executive summary*. https://iris.who.int/handle/10665/345334

Wyatt, L. H., Weaver, A. M., Moyer, J., Schwartz, J. D., Di, Q., Diaz-Sanchez, D., Cascio, W. E., & Ward-Caviness, C. K. (2022). Short-term PM2.5 exposure and early-readmission risk: a retrospective cohort study in North Carolina heart failure patients. *American Heart Journal*, *248*, 130-138. https://doi.org/10.1016/j.ahj.2022.02.015

Yang, J., Fan, X., Zhang, H., Zheng, W., & Ye, T. (2023). A review on characteristics and mitigation strategies of indoor air quality in underground subway stations. *Science of The Total Environment*, *869*, 161781. https://doi.org/10.1016/j.scitotenv.2023.161781

Yang, J., Liu, M., Zhang, H., Zheng, W., & Ye, T. (2024). Numerical analysis and improvement strategy of particulate matter concentration in underground subway stations. *Tunnelling and Underground Space Technology*, *146*, 105662. https://doi.org/10.1016/j.tust.2024.105662

Yuval, Levi, Y., Dayan, U., Levy, I., & Broday, D. M. (2020). On the association between characteristics of the atmospheric boundary layer and air pollution concentrations. *Atmospheric Research*, *231*, 104675. https://doi.org/10.1016/j.atmosres.2019.104675

Zhang, Y., Ding, Z., Xiang, Q., Wang, W., Huang, L., & Mao, F. (2020). Short-term effects of ambient PM1 and PM2.5 air pollution on hospital admission for respiratory diseases: Case-crossover evidence from Shenzhen, China. *International Journal of Hygiene and Environmental Health*, *224*, 113418. https://doi.org/10.1016/j.ijheh.2019.11.001

Zhao, B., & Wu, J. (2006). Modeling particle deposition from fully developed turbulent flow in ventilation duct. *Atmospheric Environment*, *40*(3), 457-466. https://doi.org/10.1016/j.atmosenv.2005.09.043

Zhao, L., Wang, J., Gao, H. O., Xie, Y., Jiang, R., Hu, Q., & Sun, Y. (2017). Evaluation of particulate matter concentration in Shanghai's metro system and strategy for improvement. *Transportation Research Part D: Transport and Environment*, *53*, 115-127. https://doi.org/10.1016/j.trd.2017.04.010

Zheng, C., Duan, D., Chang, Q., Liu, S., Yang, Z., Liu, X., Weng, W., & Gao, X. (2019). Experiments on Enhancing the Particle Charging Performance of an Electrostatic Precipitator. *Aerosol and Air Quality Research*, *19*(6), 1411-1420. https://doi.org/10.4209/aaqr.2018.11.0400

Zuur, A. F., Ieno, E. N., Walker, N., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R*. Springer. https://doi.org/10.1007/978-0-387-87458-6

ANNEXES

Annexe 1 : A novel method for establishing typical profile in underground railway stations

Valisoa M. Rakotonirinjanahary, Suzanne Crumeyrolle, Mateusz Bogdan, Benjamin Hanoune

Indoor Environments, Volume 1, Issue 3, October 2024, 100040

https://doi.org/10.1016/j.indenv.2024.100040

Résumé :

La qualité de l'air dans les enceintes ferroviaires souterraines (EFS) représente un enjeu majeur de santé publique en raison des concentrations extrêmement élevées de particules fines, notamment les PM_{10} et les $PM_{2.5}$. En effet, les sources de particules sont nombreuses et importantes, telles que le freinage des trains et l'effet de tunnel, et les EFS sont souvent des espaces confinés avec des taux de renouvellement d'air faibles. Malgré de multiples mesures de particules dans ces environnements, la variabilité des concentrations d'une gare à l'autre reste mal comprise.

Nous présentons ici une méthodologie pour établir un profil journalier des concentrations de masse de particules, basée sur une série de mesures de cinq ans dans une gare souterraine parisienne. Cette approche intègre un processus approfondi de nettoyage des données, incluant l'identification des périodes d'activité de la gare et la détection de données physiquement incohérentes ou mathématiquement aberrantes, ainsi qu'un modèle de régression linéaire. Cette méthodologie produit trois résultats utilisables : un profil typique pour les jours de semaine, un profil typique pour les weekends, et un coefficient d'amplitude journalière (DAC) des concentrations de particules pour la période considérée. Le DAC est une métrique quotidienne des niveaux de pollution, permettant d'analyser les tendances temporelles et de faciliter la comparaison avec d'autres données ayant des fréquences d'acquisition différentes.

La méthodologie développée ici dans une GFS spécifique pour les mesures de PM_{10} peut être facilement appliquée à d'autres fractions de taille de particules ou à d'autres paramètres mesurés présentant un profil journalier. Les concentrations de PM_{10} en semaine montrent deux pics distincts correspondant aux heures de pointe du matin et du soir, avec une concentration moyenne diurne de 193 $\mu g/m^3$. Ces pics sont retardés d'environ 1 heure par rapport au trafic des trains. Les weekends présentent des niveaux de particules systématiquement plus faibles, sans pics observables, avec une moyenne de 157 $\mu g/m^3$ pendant la journée.

Notre analyse révèle l'évolution temporelle à long terme des concentrations de particules dans la GFS, mettant en évidence des variations saisonnières avec des concentrations de PM_{10} plus élevées en été (jusqu'à 400 $\mu g/m^3$) et plus faibles en hiver (jusqu'à 250 $\mu g/m^3$). Cette évolution saisonnière en milieu intérieur n'est pas corrélée avec celle observée en extérieur, où les concentrations sont généralement plus élevées en hiver. Par ailleurs, nos résultats montrent que la période optimale (DAC ~1) pour réaliser des expérimentations et obtenir des profils fiables se situe au printemps (avril, mai, juin).

Indoor Environments 1 (2024) 100040

Contents lists available at ScienceDirect

Indoor Environments

journal homepage: www.sciencedirect.com/journal/indoor-environments

A novel method for establishing typical daily profile of PM concentrations in underground railway stations

Valisoa M. Rakotonirinjanahary[a,d,*], Suzanne Crumeyrolle[b], Mateusz Bogdan[c], Benjamin Hanoune[a]

[a] Univ. Lille, CNRS, UMR 8522 – PC2A – Physicochimie des Processus de Combustion et de l'Atmosphère, Lille F-59000, France
[b] Univ. Lille, CNRS, UMR 8518 – LOA – Laboratoire d'Optique Atmosphérique, Lille F-59000, France
[c] AREP L'hypercube, 16 avenue d'Ivry, Paris 75013, France
[d] SNCF Holding, 2 Place aux Etoiles, Saint-Denis 93210, France

ARTICLE INFO

Keywords:
Indoor air quality
Underground railway stations
PM concentrations
Daily profile
Temporal evolution

ABSTRACT

The air quality in underground railway stations (URS) poses a significant public health concern due to extremely high concentrations of particulate matter: PM_{10} and $PM_{2.5}$. Indeed, PM sources are strong and numerous, such as train braking and tunnel effect and URS are often confined spaces with low air change rates. Despite multiple PM measurements within URS, the variability of those concentrations from stations to stations is still poorly understood. We present here a methodology for establishing a daily profile of particle mass concentrations, based on a 5-year long measurement series in a Parisian URS. This approach incorporates an extensive data cleaning process based on the identification of URS operation periods and physically inconsistent or mathematically aberrant data, together with a linear regression model. This methodology delivers three usable outcomes: a typical profile for weekdays, a typical profile for weekends, and a PM concentration Daily Amplitude Coefficient (DAC) for the considered period. The DAC is a daily metric of the pollution levels, that enables the analysis of temporal trends and facilitates the comparison with other data with other acquisition frequency. The methodology developed here in a specific URS for PM_{10} measurements can be easily applied to different particle size fractions or to other measured parameters exhibiting a daily profile. Weekdays PM_{10} concentrations exhibit two distinct peaks corresponding to morning and evening rush hours, with an average daytime concentration of 193 $\mu g/m^3$. These peaks are delayed by ~1 hour compared to the train traffic. Weekends show consistently lower PM levels with no observable peaks, averaging 157 $\mu g/m^3$ during the day. Our analysis reveals the long-term temporal evolution of PM concentration within the URS, highlighting seasonal patterns with higher PM_{10} concentrations observed in summer (up to 400 $\mu g/m^3$) and lower values in winter (down to 250 $\mu g/m^3$). This indoor seasonal evolution is not correlated with the outdoor temporal evolution, showing higher concentrations during the winter. Furthermore, our results show that the optimal period (DAC ~ 1) for conducting experiments to obtain reliable profiles is during the spring months (April, May, June).

1. Introduction

Currently, trains represent the most widely utilized means of transportation, increasingly establishing themselves as indispensable components of daily commutes within major urban centers. The emphasis on air quality within trains and URS has grown significantly due to inherent factors like confined spaces, limited ventilation with high passenger number. Numerous investigations have shown that PM concentrations within URS consistently and largely exceed the air quality guidelines set by the World Health Organization [1], which stipulate daily outdoor thresholds of 45 and 15 $\mu g/m^3$ for PM_{10} (particles with an aerodynamic diameter up to 10 μm) and $PM_{2.5}$ (particles with an aerodynamic diameter up to 2.5 μm), respectively. Indeed, extremely high PM concentrations are observed in platforms areas. PM_{10} levels can reach up to 440 $\mu g/m^3$ in Nanjing, China [2], and as low as 51 $\mu g/m^3$ in Taipei, Taiwan [3]. For $PM_{2.5}$, concentrations vary from 18 $\mu g/m^3$ in Vancouver, Canada [4], to 195 $\mu g/m^3$ in Athens, Greece [5]. This variability is also noted in cities such as Stockholm, Sweden (204 $\mu g/m^3$ for PM_{10} and 102 $\mu g/m^3$ for $PM_{2.5}$) [6], and Los Angeles, USA (78 $\mu g/m^3$ for PM_{10} and 57 $\mu g/m^3$ for $PM_{2.5}$) [7]. A comprehensive list of PM

* Corresponding author at: Univ. Lille, CNRS, UMR 8522 – PC2A – Physicochimie des Processus de Combustion et de l'Atmosphère, Lille F-59000, France
E-mail address: miadanavalisoa.rakotonirinjanahary@univ-lille.fr (V.M. Rakotonirinjanahary).

https://doi.org/10.1016/j.indenv.2024.100040
Received 24 May 2024; Received in revised form 26 July 2024; Accepted 7 August 2024
2950-3620/© 2024 The Author(s). Published by Elsevier Inc. on behalf of International Society of Indoor Air Quality and Climate. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

concentrations in subway metro station platforms worldwide is provided by Passi et al. [8]. Additionally, elevated PM concentrations are also prevalent in the interior of trains, as observed in research carried out in Prague, Czech Republic (126 μg/m^3 for PM_{10}) [9], Seoul, Korea (142 μg/m^3 for PM_{10}) [10], or Lisbon, Portugal (80 μg/m^3 and 30 μg/m^3 for PM_{10} and $PM_{2.5}$) [11]. Moreover, the particle mass concentrations (PM_{10} and $PM_{2.5}$) within underground railway stations are several times higher than those in typical urban outdoor air. For instance, Raut et al. [12] reported that the average daytime PM_{10} and $PM_{2.5}$ concentrations of Paris URS were about 5–30 times higher than PM observed outdoor. Similarly, Carteni et al. [13] found that PM concentrations in Naples URS were 2–14 than those outdoors, and Smith et al. [14] observed the PM levels in London URS were approximately three times higher than those in the surrounding outdoor areas.

While passengers spend only brief periods (on average 7 minutes) [15] within these stations, even short-term exposure to heightened levels of air pollutants can lead to various adverse health consequences affecting the cardiovascular system, such as myocardial infarction and cardiac failure [16–18], as well as the respiratory system, with conditions including arrhythmia, deep vein thrombosis, stroke, wheeze, asthma, bronchiolitis, and other respiratory infections [19–23]. Furthermore, recent publications have connected this exposure to health issues like type 2 diabetes [24], dementia [25], and cognitive function decline [26,27]. In addition to its role as an air pollutant, PM can serve as a carrier for toxic and harmful substances such as bacteria or viruses, as evidenced by the global COVID-19 pandemic [28], thereby presenting a significant health risk to exposed individuals. Consequently, addressing air quality concerns within URS emerges as a critical imperative for public health.

According to Fortain [29], we can distinguish different sources of particles in an URS. First, railway activity generates a large quantity of particles. They can have various origins: equipment wear (brakes, wheel/rail and pantograph/catenary friction, ballast) primarily producing metallic particles, network wear (wear of walls, seats, and especially floor coverings), and nightly maintenance work specific to railway activity [30]. Moreover, the mechanism of the passenger-flow source is the resuspension of airborne particles by passenger movements [31]. Finally, as station openings and ventilation systems are often located close to traffic routes, they serve as pathways for outside air to enter the station, thereby introducing pollutants into the URS.

Several studies, including those conducted by Barmparesos et al. [5] and Kam et al. [7], have put forth measurements of PM concentrations in URS. Findings show that there are significant variations from one URS to another, and from one day to another within a given URS. Nevertheless, Raut et al. [12], Birenzvige et al. [32], Querol et al. [33], and Walther and Bogdan [34] have demonstrated different diurnal cycle patterns between weekdays and weekend periods. On weekdays, a regular bimodal pattern of train traffic is observed with two rush periods (morning and late afternoon) and minimum levels during the night leading to the similar PM concentration daily pattern. During weekends, the concentrations remain relatively constant during daytime with lower levels in comparison to weekdays. Furthermore, they found that the daily PM concentration pattern and the timing of the diurnal cycle are remarkably consistent from one day to the next, despite observing day-to-day fluctuations in PM concentration levels.

According to the actual knowledge of the authors, no study has addressed the establishment of a baseline daily profile of particle concentration reflecting the weekday and weekend patterns. The present work introduces a clear methodology to provide a typical daily profile of particle concentrations within a URS based on long term measurements performed in different URS, aiming to surpass simple profile averaging.

2. Experimental section

This study relies on the existence of a unique dataset spanning over five years, recorded at quarter-hourly intervals. This extensive and detailed dataset allows for precise analysis, setting the foundation for the insights presented in our study.

Table 1
URS characteristics.

	Characteristics
RER line	Line C
Mechanical ventilation system	No
Platform depth	~10 m below ground
Operation start	1854
Ballast	Yes
Passengers number (#/year)	982 800
Tracks	2
Platform	1
Length	250 m
Platform screen door	No
Train type	Z5600, Z8800, Z20500, Z20900

Note: RER line C is one of the lines of the Réseau Express Régional (RER) in Paris, which serves the Paris metropolitan area and its suburbs.

2.1. Measurement site

This study is based on experimental data in one Parisian URS (Table 1) within the SNCF network. This specific station has been selected for the development of the methodology because it is one of the three URS in the SNCF network that has been equipped with instrumentation (**Section 2.2**) for continuous data measurement since April 2018. This extended period of data collection allows a comprehensive analysis of air quality trends over time. 24 other stations were equipped for shorter periods ranging from one to three weeks. This specific station is also among the SNCF URS with the highest PM concentrations. Finally, this station has been the subject of experimental interventions, including the installation of filtration systems in 2019 and 2020. The methodology developed within this study will eventually allow to assess the effectiveness of these filtration systems.

No assumption is made that this station is representative of the URS within the Paris urban train stations network. This will be examined in a future global analysis of all PM levels inside the 27 stations.

The schematic plan of the URS is shown on the Fig. 1 with the main characteristics described in Table 1. The station consists of two train tracks (Northbound, from left to right; Southbound, from right to left; and a central platform) and the length of the central platform is approximately 250 m.

2.2. Instrumentation

The particle measurements were conducted using two TEOM FDMS 1405 F particle analyzers, to measure PM_{10} and $PM_{2.5}$ according to standard NF EN 12341 by FDMS (measurement by micro-balance, considering the volatile fraction of particles). These analyzers utilize a tapered element oscillating microbalance (TEOM) coupled with a filter dynamic measurement system (FDMS), ensuring high short-term precision (mean concentration values at 15-minute intervals). One of the two analyzers is equipped with a PM_{10} US EPA sampling head, while the other includes a PM_{10} US EPA sampling head in addition to a Sharp Cut Cyclone (SCC) for $PM_{2.5}$ sampling. Both instruments were positioned in the middle of the platform (Fig. 1) and were sampling at 2.2 m above the platform floor.

In addition to the TEOMs, a TSI Q-TRAK device was installed adjacent to them, enabling the measurement of temperature, relative humidity, and CO_2 concentrations. Similar to the PM concentrations, measurements of these parameters were taken at a quarter hourly time step.

AIRPARIF (https://www.airparif.asso.fr/, last access: May 16, 2024), a French Air Quality Monitoring Association in the Île-de-France region, conducted the data collection since April 2018. Regular

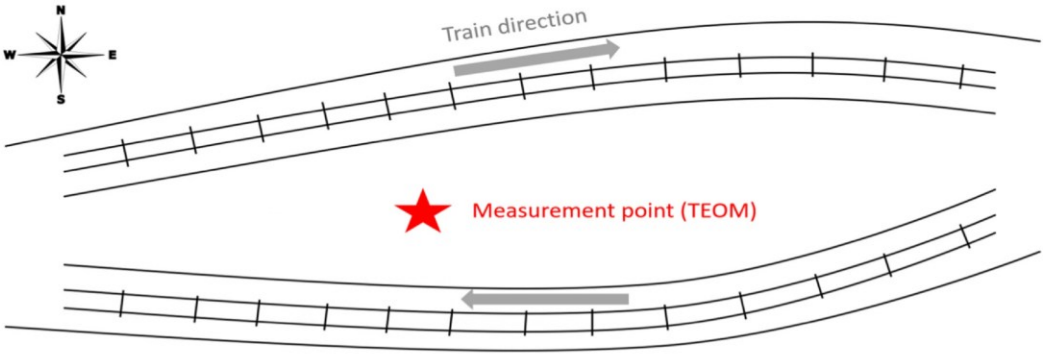


Fig. 1. Plan of the considered URS.

verification, maintenance, and calibration operations ensure that the collected data are of satisfactory precision, accuracy, completeness, comparability, and representativeness. A validation process by qualified personnel consists of two mandatory steps: a technical validation (conducted daily), and an environmental validation (conducted on a weekly basis). Invalidation may result from technical issues with the analyzer, external events (such as power outages) rendering the data unrepresentative, etc.

Data exploitation is based on validated readings. A data point is considered valid if at least 75 % of its constituent elements are valid. For instance, an hourly average is computable if at least 75 % of the 15-minute data points are valid, whether consecutive or not within the hour [35].

2.3. Dataset

The dataset for the considered URS covers the period from April 2018 to the present day. For the purposes of this study, we focus on the data spanning from April 2018 to December 2022. In this study, we focus on PM_{10} fraction, but similar work has been performed on $PM_{2.5}$ fraction, giving similar results.

Fig. 2 illustrates the timeseries of PM_{10} concentrations at a quarter hourly time steps within the specified URS, highlighting notable periods that have influenced the dynamics of the RER C line, including strikes (green band), COVID-19 lockdowns (orange band), summer works (blue band), and particle filtration experiment periods (grey band) (Table S1, supplemental material). The descriptive statistics for the PM_{10}, $PM_{2.5}$, and comfort parameters (CO_2, relative humidity, temperature) are presented in Table 2.

3. Establishment of typical daily profile PM concentration

Upon examining one week of data, Fig. 3 clearly illustrates the diurnal cycle patterns of PM_{10} between weekdays and weekend periods in the URS as discussed by Raut et al. [12] and Walther and Bogdan [34]. Throughout Monday to Friday (grey area), the daily PM_{10} concentration during daytime follows a similar trend and decreases during the weekend (green area). During weekdays, two peaks of PM_{10} concentration occurring generally during rush periods (in average $\sim 300\ \mu g/m^3$ against $\sim 320\ \mu g/m^3$ [12]) are observed followed by reaching its minimum levels ($10\ \mu g/m^3$) during the night after traffic stop. The PM_{10} concentration begins to increase at 5:30 am corresponding to the train operation start, reaching its first peak at 10:00 am. It then decreases at 01:00 pm (around $250\ \mu g/m^3$), followed by a subsequent increase, reaching its second peak at 07:00 pm, and finally decreases during the night. During weekends, the concentrations remain almost constant during daytime (in average $\sim 150\ \mu g/m^3$, against $\sim 120\ \mu g/m^3$ [12]).

Moreover, the one-week dataset displayed on Fig. 3 reveals deeper insights, as previously highlighted [12,34]. For instance, the overnight peaks circled in blue are attributed to nocturnal work within the station and the passage of diesel trains, while the daytime peaks circled in red remain unexplained.

Typically, discussions about the typical profile of PM concentration involve calculating only the average, often overlooking the fact that some profiles deviate from this pattern. This oversight can lead to an inaccurate typical profile, especially when working with limited datasets. In this section, we provide a concise overview of the rigorous methodology used to establish the typical profile of PM concentration,

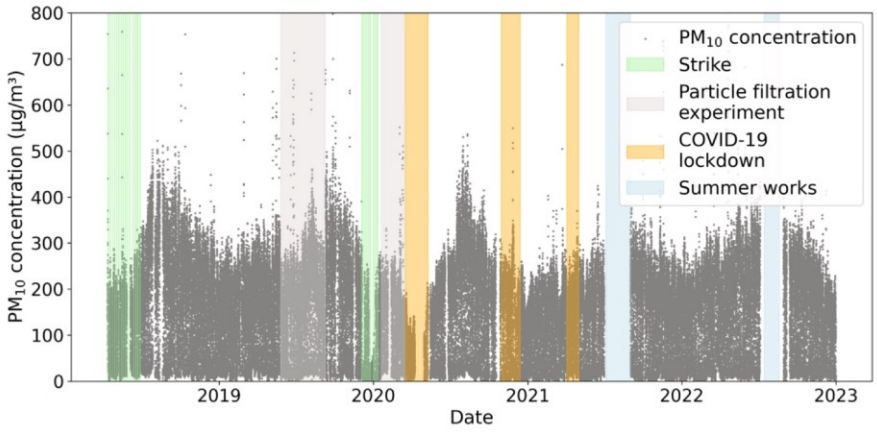


Fig. 2. Timeseries of PM10 concentrations (2018–2022). Shaded area corresponds to notable periods: strikes (green bands), COVID-19 lockdowns (orange bands), summer works (blue bands) and particle filtration experiments (grey bands).

Table 2
Summary statistics of collected data during the considered period.

Variables	Data points	Unit	Range (min, max)	Range (25th, 75th percentile)	Mean	SD
PM_{10}	141 428	$\mu g/m^3$	(0.1, 4492.9)	(60.1, 214.8)	146.6	102.2
$PM_{2.5}$	138 741	$\mu g/m^3$	(0.1, 2684.2)	(29, 84.3)	60.2	42.2
CO_2	143 817	ppm	(351, 1456)	(436.3, 526)	494.9	83.1
RH	142 767	%	(26.6, 90.4)	(51, 66)	58.5	10.3
T	142 691	°C	(8.6, 31.2)	(14.4, 22.9)	18.3	4.9

Note: SD = standard deviation; RH = relative humidity; T = temperature

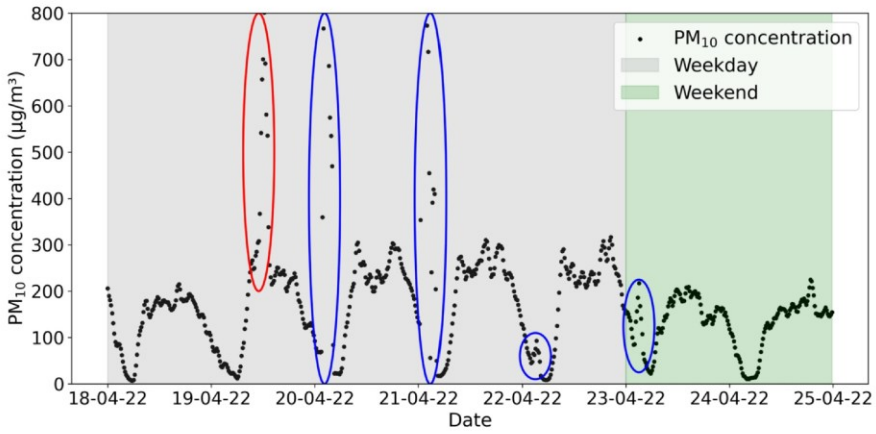


Fig. 3. PM_{10} concentration time profile for one week (18/04/2022–25/04/2022) inside the considered URS.

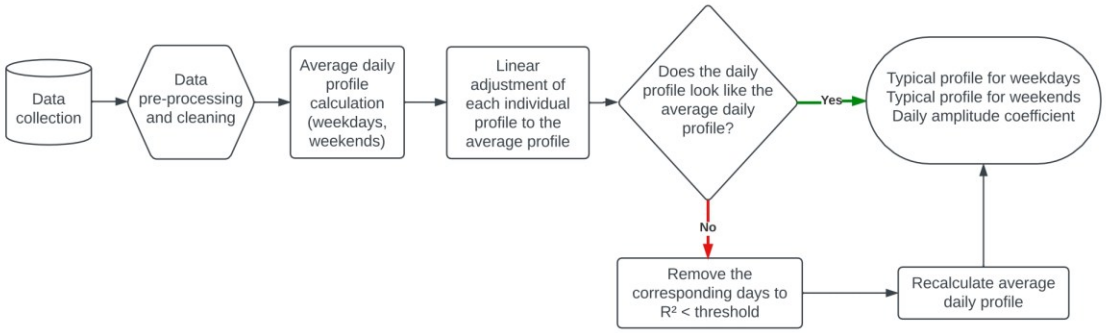


Fig. 4. Flowchart for the establishment of typical profile PM concentration.

visually presented on the flowchart in Fig. 4. Weekdays and weekends are treated separately due to different train schedules during these two periods, leading to different PM_{10} profiles (Fig. 3).

The methodology involves three main steps (Fig. 4): data preprocessing and cleaning (**Section 3.1**), profile fitting (**Section 3.2**), and statistical profile exclusion (**Section 3.3**). Thus, this methodology yields three usable outputs: a typical profile for weekdays, a typical profile for weekends, and a Daily Amplitude Coefficient (hereafter referred to as "DAC") associated with each daily profile within the considered period. The following subsections explain each step of our methodology in details.

3.1. Data preprocessing and cleaning

Dataset often presents gaps and noise, especially concerning PM concentration in URS. These issues may arise from missing values, outliers, and uncertainties in measurement, alongside exceptional events that could disrupt the data analysis process. Addressing these challenges is crucial as missing values and errors can greatly impact the accuracy of analytical techniques and the interpretation of the results. Consequently, the first step of our methodology involves data preprocessing, which encompasses the exclusion of notable periods and the identification and removal of outliers from our initial dataset.

Between April 2018 and December 2022, normal train operations of the considered URS have been disrupted several times (Fig. 2, Table S1 supplemental materials). During those periods, PM_{10} concentrations are not representative of URS regular functioning and normal concentration levels. Consequently, these periods were completely removed from our database. This step resulted in the removal of 21 % of the initial data (Table 3).

As previously noted, PM concentrations within URS tend to be higher than those outdoors. This circumstance results in a rapid degradation of instrument response over time. Moreover, the measurements of both particle fractions are carried out by two distinct TEOMs. Consequently, in our database, instances occur where $PM_{2.5}$ concentrations are higher than those of the PM_{10} fraction, especially during early morning when the concentrations are low (down to 5 $\mu g/m^3$) due to the absence of train operations. Therefore, all data points corresponding to $PM_{2.5}$ larger than PM_{10} concentrations were excluded from the initial database. This step leads to the removal of 4 % invalidated data (Table 3).

Table 3
Summary of data cleaning steps and proportion of data points removed.

Data cleaning steps	Proportion of data points removed from initial database
Strikes	4 %
COVID − 19 lockdowns	7 %
Summer works	1 %
Particle depollution experiment	9 %
Invalidated data	4 %
Outliers	1 %
Total	26 %

Outlined in Fig. 3, the dataset highlights outliers during nighttime hours (blue circle), accounting for 0.25 % of the initial database, and during daytime (red circle), accounting for 0.75 %. To address these outliers, various statistical-based approaches were tested, including descriptive statistics (histogram, boxplot: Interquartile Range), statistical tests (Z score), proximity-based methods (K-Nearest neighbor, Local Outlier Factor), and tree-based approaches (Isolation Forest). Among these, Isolation Forest proved to be the most satisfactory solution for our dataset. It successfully detects outliers not only during daytime (as shown by the peaks circled in red in Fig. 3) but also during nighttime (as evidenced by the peaks circled in blue in Fig. 3). In this study, we employed this approach, an unsupervised machine learning algorithm pioneered by Liu et al. [36]. This algorithm specializes in identifying anomalies within a dataset, specifically outliers that significantly differ from most of other data points. We used a dataset with specific characteristics: size of $n = 141428$ data points, dimensionality of $d = 1$, and contamination rate of 0.004. The contamination rate was set based on our data volume, as it reflects the proportion of data points expected to be anomalies. This step leads to the removal of 1 % of outliers (Table 3).

One can also note that holidays may lead to a reduced train and traveler frequency within the considered URS. To understand the impact of those breaks in comparison to normal operation, we selected the month when holidays occur (winter – February, spring – April, summer – July, fall – October breaks) across the entire period. During the months, we segregated the observed PM_{10} concentrations into two categories: those recorded during the holidays and those during the remainder of the month. Comparing the PM_{10} levels between these two periods revealed almost no difference, with an average variation of about 5 %, except for summer due to insufficient July data availability. Therefore, holiday periods are considered as suitable for retention in the dataset. To make this determination, we referred to the data shown in the supplementary material section (Figure S1), which compares periods with and without breaks at the considered URS.

Table 3 presents a comprehensive overview of the data cleaning result. It is important to note that the majority of removed data points are attributed to station operation, accounting for approximately 21 % of the initial database, while only 5 % are related to raw data. Overall, 26 % (36 089) of data points have been excluded from the initial database, leading our clean database with 105 339 data points of PM_{10}.

3.2. Profile fitting

In this section, our goal is to go beyond simple profile averaging. Indeed, it is imperative to refine the dataset by excluding profiles that are less representative. This entails aligning individual daily profiles with the overall daily average profile and removing those that significantly deviate from it. In essence, this step involves the identification of individual daily profiles that do not conform to the average daily profile.

An initial average daily profile (red dotted line for weekdays, Fig. 5a) is calculated, point by point at a quarterly hourly time step, using all daily profiles remaining after the data preprocessing and cleaning (**Section 3.1**).

Each individual day is then compared to this initial average daily profile. For this comparison, we perform a simple linear regression constrained to zero. This choice of constraint is based on the fact that all data points fall within a range of higher concentration, approximately around 200 $\mu g/m^3$ (Fig. 5). This model helps to establish the relationship between each time step (quarterly time step) of the individual profile (the dependent variable) to the corresponding time step of the average daily profile (the independent variable). The resulting slope, referred to hereafter as the daily amplitude coefficient (DAC), is calculated over 96 (4*24) data points for each individual profile. This DAC quantifies the amplitude of an individual profile relative to the average daily profile. When DAC is greater than 1 (lower than 1), the concentration during the specific day is higher (lower) than the average PM_{10} concentration profile.

As an example, Fig. 5b illustrates the linear regression model applied to the profile on November 10, 2022. The resulting DAC for this specific day is calculated to be 1.33, indicating that the amplitude of the profile on that day exceeds the initial average daily profile. Furthermore, the high coefficient of determination ($R^2 = 0.96$) for this individual day suggests a strong fit to the initial average daily profile.

3.3. Statistical profile exclusion

Once each individual daily profile was adjusted to the initial average daily profile, we employed a filtering concept based on R^2 to exclude profiles that deviated significantly from the average. R^2 measures the agreement between each individual daily profile and the average daily profile, with a maximum value of 1 indicating a perfect match.

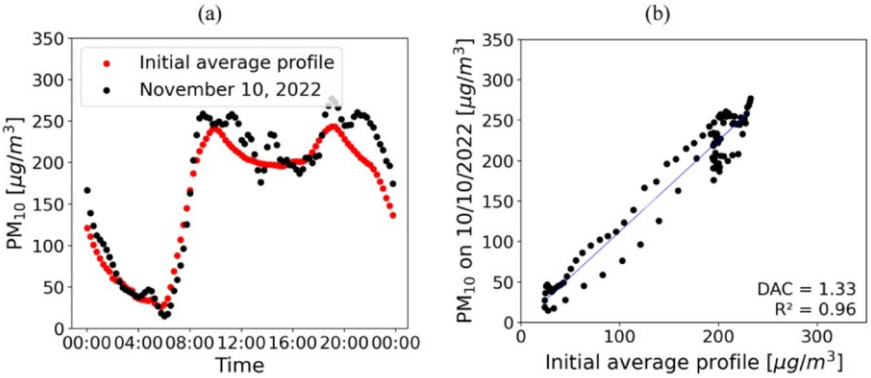


Fig. 5. (a) Profiles comparison: initial average profile (red dotted line) and the profile on November 10, 2022 (black dotted line), (b) Linear regression model, based on each PM measurements of the daily profile, for PM_{10} average profile and PM_{10} profile on that day.

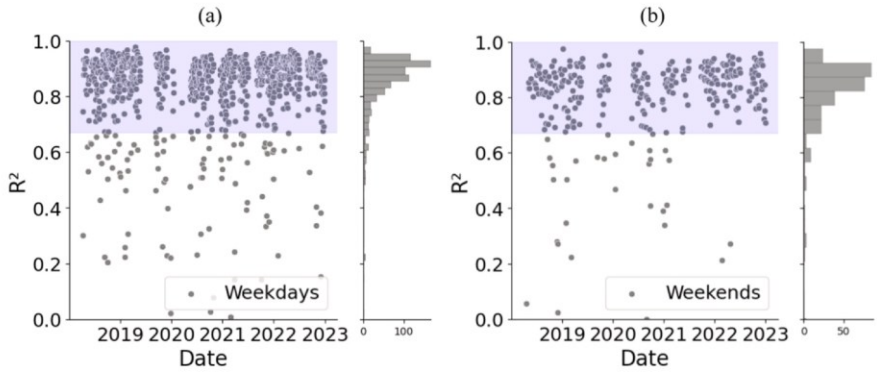


Fig. 6. Temporal variation of the daily goodness of fit R^2 for (a) weekdays and (b) weekends over the considered period. The joint plots along margins show its distribution.

In this study, individual daily profiles with $R^2 \leq 0.67$ were excluded, for both weekdays (Fig. 6a) and weekends (Fig. 6b), from the clean database. This threshold of 0.67 was chosen based on the distribution of R^2 values represented by the histogram on the right of each figure, to ensure that daily profiles closely match the average daily profile (Fig. 6). Through extensive testing, we determined that profiles with an R^2 below this value did not sufficiently align with the initial average profile. As with the DAC, this process was systematically repeated for each day within the study period. In the end of this process, 88 % (72 576 data points) and 90 % (25 632 data points) of profiles were retained (blue band, Fig. 6a and b) for weekdays and weekends, respectively, and two new average daily profiles were recalculated, representing the ultimate typical daily profiles (weekdays and weekends), as outlined in the flowchart in Fig. 4.

4. Results and discussions

4.1. Daily profiles

As previously mentioned, the aim of this study is to establish a clean daily profile. By meticulously applying all steps outlined in our methodology (**Section 3**), Fig. 7(a and b for weekdays and weekends, respectively) illustrates our reference daily profile (blue shaded areas represent one standard deviation), alongside initial average profile (red dotted line).

Weekdays and weekends exhibit distinct patterns, as discussed by Raut et al. [12] and Walther and Bogdan [34]. On weekdays, two peaks are observed during the morning (10:00 am) and evening (07:00 pm) rush hours, with the maximum amplitude of the profile reaching around 250 µg/m^3 for the considered URS. The average daytime PM concentration on weekdays is 193 µg/m^3. On the weekend, lower PM concentrations are observed, with an average daytime PM concentration of 157 µg/m^3 and no observable peaks, remaining almost constant during daytime (around 170 µg/m^3). Indeed, train operation differs between these two periods of the week, with trains operating from 05:00 am until 11:30 pm on weekdays and from 05:00 am until 01:00 am the following day on weekends. Moreover, during weekdays, trains run approximately every 8 minutes during rush hour and every 15 minutes outside of peak times explaining the two observed peaks during weekdays. While, during weekends, the frequency remains all daylong to approximately every 1 per 15 minutes, leading to alterations in the dynamics of the URS.

At the start of train operations (05:00 am), there is a one-hour delay before PM_{10} concentrations begin to rise (06:00 am) for both weekdays and weekends. This observation aligns with the findings of Raut et al. [12] and Birenzvige et al. [32], who noted that concentrations closely track traffic intensity but with a delay of approximately an hour. Indeed, their findings emphasize that this delay is primarily attributed to the influence of particulate dispersion by resuspension caused by traffic intensity and an increasing number of passengers.

Regarding the difference between the final reference profile and the initial average profile, we computed the Mean Absolute Error (MAE). The MAE values were 2.9 µg/m^3 for weekdays and 3.7 µg/m^3 for weekends. These results underscore the varying degrees of divergence between the initial average and reference daily profiles. The smaller difference between these profiles (weekdays and weekends) can be attributed to the volume of data.

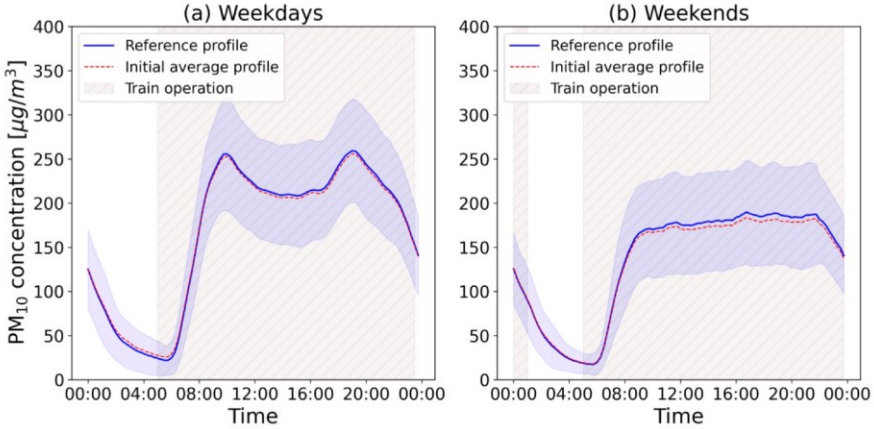


Fig. 7. Comparison between the reference profile (blue line) and the initial average profile (red dashed line) for (a) weekdays, and (b) weekends. Blue shaded areas represent one standard deviation ($\sigma = 1$).

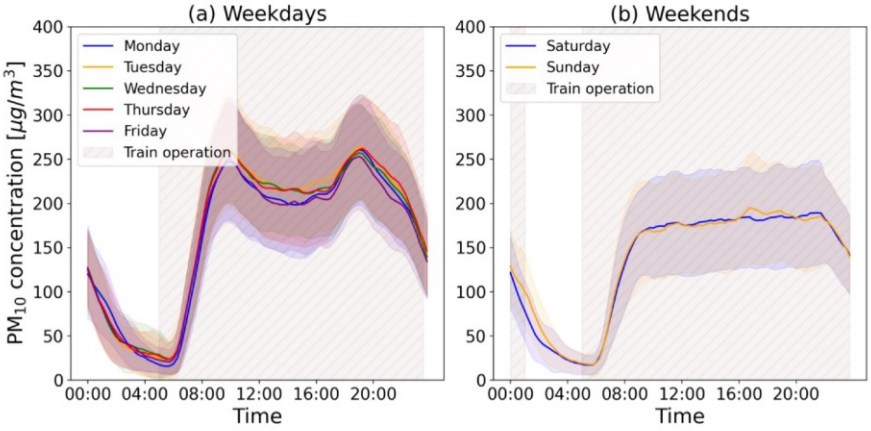


Fig. 8. Typical profiles at the considered URS for (a) weekdays and (b) weekends. Hatched shaded area indicates the train operation. Shaded areas represent one standard deviation (σ=1).

Examining the day-of-week profiles shown in Fig. 8, it is notable that the profiles for Monday through Friday exhibit significant similarities, as do those for Saturday and Sunday. All profiles depicted in Fig. 8(a) and (b) fall within the shaded region representing one standard deviation. Moreover, the typical profiles for Monday to Friday share similarities due to consistent train operations from 05:00–23:00 (hatched shaded area, Fig. 8a), as do those for Saturday and Sunday, operating from 05:00–01:00 (hatched shaded area, Fig. 8b). Across all these profiles, concentrations decrease from the end of service until reaching approximately 10 μg/m^3 in the morning around 05:00 am, clearly indicating that particle presence is strongly associated with station operation. Indeed, the MAE between PM_{10} final profiles observed on Tuesdays and Fridays, representing the highest and lowest concentration profiles respectively, is only 10 μg/m^3. Similarly, the MAE between PM_{10} final profiles observed on Saturdays and Sundays on weekends is merely 4 μg/m^3. These observations attest that our reference profile is sufficiently robust to serve as a baseline reference for assessing air quality in URS. It is worth mentioning that the concentrations values observed are consistent with the published concentrations of particles on the platforms of underground train or subway stations in various cities, as given in the global list compiled by Passi et al. [8] in 2021.

4.2. Monthly analysis

In this section, we will initially verify whether the profile remains consistent to our reference profile (blue line, Fig. 7) across all months. Subsequently, we will move into the analysis of the monthly variations of the reference profile. Fig. 9 illustrates the profile calculated for each month over the entire period. These typical daily profiles for each month were derived by averaging all the data observed during the specific month of the full dataset, for weekdays and weekends. PM concentration levels notably increase during the months of July, August, September, and October for both weekdays and weekends, followed by a decrease during the months of December, January, February, and March.

During weekdays, the absolute maximum PM concentration level (~320 μg/m^3) is observed during the morning peak (10:00 am) in August, while the daytime minimum (~175 μg/m^3) occurs between the two peaks of the day (02:00 pm) in January. On weekends, however, the PM concentration remains relatively constant during daytime, with pronounced levels reaching 275 μg/m^3 and 250 μg/m^3 for August and July, respectively.

While the amplitude varies between each period due to seasonal variation, our analysis reveals that the reference profile pattern generally remains consistent throughout the months. Table 4 presents the statistical analysis of the DAC during each month (median value, 25th percentile, and the 75th percentile), as well as for the full period (2018–2022), so as to check for the stability of the daily profiles on the long term. Blanks corresponds to instances where no data were available, or data were removed during preprocessing and cleaning (**Section 3.1**), or excluded from statistical profile exclusion criteria (**Section 3.3**) or not representative due to having less than 50 % of the data points available.

It is worth noting that during August and July, the concavity between the two peaks of the day (10:00 am to 07:00 pm) on weekdays is less pronounced. This observation can be tentatively explained by changes in passenger routines during these months or may be artificially caused by missing data points. Specifically, data for July and August were only available in 2018 and 2019 with more than 50 % of the data points available (Table 4). Moreover, the reduced number of data points, which consequently leads to increased variability, may also explain the high amplitude observed during weekends (grey and cyan curves, Fig. 9b).

4.3. Long term evolution of PM concentration in the URS

Fig. 10 shows the original variation of quarter hourly PM_{10} concentration together with the DAC within the considered period. With access to a unique database covering approximately five years data, we can now quantify the impact of seasonality on PM_{10} concentrations. A relatively constant trend is observed throughout the observed period, with higher values during the summer (up to 400 μg/m^3) and lower values during winter (down to 250 μg/m^3). These findings are also evident from the temporal variation of DAC (blue dots, Fig. 10) and the monthly distribution of DAC (Fig. S3, supplemental material).

During summer, maximum DAC values are approximately 1.5, whereas during winter, these values drop to around 0.5. This suggests a discernible departure from the typical trend of PM concentration based on the season. The typical seasonal profile at the considered URS is illustrated in Fig. S2 (supplementary material).

Based on these findings, we can infer approximately a 1.5-fold difference between summer and winter, implying that knowledge of the typical profile of one season might enable us to approximate the profile of the other.

Based on Table 4, we followed a similar process for the PM_{10} outdoor data (as referenced in Table S2, supplemental material). The PM_{10} outdoor data were collected from the measurement station closest to the considered URS. Fig. 11 illustrates the radar plot of monthly distribution of the DAC mean value of PM_{10} within the considered URS

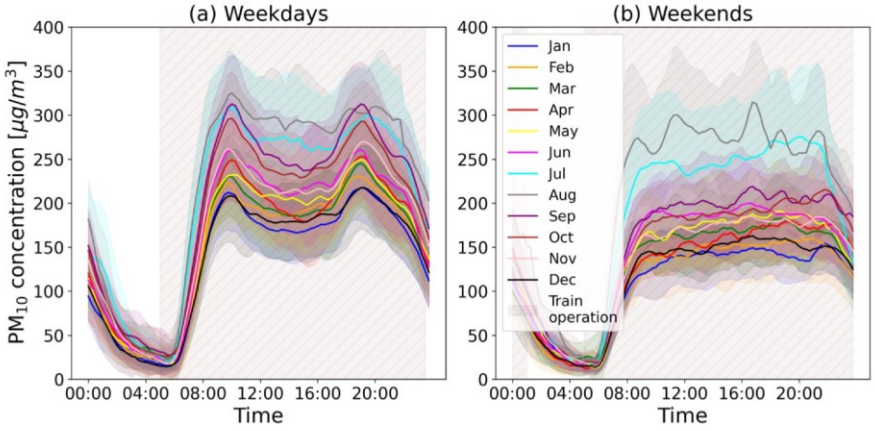


Fig. 9. Typical monthly profile for (a) weekdays and (b) weekends. Shaded areas represent one standard deviation (σ = 1).

Table 4
DAC averaged statistics for each month (median, 25th percentile, and 75th percentile).

DAC Median (25th -75th percentile)	2018	2019	2020	2021	2022	2018–2022
January	-	0.88 (0.84–0.99)	-	0.72 (0.69–0.79)	0.84 (0.80–0.90)	0.83 (0.74–0.89)
February	-	0.96 (0.87–1.08)	-	0.81 (0.72–0.89)	0.94 (0.87–0.98)	0.89 (0.80–0.96)
March	-	0.95 (0.87–1.06)	-	0.82 (0.74–0.93)	1.02 (0.94–1.15)	0.95 (0.82–1.06)
April	-	1.00 (0.94–1.09)	-	-	0.98 (0.90–1.05)	0.99 (0.93–1.08)
May	-	1.13 (0.90–1.18)	0.69 (0.62–0.81)	0.87 (0.76–0.91)	1.14 (1.07–1.26)	0.99 (0.81–1.16)
June	0.95 (0.86–0.99)	-	0.96 (0.89–1.13)	0.98 (0.84–1.04)	1.17 (1.09–1.30)	1.01 (0.90–1.17)
July	1.31 (1.14–1.61)	-	1.23 (1.00–1.48)	-	-	1.27 (1.07–1.51)
August	1.43 (1.32–1.55)	-	1.35 (1.26–1.50)	-	-	1.38 (1.31–1.54)
September	1.29 (1.18–1.40)	-	1.16 (1.06–1.26)	1.10 (1.05–1.15)	1.15 (1.09–1.21)	1.16 (1.09–1.26)
October	1.20 (1.11–1.29)	1.27 (1.20–1.43)	-	0.99 (0.96–1.06)	1.12 (1.03–1.21)	1.14 (1.02–1.27)
November	1.11 (0.96–1.23)	1.10 (1.04–1.19)	-	0.96 (0.86–1.01)	0.97 (0.91–1.03)	1.02 (0.92–1.11)
December	0.98 (0.97–1.06)	-	-	0.88 (0.81–0.95)	0.72 (0.66–0.78)	0.87 (0.74–0.98)

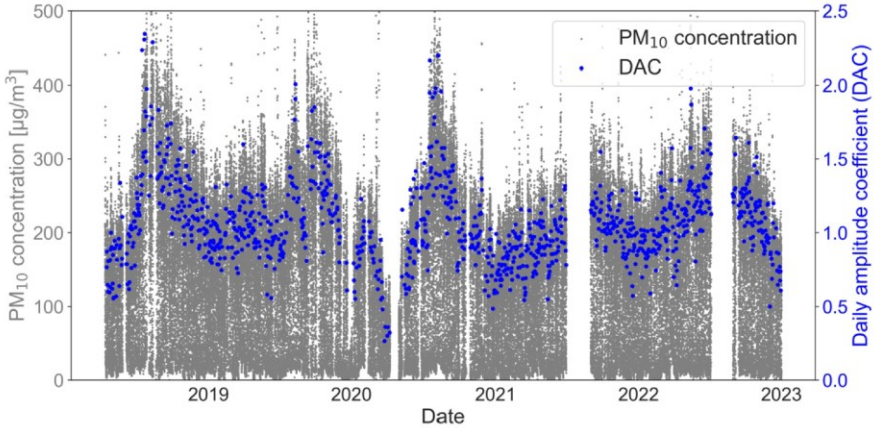


Fig. 10. Temporal evolution of DAC (blue dots) with original PM_{10} concentration (grey dots) (2018–2022).

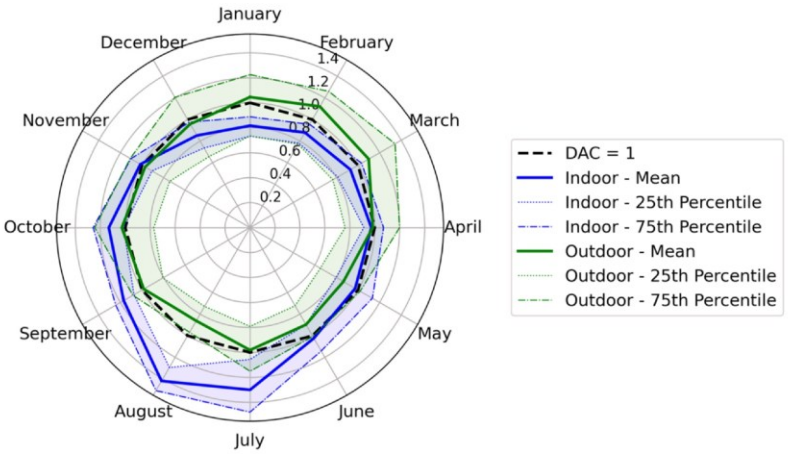


Fig. 11. Radar plot showing monthly distribution of the DAC mean value of PM_{10} within URS (blue) and the DAC mean value of outdoor PM_{10} (green) from the nearest station of the considered URS. The solid line represents the mean DAC value, the dotted line represents the 25th percentile, and the dash-dotted line represents the 75th percentile of DAC for the period 2018–2022.

(blue) for the entire period (2018–2022), alongside the DAC mean value of outdoor PM_{10} data (green) from the nearest measurement station for the same period. The black dashed line represents the reference level of PM_{10} ($\overline{DAC} = 1$). As already pointed out, PM_{10} concentrations (blue curves) are higher than the average level of the profile during the summer months (July, August, September), peaking in August ($\overline{DAC} \approx 1.3$), whereas they are lower in winter (December, January, February), with the lowest levels in January ($\overline{DAC} \approx 0.6$). This is not the case for the outdoor PM_{10} concentrations, exhibiting higher levels between January and March, peaking in February ($\overline{DAC} \approx 1.1$), and lower levels in the summer, reaching their minimum in August ($\overline{DAC} \approx 0.8$). For a detailed view of daily PM_{10} outdoor, see Fig. S4 (supplemental materials).

Furthermore, Fig. 11 shows that during the spring months (April, May, June), the mean DAC value of indoor PM_{10} consistently remains at 1. This suggests that the ideal period for conducting experiments within the considered URS to obtain a representative reference profile is during these months, as the indoor PM_{10} levels precisely match the reference profile level. It is worth to note that while this observation holds true for this specific URS, its applicability to other URSs requires further generalization.

These observations show that PM seasonal pattern inside a URS is entirely different from what is observed outdoors, as seen in the studies of H. Luo et al. [37] in China or Gokul et al. [38] in India, who stipulated that the amplitude of PM concentration in summer is smaller than that in winter. Additionally, contrary to the results of Kwon et al. [39] in Seoul, Korea or of Van Ryswyk et al. [4] in Montreal and Vancouver, Canada, these findings suggest higher concentrations during the warmer period.

Higher concentrations during the warmer period could be attributed, at least in part, to the train braking system and impacts of meteorological parameters. Walia et al. [40] stated that brake energy for the cast iron brake blocks for driving trailer during winter is lower as compared to summer.

5. Conclusions

In this paper, we presented a novel methodology for establishing a reference PM profile within an URS. In the first step, a standardized data preprocessing and cleaning process is performed (excluding specific operational periods, inconsistent data, and outliers). A pivotal component of this methodology is robust data cleaning, where we introduced an AI-based approach, to effectively detect outliers. Our results demonstrated the high accuracy of this tool, particularly in identifying daytime and nighttime outliers caused by station exploitation and nocturnal work within the station or the passage of diesel trains, respectively. Only 26 % of the initial database were removed, with approximately 21 % related to URS operation and only 5 % related to raw data.

In the second step, an iterative averaging method is implemented to obtain the reference profile over the entire considered period. It is mainly based on performing a simple yet effective linear regression model to identify individual daily profiles that deviate from the reference profile pattern and transform into a daily amplitude coefficient (DAC) for easier analysis. The DAC, as a daily metric of the pollution levels, enables the analysis of temporal trends, and facilitates the comparison with other data with other acquisition frequency. This model leverages the goodness of fit, allowing us to move beyond mere profile averaging and identify those profiles that do not align with the initial average profile. Profiles corresponding to R^2 values larger than 0.67 were selected from our dataset.

We observed two distinct peaks in PM concentration during weekdays, mainly occurring one hour after rush hours, with an average daytime PM concentration of 193 $\mu g/m^3$. In contrast, weekends showed lower PM concentration levels, averaging 157 $\mu g/m^3$, with no observable peaks. Additionally, this work relies on the utilization of a vast database spanning approximately five years, with data collected at quarter-hourly intervals. This extensive dataset allowed us to observe long-term trends, such as higher PM concentration values in summer and lower values in winter, with a noticeable stability in PM profile at both monthly and yearly scales. Furthermore, results suggest that the optimal period for conducting experiments within the considered URS to obtain a representative reference profile is during the spring months (April, May, June). Moreover, we found the indoor PM concentrations are decorrelated from the outdoor concentrations, highlighting different dynamics within the URS compared to the external environment. It should be emphasized that this versatile methodology, offering flexibility and applicability; should be applied to different particle classes at various URS. What sets this work apart is our departure from the traditional approach of solely averaging daily profiles to establish a reference at a given URS. Instead, our methodology rigorously filters out profiles that do not conform to the reference profile pattern. Our results highlight the effectiveness of our methodology in establishing a reference profile at the considered URS across various periods. The application of our methodology to other URS can be a valuable direction for future research or extension of the current methodology.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

Acknowledgements

The authors greatly acknowledge SNCF for funding the PhD of Valisoa Rakotonirinjanahary (ANRT CIFRE PhD grant n°2021/0856, managed by the National Research Technology Agency). Authors are also grateful to AIRPARIF for conducting data collection. This research is part of the CaPPA project (Chemical and Physical Properties of the Atmosphere), funded by the French National Research Agency (ANR) through the PIA (Programme Investissement d'Avenir) with the contract ANR-11-LABX-005–001. Additionally, it is affiliated with the Institute of Multidisciplinary Research in Environmental Sciences (IRePSE Fed 4129) and the CPER ECRIN research project, funded by the French Ministry of Higher Education and Research, the Hauts-de-France Region, and the European Regional Development Fund.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.indenv.2024.100040.

References

[1] World Health Organization, WHO global air quality guidelines: particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide: executive summary, World Health Organization, 2021 accessed December 8, 2023<https://iris.-who.int/handle/10665/345334>.
[2] P. Mao, J. Li, L. Xiong, R. Wang, X. Wang, Y. Tan, H. Li, Characterization of Urban Subway Microenvironment Exposure—A Case of Nanjing in China, Int. J. Environ. Res. Public Health 16 (2019) 625, https://doi.org/10.3390/ijerph16040625.
[3] Y.-H. Cheng, Z.-S. Liu, J.-W. Yan, Comparisons of PM10, PM2.5, Particle Number, and CO2 Levels inside Metro Trains between Traveling in Underground Tunnels and on Elevated Tracks, Aerosol Air Qual. Res. 12 (2012) 879–891, https://doi.org/10.4209/aaqr.2012.05.0127.
[4] K. Van Ryswyk, A.T. Anastasopolos, G. Evans, L. Sun, K. Sabaliauskas, R. Kulka, L. Wallace, S. Weichenthal, Metro Commuter Exposures to Particulate Air Pollution and PM2.5-Associated Elements in Three Canadian Cities: The Urban Transportation Exposure Study, Environ. Sci. Technol. 51 (2017) 5713–5720, https://doi.org/10.1021/acs.est.6b05775.
[5] N. Barmparesos, V. D. Assimakopoulos, M.N. Assimakopoulos, E. Tsairidi, Particulate matter levels and comfort conditions in the trains and platforms of the Athens underground metro, AIMSES 3 (2016) 199–219, https://doi.org/10.3934/environsci.2016.2.199.
[6] M. Tu, U. Olofsson, PM levels on an underground metro platform: A study of the train, passenger flow, urban background, ventilation, and night maintenance effects, Atmos. Environ.: X 12 (2021) 100134, https://doi.org/10.1016/j.aeaoa.2021.100134.
[7] W. Kam, K. Cheung, N. Daher, C. Sioutas, Particulate matter (PM) concentrations in underground and ground-level rail systems of the Los Angeles Metro, Atmos. Environ. 45 (2011) 1506–1516, https://doi.org/10.1016/j.atmosenv.2010.12.049.
[8] A. Passi, S.M.S. Nagendra, M.P. Maiya, Characteristics of indoor air quality in underground metro stations: A critical review, Build. Environ. 198 (2021) 107907, https://doi.org/10.1016/j.buildenv.2021.107907.
[9] M. Braniš, The contribution of ambient sources to particulate pollution in spaces and trains of the Prague underground transport system, Atmos. Environ. 40 (2006) 348–356, https://doi.org/10.1016/j.atmosenv.2005.09.060.
[10] S.-B. Kwon, D. Park, Y. Cho, E.-Y. Park, Measurement of natural ventilation rate in Seoul metropolitan subway cabin, Indoor Built Environ. 19 (2010) 366–374, https://doi.org/10.1177/1420326X10367305.
[11] M.J. Ramos, A. Vasconcelos, M. Faria, Comparison of Particulate Matter Inhalation for Users of Different Transport Modes in Lisbon, Transp. Res. Procedia 10 (2015) 433–442, https://doi.org/10.1016/j.trpro.2015.09.093.
[12] J.-C. Raut, P. Chazette, A. Fortain, Link between aerosol optical, microphysical and chemical measurements in an underground railway station in Paris, Atmos. Environ. 43 (2009) 860–868, https://doi.org/10.1016/j.atmosenv.2008.10.038.
[13] A. Cartenì, F. Cascetta, S. Campana, Underground and ground-level particulate matter concentrations in an Italian metro system, Atmos. Environ. 101 (2015) 328–337, https://doi.org/10.1016/j.atmosenv.2014.11.030.
[14] J.D. Smith, B.M. Barratt, G.W. Fuller, F.J. Kelly, M. Loxham, E. Nicolosi, M. Priestman, A.H. Tremper, D.C. Green, PM2.5 on the London Underground, Environ. Int 134 (2020) 105188, https://doi.org/10.1016/j.envint.2019.105188.
[15] H. Marcus, Waiting experience at train stations, Journal of Medical Internet Research - J MED INTERNET RES (2011).
[16] R. Luo, H. Dai, Y. Zhang, P. Wang, Y. Zhou, J. Li, M. Zhou, L. Qiao, Y. Ma, S. Zhu, S. Hu, C. Huang, H. Shi, Association of short-term exposure to source-specific PM2.5 with the cardiovascular response during pregnancy in the Shanghai MCPC study, Sci. Total Environ. 775 (2021) 145725, https://doi.org/10.1016/j.scitotenv.2021.145725.
[17] M. Renzi, M. Stafoggia, P. Michelozzi, M. Davoli, F. Forastiere, A.G. Solimini, Short-term exposure to PM2.5 and risk of venous thromboembolism: A case-crossover study, Thromb. Res 190 (2020) 52–57, https://doi.org/10.1016/j.thromres.2020.03.008.
[18] L.H. Wyatt, A.M. Weaver, J. Moyer, J.D. Schwartz, Q. Di, D. Diaz-Sanchez, W.E. Cascio, C.K. Ward-Caviness, Short-term PM2.5 exposure and early-readmission risk: a retrospective cohort study in North Carolina heart failure patients, Am. Heart J. 248 (2022) 130–138, https://doi.org/10.1016/j.ahj.2022.02.015.
[19] T. Chen, F. Chen, K. Wang, X. Ma, X. Wei, W. Wang, P. Huang, D. Yang, Z. Xia, Z. Zhao, Acute respiratory response to individual particle exposure (PM1.0, PM2.5 and PM10) in the elderly with and without chronic respiratory diseases, Environ. Pollut. 271 (2021) 116329, https://doi.org/10.1016/j.envpol.2020.116329.
[20] R. Lei, F. Zhu, H. Cheng, J. Liu, C. Shen, C. Zhang, Y. Xu, C. Xiao, X. Li, J. Zhang, R. Ding, J. Cao, Short-term effect of PM2.5/O3 on non-accidental and respiratory deaths in highly polluted area of China, Atmos. Pollut. Res. 10 (2019) 1412–1419, https://doi.org/10.1016/j.apr.2019.03.013.
[21] L. Pini, J. Giordani, G. Gardini, C. Concoreggi, A. Pini, E. Perger, E. Vizzardi, D. Di Bona, C. Cappelli, M. Ciarfaglia, C. Tantucci, Emergency department admission and hospitalization for COPD exacerbation and particulate matter short-term exposure in Brescia, a highly polluted town in northern Italy, Respir. Med 179 (2021) 106334, https://doi.org/10.1016/j.rmed.2021.106334.
[22] B. Sun, J. Song, Y. Wang, J. Jiang, Z. An, J. Li, Y. Zhang, G. Wang, H. Li, N.E. Alexis, I. Jaspers, W. Wu, Associations of short-term PM2.5 exposures with nasal oxidative stress, inflammation and lung function impairment and modification by *GSTT1*-null genotype: A panel study of the retired adults, Environ. Pollut. 285 (2021) 117215, https://doi.org/10.1016/j.envpol.2021.117215.
[23] Y. Zhang, Z. Ding, Q. Xiang, W. Wang, L. Huang, F. Mao, Short-term effects of ambient PM1 and PM2.5 air pollution on hospital admission for respiratory diseases: Case-crossover evidence from Shenzhen, China, Int. J. Hyg. Environ. Health 224 (2020) 113418, https://doi.org/10.1016/j.ijheh.2019.11.001.
[24] C. Liu, G. Cao, J. Li, S. Lian, K. Zhao, Y. Zhong, J. Xu, Y. Chen, J. Bai, H. Feng, G. He, X. Dong, P. Yang, F. Zeng, Z. Lin, S. Zhu, X. Zhong, W. Ma, T. Liu, Effect of long-term exposure to PM2.5 on the risk of type 2 diabetes and arthritis in type 2 diabetes patients: Evidence from a national cohort in China, Environ. Int. 171 (2023) 107741, https://doi.org/10.1016/j.envint.2023.107741.
[25] D. Wood, D. Evangelopoulos, S. Beevers, N. Kitwiroon, K. Katsouyanni, Exposure to Ambient Air Pollution and the Incidence of Dementia in the Elderly of England: The ELSA Cohort, Int J. Environ. Res Public Health 19 (2022) 15889, https://doi.org/10.3390/ijerph192315889.
[26] E. Dimakakou, H.J. Johnston, G. Streftaris, J.W. Cherrie, Exposure to environmental and occupational particulate air pollution as a potential contributor to neurodegeneration and diabetes: a systematic review of epidemiological research, Int. J. Environ. Res. Public Health 15 (2018) 1704, https://doi.org/10.3390/ijerph15081704.
[27] J. Duchesne, L.-A. Gutierrez, I. Carrière, T. Mura, J. Chen, D. Vienneau, K. de Hoogh, C. Helmer, B. Jacquemin, C. Berr, M. Mortamais, Exposure to ambient air pollution and cognitive decline: Results of the prospective Three-City cohort study, Environ. Int. 161 (2022) 107118, https://doi.org/10.1016/j.envint.2022.107118.
[28] L. Chang, W.T. Chong, Y.H. Yau, T. Cui, X.R. Wang, F. Pei, Y.Q. Liu, S. Pan, An investigation of the PM2.5 concentrations and cumulative inhaled dose during subway commutes in Changchun, China, Int J. Environ. Sci. Technol. (Tehran) (2023) 1–14, https://doi.org/10.1007/s13762-023-04994-7.
[29] A. Fortain, Caractérisation des particules en gares souterraines, These de doctorat, La Rochelle, 2008. https://www.theses.fr/2008LAROS231 (accessed June 8, 2022).
[30] L. Chang, W.T. Chong, X. Wang, F. Pei, X. Zhang, T. Wang, C. Wang, S. Pan, Recent progress in research on PM2.5 in subways, Environ. Sci.: Process. Impacts 23 (2021) 642–663, https://doi.org/10.1039/D1EM00002K.
[31] C. Johansson, P.-Å. Johansson, Particulate matter in the underground of Stockholm, Atmos. Environ. 37 (2003) 3–9, https://doi.org/10.1016/S1352-2310(02)00833-6.
[32] A. Birenzvige, J. Eversole, M. Seaver, S. Francesconi, E. Valdes, H. Kulaga, Aerosol Characteristics in a Subway Environment, Aerosol Sci. Technol. 37 (2003) 210–220, https://doi.org/10.1080/02786820300941.
[33] X. Querol, T. Moreno, A. Karanasiou, C. Reche, A. Alastuey, M. Viana, O. Font, J. Gil, E. de Miguel, M. Capdevila, Variability of levels and composition of PM_{10} and $PM_{2.5}$ in the Barcelona metro system, Atmos. Chem. Phys. 12 (2012) 5055–5076, https://doi.org/10.5194/acp-12-5055-2012.
[34] E. Walther, M. Bogdan, A novel approach for the modelling of air quality dynamics in underground railway stations, Transp. Res. Part D: Transp. Environ. 56 (2017) 33–42, https://doi.org/10.1016/j.trd.2017.07.014.
[35] AIRPARIF, Mesures de la qualité de l'air intérieur sur les quais des gares souterraines franciliennes de la SNCF: Rapport final, AIRPARIF, 2020. https://www.airparif.asso.fr/etudes/2021/qualite-de-lair-interieur-quais-gares-sncf-rapport-final-2020 (accessed June 13, 2022).
[36] F.T. Liu, K.M. Ting, Z.-H. Zhou, Isolation Forest, 2008 Eighth IEEE Int. Conf. Data Min. (2008) 413–422, https://doi.org/10.1109/ICDM.2008.17.
[37] H. Luo, Y. Han, X. Cheng, C. Lu, Y. Wu, Spatiotemporal Variations in Particulate Matter and Air Quality over China: National, Regional and Urban Scales, Atmosphere 12 (2021) 43, https://doi.org/10.3390/atmos12010043.
[38] P.R. Gokul, A. Mathew, A. Bhosale, A.T. Nair, Spatio-temporal air quality analysis and PM2.5 prediction over Hyderabad City, India using artificial intelligence techniques, Ecol. Inform. 76 (2023) 102067, https://doi.org/10.1016/j.ecoinf.2023.102067.
[39] S.-B. Kwon, W. Jeong, D. Park, K.-T. Kim, K.H. Cho, A multivariate study for characterizing particulate matter (PM10, PM2.5, and PM1) in Seoul metropolitan subway stations, Korea, J. Hazard. Mater. 297 (2015) 295–303, https://doi.org/10.1016/j.jhazmat.2015.05.015.
[40] M.S. Walia, T. Vernersson, R. Lundén, F. Blennow, M. Meinel, Temperatures and wear at railway tread braking: Field experiments and simulations, Wear 440–441 (2019) 203086, https://doi.org/10.1016/j.wear.2019.203086.

Annexe 2 : Comparison of three methods for evaluating filtration system performance in underground railway stations

Valisoa M. Rakotonirinjanahary, Maryam Rhamani, Suzanne Crumeyrolle, Mateusz Bogdan, Benjamin Hanoune

Résumé :

La qualité de l'air dans les enceintes ferroviaires souterraines (EFS) constitue un défi de santé publique en raison des concentrations élevées de particules fines (PM). Diverses stratégies d'atténuation ont été mises en œuvre, notamment l'utilisation de matériaux non métalliques dans les systèmes de freinage des trains, la ventilation forcée dans les tunnels, les portes palières et les systèmes de filtration de l'air. Cette étude se concentre sur l'évaluation de la performance des systèmes de filtration de l'air dans une gare souterraine parisienne, en utilisant trois méthodes distinctes : le coefficient d'amplitude journalière (DAC), les modèles additifs généralisés (GAM) et les réseaux de neurones à mémoire à long et court terme (LSTM). Deux technologies de filtration ont été évaluées — Expérience 1 : ionisation positive et Expérience 2 : système de filtration à base d'eau — en se focalisant sur la réduction des concentrations de PM_{10} et de $PM_{2.5}$.

Les méthodes DAC et GAM fournissent des résultats de réduction sur l'ensemble de la période expérimentale, tandis que LSTM analyse uniquement les trois premiers jours. Le DAC capture efficacement les tendances à long terme, mais il peut être affecté par la disponibilité des données. Le GAM est idéal pour analyser les relations non linéaires et évaluer la performance des systèmes, bien qu'il ne révèle pas de tendances de réduction pendant la période expérimentale. Le LSTM excelle dans la prévision à court terme, capturant efficacement les modèles temporels complexes, mais il nécessite des données d'entrée substantielles et montre une incertitude croissante dans les prédictions à long terme.

Les résultats des trois méthodes indiquent que l'Expérience 1 surpasse l'Expérience 2 pour les deux classes de particules. Le DAC et le GAM montrent que les deux technologies de filtration sont plus efficaces pour réduire les concentrations de $PM_{2.5}$ que celles de PM_{10}, tandis que le LSTM indique l'inverse. Plus précisément, le DAC enregistre une réduction de 17,8 % ± 3,2 % pour l'Expérience 1 concernant les PM_{10}, contre seulement 1,6 % ± 0,4 % pour l'Expérience 2. Le GAM montre une réduction de 13,1 % ± 0,8 % pour l'Expérience 1, tandis que le LSTM indique une réduction de 26,8 % ± 6,1 %. Pour les $PM_{2.5}$, le DAC enregistre une réduction de 25,9 % ± 6,6 % pour l'Expérience 1, avec des réductions de 26 % ± 0,8 % et 22,6 % ± 4,8 % rapportées respectivement par le GAM et le LSTM.

Sur la base de ces résultats, nous recommandons une approche hybride : utiliser le DAC pour l'évaluation des tendances à long terme, le GAM pour l'analyse approfondie des interactions et l'évaluation des tendances étendues, et le LSTM pour la prévision à court terme. Cette stratégie intégrée offre une compréhension complète de l'efficacité des systèmes de filtration, soutenant ainsi une prise de décision éclairée pour améliorer la qualité de l'air dans les gares souterraines

Comparison of three methods for evaluating the performance of particle filtration systems in underground railway stations

Valisoa M. Rakotonirinjanahary[1,5,*], Suzanne Crumeyrolle[2], Maryam Rahmani[3], Mateusz Bogdan[4], and Benjamin Hanoune[1]

[1]Univ. Lille, CNRS, UMR 8522 – PC2A – Physicochimie des Processus de Combustion et de l'Atmosphère, F-59000 Lille, France

[2]Univ. Lille, CNRS, UMR 8518 – LOA – Laboratoire d'Optique Atmosphérique, F-59000 Lille, France

[3]Univ. Lille, Inria, CNRS, UMR 9189 CRIStAL, France

[4]AREP L'hypercube, 16 avenue d'Ivry, 75013 Paris, France

[5]SNCF Holding, 2 Place aux Etoiles, 93210 Saint-Denis, France

[*]Corresponding author: miadanavalisoa.rakotonirinjanahary@univ-lille.fr

Abstract

The air quality in underground railway stations (URS) is a public health challenge due to elevated concentrations of particulate matter (PM). Various mitigation strategies have been implemented, including the use of non-metallic materials in train braking systems, forced ventilation in tunnels, platform screen doors (PSD), and air filtration systems. This study focuses on evaluating the performance of air filtration systems in a Parisian underground railway station using three distinct methods: daily amplitude coefficient (DAC), generalized additive models (GAM), and long short-term memory (LSTM) neural networks. Two filtration technologies — Experiment 1: positive ionization and Experiment 2: water-based filtration system — were assessed, focusing on reductions in PM_{10} and $PM_{2.5}$ concentrations.

The DAC and GAM methods provide reductions over the entire experimental period, while LSTM analyzes only the first three days. DAC effectively captures long-term trends, but it can be affected by data availability. GAM is ideal for analyzing nonlinear relationships and assessing system performance, though it does not reveal reduction trends during the experimental period. LSTM excels in short-term forecasting, effectively capturing complex temporal patterns, but requires substantial input data and shows increasing uncertainty in long-term predictions.

The results of all three methods indicate that Experiment 1 outperformed Experiment 2 for both PM classes. DAC and GAM showed that both filtration technologies were more effective at reducing $PM_{2.5}$ concentrations than PM_{10}, while LSTM indicated the opposite. Specifically, DAC recorded a PM_{10} reduction of $17.8\% \pm 3.2\%$

for Experiment 1, compared to only 0.4% ± 0.1% for Experiment 2. GAM showed a reduction of 13.1% ± 0.8% for Experiment 1, while LSTM indicated a reduction of 26.8% ± 6.1%. For $PM_{2.5}$, DAC recorded a reduction of 25.9% ± 6.6% in Experiment 1, with GAM and LSTM reporting reductions of 26% ± 0.8% and 22.6% ± 4.8%, respectively. Based on these findings, we recommend a hybrid approach: using DAC for long-term trend evaluation, GAM for in-depth interaction analysis and extended trend assessment, and LSTM for short-term forecasting. This integrated strategy offers a comprehensive understanding of the efficiency of the filtration system, supporting informed decision making to improve air quality at underground stations.

Keywords: air quality, underground train stations, PM concentrations, filtration system performance.

1. Introduction

Particulate matter (PM) poses a significant threat to human health. Even short-term exposure, ranging from an hour to several days, can cause adverse effects, such as respiratory and cardiovascular diseases, and an increased risk of premature mortality (WHO 2021). In underground railway stations (URS), the situation is even more alarming, raising concerns about the health risks for both passengers and workers in these enclosed spaces. In fact, PM concentrations not only exceed those found in outdoor air (Passi, Nagendra & Maiya 2021), but also far exceeds the thresholds set by the World Health Organization (WHO 2021), which are intended primarly for outdoor air. Numerous studies across different regions of the world have shown that particle concentrations in URS can be 3 to 10 times higher than outdoor levels, and sometimes even more elevated (Raut, Chazette & Fortain 2009; Tu & Olofsson 2021; C.-Y. Wang et al. 2021; Passi, Nagendra & Maiya 2021; Karim et al. 2024).

PM concentration within URS exhibits significant variability, influenced by factors such as outdoor PM concentration, operating hours, piston effect, braking systems, train frequency, train speed, station depth, station design, passenger density, weather conditions, ventilation systems, system age, number of line interchanges, Heating Ventilation and Air Conditioning (HVAC) system, maintenance and construction, cleaning activities, and outdoor traffic (Passi, Nagendra & Maiya 2021; Chang et al. 2021; S. Wang et al. 2024).

Efforts to improve air quality in URS have attracted attention. Various mitigation strategies have been implemented, such as the use of non-metallic materials in the train drilling system (Salma, Weidinger & Maenhaut 2007; Karlsson et al. 2006), forced mechanical ventilation in tunnels (T. Moreno et al. 2014; Teresa Moreno et al. 2017), platform screen doors (PSDs) (K.-H. Kim et al. 2012; Son et al. 2013; Teodosiu & Kubinyecz 2022), and air filtration systems (J.-B. Kim et al. 2014; Tokarek & Bernis 2006; Sim et al. 2018; Pretot et al. 2022).

This study will specifically focus on performances of air filtration systems. Assessing the effectiveness of air filtration systems in environments with high variability presents challenges, as measuring the extent of air quality improvement and determining whether the systems achieve their intended objectives can be complex. To our knowledge, only the study conducted by Pretot et al. 2022 addressed this issue by comparing the daily mean profiles of PM_{10} (particles with a diameter below 10 µm) and $PM_{2.5}$ (particles matter with a diameter below 10 µm) during periods with and without filtration systems, employing a method that averaged hourly data to obtain daily profiles for comparison. However, simply averaging the data may not suffice, given the significant temporal variability in PM concentrations. Relying solely on averaged profiles can lead to conclusions that do not capture the true dynamics of air quality in these confined environments.

This study aims to address these challenges by developing three comprehensive approaches to assess the performances of filtration systems in URS. The first approach uses the Daily Amplitude Coefficient (DAC) metric, which was developed and detailed in a previous paper (Rakotonirinjanahary et al. 2024). The second approach applies the Generalized Additive Model (GAM), while the third employs the LSTM model (Rahmani et al. 2024). Along with assessing the air quality improvements offered by filtration systems, the paper examines the pros and cons of each analytical method, considering factors such as precision, reliability, and ease of implementation.

2. Materials and methods

2.1 Dataset

The study used PM measurements taken on the platform of a Parisian URS on the RER C line, part of the SNCF network. PM_{10} and $PM_{2.5}$ concentrations were measured every 15 minutes by two TEOM 1405-D devices (Tapered Element Oscillating Microbalance) located near the midpoint of the platform since April 2018. The timeseries are presented in Figure 1. In addition to PM data, CO_2 concentrations were measured at the same 15-minute intervals using a CO_2 analyzer model 410i with infrared correlation, while temperature and relative humidity were measured using an analog TRA-102D sensor (0–10 Volt). Moreover, external environmental data from the nearest AIRPARIF monitoring station were also incorporated, including measurements of external PM_{10}, $PM_{2.5}$, NO_2, NO, and NOx, enhancing the analysis of air quality variations influenced by both internal and external sources. These data are available online at `https://data-airparif-asso.opendata.arcgis.com/`. Further details regarding the URS and statistical description of the dataset can be found in Rakotonirinjanahary et al. 2024. During this time, SNCF implemented two different filtration systems. The first experiment (hereafter referred to as Experiment 1), conducted from May to September 2019 (highlighted by the red band in Figure 1), tested a positive ionization technology. Six devices were installed around the measurement bay, each using a different dust collection system: vacuum-based,

water-based collection bins, vibration-based, and mist spray systems. Positive ionization electrostatics works by generating charged particles (ions) that interact with airborne dust and particulates. In this process, positive ions are emitted into the air, where they attach to particulate matter. This causes the particles to become electrically charged, promoting their aggregation or attraction to nearby collection plates. The cumulative air purification capacity of the six depollution modules is 52 000 m^3/h (AIRLAB 2022). The second experiment (hereafter referred to as Experiment 2), which ran from January to March 2020 (represented by the blue band in Figure 1), employed a water-based filtration system. This system operates by drawing air through a depression mechanism, forcing it into water. The resulting bubbling effect helps transfer airborne pollutants into the water. To enhance filtration efficiency, calcium chloride is added to the water, aiding in the capture of organic molecules. Similar to Experiment 1, six devices were installed around the measurement bay, with a cumulative air purification capacity of 15000 m^3/h (AIRLAB 2022).

In both cases, filtration systems were placed near the TEOMs and their impact is expected to be confined to their immediate vicinity.

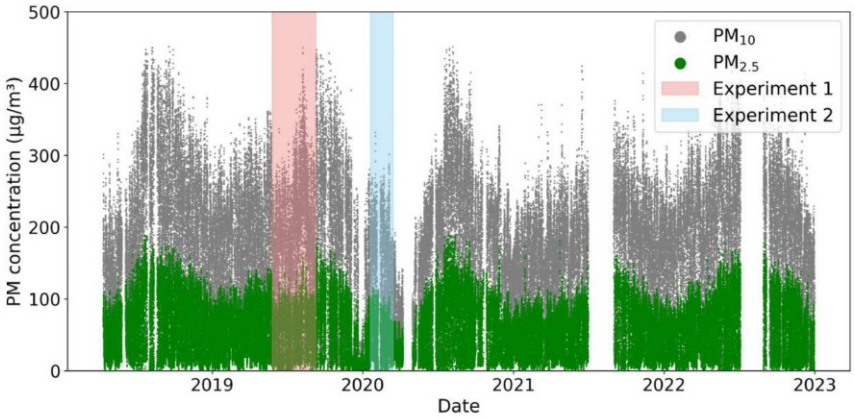


Figure 1: Temporal evolution of PM_{10} (gray dots) and $PM_{2.5}$ (green dots)

2.2 Methodologies for assessing the performance of filtration systems

In this section, we provide a detailed overview of three distinct methods: the Daily Amplitude Coefficient (DAC) method, the Generalized Additive Model (GAM), and the Long Short-Term Memory (LSTM) model. Each of these approaches offers a unique perspective on system efficiency, facilitating a comprehensive assessment of the filtration technologies under investigation.

2.2.1 Ratio of Daily amplitude coefficient (DAC)

In a previous paper (Rakotonirinjanahary et al. 2024), we presented a methodology for establishing typical time profiles of PM concentrations in URS, three key outcomes were identified: two distinct typical profiles of PM concentrations—one for weekdays and one for weekends (see supplementary materials for typical for PM_{10} and $PM_{2.5}$ profiles)—and a daily metric termed DAC (Figure 2). The DAC quantifies the amplitude of the daily PM profile relative to the overall average profile across the entire measurement period. A DAC value greater than 1 indicates that the daily profile exceeds the average, while a value below 1 suggests it falls below. As outlined in Rakotonirinjanahary et al. 2024, the data preprocessing required for this method involves a manual cleaning process to exclude periods affected by significant events (such as COVID-19, strikes, and summer maintenance works) (see Table S1 and Figure S1 in supplementary materials), followed by an automatic outlier detection.

This metric demonstrated a significant seasonal evolution for both PM_{10} and $PM_{2.5}$, as illustrated by the gray and green dots in Figure 2, respectively. Both PM fractions exhibited similar annual patterns, with peaks occurring during the summer months ($DAC \approx 1.5$) and declines in the winter ($DAC \approx 0.5$), reflecting their concurrent behavior throughout the years.

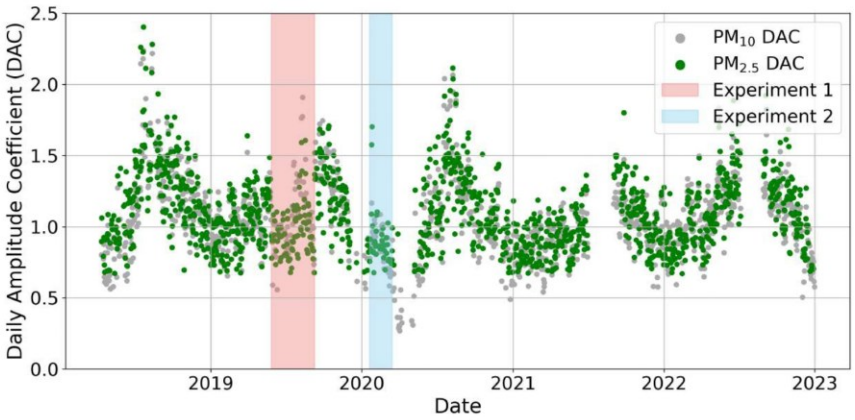


Figure 2: DAC temporal evolution of PM_{10} (gray dots) and $PM_{2.5}$ (green dots)

To account for both seasonal and annual variations, each filtration experiment incorporates data from the entire measurement year in addition to the filtration period (2019 for Experiment 1 and 2020 for Experiment 2), which is then compared on a day-by-day basis with the averaged data from the remaining years (excluding the experimental years). Such an approach enables a more comprehensive and robust comparison across multiple baseline conditions, thereby enhancing the accuracy of the analysis. Specifically, for Experiment 1, the average excludes years 2019, while for Experiment 2, years 2020 are excluded. These averaged years serve as the reference to compare daily data against the experimental periods. DACs from both periods were compared on a day-by-day basis,

matching corresponding days of the week (e.g., Monday to Monday, Tuesday to Tuesday), to reflect typical daily variations in PM concentrations.

To quantify the effectiveness of filtration system, we calculated the ratio of the DACs over the full year when filtration system was settled (DAC_{Y_filt}) and over the full year without it (DAC_{Y_nofilt}):

$$Ratio = \frac{DAC_{Y_filt}}{DAC_{Y_nofilt}} \tag{1}$$

A ratio greater than 1 indicates that the PM concentration level during the year when the experiment was conducted within the considered URS is higher than that of the year when no experiment took place.

From this, the percentage reduction in PM concentrations, representing the performance of the filtration system, can be derived as:

$$Reduction(\%) = \left(1 - \frac{\overline{Ratio}_{in_EXP}}{\overline{Ratio}_{out_EXP}}\right) \times 100 \tag{2}$$

where $\overline{Ratio}_{in_EXP}$ is the DAC ratio average during the experimentation period, and $\overline{Ratio}_{out_EXP}$ is the DAC ratio average outside the experimentation period.

To assess the variability of these reductions, we also calculated the standard deviation ($\sigma_{reduction}$) of the DAC ratios. This gives an indication of the consistency of the filtration system's performance across the experimental period and is derived as follows:

$$\sigma_{reduction}(\%) = Reduction(\%) \times \sqrt{\left(\frac{\sigma_{in_EXP}}{\overline{Ratio}_{in_EXP}}\right)^2 + \left(\frac{\sigma_{out_EXP}}{\overline{Ratio}_{out_EXP}}\right)^2} \tag{3}$$

where σ_{in_EXP} and σ_{out_EXP} are the standard deviations of the DAC ratios during and outside the experimental period, respectively.

2.2.2 Generalized additive model (GAM) method

GAM models, introduced by Hastie & Tibshirani 1986, extend generalized linear models by incorporating both parametric and non-parametric components. GAM flexibly captures linear and nonlinear relationships between explanatory and response variables, reducing the risks of incorrect linear assumptions. GAM techniques have been previously applied to explore complex non-linear correlations with air pollution (Dominici et al. 2002; Camalier, Cox & Dolwick 2007; Pearce et al. 2011; Bertaccini, Dukic & Ignaccolo 2012; Gong, Kaulfus et al. 2017; Gong, Hong & Jaffe 2018; Zeng et al. 2020). The general form can be written as follow:

$$g(\mathbb{E}(y_i)) = \beta_0 + \sum_{j=1}^{m} \alpha_j C_j + \sum_{k=1}^{p} S_k(x_{ik}) + \epsilon_i, \tag{4}$$

where i varies from 1 to n is the number of observations; j varies from 1 to m, where m is the number of categorical variables in the model; and k varies from 1 to p, where p is the number of continuous variables in the model. $\mathbb{E}(y_i)$ represents the expected value of the response variable y_i, $g(\cdot)$ is the link function that connects the expected value of the response to the predictors. β_0 is the intercept, representing the baseline value of the response variable when all explanatory variables are held constant; in other words, it is the typical concentration when filtration systems are not installed. α_j are the coefficients associated with each categorical variables C_j; $S_j(x_{ij})$ are smooth functions of the continuous predictors x_{ij}; and ϵ_i is the residual.

Given our goal of accurately capturing the influence of environmental factors on particulate matter concentrations, we employ the identity link function ($g(z) = z$) in our GAM model. This choice is particularly suitable as it allows the model outputs to remain on the original scale of the response variable. Consequently, we can directly interpret the effects of predictors such as Experiment 1 and Experiment 2 on PM levels.

In this study, we considered quarter-hourly PM concentrations as the response variable. The explanatory categorical variables included both experiments (Experiment 1 and Experiment 2) as binary (boolean) variables, indicating whether they were OFF or ON, and the temporal factor (Day of the Week), which is counted from 1 to 7, representing each day from Monday to Sunday. The explanatory continuous variables included temperature, relative humidity, carbon dioxide, outdoor pollutants (PM_{10}, $PM_{2.5}$, NO_2, NO, NOx), and other temporal factors (time of day, week number, and day of the week).The inclusion of these temporal factors is based on their relevance to station and train operations (Time), variability in PM concentrations across different weeks (Week), and day-to-day fluctuations (Day of the Week). Similar to the DAC method, preliminary data cleaning was required. Significant periods, such as the COVID-19 pandemic, strikes, and summer maintenance works, along with outliers, were removed from the dataset to ensure accuracy.

Table 1: VIF values of explanatory variables in GAM.

Iteration number	**1**	**2**	**3**	**4**	**5**
PM_{10}	12.9	12.9	1.9	1.8	1.8
$PM_{2.5}$	12.5	12.5	Removed	Removed	Removed
CO_2	1.5	1.5	1.5	1.5	1.5
RH	1.7	1.7	1.7	1.7	1.7
T	2.0	2.0	1.9	1.7	1.6
PM_{10_out}	3.8	3.8	3.8	3.8	1.9
$PM_{2.5_out}$	2.9	2.9	2.9	2.9	Removed
NO_{2_out}	4.2	4.2	4.1	2.2	2.1
NO_{out}	Inf	4.9	4.7	Removed	Removed
NOx_{out}	Inf	Removed	Removed	Removed	Removed
Time	1.6	1.6	1.6	1.5	1.5
Week	1.2	1.2	1.2	1.2	1.2
DOW	1.1	1.1	1.1	1.1	1.1

To assess multicollinearity among these explanatory variables and determine whether any should be excluded before analysis, we calculated the variance inflation factor (VIF). Higher VIF values indicate stronger multicollinearity. As noted by Zuur et al. 2009, a cut-off value of 5, or even 3, can be applied to remove collinear variables. We adopted a conservative threshold of 3 to identify and exclude predictors that might contribute to significant multicollinearity, which could destabilize the GAM model (Belušić, Hercog-Bulić & Klaić 2015; Zuur et al. 2009). We employed an iterative approach, removing one variable at a time, recalculating the VIF, and repeating this process until all remaining variables met the VIF threshold. Table 1 summarizes the VIF values of the explanatory variables in our GAM model following this procedure.

Consequently, $PM_{2.5}$ (or PM_{10} when $PM_{2.5}$ is the response variable), $PM_{2.5_out}$ (or PM_{10_out} when $PM_{2.5}$ is the response variable), NO_{out}, and NOx_{out} were removed from the model due to high collinearity. The final model specified for this study is presented as follows:

$$PM = \beta_0 + \underbrace{\alpha_{DOW} DOW + \alpha_{Exp1} Exp1 + \alpha_{Exp2} Exp2}_{\text{Categorical Variables}} + s(CO_2) + s(T) + s(RH) + s(PM_{out}) + \underbrace{s(NO_{2_out}) + s(\text{Time, by=DayType}) + s(\text{Week})}_{\text{Continuous Variables (Smooth terms)}} + \epsilon, \quad (5)$$

where DOW represents the day of week, which is a categorical variable ranging from 1 to 7; α_{DOW} is the coefficient associated with the day of week, reflecting its direct influence on PM concentrations. $Exp1$ and $Exp2$ are boolean variables indicating whether Experiment 1 or Experiment 2 is ON or OFF. The coefficients α_{Exp1} and α_{Exp2} measure the impact of each experiment on PM levels, where larger values indicate a stronger effect. The term $s(CO_2)$ is a smooth function representing the effect of carbon dioxide levels on PM concentrations. Similarly, s(T) is the smooth term for Temperature, s(RH) for relative humidity, s(PM_{out}) and s(NO_{2_out}) for the concentrations of particulate matter and nitrogen dioxide outdoors, respectively. The function s(Time, by=DayType) categorizes the hour of the day into weekdays and weekends, while s(Week) represents for the week number.

The effectiveness of both experiments was analyzed by examining the partial dependence plots to assess the marginal effects of each variable. The significance of each variable was derived from the summary of the parametric coefficients and the approximate significance of the smooth terms within the GAM model. Key statistical terms presented in the results are summerized in Table S2 (supplementary materials).

To quantify the effectiveness of the filtration system, the percentage reduction in PM concentrations, can be derived as follows for Experiment 1:

$$Reduction(\%) = \left(\frac{\alpha_{Exp1}}{\beta_0}\right) \times 100 \quad (6)$$

where: $Reduction(\%)$ represents the percentage reduction in PM concentrations due to the filtration system effect, α_{Exp1} the coefficient associated with the experiment indicating the change in PM concentrations when the filtering systems are ON and β_0, the intercept, is the average PM concentration without the filtration system, serving as the baseline for comparison. Thus, $\frac{\alpha_{Exp1}}{\beta_0}$ represents the fractional reduction in pollution concentration caused by the filtration system compared to the baseline pollution concentration.

And the standard deviation ($\sigma_{reduction}$) is derived as follows:

$$\sigma_{reduction}(\%) = Reduction(\%) \times \sqrt{\left(\frac{\sigma_{Exp1}}{\alpha_{Exp1}}\right)^2 + \left(\frac{\sigma_{\beta_0}}{\beta_0}\right)^2} \quad (7)$$

where: $\sigma_{reduction}(\%)$ represents the standard deviation of the percentage reduction, providing a measure of uncertainty in the reduction estimate. σ_{Exp1} denotes the standard deviation of the coefficient α_{Exp1}, reflecting the variability in the PM concentration changes associated with the filtration system. σ_{β_0} gives the standard deviation of the intercept, indicating the variability in the baseline PM concentration.

2.2.3 Long short-term memory (LSTM) method

LSTM networks, introduced by Hochreiter & Schmidhuber 1997, represent a cutting-edge approach in artificial neural networks. These networks are particularly effective for handling sequential data. The key feature of LSTM networks is their ability to maintain the memory state over long periods, thanks to the presence of the memory cell. The architecture of an LSTM cell includes input, output, and forget gates that control the flow of information.

- The input gate controls the extent to which new information flows into the memory cell.
- The forget gate determines how much of the previous data will be discarded and how much will be retained for future steps.
- The output gate regulates the information sent from the memory cell to the output layer.

This design allows the model to retain information over multiple time steps, which is essential for assessing air quality, as pollutant concentrations are influenced by their prior states (Li et al. 2017; Tran et al. 2023).

The LSTM model developed in this study is based on the predictive framework PM-Forecast, which employs a self-adaptive LSTM architecture established by Rahmani et al. 2024. Unlike their study, we opted for a univariate LSTM model, which is trained exclusively on historical PM concentration data. This decision was primarily influenced

by the insufficient availability of relevant explanatory variables, as well as evidence from previous studies in different domains (Salehi et al. 2024; Assad, Cara & Ortega-Mier 2022) indicating that univariate LSTM models can achieve robust performance.

In this artificial neural network-based analysis, we compare PM concentrations from the first five days of the experiment to actual measurements. As with the other two methods, data preprocessing was necessary, starting with resampling the data to one point per hour. This approach allows us to simplify the dataset while retaining the essential temporal information. Instead of predicting 480 time steps (96 data points per day $\times$ 5), we focus on predicting 120 time steps (24 data points per day $\times$ 5), looking one hour ahead. Then we did data cleaning by removing notable periods and outliers from the initial database.

Based on the standard 80/20 data split ratio, we used 8 months of data for training set and 2 months for testing set in experiment 1. For Experiment 2, 16 months of data were allocated for training and 4 months for testing. For Experiment 1, we had 1 year of data before the experiment began, while for Experiment 2, we had 20 months of pre-experiment data. During the training phase, we selected the following hyperparameters to maximize the performance of the model:

- Learning rate: set to 10^{-3} after evaluating the range $[10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}]$, determining the step size during model optimization.
- Batch size: chosen as 64 after testing values in the range [16, 32, 64, 128], indicating number of samples processed per iteration.
- Epochs: fixed at 200 after considering values between [50, 100, 200], representing the number of complete passes through the dataset.

To prevent overfitting and promote model generalization, as in Rahmani et al. 2024, we employed Early Stopping, which interrupts training once performance stops improving or begins to degrade, avoiding unnecessary further training.

Model performance is evaluted using several metrics: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Square Error (MSE), R-squared (R^2), and Weighted Mean Absolute Percentage Error (WMAPE). RMSE measures the average magnitude of prediction errors, while MSE focuses on larger errors without the square root. MAE measures the average absolute difference between predicted and actual values, reflecting precision. R^2 shows the proportion of correctly predicted instances, indicating model effectiveness. WMAPE provides a nuanced perspective by weighting absolute percentage errors according to their significance across different scenarios (Rahmani et al. 2024).

To quantify the effectiveness of the filtration system, the percentage reduction in PM concentrations, can be derived as :

$$\overline{Reduction}(\%) = \frac{1}{N}\sum_{t=1}^{N}\left(1 - \frac{PM_{LSTM,t}}{PM_{GT,t}}\right) \times 100 \tag{8}$$

where: $\overline{Reduction}(\%)$ is the mean percentage reduction, representing the average difference between the predicted PM values from the LSTM model and the ground truth PM values. N denotes the total number of time steps. $PM_{LSTM,t}$ is the predicted PM concentration at time t, and $PM_{GT,t}$ represents the ground truth PM concentration at time t.

The standard deviation ($\sigma_{reduction}$) is derived as follows:

$$\sigma_{reduction}(\%) = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left((1-\frac{PM_{LSTM,t}}{PM_{GT,t}})*100-\overline{Reduction}(\%)\right)^2} \quad (9)$$

3. Results and discussions

3.1 DAC-based analysis results

As outlined in Section 2, the effectiveness of the filtration systems was assessed by comparing the DAC across multiple baseline conditions. For Experiment 1, DAC data from the year 2019 (when Experiment 1 was conducted at the station) was compared against an average of the years 2018, 2020, 2021, and 2022. For Experiment 2, conducted in 2020, the DAC data was compared with an average of the years 2018, 2019, 2021, and 2022. This averaging approach allows for a robust comparison by mitigating the effects of annual variations and atypical disruptions, ensuring a more consistent and comprehensive analysis of the filtration system's effectiveness.

Figure 3 illustrates the evolution of the DAC ratios: (a) shows the comparison for PM_{10} during 2019, the year of Experiment 1, against the averaged data from 2018, 2020, 2021, and 2022; the red band indicates the duration of Experiment 1. Panel (b) depicts the comparison for PM_{10} in 2020, the year of Experiment 2, against the averaged data from 2018, 2019, 2021, and 2022; the blue band represents the period of Experiment 2. Panels (c) and (d) provide similar comparisons for $PM_{2.5}$. The dotted blue line represents the mean DAC ratio of the years averaged as references during the filtration periods, while the dotted red line shows the mean DAC ratio for the experimental years.

In Experiment 1, both PM_{10} (Figure 3a) and $PM_{2.5}$ (Figure 3c) exhibit decreases in the DAC ratio, dropping from approximately 1.1 to 0.9 for PM_{10}, and from 1.1 to 0.8 for $PM_{2.5}$. This corresponds to a reduction of 17.8% ± 3.2% for PM_{10} and 25.9% ± 6.6% for $PM_{2.5}$ (Eq. 2 and 3). These results suggest that the filtration system installed during Experiment 1 led to a significant reduction in particulate matter concentrations and was more effective at filtering $PM_{2.5}$ than PM_{10}. For both particle size classes, there appears to be a decrease in the DAC ratio during the first month of the experiment. However, between August and September—before the end of the experiment—an increase in the DAC ratio is noted. These observations might be due to annual variability of the PM concentration, to technical intervention on the filtration system, to stabilization of its

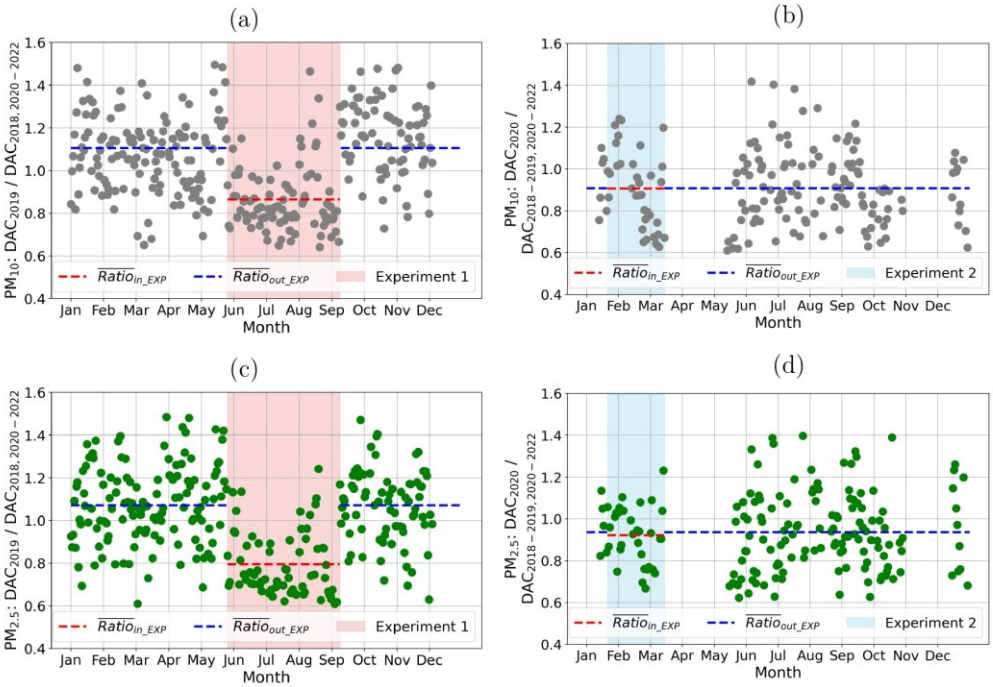


Figure 3: Evolution of DAC ratios for PM_{10} and $PM_{2.5}$: (a) 2019 vs. 2018, 2020, 2021, 2022: PM_{10}; (b) 2020 vs. 2018, 2019, 2021, 2022: PM_{10}; (c) 2019 vs. 2018, 2020, 2021, 2022: $PM_{2.5}$; and (d) 2020 vs. 2018, 2019, 2021, 2022: $PM_{2.5}$.

efficacity; or might even be just an artifact caused by the number of data points retained for the analysis. In the absence of additional information, and in the framework of this global method of analyzing the performance of the system, we will not discuss specifically this effect.

In contrast, Figures 3b and 3d reveal no significant changes during Experiment 2, with the DAC ratio fluctuating around 1. This suggests that Experiment 2 had little observable effect, resulting in reductions of only $0.4\% \pm 0.1\%$ for PM_{10} and $3.7\% \pm 1.0\%$ for $PM_{2.5}$ (Eq. 2 and 3). It is important to consider that 2020 was a particularly unique year, marked by strikes occurring just before the experiment and the onset of the COVID-19 pandemic shortly thereafter. These events limited the data available for analysis, making it difficult to draw definitive conclusions about the effectiveness of Experiment 2.

3.2 GAM-based analysis results

In this section, we discuss the findings from the GAM model applied to PM_{10} and $PM_{2.5}$. Tables 2 and 3 summarize the results of the parametric coefficients and the approximate significance of the smooth terms on PM_{10}, respectively. Figure 4 illustrates the effect of these influencing factors on PM_{10} concentrations as modeled by the GAM. Note that the y-axis of the partial dependence plot represents the estimated average effect of corresponding predictor variable on the response variable (PM concentrations), while holding all other

predictors constant. Adjacent to it, the effective degree of freedom (EDF) is indicated.

Table 2: Parametric Coefficients on PM_{10}

Term	Estimate	Std. Error	t value	Pr(>|t|)
β_0	156.8	0.8	204.9	< 2e-16 ***
DOW2	6.8	1.1	6.4	1.98e-10 ***
DOW3	5.8	1.1	5.3	9.79e-08 ***
DOW4	4.2	1.1	3.9	9.84e-05 ***
DOW5	0.0	1.1	0.0	1.000
DOW6	-32.4	1.1	-30.2	< 2e-16 ***
DOW7	-28.9	1.1	-26.3	< 2e-16 ***
Exp1-On	-20.5	1.3	-15.5	< 2e-16 ***
Exp2-On	1.3	2.0	0.7	0.512

Sig. codes: [0, 0.001]:'***']0.1, 1]:' '
Note: Term - Predictor variable; Estimate - Effect of the predictor variable on PM_{10} concentration. A positive value indicates an increase in PM_{10}, while a negative value indicates a decrease.; Std. Error - Variability of the estimate; t value - Test statistic; Pr(>|t|) - p-value indicating significance.

Table 3: Approximate Significance of Smooth Terms on PM_{10}

Smooth Term	Edf	Ref.df	F	p-value
s(CO_2)	8.0	8.7	163.2	< 2e-16 ***
s(RH)	7.2	8.2	23.7	< 2e-16 ***
s(T)	8.0	8.7	99.8	< 2e-16 ***
s(PM_{10_out})	1.0	1.1	43.9	< 2e-16 ***
s(NO_{2_out})	4.2	5.2	3.8	0.00152 **
s(Time):Weekdays	9.0	9.0	3331.3	< 2e-16 ***
s(Time):Weekends	8.9	9.0	880.1	< 2e-16 ***
s(Week)	4.0	4.0	367.2	< 2e-16 ***

Sig. codes: [0, 0.001]:'***']0.001, 0.01]:'**'
Note: Smooth Term - Non-linear predictor variable; Edf - Effective degrees of freedom, indicating smoothness; Ref.df - Reference degrees of freedom, related to model complexity; F - F-statistic for significance of the smooth term; p-value - Indicates statistical significance of the smooth term.

From Table 2, we observe the effects of various categorical factors, including both filtration experiments. DOW1 (which represents Monday), Exp1-Off, and Exp2-Off are not included in the table because they serve as reference categories in our model. This means that the effects of all other DOW are measured relative to Monday, and the effects of turning ON Experiment 1 or Experiment 2 are compared to when both experiments are OFF. Experiment 1 shows a significant reduction in PM_{10} concentrations, with a coefficient estimate of -20.5 $\mu g/m^3$ and a standard error of 1.3 $\mu g/m^3$ (see Figure 4). This translates to a 13.1% reduction (Eq. 6) in PM_{10} concentrations during Experiment 1, with an error margin of $\pm$ 0.8% (Eq. 7). Thus, the reduction in PM_{10} concentrations during Experiment 1 is approximately 13.1% $\pm$ 0.8%, highlighting the strong and statistically significant impact of the intervention.

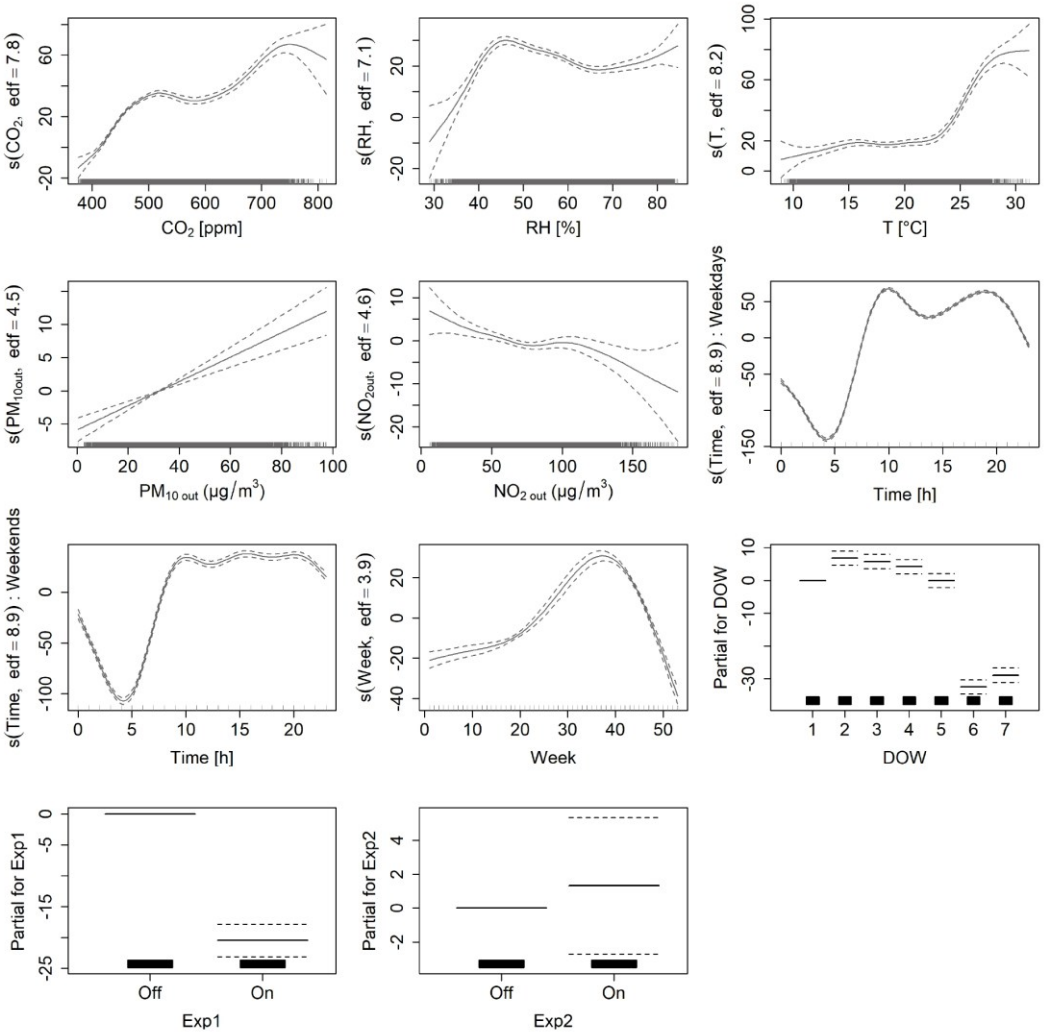


Figure 4: Effect of influencing factors [CO_2, RH, T, PM_{10_out}, NO_{2_out}, Time, Week, DOW, Experiment 1, and Experiment 2] on PM_{10} concentration in the GAM. The y-axis represents the smoothing function values; for example, s(T) shows the trend in PM_{10} concentration when air temperature changes, and df is the degree of freedom for the trend. The x-axis represents the measured values of the influencing factor, the solid black curve indicates the trend in PM_{10} concentration with the change of influencing factors, and the dotted line around the solid black line indicates 95% Confidence Interval (lower and upper limits) of PM_{10} concentration.

As illustrated in Figure 4, Experiment 2 results in a less pronounced increase, which was not statistically significant ($p = 0.5$, Table 2), with a positive coefficient estimate of 1.3 µg/m^3 and a standard error of 2.0 µg/m^3. This leads to an approximate increase of 1.3% ± 2% in PM_{10} concentrations during Experiment 2 (according to Eq. 6 and 7).

Experiment 1 shows strong statistical significance (p-values $< 2e^{-16}$) while Experiment 2 was not statistically significant. Experiment 1 had a more substantial effect on PM_{10}

concentrations compared to Experiment 2, with 13.1% ± 0.8%. Regarding the DOW, the GAM indicates differences in PM_{10} concentrations across different days of the week (Figure 4), which were not apparent in our previous study using the DAC method. Weekdays (DOW2 to DOW4) exhibit higher PM_{10} levels compared to weekends (DOW6, and DOW7), as shown by the parametric coefficients in 2, likely due to variations in human activity, or operational schedules. Specifically, the coefficients for DOW2 (6.8), DOW3 (5.8), and DOW4 (4.2) are positive and significantly different from zero ($p < 2e^{-16}$), while DOW6 and DOW7 show negative and significant values (-32.4 and -28.9, respectively), suggesting a decrease in PM_{10} levels compared to weekdays.These findings align with those presented in our previous paper (Rakotonirinjanahary et al. 2024). Note that the coefficient for DOW5 is 0, with a maximum p-value (1), indicating that the PM concentration on this day is statistically indistinguishable from that on DOW1. This means there is no significant difference between DOW5 and DOW1 as shown in Figure 4.

Table 3 details the significance of smooth terms related to CO_2, relative humidity (RH), temperature (T), outdoor PM_{10}, outdoor NO_2, Time, and Week. All smooth terms are highly significant (p-values $< 2e^{-16}$), reinforcing the importance of these variables in influencing PM_{10} concentrations. However, the term for NO_2, suggests a lower level of significance ($p < 0.01$), pointing to a less pronounced but still notable impact on PM_{10} levels. The smooth term for Time, with an F-value of 3331.3 for weekdays and 880.1 for weekends, exhibits the strongest influence, indicating that this variable was the main driving factor for PM_{10} concentrations. Indeed hour of the day plays a critical role, with the smooth term for time showing a highly significant and nonlinear pattern (Figure 4). The variation over time likely reflects daily commuting patterns, which are closely related to train frequency (Birenzvige et al. 2003; Raut, Chazette & Fortain 2009; Şahin et al. 2012; Querol et al. 2012; Walther & Bogdan 2017), with spikes in PM_{10} concentrations during peak travel times on weekdays and more stable levels on weekends (Rakotonirinjanahary et al. 2024).

The second most important factor is the Week, with an F-value of 367.2 (Table 3). The seasonal variation represented by the Week variable significantly influences PM concentrations within the station (Rakotonirinjanahary et al. 2024). The GAM captures variability across the year, which may reflect long-term changes in environmental conditions or usage patterns of the underground stations. Notably, PM concentrations peak in summer and decrease in winter (Figure 4), as highlighted in Rakotonirinjanahary et al. 2024.However, the underlying causes of this seasonal trend are still not fully understood, especially considering that similar trends are not observed everywhere, such as in Stockholm, Sweden (Tu & Olofsson 2024), and Seoul, South Korea (Jung Park et al. 2024).

The third most important factor is the CO_2, which is associated to the number of people present in the station at the time (Yang et al. 2023). As illustrated in Figure 4, a significant positive nonlinear relationship is observed between CO_2 levels and PM_{10} concentrations. As CO_2 levels increase, PM_{10} concentrations rise as well, particularly at

higher levels of CO_2.

On the contrary, T, RH, and outdoor PM_{10}, even more so outdoor NO_2, do not exhibit a strong impact on PM_{10} concentrations. However, Figure 4 illustrates that T, RH, and outdoor PM_{10} each have a positive nonlinear relationship with PM_{10} concentrations, though these effects are less pronounced compared to those of CO_2.

Table 4: Parametric coefficients on $PM_{2.5}$

Term	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**
β_0	63.1	0.3	204.1	< 2e-16 ***
DOW2	3.1	0.4	7.2	4.93e-13 ***
DOW3	2.4	0.4	5.6	2.31e-08 ***
DOW4	2.0	0.4	4.6	5.10e-06 ***
DOW5	0.4	0.4	1.0	0.316679
DOW6	-12.0	0.4	-27.5	< 2e-16 ***
DOW7	-10.9	0.4	-24.5	< 2e-16 ***
Exp1-On	-16.4	0.5	-30.7	< 2e-16 ***
Exp2-On	-2.9	0.8	-3.5	0.000464 ***

Sig. codes: [0, 0.001]:'***']0.1, 1]:' '

Table 5: Approximate significance of smooth terms on $PM_{2.5}$

Smooth Term	**Edf**	**Ref.df**	**F**	**p-value**
s(CO_2)	7,8	8,6	176,3	< 2e-16 ***
s(RH)	7,1	8,1	20,5	< 2e-16 ***
s(T)	8,2	8,8	88,1	< 2e-16 ***
s(PM_{25_out})	4,5	5,5	32,8	< 2e-16 ***
s(NO_{2_out})	4,6	5,7	4,2	0.000486 ***
s(Time):Weekdays	9,0	9,0	2541,3	< 2e-16 ***
s(Time):Weekends	8,9	9,0	647,0	< 2e-16 ***
s(Week)	4,0	4,0	301,5	< 2e-16 ***

Sig. codes: [0, 0.001]:'***'

As observed with PM_{10}, the results for $PM_{2.5}$ in Table 4 show a significant reduction in particle concentrations during Experiment 1. The decrease of 16.4 µg/m^3 corresponds to a reduction of approximately 26% ± 0.8%. However, there is a marked difference in the effectiveness of Experiment 2. For $PM_{2.5}$, Experiment 2 leads to a reduction of only 2.9 µg/m^3, corresponding to a 4.6% ± 1.3% decrease. Although statistically significant, the smaller reduction suggests again that this method is less effective compared to the Experiment 1. This contrasts with the non significant result observed for PM_{10}. These results indicate that while Experiment 1 was effective for both $PM_{2.5}$ and PM_{10}, Experiment 2 had a negligible and statistically insignificant effect on PM_{10} concentrations.

Nonetheless, as illustrated in Table 4, 5, and Figure 5, the GAM model continues to reveal a complex and nonlinear relationship between $PM_{2.5}$ concentrations and the influencing variables (CO_2, RH, T, PM_{10_out}, NO_{2_out}, Time, Week, DOW, Experiment 1, and Experiment 2). This aligns with the findings for PM_{10}, emphasizing that both

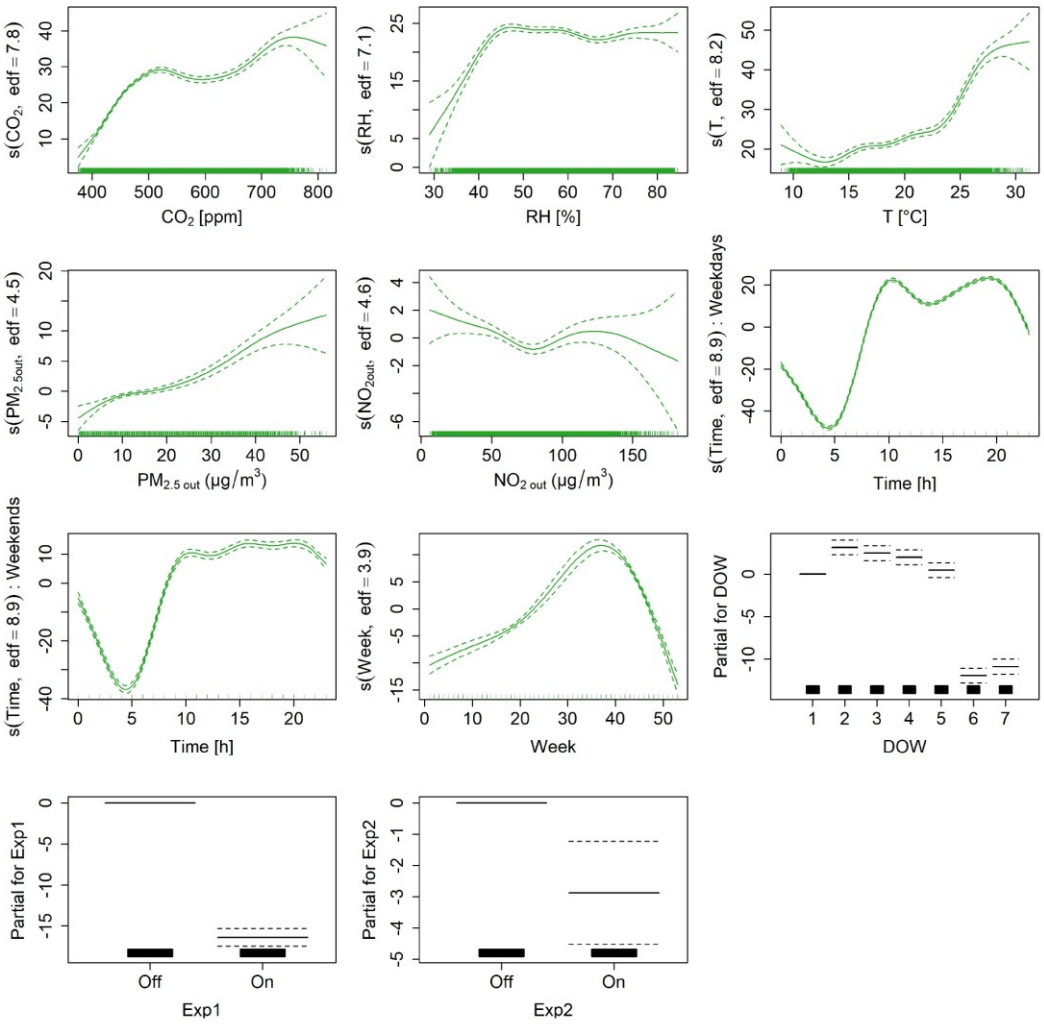


Figure 5: Effect of influencing factors [CO_2, RH, T, PM_{10_out}, NO_{2_out}, Time, Week, Weekday, Year, Exp1, and Exp2] on $PM_{2.5}$ concentration in the GAM.

particulate matter concentrations exhibit intricate dependencies on the considered factors (Figure 5).

3.3 LSTM-based analysis results

In this study, we used the LSTM model proposed by Rahmani et al. 2024 to forecast PM concentrations. For both experiments, data from the period prior to the start of each experiment was used to train the model, with 80% of the data allocated for training and 20% for testing. For Experiment 1, we used 8 months of data for training set and 2 months for testing set. For Experiment 2, 16 months of data were allocated for training and 4 months for testing. For each class of PM in the two experiments, the model was trained to consider the time points from the current time (t) to 48 hours prior (t-48) as

input to forecast one hour ahead (t+1).

Table 6 provides a comprehensive summary of the evaluation metrics for the prediction results obtained during the training and testing periods for both PM_{10} and $PM_{2.5}$ across both Experiments 1 and 2. RMSE and R^2 metrics, in particular, highlight the model's accuracy and its ability to explain the variance in the observed data.

Table 6: Evaluation metrics (RMSE, MAE, MSE, R^2, WMAPE) for prediction results during the training and testing periods for both classes of PM for each dataset used in Experiments 1 and 2.

Evaluation metrics			**RMSE** ($\mu g/m^3$)	**MAE** ($\mu g/m^3$)	**MSE** ($\mu g/m^3$)	**R^2** (%)	**WMAPE** (%)
Train set	PM_{10}	Exp. 1	29.2	20.9	853.8	90.6	13.0
		Exp. 2	29.3	21.1	856.7	90.1	13.3
	$PM_{2.5}$	Exp. 1	12.6	8.9	159.9	90.2	13.1
		Exp. 2	12.3	8.9	152.9	90.1	13.6
Test set	PM_{10}	Exp. 1	25.6	19.2	659.2	90.6	13.0
		Exp. 2	25.7	20.6	661.1	90.1	13.3
	$PM_{2.5}$	Exp. 1	10.2	7.7	105.9	90.0	13.8
		Exp. 2	10.1	7.6	102.0	89.7	14.2

For PM_{10}, the RMSE values for the test set range from 25.6 to 25.7 µg/m^3 , with the lowest error observed in Experiment 1. RMSE values found in this study align closely with those reported by Park et al. 2018, who evaluated the performance of an Artificial Neural Network (ANN) in 6 URS, where the RMSE values for PM_{10} ranged from 19.4 to 29.6 µg/m^3.

These results reflect a good model fit, supported by R^2 values exceeding 90% across all experiments, indicating that over 90% of the variance in PM_{10} concentrations is explained by the model. Similarly, for $PM_{2.5}$, RMSE values are lower, ranging from 10.1 µg/m^3 to 10.2 µg/m^3, with R^2 values close to or exceeding 89%, demonstrating robust predictive performance.

Overall, the high R^2 values indicates that the model performs well in forecasting PM concentrations. The lower RMSE observed in the test dataset compared to the training dataset suggests that overfitting has been avoided. Additionally, the consistency of R^2 values across both experiments and particle classes highlights the model's reliability in capturing the underlying patterns in the data.

Figure 6 evaluates the prediction accuracy of PM levels for both training and testing sets, across forecasting horizons ranging from 1 to 72 hours. Figures 6a and 6b show the average R^2 and RMSE scores over various prediction windows [1, 6, 12, 24, 48, 72] hours. To ensure accurate forecasting, each time a prediction was made, the predicted value was re-injected into the historical dataset. This process involved removing the oldest data point and adding the new prediction as the most recent entry. For instance, when forecasting for the next 12 hours, the model generates hourly predictions iteratively, which are then averaged to compute the overall R^2 for that period. A decline in accuracy is observed as the prediction window increases, with the R^2 score dropping below 0.5 beyond the 72-hour mark, indicating the model's limitations to predict PM levels over periods longer than 3 days.

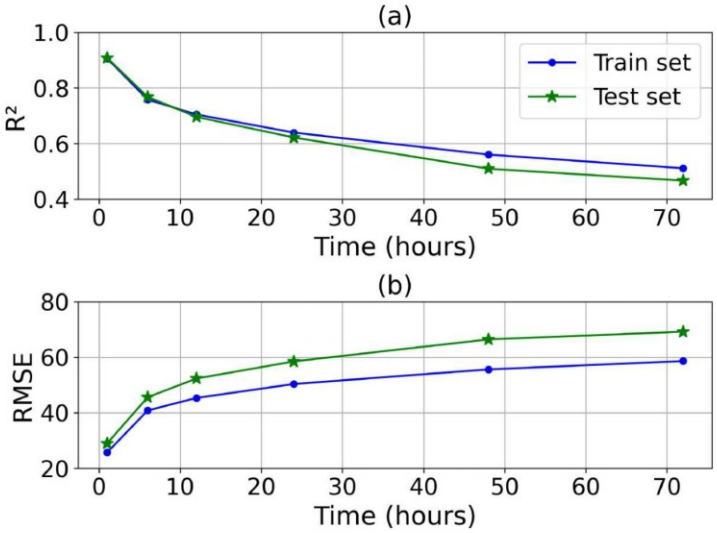


Figure 6: Performance evaluation of our LSTM model: (a) Accuracy assessed by R^2 (%), and (b) RMSE ($\mu g/m^3$). The blue lines denote the performance on the training sets, while the green lines represent the performance on the test sets.

The reduction effectiveness was then computed by comparing the first three days of predictions to the ground truth PM concentrations during this period, as these provided the most reliable results (Figure 6) between 06:00 am to 23:00 pm (train operation). However, we extended the forecast to five days to gain a broader understanding of PM concentration trends across a full weekday cycle. This allowed for a more comprehensive comparison of predicted and observed PM concentrations, particularly in assessing variations driven by weekday activities and environmental conditions.

Figure 7 illustrates the ground truth PM concentrations (blue line) and the predicted PM concentrations from the LSTM model (red line) for PM_{10}: (a) experiment 1, (b) experiment 2; and for $PM_{2.5}$: (c) experiment 1, (d) experiment 2 during the first five days of each experiment.

In general, the model effectively reproduces the daily patterns of PM concentrations for both particle fractions across the two configurations, Experiment 1 and Experiment 2. For clarity, the five-day comparison corresponds to weekdays. As shown in Rakotonirinjanahary et al. 2024, two peaks in PM concentrations are typically observed during the morning (10:00 am) and evening (07:00 pm) rush hours, alongside an increase in PM levels at the start of train operations and a decrease toward the end. Additionally, we observed a similar weekly trend as found with the GAM model, where PM concentrations increase from Monday to Wednesday, then decrease from Wednesday to Friday.

In Experiment 1, a significant reduction in both PM_{10} and $PM_{2.5}$ concentrations is observed, particularly from the second day onward, with total reductions of 26.8% $\pm$ 6.1% for PM_{10} and 22.6% $\pm$ 4.8% for $PM_{2.5}$ (Eq. 8 and 9) during the first three days. On the first day, there is no difference between measured and predicted values due to the initial activation phase of the filtration system, which may require time to reach full

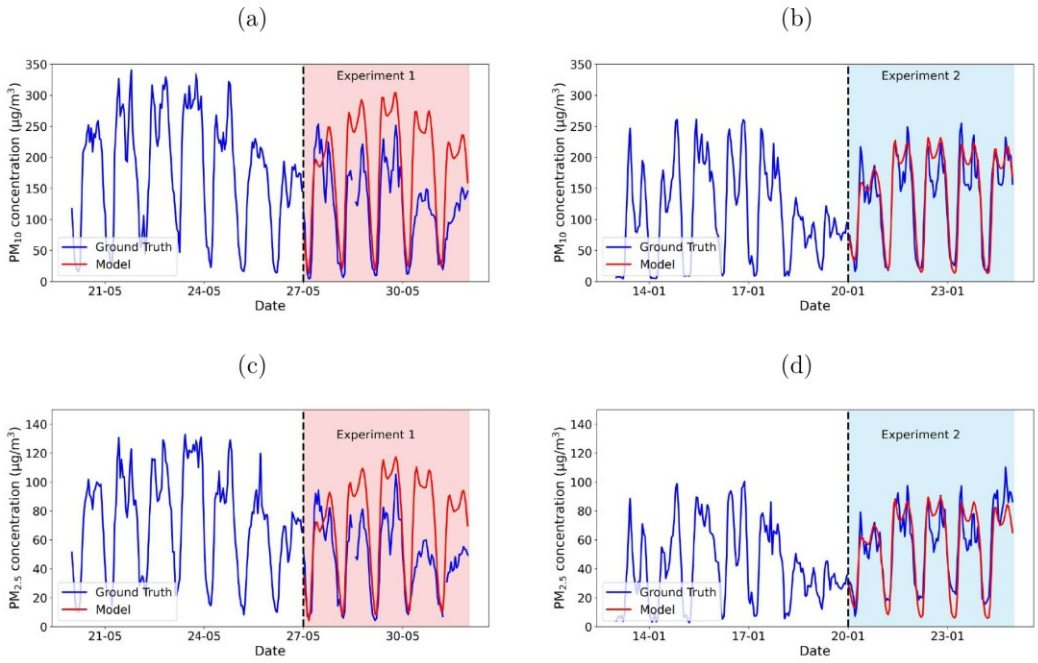


Figure 7: Ground Truth PM Concentrations (Blue Line) and Predicted PM Concentrations from the LSTM Model (Red Line) for PM_{10}: (a) Experiment 1, (b) Experiment 2; and for $PM_{2.5}$: (c) Experiment 1, (d) Experiment 2.

operational efficiency after installation (Figure 7a and 7c).

In contrast, Experiment 2 displays a similar trend in both the magnitude and hourly pattern of PM_{10} concentrations (Figure 7b), with a minimal reduction of $5.3\% \pm 1.2\%$ over the first three days. For $PM_{2.5}$, Figure 7d reveals comparable results during the daytime, with only a a $2.7\% \pm 0.7$ % decrease during the first three days of the experiment. At night, the LSTM model predictions are consistently less than the ground truth values. This discrepancy may be attributed to the deposition behavior of fine particles, which tend to accumulate during nighttime hours, as highlighted by Walther, Bogdan & Cohen 2017.

3.4 Evaluating PM concentration reductions through three approches

Table 7 presents a comparative summary of the reductions in PM_{10} and $PM_{2.5}$ concentrations achieved through three different methods, evaluated over two experimental setups.

The results from all three methods demonstrate that Experiment 1 outperformed Experiment 2 for both PM_{10} and $PM_{2.5}$ particle size classes. This could be attributed to the differing cumulative filtration capacities of the systems, or possibly because Experiment 2 was conducted during periods of significant disruption, such as the strikes before and the COVID pandemic after the experiment. Specifically, DAC and GAM consistently

indicated that Experiment 1 captured $PM_{2.5}$ more effectively than PM_{10} across both experimental setups, reflecting a trend where finer particles were more efficiently filtered. However, there is a notable difference in terms of reduction levels. This discrepancy may be attributed to the fact that GAM provides a clearer picture of the true effect of the filtration system by isolating the impact of the filters from other variables. In contrast, the LSTM model revealed a stronger reduction in PM_{10} compared to $PM_{2.5}$ during the first three days of both experiments.

Table 7: Summary of PM concentration reductions (%) via three methods: DAC, GAM, and LSTM

Method		DAC	GAM	LSTM
PM_{10}	Exp. 1	17.8 ± 3.2	13.1 ± 0.8	26.8 ± 6.1
	Exp. 2	0.4 ± 0.1	Not significant	5.3 ± 1.2
$PM_{2.5}$	Exp. 1	25.9 ± 6.6	26 ± 0.8	22.6 ± 4.8
	Exp. 2	3.7 ± 1.0	4.6 ± 1.3	2.7 ± 0.7

3.5 Comparison of filtration efficiency assessment methods

In this section, we compare all three methods used to assess the performance of the filtration systems—DAC, GAM, and LSTM—and provide recommendations based on their strengths, limitations, and potential applications. Table 8 offers an overview and comparison of these three methods.

Table 8: Comparison of three methods: DAC, GAM, and LSTM

Criteria	DAC	GAM	LSTM
Objective	Compares the DAC between a year with a filtration system and a year without	Determines the marginal effect of the considered experiment	Compares PM concentrations between ground truth and predicted
Method type	Statistical	Statistical	Machine learning
Data	1/4 hourly PM concentrations	1/4 hourly PM concentrations with additional explanatory variables	Hourly PM concentrations
Data cleaning	Required	Required	Required
Computational Complexity	High	High	High
Interpretability	Moderate	High	Low
Temporal analysis	Long-term	Long-term	Short-term
Prediction accuracy	-	Moderate	High
Robustness	Moderate	High	Moderate
Precision	High	High	Low

The DAC method offers a straightforward approach for evaluating PM long-term

trends. By comparing different periods, such as years with and without filtration systems, it allows clear and fast visualization of filtration system performance, seasonal and temporal patterns in PM reduction. Prior to this, careful data preprocessing is necessary to avoid the impact of abnormal periods and establish a reliable average profile, enabling accurate DAC calculations. This method simplifies the conversion of quarter-hourly time steps into daily metrics, facilitating faster and easier interpretation of daily PM concentrations. As shown in Figure 3, it provides an overview of the temporal evolution of the reductions. However, its effectiveness relies heavily on data availability, and external disruptions like strikes or pandemics can distort the results. Nevertheless, it remains a valuable tool, particularly for long-term assessments. A reference year with sufficient, uninterrupted data is essential for reliable evaluations across the experimental period, especially like during Experiment 2 when data gaps exist.

In contrast, the GAM method offers a more flexible and statistically robust means of analysis. Unlike DAC, GAM isolates the effects of the experiments by adjusting for multiple explanatory variables, such as the day of the week (DOW) and seasonal factors, providing a clearer understanding of the filtration system's impact on PM concentrations. The GAM reveals variations in PM_{10} levels across different days, which were not evident in our previous study using DAC method (Rakotonirinjanahary et al. 2024). Additionally, GAM models nonlinear relationships, capturing interactions between PM and multiple predictors. This capability allows for the quantification of marginal effects for individual predictors, making the model highly interpretable and offering deeper insights. The GAM results represent the marginal effects of each predictor, holding all others predictors constant, which aligns with our objective of evaluating the specific impacts of experimental conditions. Despite its advantages, the GAM approach does present some challenges, such as evaluating the reduction evolation throughout the entire period of the experiment, identifying relevant explanatory variables, and ensuring data availability.

The LSTM model provides an advanced machine learning approach, excelling in capturing complex temporal dependencies within the data. It is particularly effective for short-term forecasting, making it ideal for real-time applications where sudden shifts in PM concentrations need to be accounted for. The LSTM model offers rapid and valuable insights into the initial performance of the filtration system. However, LSTM comes with certain drawbacks. As the prediction horizon extends, uncertainties increase, making long-term forecasts less reliable. The model is computationally intensive, requires large datasets, and lacks the interpretability of GAM, which makes it more difficult to understand how specific variables contribute to the results. The model's predictive power decreases significantly beyond a 72-hour forecast horizon (Figure 6).

4. Conclusions

This study evaluated the performance of filtration systems in underground train stations using three distinct analytical approaches: the Daily Amplitude Coefficient (DAC), Generalized Additive Models (GAM), and Long Short-Term Memory (LSTM) neural networks. Each method provided unique insights into the behavior of particulate matter (PM_{10} and $PM_{2.5}$) concentrations in response to different filtration technologies.

The DAC method is recommended for long-term trend analysis, particularly when data consistency and minimal external disturbances are ensured. Its ability to condense time-series data into easily interpretable daily metrics makes it valuable for understanding broad trends over extended periods.

The GAM method excels at offering statistically robust insights by accounting for multiple influencing factors such as day of the week and seasonal effects. Its flexibility in modeling complex, nonlinear relationships also makes it well-suited for long-term trend evaluation, particularly when the goal is to isolate specific influences like the experimental conditions.

The LSTM model is ideal for real-time monitoring and short-term evaluation of filtration system performance. It can rapidly detect changes in PM concentrations, offering immediate feedback on filtration effectiveness, though it shows some limitations in long-term predictability due to increasing uncertainty over extended horizons.

Results from all three methods confirmed that Experiment 1, which used positive ionization technology, was more effective in reducing both PM_{10} and $PM_{2.5}$ concentrations compared to Experiment 2, which employed water-based technology. DAC and GAM methods indicated that $PM_{2.5}$ was more efficiently captured in both experiments. However, the LSTM model showed a stronger reduction in PM_{10} compared to $PM_{2.5}$, highlighting its sensitivity to short-term fluctuations in coarser particles. It is important to note that these findings are localized to the measurement point and may not fully represent conditions across the entire platform, given the proximity of the filtration systems to the PM monitoring location.

While each method has distinct strengths, the choice of approach should be guided by the specific goals of the analysis. A combined approach—using DAC for long-term trends, GAM for detailed variable interactions, and LSTM for short-term forecasting—offers the most comprehensive understanding of filtration system performance.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author used GPT-3.5 to enhance readability and language. After using this tool, the authors thoroughly reviewed and edited the content as necessary and assume full responsibility for the final version of the publication.

CRediT authorship contribution statement

Valisoa Rakotonirinjanahary: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Suzanne Crumeyrolle**: Writing – review & editing, Validation, Methodology. **Maryam Rahmani**: Methodology, Software. **Mateusz Bogdan**: Writing – review & editing, Validation, Software. **Benjamin Hanoune**: Writing – review & editing, Supervision, Methodology, Conceptualization, Validation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors greatly acknowledge SNCF, managed by the National Research Technology Agency, for funding the PhD of Valisoa Rakotonirinjanahary (ANRT CIFRE PhD grant n°2021/0856). Authors are also grateful to AIRPARIF for conducting data collection and providing valuable insights as part of their ongoing collaboration of SNCF DexGIF. This research is part of the CaPPA project (Chemical and Physical Properties of the Atmosphere), funded by the French National Research Agency (ANR) through the PIA (Programme Investissement d'Avenir) with the contract ANR-11-LABX-005-001. Additionally, it is affiliated with the Institute of Multidisciplinary Research in Environmental Sciences (IRePSE Fed 4129) and the CPER ECRIN research project, funded by the French Ministry of Higher Education and Research, the Hauts-de-France region, and the European Regional Development Fund.

References

AIRLAB (Feb. 2022). *Expérimentations "Innovons pour l'air de nos stations" synthèse et avis d'AIRLAB*. Tech. rep.

Assad, Daniel Bouzon Nagem, Javier Cara & Miguel Ortega-Mier (Dec. 2022). 'Comparing Short-Term Univariate and Multivariate Time-Series Forecasting Models in Infectious Disease Outbreak'. *Bulletin of Mathematical Biology* 85.1, p. 9. ISSN: 1522-9602. DOI: 10.1007/s11538-022-01112-5.

Belušić, Andreina, Ivana Herceg-Bulić & Zvjezdana Bencetić Klaić (Jan. 2015). 'Using a generalized additive model to quantify the influence of local meteorology on air quality in Zagreb'. *Geofizika* 32.1, pp. 47–77. ISSN: 1846-6346. DOI: 10.15233/gfz.2015.32.5.

Bertaccini, Pancrazio, Vanja Dukic & Rosaria Ignaccolo (2012). 'Modeling the Short-Term Effect of Traffic and Meteorology on Air Pollution in Turin with Generalized Additive Models'. *Advances in Meteorology* 2012.1, p. 609328. ISSN: 1687-9317. DOI: 10.1155/2012/609328.

Birenzvige, A. et al. (Mar. 2003). 'Aerosol Characteristics in a Subway Environment'. *Aerosol Science and Technology* 37.3, pp. 210–220. ISSN: 0278-6826. DOI: 10.1080/02786820300941.

Camalier, Louise, William Cox & Pat Dolwick (Oct. 2007). 'The effects of meteorology on ozone in urban areas and their use in assessing ozone trends'. *Atmospheric Environment* 41.33, pp. 7127–7137. ISSN: 1352-2310. DOI: 10.1016/j.atmosenv.2007.04.061.

Chang, Li et al. (May 2021). 'Recent progress in research on PM2.5 in subways'. *Environmental Science: Processes & Impacts* 23.5, pp. 642–663. ISSN: 2050-7895. DOI: 10.1039/D1EM00002K.

Dominici, Francesca et al. (Aug. 2002). 'On the use of generalized additive models in time-series studies of air pollution and health'. *American Journal of Epidemiology* 156.3, pp. 193–203. ISSN: 0002-9262. DOI: 10.1093/aje/kwf062.

Gong, Xi, Song Hong & Daniel A. Jaffe (2018). 'Ozone in China: Spatial Distribution and Leading Meteorological Factors Controlling O3 in 16 Chinese Cities'. *Aerosol and Air Quality Research* 18.9, pp. 2287–2300. ISSN: 2071-1409. DOI: 10.4209/aaqr.2017.10.0368.

Gong, Xi, Aaron Kaulfus et al. (Nov. 2017). 'Quantifying O3 Impacts in Urban Areas Due to Wildfires Using a Generalized Additive Model'. *Environmental Science & Technology* 51.22, pp. 13216–13223. ISSN: 1520-5851. DOI: 10.1021/acs.est.7b03130.

Hastie, Trevor & Robert Tibshirani (1986). 'Generalized Additive Models'. *Statistical Science* 1.3, pp. 297–310. ISSN: 0883-4237.

Hochreiter, Sepp & Jürgen Schmidhuber (Nov. 1997). 'Long Short-Term Memory'. *Neural Computation* 9.8, pp. 1735–1780. ISSN: 0899-7667. DOI: 10.1162/neco.1997.9.8.1735.

Jung Park, Su et al. (Oct. 2024). 'Nationwide distribution and factors related to indoor fine particulate matter concentrations in subway stations in Korea: Results of first-year measurement'. *Environment International*, p. 109078. ISSN: 0160-4120. DOI: 10.1016/j.envint.2024.109078.

Karim, Anjum Shahina et al. (Aug. 2024). 'Assessment of air quality in the Philadelphia, Pennsylvania subway'. *Journal of Exposure Science & Environmental Epidemiology*, pp. 1–9. ISSN: 1559-064X. DOI: 10.1038/s41370-024-00711-9.

Karlsson, Hanna L. et al. (Sept. 2006). 'Comparison of genotoxic and inflammatory effects of particles generated by wood combustion, a road simulator and collected from street and subway'. *Toxicology Letters* 165.3, pp. 203–211. ISSN: 0378-4274. DOI: 10.1016/j.toxlet.2006.04.003.

Kim, Ki-Hyun et al. (Mar. 2012). 'A noticeable shift in particulate matter levels after platform screen door installation in a Korean subway station'. *atmospheric environment* 49, pp. 219–223. DOI: 10.1016/j.atmosenv.2011.11.058.

Kim, Jong-Bum et al. (Aug. 2014). 'Status of PM in Seoul metropolitan subway cabins and effectiveness of subway cabin air purifier (SCAP)'. *Clean Technologies and Environmental Policy* 16.6, pp. 1193–1200. ISSN: 1618-954X, 1618-9558. DOI: 10.1007/s10098-013-0708-1.

Li, Xiang et al. (Dec. 2017). 'Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation'. *Environmental Pollution* 231, pp. 997–1004. ISSN: 0269-7491. DOI: 10.1016/j.envpol.2017.08.114.

Moreno, T. et al. (Aug. 2014). 'Subway platform air quality: Assessing the influences of tunnel ventilation, train piston effect and station design'. *Atmospheric Environment* 92, pp. 461–468. ISSN: 1352-2310. DOI: 10.1016/j.atmosenv.2014.04.043.

Moreno, Teresa et al. (Apr. 2017). 'The effect of ventilation protocols on airborne particulate matter in subway systems'. *The Science of the Total Environment* 584-585, pp. 1317–1323. ISSN: 1879-1026. DOI: 10.1016/j.scitotenv.2017.02.003.

Park, Sechan et al. (Jan. 2018). 'Predicting PM10 concentration in Seoul metropolitan subway stations using artificial neural network (ANN)'. *Journal of Hazardous Materials* 341, pp. 75–82. ISSN: 0304-3894. DOI: 10.1016/j.jhazmat.2017.07.050.

Passi, Amit, S. M. Shiva Nagendra & M. P. Maiya (July 2021). 'Characteristics of indoor air quality in underground metro stations: A critical review'. *Building and Environment* 198, p. 107907. ISSN: 0360-1323. DOI: 10.1016/j.buildenv.2021.107907.

Pearce, John L. et al. (Feb. 2011). 'Quantifying the influence of local meteorology on air quality using generalized additive models'. *Atmospheric Environment* 45.6, pp. 1328–1336. ISSN: 1352-2310. DOI: 10.1016/j.atmosenv.2010.11.051.

Pretot, Pierre-Emmanuel et al. (Apr. 2022). 'Fast method to design air filtration solution at low energy cost in subterranean train stations'. *7th World Congress on Civil, Structural, and Environmental Engineering (CSEE'22)*. Lisbon, Portugal.

Querol, X. et al. (June 2012). 'Variability of levels and composition of PM_{10} and $PM_{2.5}$ in the Barcelona metro system'. *Atmospheric Chemistry and Physics* 12.11, pp. 5055–5076. ISSN: 1680-7316. DOI: 10.5194/acp-12-5055-2012.

Rahmani, Maryam et al. (Aug. 2024). 'PMFORECAST: leveraging temporal LSTM to deliver in situ air quality predictions'. *Environmental Science and Pollution Research International*. ISSN: 1614-7499. DOI: 10.1007/s11356-024-34623-w.

Rakotonirinjanahary, Valisoa M. et al. (Oct. 2024). 'A novel method for establishing typical daily profile of PM concentrations in underground railway stations'. *Indoor Environments* 1.3, p. 100040. ISSN: 2950-3620. DOI: 10.1016/j.indenv.2024.100040.

Raut, J. -C., P. Chazette & A. Fortain (Feb. 2009). 'Link between aerosol optical, microphysical and chemical measurements in an underground railway station in Paris'.

Atmospheric Environment 43.4, pp. 860–868. ISSN: 1352-2310. DOI: 10.1016/j.atmosenv.2008.10.038.

Şahin, Ülkü Alver et al. (Jan. 2012). 'PM10 concentrations and the size distribution of Cu and Fe-containing particles in Istanbul's subway system'. *Transportation Research Part D: Transport and Environment* 17.1, pp. 48–53. ISSN: 1361-9209. DOI: 10.1016/j.trd.2011.09.003.

Salehi, Sajad et al. (May 2024). 'Comparative study of univariate and multivariate strategy for short-term forecasting of heat demand density: Exploring single and hybrid deep learning models'. *Energy and AI* 16, p. 100343. ISSN: 2666-5468. DOI: 10.1016/j.egyai.2024.100343.

Salma, Imre, Tamás Weidinger & Willy Maenhaut (Dec. 2007). 'Time-resolved mass concentration, composition and sources of aerosol particles in a metropolitan underground railway station'. *Atmospheric Environment* 41.37, pp. 8391–8405. ISSN: 1352-2310. DOI: 10.1016/j.atmosenv.2007.06.017.

Sim, Jung-Bo et al. (Mar. 2018). 'Baffle dust collector for removing particles from a subway tunnel during the passage of a train'. *Journal of Mechanical Science and Technology* 32.3, pp. 1415–1421. ISSN: 1976-3824. DOI: 10.1007/s12206-018-0245-2.

Son, Youn-Suk et al. (Mar. 2013). 'The Effect of Platform Screen Doors on PM10 Levels in a Subway Station and a Trial to Reduce PM10 in Tunnels'. *Asian Journal of Atmospheric Environment* 7, pp. 38–47. DOI: 10.5572/ajae.2013.7.1.038.

Teodosiu, Catalin Ioan & Vladimir Francisc Kubinyecz (Jan. 2022). 'Numerical Study on the Impact of Platform Screen Doors in a Subway Station with a Train on Fire'. *Applied Sciences* 12.16, p. 8296. ISSN: 2076-3417. DOI: 10.3390/app12168296.

Tokarek, S. & A. Bernis (Nov. 2006). 'An exemple of particle concentration reduction in Parisian subway stations by electrostatic precipitation'. *Environmental Technology* 27.11, pp. 1279–1287. ISSN: 0959-3330. DOI: 10.1080/09593332708618746.

Tran, Huynh Duy et al. (Dec. 2023). 'Forecasting hourly PM2.5 concentration with an optimized LSTM model'. *Atmospheric Environment* 315, p. 120161. ISSN: 1352-2310. DOI: 10.1016/j.atmosenv.2023.120161.

Tu, Minghui & Ulf Olofsson (June 2021). 'PM10 in underground stations from different types of trains'. *Transportation Research Part D: Transport and Environment* 95, p. 102867. ISSN: 1361-9209. DOI: 10.1016/j.trd.2021.102867.

Tu, Minghui & Ulf Olofsson (Apr. 2024). 'Estimating PM levels on an underground metro platform by exploring a new model-based factor research'. *Atmospheric Environment: X* 22, p. 100261. ISSN: 2590-1621. DOI: 10.1016/j.aeaoa.2024.100261.

Walther, E. & M. Bogdan (Oct. 2017). 'A novel approach for the modelling of air quality dynamics in underground railway stations'. *Transportation Research Part D: Transport and Environment* 56, pp. 33–42. ISSN: 1361-9209. DOI: 10.1016/j.trd.2017.07.014.

Walther, E., M. Bogdan & R. Cohen (Dec. 2017). 'Modelling of airborne particulate matter concentration in underground stations using a two size-class conservation model'.

Science of The Total Environment 607-608, pp. 1313–1319. ISSN: 0048-9697. DOI: 10.1016/j.scitotenv.2017.07.090.

Wang, Cheng-Yi et al. (Mar. 2021). 'Identification of High Personal PM2.5 Exposure during Real Time Commuting in the Taipei Metropolitan Area'. *Atmosphere* 12.3, p. 396. ISSN: 2073-4433. DOI: 10.3390/atmos12030396.

Wang, Shunyao et al. (Aug. 2024). 'Indoor air quality in subway microenvironments: Pollutant characteristics, adverse health impacts, and population inequity'. *Environment International* 190, p. 108873. ISSN: 0160-4120. DOI: 10.1016/j.envint.2024.108873.

WHO, World Health Organization (2021). 'WHO Global Air Quality Guidelines: Particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide: executive summary'.

Yang, Junbin et al. (Apr. 2023). 'A review on characteristics and mitigation strategies of indoor air quality in underground subway stations'. *Science of The Total Environment* 869, p. 161781. ISSN: 0048-9697. DOI: 10.1016/j.scitotenv.2023.161781.

Zeng, Yingying et al. (2020). 'Prediction of Potentially High PM2.5 Concentrations in Chengdu, China'. *Aerosol and Air Quality Research* 20.5, pp. 956–965. ISSN: 2071-1409. DOI: 10.4209/aaqr.2019.11.0586.

Zuur, Alain F. et al. (2009). *Mixed effects models and extensions in ecology with R*. Statistics for Biology and Health. New York, NY: Springer. ISBN: 978-0-387-87457-9 978-0-387-87458-6. DOI: 10.1007/978-0-387-87458-6.

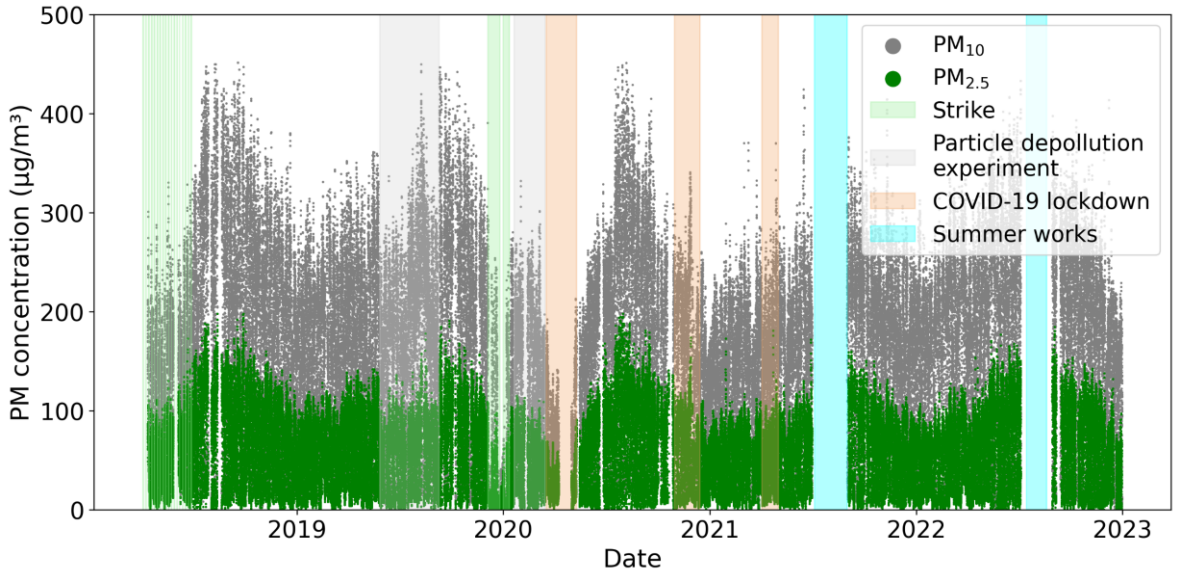
Annexe 3 : Périodes notables et élimination des profils atypiques dans les gares avec mesures longue durée

Cette annexe présente un récapitulatif des périodes notables enregistrées dans les trois gares disposant de mesures de longue durée. Elle met en évidence les variations temporelles des concentrations en PM_{10} et $PM_{2.5}$ dans ces gares, en particulier les tendances saisonnières et les événements spécifiques qui ont influencé les résultats (Figure A. 1).

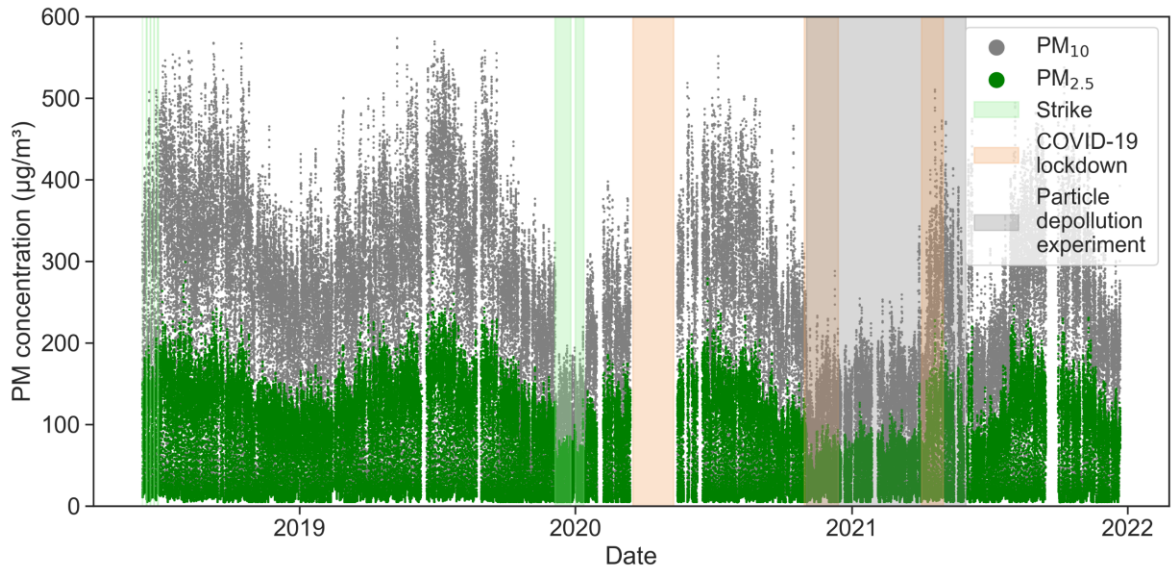
- **Périodes notables enregistrées dans chaque gare**

La première section de l'annexe est dédiée à l'évolution temporelle des concentrations en PM_{10} et $PM_{2.5}$ mesurées à la gare d'Avenue Foch entre 2018 et 2022, à Sevran Beaudottes entre 2018 et 2021, et à Magenta entre 2017 et 2021. Ces périodes ont été identifiées comme étant particulièrement significatives (Figure A. 1).

(a)



(b)



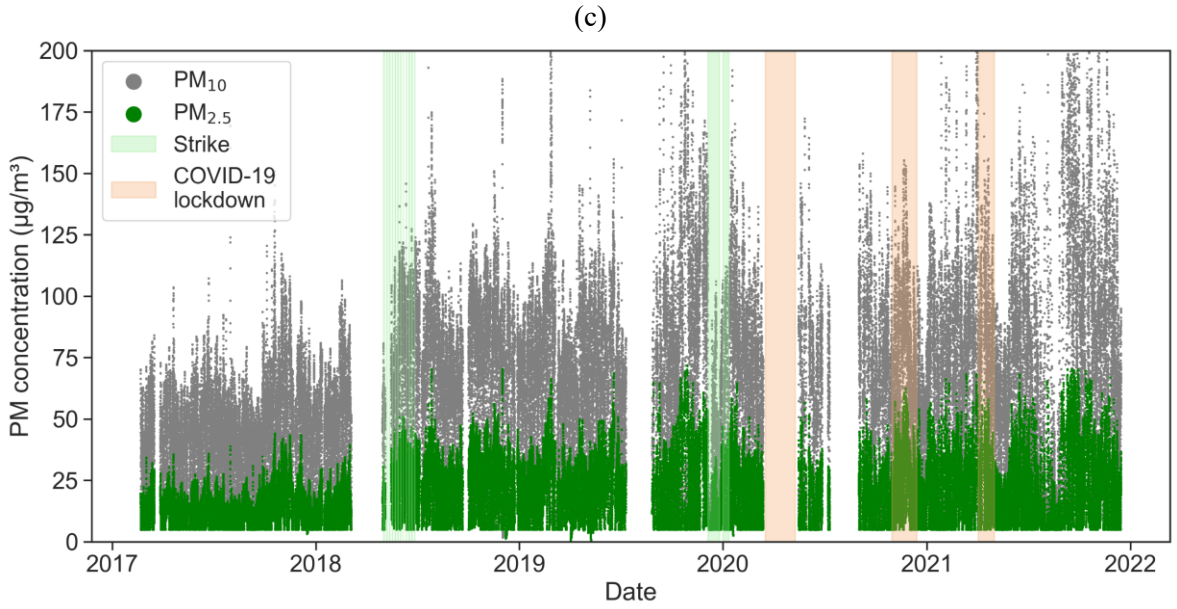


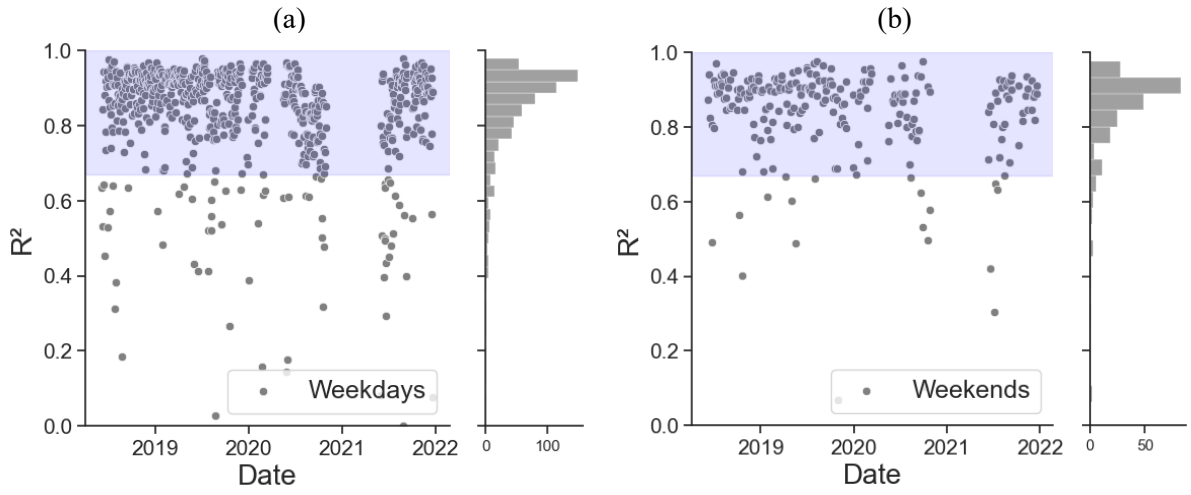
Figure A. 1 : Évolution temporelle des concentrations en PM_{10} et $PM_{2.5}$ mesurées à la gare de (a) Avenue Foch entre 2018 et 2022, (b) Sevran Beaudottes entre 2018 et 2021 et (c) Magenta entre 2017 et 2021 mettant en évidence les périodes notables

- **Elimination des profils atypiques**

En complément des analyses réalisées à la gare d'Avenue Foch, cette section présente les résultats relatifs à l'élimination des profils atypiques aux gares de Sevran Beaudottes et Magenta.

  - **Cas de Sevran Beaudottes**

Pour la gare de Sevran-Beaudottes, nous avons également appliqué un seuil de $R^2 = 0,67$ pour filtrer les profils non représentatifs, pour les deux classes de particules. Ce seuil, déjà validé pour les données de la gare d'Avenue Foch, s'est à nouveau révélé parfaitement adapté aux caractéristiques du jeu de données de la gare de Sevran-Beaudottes (Figure A. 2).



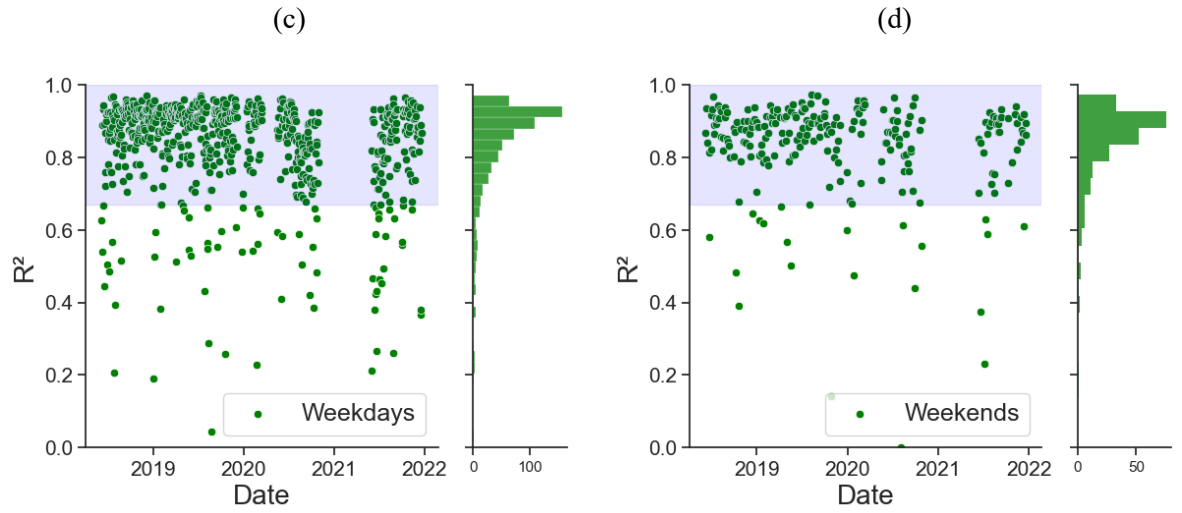


Figure A. 2 : Variation temporelle de R² pour les PM_{10} (en gris) et les $PM_{2.5}$ (en vert) à la gare de Sevran-Beaudottes : (a) jours ouvrés pour PM_{10}, (b) weekends pour PM_{10}, (c) jours ouvrés pour $PM_{2.5}$, et (d) weekends pour $PM_{2.5}$

- **Cas de Magenta**

Enfin, pour la gare de Magenta, une approche similaire a été adoptée. Sur la base de l'analyse des distributions des coefficients de détermination, un seuil de $R^2 = 0,6$ a été retenu pour filtrer les profils atypiques. Les résultats sont présentés dans la Figure A. 3, qui montre les variations temporelles du coefficient de détermination pour les PM_{10} et $PM_{2.5}$, séparées en jours ouvrés et weekends.

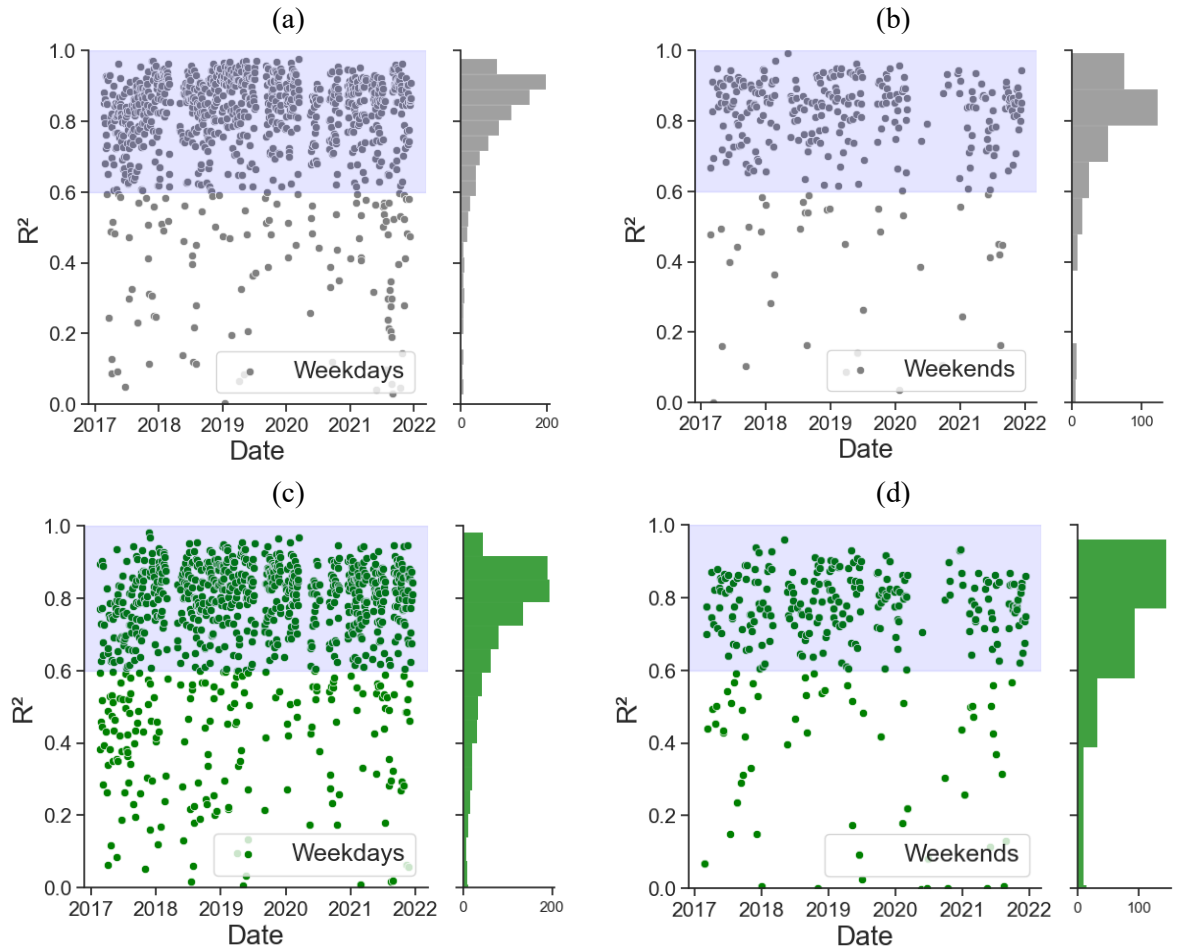


Figure A. 3 : Variation temporelle de R² pour les PM_{10} (en gris) et les $PM_{2.5}$ (en vert) à la gare de Magenta : (a) jours ouvrés pour PM_{10}, (b) weekends pour PM_{10}, (c) jours ouvrés pour $PM_{2.5}$, et (d) weekends pour $PM_{2.5}$

Annexe 4 : En EFS – Profils journaliers types du lundi au vendredi (jours ouvrés) et du samedi au dimanche (weekends) et analyse du temps de déposition ($PM_{2.5}$ – $PM_{2.5-10}$)

Cette annexe présente une analyse détaillée des profils journaliers types des concentrations de PM_{10} et $PM_{2.5}$ dans les gares avec des mesures de longue durée, du lundi au vendredi (jours ouvrés) et du samedi au dimanche (weekends), ainsi que des résultats relatifs aux temps de déposition des fractions $PM_{2.5}$ et $PM_{2.5-10}$.

- **Profils journaliers types des PM_{10}**

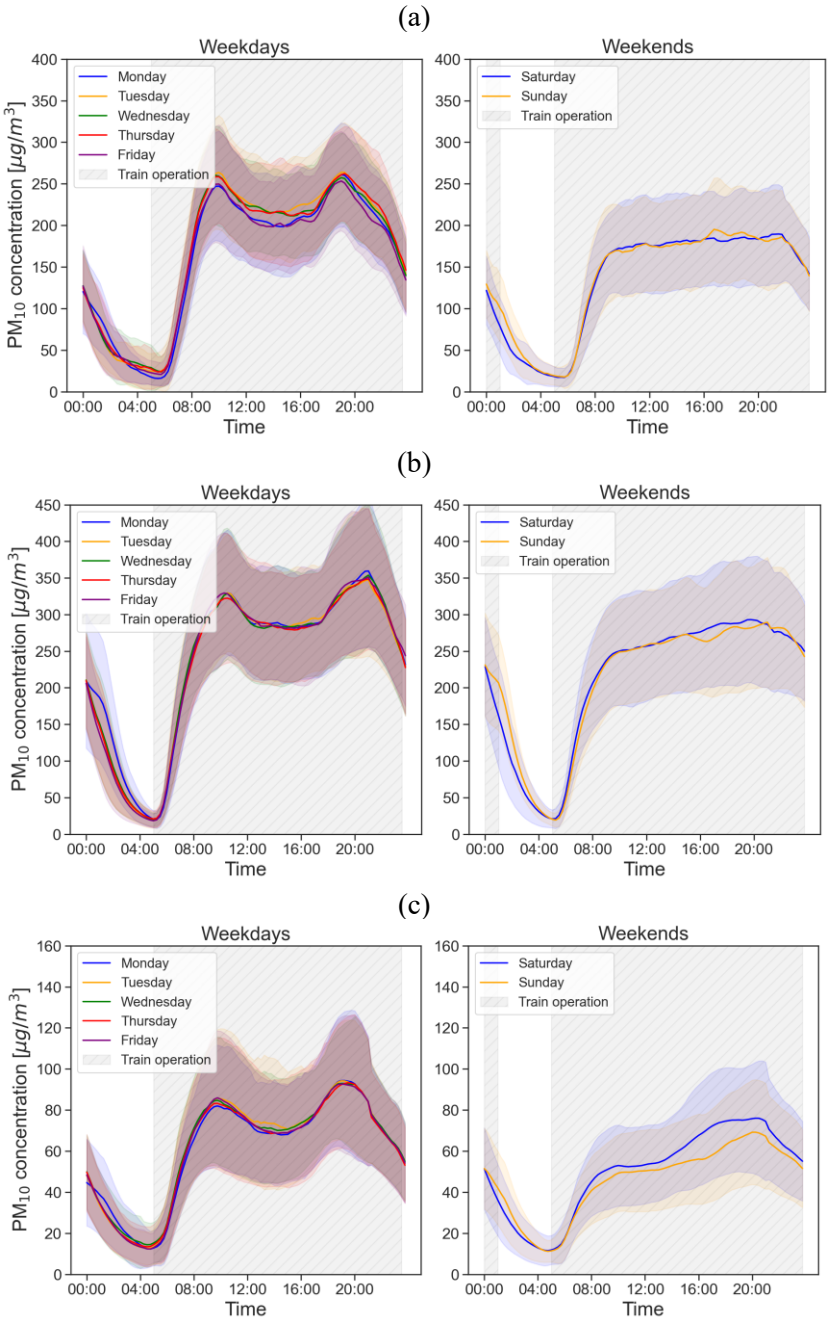


Figure A. 4 : Profils journaliers types des concentrations de PM_{10} du lundi au vendredi (à gauche) et du samedi au dimanche (à droite) dans les gares avec mesures longue durée : (a) Avenue Foch, (b) Sevran Beaudottes et (c) Magenta. La zone hachurée indique la période d'exploitation des trains. Les zones ombrées représentent un écart type (σ=1) sur les différents jours.

- **Profils journaliers types des $PM_{2.5}$**

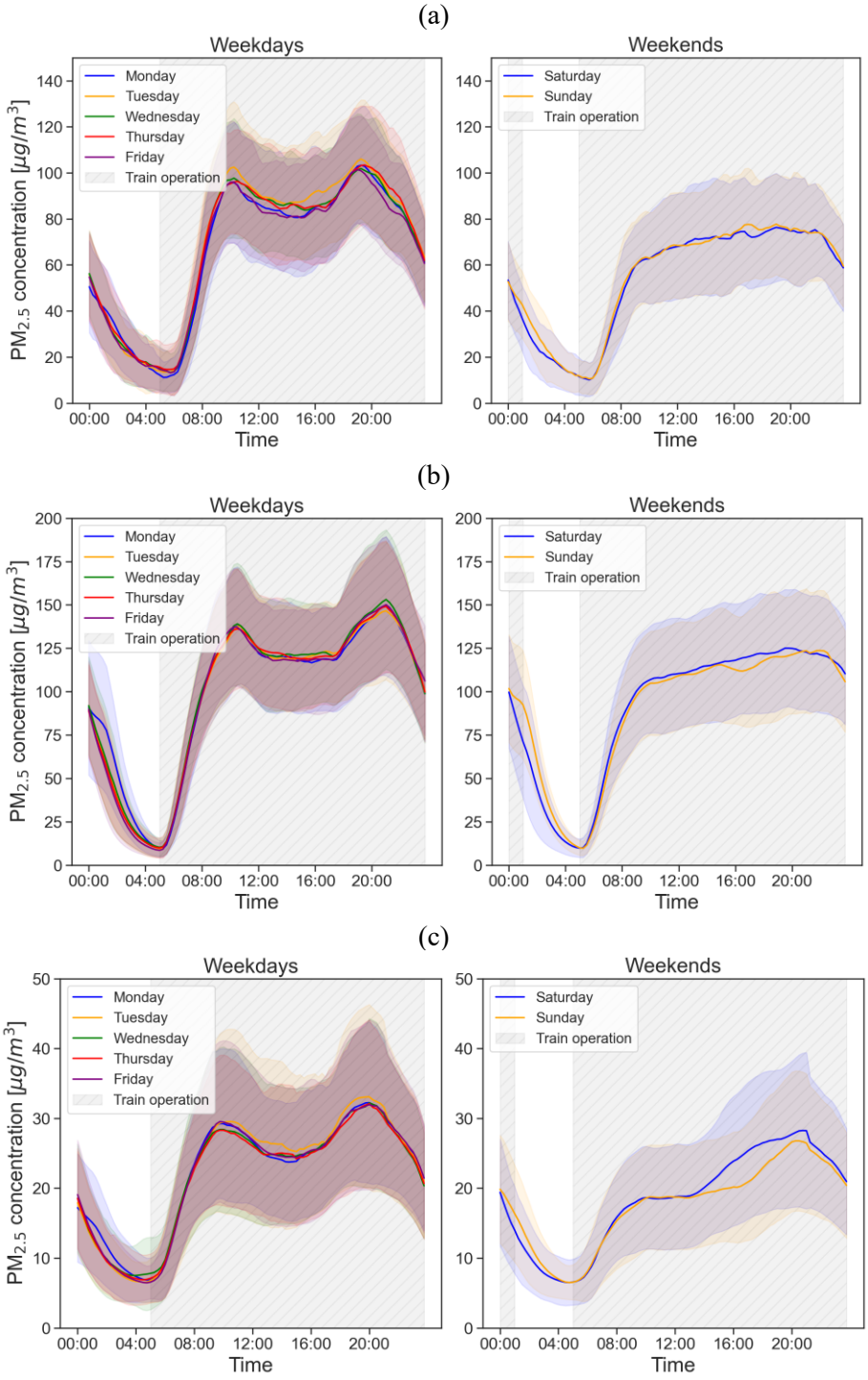


Figure A. 5 : Profils typiques des concentrations de $PM_{2.5}$ du lundi au vendredi (à gauche) et du samedi au dimanche (à droite) dans les gares avec mesures longue durée : (a) Avenue Foch, (b) Sevran Beaudottes et (c) Magenta. La zone hachurée indique la période d'exploitation des trains. Les zones ombrées représentent un écart type (σ=1) sur les différents jours.

- **Profils journaliers types des $PM_{2.5}$ dans les gares avec mesures de courte durée**

La Figure A. 6 présente les profils journaliers types des concentrations en $PM_{2.5}$ pour les jours ouvrés et les weekends dans les stations temporaires (campagnes de 2 à 3 semaines).

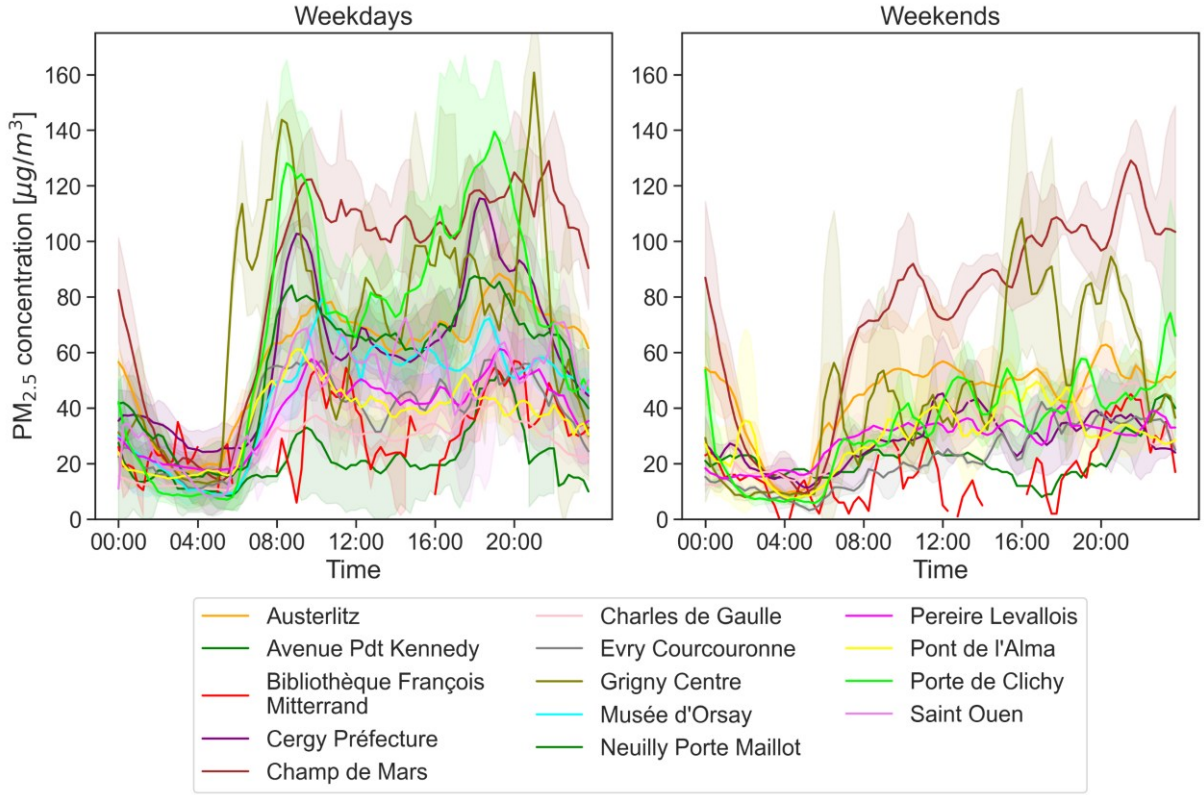


Figure A. 6 : Profils journaliers et écarts-types des concentrations en $PM_{2.5}$ pour les jours ouvrés et les weekends dans les stations temporaires (campagnes de 2 à 3 semaines)

- **Temps de dépôt pour les $PM_{0\text{-}2.5}$ et $PM_{2.5\text{-}10}$**

Pour des raisons de conformité avec les résultats obtenus par Walther et al. (2017), nous avons effectivement examiné le temps de dépôt relatif à la fraction $PM_{2.5\text{-}10}$. La Figure A. 7 montre la variation des concentrations de $PM_{2.5\text{-}10}$ et $PM_{0\text{-}2.5}$ dans les trois gares ayant des mesures de longue durée.

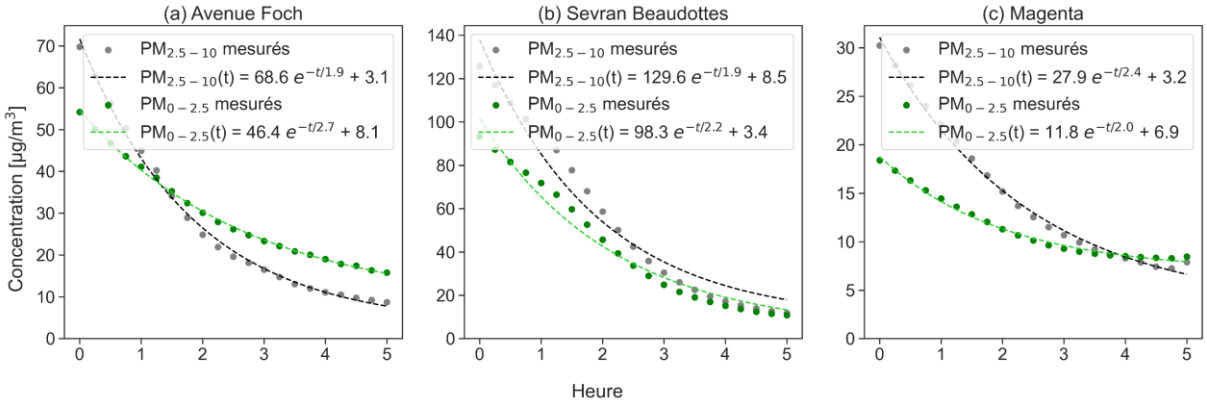


Figure A. 7 : Analyse du temps de dépôt nocturne des concentrations de $PM_{2.5\text{-}10}$ (en gris) et $PM_{0\text{-}2.5}$ (en vert) dans les gares de (a) Avenue Foch, (b) Sevran Beaudottes et (c) Magenta. Les lignes pointillées représentent les ajustements exponentiels des courbes de concentration

Annexe 5 : En extérieur – Profils journaliers types du lundi au vendredi (jours ouvrés) et du samedi au dimanche (weekends)

Cette annexe présente les profils journaliers types des concentrations de PM_{10} et de $PM_{2.5}$ mesurées dans les stations de surveillance extérieures : Porte d'Auteuil, Bobigny et Paris 18ème. Les données sont comparées pour les jours ouvrés (lundi au vendredi) et les weekends (samedi et dimanche).

- **Profils journaliers types des PM_{10}**

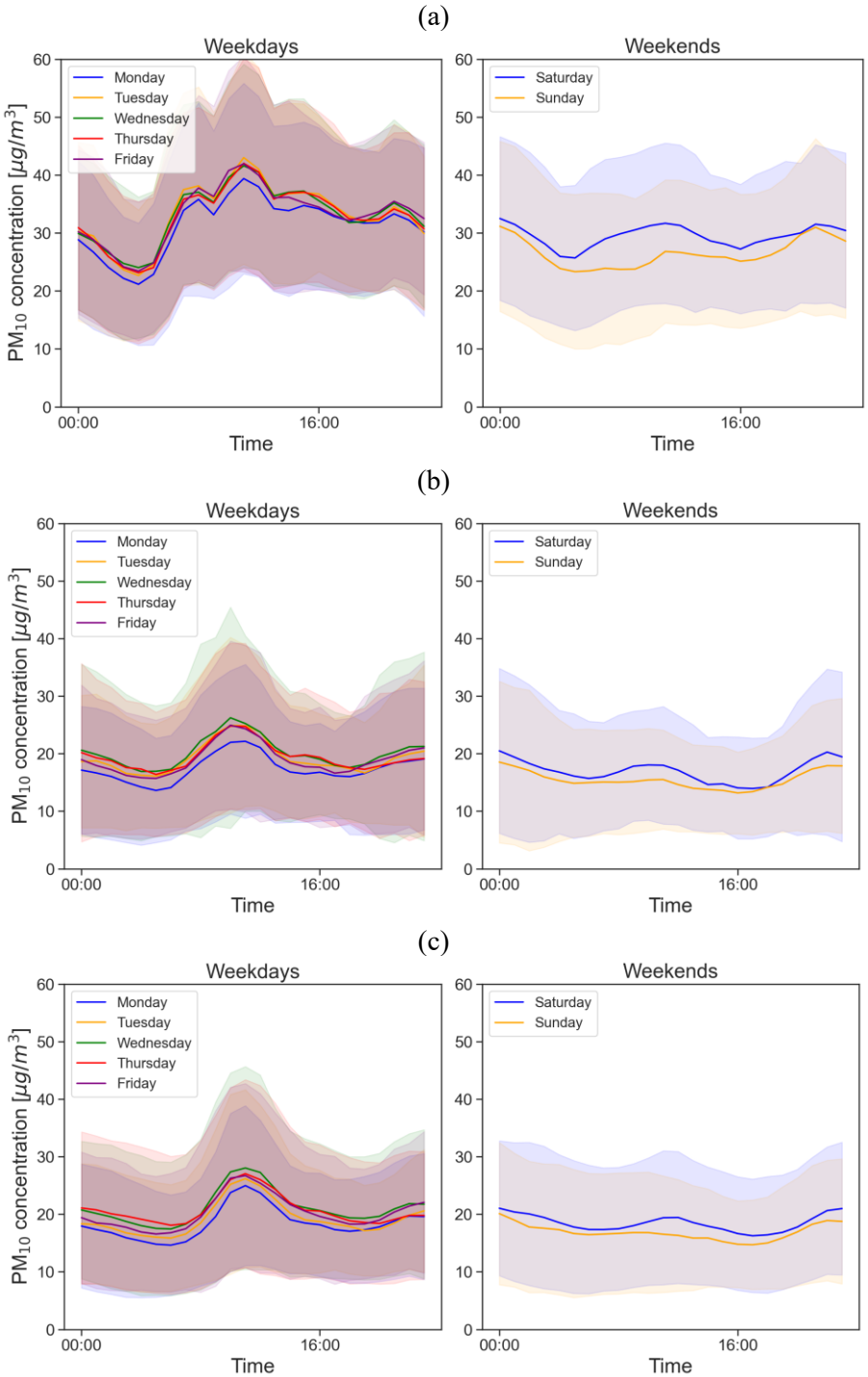


Figure A. 8 : Profils typiques des concentrations de PM_{10} du lundi au vendredi (à gauche) et du samedi au dimanche (à droite) dans les stations de mesure extérieur : (a) Porte d'Auteuil, (b) Bobigny et (c) Paris 18ème. Les zones ombrées représentent un écart type (σ=1)

- **Profils journaliers types des $PM_{2.5}$**

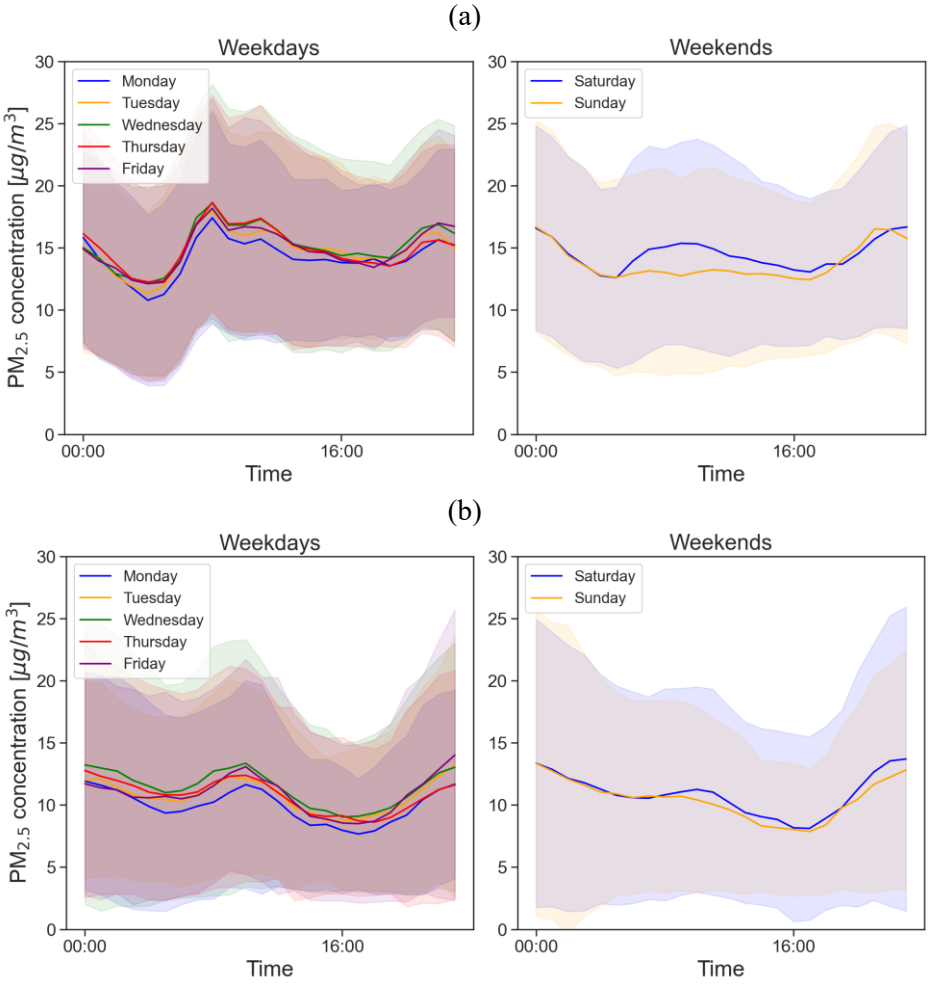


Figure A. 9 : Profils types des $PM_{2.5}$ pour les jours de semaine (à gauche) et les weekends (à droite) dans les stations de mesure extérieur : (a) Porte d'Auteuil et (b) Bobigny. Les zones ombrées représentent un écart type (σ=1)

- **Profils journaliers types des $PM_{2.5}$ les jours ouvrés et les weekends dans les stations de mesures extérieures : Porte d'Auteuil et Bobigny**

Les profils journaliers types des $PM_{2.5}$ (Figure A. 10) sont obtenus en calculant la moyenne des profils journaliers pour chaque catégorie (Figure A. 9) : du lundi au vendredi pour les jours ouvrés, et du samedi au dimanche pour les weekends. Les écarts-types associés, représentés sur les figures, sont calculés en appliquant les lois de propagation des incertitudes aux écarts-types individuels des jours ouvrés et des weekends (Equation 4.19).

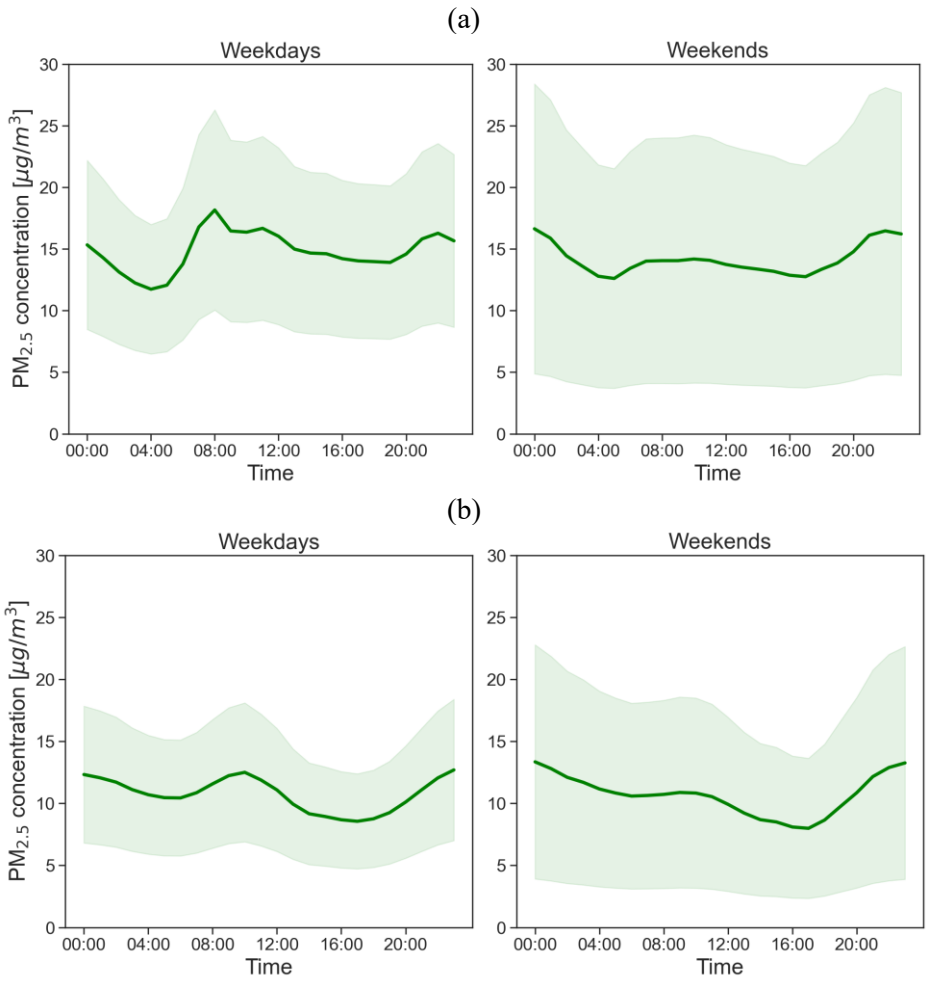


Figure A. 10 : Profils types des $PM_{2.5}$ pour les jours de semaine (à gauche) et les weekends (à droite) dans les stations de mesure extérieur : (a) Porte d'Auteuil et (b) Bobigny. Les zones ombrées représentent un écart type (σ=1)

RÉSUMÉ

Les enceintes ferroviaires souterraines (EFS) présentent des concentrations en particules (PM_{10} et $PM_{2.5}$) souvent supérieures aux seuils d'exposition recommandés par l'OMS, ainsi qu'à celles mesurées dans l'air extérieur. Cette situation peut avoir des impacts négatifs, tant sur la santé des voyageurs et des agents que sur l'image et l'attractivité des transports ferroviaires. Consciente de ces enjeux, la SNCF mène depuis 2016 des campagnes de mesure pour suivre les $PM_{2.5}$ et PM_{10} dans l'ensemble de ses gares souterraines franciliennes. Parmi ces gares, trois sont instrumentées en continu depuis plusieurs années. De plus, des expérimentations de filtration de l'air au niveau des quais ont été réalisées pour améliorer la qualité de l'air dans ces environnements spécifiques. Cependant, l'analyse et l'interprétation de ces mesures représentent aujourd'hui encore un vrai défi, notamment la définition de valeurs de référence. C'est la raison pour laquelle la SNCF a souhaité explorer et développer de nouvelles approches.

Les travaux présentés dans cette thèse s'attèlent donc au développement de nouvelles méthodes d'analyse, adaptées aux dynamiques spécifiques des EFS, notamment les très fortes fluctuations des concentrations en particules. Pour ce faire, à partir des mesures longue durée (plusieurs années), nous avons développé une méthodologie rigoureuse et robuste pour déterminer les profils journaliers types dans chaque gare et un coefficient d'amplitude journalier (DAC – daily amplitude coefficient), en distinguant jours de semaine et weekends. Nos résultats révèlent une évolution saisonnière des concentrations, avec des niveaux plus élevés en été et plus faibles en hiver, qui diffère donc de l'évolution rapportée dans la littérature dans les EFS et dans l'air extérieur. L'approche DAC permet de s'affranchir des fluctuations saisonnières pour analyser l'évolution des concentrations de particules sur le long terme. Elle facilite également les comparaisons entre les différentes gares et contribue à identifier quelques paramètres influençant les concentrations en particules, tels que la fréquence des trains, la présence de systèmes de ventilation, la typologie de la gare (souterraine ou partiellement souterraine) ou la ligne ferroviaire concernée.

Une analyse plus poussée a été réalisée à l'aide de Modèles Additifs Généralisés (GAM). Elle a permis d'identifier les facteurs les plus importants et de quantifier leur impact, parmi lesquels l'heure de la journée, reflétant la fréquence des trains, la période de l'année et les niveaux de CO_2. Les effets de la température et de l'humidité, mesurés au niveau du quai de la gare, sont moins marqués, et la contribution de l'air extérieur est très faible, ce qui suggère que les principaux facteurs influençant les concentrations de particules proviennent principalement de l'intérieur des EFS.

Enfin, trois expérimentations de dépollution ont été évaluées : une technologie d'ionisation positive, une technologie par voie humide et une filtration classique. Les résultats obtenus via DAC et GAM montrent une efficacité variable selon les technologies et les classes de particules (PM_{10} et $PM_{2.5}$), avec des taux d'efficacité allant de 0 à 34 % selon la technologie utilisée. Nous avons également mis en œuvre un modèle de réseaux de neurones récurrents (LSTM), qui permet d'analyser l'efficacité de ces systèmes de filtration de manière quasi immédiate et non a posteriori des concentrations.

Cette thèse propose ainsi une méthodologie rigoureuse et adaptable pour comprendre la dynamique des concentrations en particules dans les EFS, applicable aux particules mais aussi à d'autres polluants. Les protocoles développés fournissent des outils immédiatement exploitables par SNCF, qui pourra les utiliser pour améliorer de manière durable la qualité de l'air dans ces environnements complexes. Des recommandations concernant les futures campagnes de mesures ont également été émises.

Mots clés :

Qualité de l'air, pollution intérieure, particules, enceintes ferroviaires souterraines, systèmes de filtration

ABSTRACT

Underground railway stations (URS) often exhibit concentrations of particulate matter (PM_{10} and $PM_{2.5}$) that exceed the exposure limits recommended by the World Health Organization, as well as those measured in outdoor air. This situation can have negative impacts, both on the health of travelers and employees and on the image and attractiveness of rail transport. Aware of these issues, since 2016 SNCF has been conducting measurement campaigns to monitor PM2.5 and PM10 across its underground stations in the Île-de-France region. Among these stations, three have been continuously instrumented for several years. Moreover, air filtration experiments at platform levels have been carried out to improve air quality in these specific environments. However, analyzing and interpreting these measurements remains challenging, particularly in defining reference values. This is why SNCF wished to explore and develop new approaches.

The work presented in this thesis focuses on the development of new analysis methods, adapted to the specific dynamics of URS, especially the very strong fluctuations in particle concentrations. Using long-term measurements (several years), we have developed a rigorous and robust methodology to determine typical daily profiles in each station and a daily amplitude coefficient (DAC), distinguishing weekdays from weekends. Our results reveal a seasonal evolution of concentrations, with higher levels in summer and lower in winter, which differs from the trends reported in the literature in URS and outdoor air. The DAC approach allows us to disregard seasonal fluctuations to analyze the long-term evolution of particle concentrations. It also facilitates comparisons between different stations and contributes to identifying some parameters influencing particle concentrations, such as train frequency, the presence of ventilation systems, the type of station (underground or partially underground), and the railway line concerned.

A more detailed analysis was carried out using Generalized Additive Models (GAM). It helped identify the most significant factors and quantify their impact, including the time of day, reflecting train frequency, the period of the year, and CO_2 levels. The effects of temperature and humidity, measured at the platform level, are less pronounced, and the contribution of outdoor air is very low, suggesting that the main factors influencing particle concentrations mainly originate from within the URS.

Finally, three depollution experiments were evaluated: a positive ionization technology, a wet technology, and traditional filtration. The results obtained via DAC and GAM show variable effectiveness depending on the technologies and particle classes (PM_{10} and $PM_{2.5}$), with efficiency rates ranging from 0 to 34% depending on the technology used. We also implemented a model of recurrent neural networks (LSTM), which allows for analyzing the effectiveness of these filtration systems almost immediately and not a posteriori of the concentrations.

This thesis thus proposes a rigorous and adaptable methodology for understanding the dynamics of particle concentrations in URS, applicable to particles but also to other pollutants. The protocols developed provide immediately usable tools for SNCF, which can use them to sustainably improve air quality in these complex environments. Recommendations for future measurement campaigns have also been outlined.

Keywords:

Air quality, indoor pollution, particulate matter, underground railway stations, filtration systems