



**HAL**  
open science

# Exposition individuelle à la pollution de l'air : mesure par capteurs miniatures, modélisation et évaluation des risques sanitaires associés

Boris Dessimond

► **To cite this version:**

Boris Dessimond. Exposition individuelle à la pollution de l'air : mesure par capteurs miniatures, modélisation et évaluation des risques sanitaires associés. Santé publique et épidémiologie. Sorbonne Université, 2021. Français. NNT : 2021SORUS297 . tel-03554640

**HAL Id: tel-03554640**

**<https://theses.hal.science/tel-03554640v1>**

Submitted on 3 Feb 2022

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

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

# THÈSE DE DOCTORAT

ÉCOLE DOCTORALE INFORMATIQUE,  
TÉLÉCOMMUNICATIONS ET ÉLECTRONIQUE

*Laboratoire d'Informatique de Paris 6  
Équipe Networks and Performance Analysis*

## EXPOSITION INDIVIDUELLE À LA POLLUTION DE L'AIR MESURE PAR CAPTEURS MINIATURES, MODÉLISATION ET ÉVALUATION DES RISQUES SANITAIRES ASSOCIÉS

*Présentée et soutenue publiquement le 22 décembre 2021 par*

BORIS DESSIMOND

*Pour obtenir le titre de*

DOCTEUR DE SORBONNE UNIVERSITÉ

### RAPPORTEURS

YAMIN LEPRINCE

*Professeur, Université Gustave Eiffel, Laboratoire ESYCOM*

MOHAMED BENBOUZID

*Professeur, Université de Bretagne Occidentale, IRDL*

### EXAMINATEURS

JEAN-FRANÇOIS BERNAUDIN

*Professeur, Sorbonne Université, Laboratoire CSPBAT*

YOUCEF AIT-AMIRAT

*Maître de conférences, Université de Franche-Comté, FEMTO-ST*

### THÈSE DIRIGÉE PAR

GIOVANNI PAU

*Professeur, Sorbonne Université, Université de Bologne*

MARIA POTOP-BUTUCARU

*Professeur, Sorbonne Université, Laboratoire LIP6*

### CO-ENCADRÉE PAR

ISABELLA ANNESI-MAESANO

*Directrice de recherche, Université de Montpellier, INSERM*

SALIM SRAIRI

*Docteur, Responsable R&D, CEREMA Île-de-France*



*Aux artistes d'aujourd'hui qui inspireront la science de demain,*

*Nanos gigantum humeris insidentes*



# REMERCIEMENTS

*Cette thèse a été produite au sein de l'Institut Pierre Louis d'Epidémiologie et de Santé Publique, sur le campus Saint-Antoine, qui m'a accueilli et dans lequel j'ai intégré pendant ces trois années, une équipe de recherche mixte. Je tiens donc à remercier spécialement Isabella Annesi-Maesano, la directrice d'équipe, qui a codirigé cette thèse et m'a énormément appris de par sa perspicacité. Merci de m'avoir offert cette opportunité et d'avoir été présente durant ces trois ans. J'ai ainsi eu la chance de collaborer avec statisticiens, géographes, pneumologues, neurologues, dont le dévouement et l'engagement ont été une source d'inspiration. Je remercie donc les membres de l'équipe et les personnes qui m'ont accompagné durant à mes travaux nommément ; Edouard Januel, Lucile Sese, Cara Maesano, Giancarlo Pesce, Julie Prudhomme, Lhote Raphael, Valentin Vinnat, ainsi que l'équipe administrative, notamment Marianne Bailly.*

*J'aimerais remercier mon équipe de direction. Particulièrement Salim Srairi, mon co-encadrant. Il m'a continuellement conseillé, a toujours été présent. J'ai pu apprécier ses qualités humaines et scientifiques. Je suis également redevable à ma directrice Maria Potop-Butucaru qui m'a supervisé durant ma dernière année, ainsi qu'à mon directeur, Giovanni Pau. Je salue par ailleurs l'équipe d'étudiants exceptionnels dont il s'est entouré, particulièrement Davide Aguiari et Furong Yang.*

*Je remercie Madame Leprince Yamin, Professeur des Universités à l'Université Gustave Eiffel, ainsi que Monsieur Benbouzid Mohamed, Professeur des Universités à l'Université de Brest pour m'avoir fait l'honneur d'examiner mon manuscrit en qualité de rapporteurs. Je remercie également Monsieur Bernaudin Jean-François, Professeur émérite à Sorbonne Université et Monsieur Ait-Amirat Youcef, Maître de conférences à l'Université de Franche-Comté, pour l'intérêt porté à mes travaux.*

*Merci à tous les membres des projets POLLUSCOPE et Pollar en particulier Hafsa El Hafyani, Robin Terrail, Basile Chaix, Jean-Marc Naude.*

*Merci à Karine Zeitouni, Professeur à l'Université de Versailles Saint-Quentin-en-Yvelines ainsi qu'à Lom Messan Hillah, Maître de conférences à Sorbonne Université, pour les précieux conseils qu'ils m'ont apportés, de par leur participation active aux comités de suivi.*

*Merci à Clive Eric Sabel, Professeur à l'Université de Aarhus, pour m'avoir convié à l'université d'été dont il a été l'organisateur ; une expérience mémorable où j'ai pu rencontrer des personnes brillantes.*

*J'aimerais également formuler la due reconnaissance à ces personnalités qui ont joué un rôle clef dans le déroulement de ma formation universitaire ; Abdenour Soualhi, Khennouf Salah, Cyril Constant.*

*J'ai une pensée pour ma mère, Viviane Tissandier, ma grand-mère Jeanine Tissandier, ma sœur, Elsa Dessimond, mon défunt grand-père Alexandre Tissandier, bien distants de moi, mais qui m'ont chacun porté et dont je porte désormais les valeurs.*

*Ma seconde famille, ces amis proches qui m'accompagnent et auxquels je dois beaucoup ; Adrian Euler, Pierre Monniot, Pierre-Yves Celle, Yoann Brunel.*

*Cette thèse fut une expérience bouleversante, qui au-delà d'obtenir comme seul fruit des compétences techniques, m'a amené à me développer sur le plan personnel. Pour cette raison, je suis infiniment redevable à mon école doctorale l'EDITE, plus largement, à Sorbonne Université et à l'Agence Nationale de la Recherche, pour avoir financé mes travaux, au travers du projet POLLUSCOPE (ANR-15-CE22-0018).*

*Très humblement,*



## RÉSUMÉ

La pollution de l'air contribue à dégrader la qualité de vie et à réduire l'espérance de vie des populations. L'organisation mondiale de la santé estime que la pollution de l'air est responsable de 7 millions de morts par an dans le monde. Elle participe à aggraver les maladies respiratoires, cause des cancers du poumon et des crises cardiaques. La pollution de l'air a donc des conséquences sanitaires importantes sur la vie humaine et la biodiversité.

Ces dernières années, des progrès considérables ont été réalisés dans le domaine des microcontrôleurs et des modules de télécommunications. Ces derniers sont de plus en plus efficaces énergétiquement, performants, abordables, accessibles et sont responsables de l'émergence des objets connectés. Parallèlement, les récents développements des microsystèmes électromécaniques et des capteurs électrochimiques ont permis la miniaturisation des technologies permettant de mesurer de nombreux paramètres environnementaux, dont la qualité de l'air. Ces avancées technologiques ont ainsi permis la conception et la production dans un cadre académique, de capteurs de la qualité de l'air ; des outils de mesure portatifs, connectés, autonomes et en capacité de réaliser des acquisitions à une fréquence temporelle élevée.

Jusqu'à récemment, l'un des principaux freins à la compréhension de l'impact de la pollution de l'air sur la santé fut l'impossibilité de connaître l'exposition réelle des individus durant leur vie quotidienne ; la pollution de l'air est complexe et varie en fonction des habitudes, des activités et environnements empruntés par les individus. Ces capteurs portatifs de la qualité de l'air permettent donc de lever entièrement ce frein ainsi qu'un nombre important de contraintes. Ils sont conçus pour être utilisables en mobilité, sur de longues périodes et produisent des données granulaires, immédiatement disponibles, décrivant l'exposition à la pollution de l'air du porteur. Bien que les modules de mesure embarqués dans ces capteurs ne soient aujourd'hui pas aussi performants que les instruments de références ou la télédétection, lorsqu'il s'agit d'évaluer l'exposition individuelle à la pollution de l'air, parce qu'ils sont au plus proche des individus, ils permettent d'obtenir l'information la plus fidèle et constituent donc un outil indispensable pour l'avenir de la recherche épidémiologique.

Dans ce contexte, nous avons participé au développement et à l'amélioration de deux capteurs de la qualité de l'air ; le CANARIN II et le CANARIN nano. Le CANARIN II est un capteur connecté communiquant par Wi-Fi, qui rapporte les concentrations de particules de diamètre 10, 2.5 et 1 micromètre, ainsi que les paramètres environnementaux de température, humidité et pression, chaque minute et les rend disponibles en temps réel. Le CANARIN nano est, quant à lui, un capteur de plus petite taille, possédant les mêmes capacités que le CANARIN II, tout en faisant en outre l'acquisition des composés organiques volatils. Il est en capacité de fonctionner de manière autonome, puisque communiquant par réseau cellulaire. Deux types de résultats ont été obtenus avec les capteurs CANARIN ; d'une part, des résultats produits à partir de leur utilisation dans des conditions de vie réelle, d'autre part, des résultats en lien avec l'interprétation et la compréhension des mesures produites par les capteurs de particules dont les CANARINs sont équipés.



Ces deux capteurs furent ainsi tous deux exploités par deux projets de recherche, au sein desquels nous avons accompagné le déploiement de plusieurs flottes de capteurs hétérogènes et réalisé l'analyse des données acquises. Premièrement au sein du projet ANR POLLUSCOPE, où 86 volontaires issus de la population générale ont porté un set de capteurs de mesure de la pollution de l'air, totalisant 101 semaines de port, dont 35 durant lesquelles les volontaires ont également été équipés de capteurs de santé. Nous avons ainsi pu analyser le lien entre la mobilité, l'évolution de la fonction respiratoire et l'exposition individuelle à la pollution de l'air. Deuxièmement au sein du projet EU EIT HEALTH POLLAR, où 43 sujets ont effectué une polysomnographie puis porté un capteur CANARIN durant une dizaine de jours, permettant d'explorer pour la première fois de façon objective le lien entre l'apnée du sommeil et l'exposition aux particules fines.

Dans l'objectif de mieux comprendre, qualifier et corriger la dispersion résultante de l'erreur de mesure au sein des capteurs de la qualité de l'air, nous avons imaginé, conçu, puis fabriqué un banc d'évaluation. Dans la thèse nous présenterons son processus de fabrication, les résultats issus de simulations numériques, puis analyserons les données produites. Grâce à un algorithme dédié ainsi qu'aux données produites par notre banc d'évaluation, nous ferons la démonstration de la réduction de la dispersion présente dans un groupe de compteurs optiques de particules.

Finalement, notre participation active au sein de ces projets exploitant ces nouvelles technologies nous a apporté un niveau de compréhension avancé des contraintes qui y sont liées, qu'elles soient techniques, humaines, budgétaires, logistiques ou organisationnelles. Nous proposons, en considération des technologies existantes et futures, une architecture de capteurs réduisant les contraintes les plus importantes, notamment en rendant centrale la relation de l'individu avec le capteur, permettant ainsi une amélioration de la quantité et qualité des données recueillies.

# ABSTRACT

## INDIVIDUAL EXPOSURE TO AIR POLLUTION, MEASUREMENT BY MINIATURE SENSORS, MODELING AND ASSESSMENT OF ASSOCIATED HEALTH RISKS

Air pollution contributes to the degradation of the quality of life and the reduction of life expectancy of populations. The World Health Organization estimates that air pollution is responsible for 7 million deaths per year worldwide. It contributes to the aggravation of respiratory diseases, causes lung cancer and heart attacks. Air pollution therefore has significant health consequences on human life and biodiversity.

Over the last few years, considerable progress has been made in the field of microcontrollers and telecommunications modules. These are more energy efficient, powerful, affordable, accessible, and are responsible for the growth of connected objects. In the meantime, the recent development of microelectromechanical systems and electrochemical sensors has allowed the miniaturization of technologies measuring many environmental parameters including air quality. These technological breakthroughs have enabled the design and production in an academic environment of portable, connected, autonomous air quality sensors capable of performing acquisitions at a high temporal frequency.

Until recently, one of the major obstacles to understanding the impact of air pollution on human health was the inability to track the real exposure of individuals during their daily lives; air pollution is complex, and varies according to the habits, activities and environments in which individuals spend their lives. Portable air quality sensors completely remove this obstacle as well as a number of other important constraints. These are designed to be used in mobility, over long periods of time, and produce immediately available granular data, which describes the exposure to air pollution of the person wearing it. Although the measurement modules embedded in these sensors are not currently as reliable as reference tools or remote sensing, when it comes to assessing individual exposure to air pollution, because they are as close as possible to the wearer, they provide the most accurate information, and are therefore an indispensable tool for the future of epidemiological research.

In this context, we have been involved in the development and improvement of two air quality sensors; the CANARIN II and the CANARIN nano. The CANARIN II is a connected sensor communicating via Wi-Fi, which reports the concentration of 10, 2.5 and 1 micrometer diameter particles, as well as the environmental parameters of temperature, humidity, and pressure, every minute, making them available in real time. The CANARIN nano is a smaller sensor which has the same capabilities as the CANARIN II, while additionally sensing volatile organic compounds levels. The CANARIN nano is able to operate autonomously, as it communicates through the cellular network. Two types of results have been obtained with the CANARIN sensors; on one hand, results produced from their use in real life conditions, and on the other hand, results related to the interpretation and understanding of the measurements produced by the particle sensors.

These two sensors were both used in two research projects, in which we have helped deploy several heterogeneous sensor fleets and analyzed the acquired data. Firstly, in the POLLUSCOPE project funded by the French National Research Agency, where 86 volunteers from the general population wore a set of air pollution sensors for a total of 101 weeks, 35 of which the volunteers were also equipped with health sensors. Secondly, in the POLLAR project, where 43 subjects underwent polysomnography and then wore one CANARIN sensor for 10 days, thus allowing for the first time to explore the link between sleep apnea and particulate matter exposure.

In order to better understand, qualify or correct the dispersion resulting from measurement errors within a batch of air quality sensors, we have imagined, designed, and produced a test bench. In the thesis, we will present this test bench, its manufacturing process, its performance, and will evaluate it by analyzing the produced data. With a dedicated algorithm, we will demonstrate the reduction of the dispersion among a batch of optical particle counters as those used in the CANARIN sensors.

Finally, our involvement in these projects using these new technologies has brought us an advanced level of understanding of the related technical, human, budgetary, logistical and organizational constraints. We will propose, in consideration of existing and future technologies, a sensor architecture that eliminates the most important constraints, in particular by making the individual's relationship with the sensor at the center of the system, thus allowing an improvement in the quality and quantity of the collected data.

# TABLE DES MATIÈRES

1. INTRODUCTION À LA QUALITÉ DE L'AIR .....	1
1.1. CHIMIE DES POLLUANTS ET RISQUES ASSOCIES .....	2
1.1.1. SOURCES DES POLLUANTS ATMOSPHÉRIQUES ET EFFETS SANITAIRES .....	2
1.1.2. PERCEPTION DU RISQUE PAR LA POPULATION .....	4
1.2. REGLEMENTATIONS DE LA POLLUTION DE L'AIR .....	5
1.3. OUTILS ET METHODES DE MESURE DE LA QUALITE DE L'AIR .....	6
1.3.1. STANDARDS ET INSTRUMENTS DE RÉFÉRENCE.....	6
1.3.2. STATIONS DE MESURE.....	7
1.3.3. MODÉLISATION ATMOSPHÉRIQUE.....	7
1.3.4. TÉLÉDÉTECTION.....	8
1.3.5. MICRO-CAPTEURS ET MODULES DE MESURE.....	9
1.3.6. CAPTEURS PORTATIFS.....	10
1.3.7. EXPOSITION INDIVIDUELLE .....	11
1.4. APPROCHE EPIDEMIOLOGIQUE ET PERSPECTIVES IoT.....	12
1.4.1. ÉTUDES ÉPIDÉMIOLOGIQUES.....	12
1.4.2. MULTIPLICATION DES SOURCES DE DONNÉES.....	12
1.4.3. LE QUANTIFIED SELF .....	13
2. OBJECTIFS ET JUSTIFICATION DE LA THÈSE.....	15
2.1. OBJECTIFS .....	15
2.2. PLAN DE THESE.....	15
3. POPULATIONS ET MÉTHODES .....	17
3.1. POPULATIONS CONSIDEREES .....	17
3.1.1. CRITÈRES D'EXCLUSION .....	17
3.2. TRAITEMENT DES DONNEES DE POLLUTION ATMOSPHERIQUE .....	17
3.3. ACQUISITION DE DONNEES CONTEXTUELLES .....	18
3.4. INDICATEURS DE SANTE.....	18
4. PROJET CANARIN.....	19
4.1. HISTORIQUE DU PROJET CANARIN.....	19
4.1. TECHNOLOGIES EMPLOYEES.....	20
4.1.1. CANARIN II.....	20
4.1.2. CANARIN NANO.....	21
4.1.3. INFRASTRUCTURE CLOUD.....	23
4.1.4. FACE AUX CAPTEURS EXISTANTS.....	23
5. PROJET POLLUSCOPE : OBSERVATOIRE PARTICIPATIF POUR LA SURVEILLANCE DE L'EXPOSITION INDIVIDUELLE À LA POLLUTION DE L'AIR EN LIEN AVEC LA SANTÉ.....	24
5.1. OBJECTIFS .....	24
5.1.1. SÉLECTION DES CAPTEURS .....	24
5.1.2. INSTRUMENTATION DES PARTICIPANTS .....	25
5.1.3. DISSÉMINATION DES DONNÉES D'EXPOSITION .....	27
5.2. RECRUTEMENT DE PARTICIPANTS.....	27
5.2.1. RECRUTEMENT AVEC ACCOMPAGNEMENT.....	28
5.2.2. RECRUTEMENT PAR VOLONTARIAT.....	29
5.2.3. DÉROULEMENT DE L'ENQUÊTE .....	29
5.3. TECHNOLOGIE DES CAPTEURS ET INSTRUMENTS EMPLOYES.....	30
5.3.1. COMPTEUR OPTIQUE DE PARTICULES.....	30
5.3.2. CAPTEUR ÉLECTROCHIMIQUE DE NO <sub>2</sub> .....	31
5.3.3. CAPTEUR DE BC.....	31
5.3.4. SPIROMÈTRE.....	31
5.3.5. OXYMÈTRE .....	32
5.4. RESULTATS.....	33

5.4.1. DESCRIPTION DES PARTICIPANTS.....	33
5.4.2. DISPARITÉS LIÉES AU RECRUTEMENT.....	33
5.5. EXPLOITATION DES DONNEES DE SPIROMETRIE.....	34
5.5.1. ÉVALUATION DU TAUX D'UTILISATION.....	35
5.5.2. SÉLECTION DES SPIROMÉTRIES.....	36
5.5.3. MISE EN RELATION DE LA FONCTION RESPIRATOIRE AVEC LES DONNÉES D'EXPOSITION À LA POLLUTION ATMOSPHÉRIQUE.....	39
5.5.4. RÉSULTATS : RELATIONS ENTRE L'EXPOSITION À LA POLLUTION ATMOSPHÉRIQUE ET LA FONCTION RESPIRATOIRE.....	41
5.6. DISCUSSION ET CONCLUSIONS.....	42
6. PROJET POLLAR : ÉTUDE DE LA RELATION ENTRE L'EXPOSITION À LA POLLUTION ATMOSPHÉRIQUE ET L'APNÉE DU SOMMEIL.....	44
6.1. OBJECTIFS DU PROJET.....	44
6.2. POPULATION ET METHODES.....	44
6.3. EXPOSITION AUX PARTICULES FINES EN RELATION AVEC L'APNÉE DU SOMMEIL.....	45
6.4. DISCUSSION ET CONCLUSIONS.....	46
7. BANC D'ESSAI MOBILE POUR L'ÉVALUATION DE LA DISPERSION.....	47
7.1. OUTILS LOGICIELS.....	48
7.2. NOMENCLATURE.....	48
COMPOSANTS PRINCIPAUX :.....	48
OUTILLAGE :.....	48
CONSOMMABLES :.....	48
7.3. PRINCIPE DE FONCTIONNEMENT ET FABRICATION.....	49
7.4. COUTS ASSOCIES.....	52
7.5. VALIDATION PAR SIMULATION NUMERIQUE.....	53
7.5.1. DIFFUSEUR.....	54
7.5.2. EFFET DES DIFFUSEURS.....	55
7.5.3. CONCLUSIONS MODÉLISATION ET PROTOTYPAGE.....	57
7.6. ÉTUDE DE LA DISPERSION.....	58
7.6.1. DESCRIPTION DES ENVIRONNEMENTS.....	58
7.6.2. VARIATIONS DE LA DISPERSION.....	59
7.6.3. POSITIONNEMENT ET VARIATIONS DE LA DISPERSION.....	60
7.6.4. CONDITIONS MÉTÉOROLOGIQUES ET VARIATIONS DE LA DISPERSION.....	62
7.6.5. ÉVÉNEMENTS DE DISPERSION INDIVIDUELS.....	65
7.7. DISCUSSION ET CONCLUSIONS.....	68
8. PROCESSUS DE RÉDUCTION DE LA DISPERSION INTRA-SET.....	69
8.1. INTRODUCTION.....	69
8.1.1. DÉFINITION DE LA PROBLÉMATIQUE.....	69
8.2. CHANGEMENT DE PARADIGME.....	69
8.3. OBJECTIFS.....	70
8.4. METHODES.....	70
8.4.1. ÉTALONNAGE DE CAPTEURS.....	70
8.4.2. VALIDATIONS DE CAPTEURS.....	71
8.4.3. LA DISPERSION PROPRE.....	72
8.4.4. LA DISPERSION INTRA-SET.....	72
8.4.5. CARACTÉRISATION ET PARAMÈTRES DE DISPERSION.....	73
8.4.6. TRANSPOSABILITÉ DE LA CARACTÉRISATION.....	74
8.5. SOURCES DE DISPERSION.....	74
8.5.1. CONDITIONS EXTÉRIEURES.....	74
8.5.2. INTÉGRATION DES MICRO-CAPTEURS.....	75
8.5.3. DURÉE DE VIE EN UTILISATION RÉELLE.....	75
8.6. JEUX DE DONNEES EXPLOITES.....	76
8.6.1. JEUX DE DONNÉES ISSUS DE VALIDATIONS :.....	76
8.6.2. JEUX DE DONNÉES ISSUS DE CAPTEURS FIXES :.....	76

8.6.3. RÉCAPITULATIF .....	76
8.7. ALGORITHME D'ESTIMATION DES CENTRES .....	77
8.7.1. DESCRIPTION DE L'ALGORITHME D'ESTIMATION DES CENTRES .....	77
8.7.2. ESTIMATION DES CENTRES SUR DONNÉES SIMULÉES.....	78
8.7.3. ESTIMATION DES CENTRES SUR DONNÉES RÉELLES .....	79
8.8. DESCRIPTION DU PROCESSUS SBDR.....	80
8.8.1. ILLUSTRATION DU PARADIGME SDBR : .....	80
8.8.2. DÉCOMPOSITION DU PROCESSUS SBDR .....	81
8.9. PARAMETRES DE L'ALGORITHME SBDR.....	84
8.9.1. POIDS DE LA CORRECTION.....	84
8.9.2. MODULATION DES CRITÈRES DE SÉLECTIVITÉ.....	85
8.9.3. PARAMÈTRES SECONDAIRES .....	85
8.10. APPLICATION DE L'ALGORITHME SBDR.....	86
8.10.1. PERFORMANCE SANS TRANSPOSITION.....	86
8.10.2. INJECTION DE DONNÉES ISSUES D'UN INSTRUMENT DE RÉFÉRENCE.....	88
8.10.3. SURVEILLANCE PRÉVENTIVE.....	89
8.10.4. TRANSPOSABILITÉ DES PROFILS DE CORRECTION.....	89
8.11. DISCUSSION ET CONCLUSIONS.....	90
8.11.1. MULTI-DIMENSIONNALITÉ.....	90
8.11.2. PERSPECTIVES .....	91
9. CONTEXTUALISATION DES DONNÉES.....	92
9.1. EXPOSITION INDIVIDUELLE EN FONCTION DE LA MOBILITE.....	92
9.2. OBSERVATIONS .....	96
9.3. DISCUSSION ET CONCLUSIONS.....	97
10. ÉVOLUTION DES ARCHITECTURES DE CAPTEURS.....	98
10.1. INTRODUCTION .....	98
10.2. ILLUSTRATION DE L'ARCHITECTURE .....	98
10.3. OBJECTIFS.....	99
10.4. SÉLECTION DU MATÉRIEL.....	100
10.5. AUTONOMIE DE L'ARCHITECTURE.....	100
10.6. OPTIMISATION DE LA CONSOMMATION ÉNERGETIQUE.....	101
10.7. INTERFAÇAGE AVEC LES SMARTPHONES.....	101
10.8. CONTRAINTES DE DÉPLOIEMENT .....	101
10.9. DISCUSSION ET CONCLUSIONS.....	103
11. CONCLUSION ET PERSPECTIVES .....	105
11.1. CONCLUSION.....	105
11.2. PERSPECTIVES.....	106
12. ANNEXES.....	107
12.1. ARTICLE : ACADEMICALLY PRODUCED AIR POLLUTION SENSORS FOR PERSONAL EXPOSURE ASSESSMENT: THE CANARIN PROJECT.....	108
12.2. POSTER ET FLYER PROJET POLLUSCOPE.....	126
13. BIBLIOGRAPHIE.....	129

L'ARTICLE « ACADEMICALLY PRODUCED AIR POLLUTION SENSORS FOR PERSONAL EXPOSURE ASSESSMENT: THE CANARIN PROJECT », A ÉTÉ PUBLIÉ LE 8 MARS 2021 DANS LE JOURNAL SENSORS.

LES CHAPITRES ET SECTIONS SUIVANTES CONTIENNENT LA TRADUCTION DE CET ARTICLE ;

4. PROJET CANARIN

5. PROJET POLLUSCOPE

6. PROJET POLLAR

9. CONTEXTUALISATION DES DONNÉES

10. ÉVOLUTION DES ARCHITECTURES DE CAPTEURS

# TABLE DES FIGURES

FIGURE 1 : STATION DE MESURE, BOULEVARD PÉRIPHÉRIQUE DE PARIS	7
FIGURE 2 : VISUALISATION DU MODÈLE CHIMERE	7
FIGURE 3 : ANTENNE DE LA MAISON DE LA TÉLÉDÉTECTION	8
FIGURE 4 : PHOTOMICROGRAPHIE DE DEUX MICRO-CAPTEURS DE LA QUALITÉ DE L'AIR	9
FIGURE 5 : MESURE DE LA CONSOMMATION SUR CARTE DE DÉVELOPPEMENT CANARIN II	20
FIGURE 6 : CAPTEUR CANARIN NANO	22
FIGURE 7 : CANARIN II AVEC UN CAPTEUR CAIRSENS NO <sub>2</sub>	26
FIGURE 8 : FLUX OPÉRATIONNEL DU PROCESSUS DE RECRUTEMENT AVEC ACCOMPAGNEMENT	28
FIGURE 9 : FLUX OPÉRATIONNEL DU PROCESSUS DE RECRUTEMENT PAR VOLONTARIAT	29
FIGURE 10 : FONCTIONNEMENT D'UN COMPTEUR DE PARTICULES OPTIQUE	30
FIGURE 11 : NIVEAU D'ÉDUCATION PAR PROCESSUS DE RECRUTEMENT	33
FIGURE 12 : LOCALISATION DES VOLONTAIRES EN FONCTION DE LA MÉTHODE DE RECRUTEMENT	33
FIGURE 13 : TAUX D'UTILISATION DURANT L'ACQUISITION	36
FIGURE 14 : REPRÉSENTATION GRAPHIQUE DE LA SÉLECTION DES SPIROMÉTRIES	38
FIGURE 15 : AGRANDISSEMENT SUR UNE SESSION D'ACQUISITION	39
FIGURE 16 : AIDE À LA SÉLECTION DU NOMBRE DE SESSIONS REQUISES	40
FIGURE 17 : VISUALISATION 3D DU MODÈLE	41
FIGURE 18 : PRÉDICTIONS DE PENTE VEMS/CVF PAR POLLUANT PAR SEMAINE D'ACQUISITION	41
FIGURE 19 : VUE EN COUPE DU MODÈLE PM <sub>10</sub>	42
FIGURE 20 : SUJET INSTRUMENTÉ AVANT POLYSOMNOGRAPHIE	44
FIGURE 21 : CONCENTRATIONS HORAIRES DE PM <sub>10</sub> SELON LE GROUPE DE L'INDICE D'APNÉE-HYPOPNÉE (IAH).	45
FIGURE 22 : ENVIRONNEMENT DE TEST EN AIR INTÉRIEUR (1 <sup>ER</sup> )	48
FIGURE 23 : ENVIRONNEMENT DE TEST EN AIR EXTÉRIEUR (2 <sup>ÈME</sup> )	49
FIGURE 24 : VUE INTÉRIEURE DU BANC D'ÉVALUATION	50
FIGURE 25 : TUBES SÉPARANT LES SECTIONS	51
FIGURE 26 : PIEDS PERMETTANT L'EXTRACTION D'AIR	51
FIGURE 27 : CONSOMMATION ÉNERGÉTIQUE DU CANARIN NANO EN FONCTION DU TEMPS	52
FIGURE 28 : VUE EN ÉCLATÉ DU BANC D'ÉVALUATION	53
FIGURE 30 : ASSEMBLAGE DIFFUSEUR ET FILTRE	54
FIGURE 29 : DIFFUSEUR DU BANC D'ÉVALUATION	54
FIGURE 31 : PRINCIPE DE FONCTIONNEMENT DU DIFFUSEUR	55
FIGURE 32 : VÉLOCITÉ DE L'AIR EN FONCTION DE LA PRÉSENCE DES DIFFUSEURS, VUE DE DESSUS	55
FIGURE 33 : VÉLOCITÉ DE L'AIR EN FONCTION DE LA PRÉSENCE DES DIFFUSEURS, VUE DE FACE	56
FIGURE 34 : POINTS D'ÉVALUATION DE LA VÉLOCITÉ	57
FIGURE 35 : ENVIRONNEMENT DE TEST EN EXTÉRIEUR (3 <sup>ÈME</sup> )	58
FIGURE 36 : VISUALISATION DE LA RELATION ENTRE ÉCART-TYPE ET MÉDIANE	59
FIGURE 37 : POSITIONNEMENT DES CAPTEURS	60
FIGURE 38 : REQM PAR CAPTEUR ET PAR PÉRIODE	61
FIGURE 39 : CORRÉLATION ENTRE LES SOURCES DE DONNÉES MÉTÉOROLOGIQUE	62
FIGURE 40 : RELATION ENTRE ÉCART-TYPE, TEMPÉRATURE ET HUMIDITÉ	63
FIGURE 41 : PROFIL DES CLASSIFICATIONS D'ÉCART-TYPE	64
FIGURE 42 : FONCTION DE LA DIRECTION DU VENT	64
FIGURE 43 : ISOLATION DES ÉVÉNEMENTS DE DISPERSION	66
FIGURE 44 : DISTRIBUTION DES ÉCHANTILLONS HORS SEUIL	67
FIGURE 45 : ÉVÉNEMENTS DE DISPERSION INDIVIDUELS PAR CAPTEUR	67
FIGURE 46 : JEU DE DONNÉES BANC2 RÉ-ÉCHANTILLONNÉ	73
FIGURE 47 : ÉVOLUTION DE LA SÉLECTION DES CENTRES EN FONCTION DE LA SÉLECTION	78
FIGURE 48 : IDENTIFICATION DE VALEURS ABERRANTES	79
FIGURE 49 : DÉMONSTRATION DU REJET DE VALEURS ABERRANTES	79
FIGURE 50 : INTRODUCTION AU RÉFÉRENTIEL SBDR	80
FIGURE 51 : ILLUSTRATION DU PROCESSUS SBDR	83
FIGURE 52 : MODULATION DU POIDS DE LA CORRECTION	84
FIGURE 53 : VISUALISATION DE LA CORRECTION DE LA DISPERSION INTRA-SET (AIRPARIF)	87
FIGURE 54 : PROFILS DE CORRECTION (AIRPARIF)	89
FIGURE 55 : RÉPARTITION DES ACTIVITÉS PAR PÉRIODE.	93
FIGURE 56 : TRAÇAGE SPATIAL DES PM <sub>10</sub> (µg/M <sup>3</sup> ) POUR LES DEUX PÉRIODES A ET B	93
FIGURE 57 : TRAÇAGE SPATIAL DES ANNOTATIONS D'ACTIVITÉ POUR LES DEUX PÉRIODES A ET B	94
FIGURE 58 : POINT D'INTÉRÊT GÉOGRAPHIQUE A	94
FIGURE 59 : POINT D'INTÉRÊT GÉOGRAPHIQUE B	95
FIGURE 60 : POINT D'INTÉRÊT B, GROS PLAN SUR « RÉSIDENCE »	95
FIGURE 61 : ILLUSTRATION DE L'ARCHITECTURE DU RÉSEAU LOCAL DE CAPTEURS	98

# TABLE DES TABLEAUX

TABLEAU 1 : COMPOSANTS DU CANARIN NANO	21
TABLEAU 2 : VÉLOCITÉ MÉDIANE EN FONCTION DE LA POSITION DES CAPTEURS	57
TABLEAU 3 : POURCENTAGE DE MESURES RÉALISÉES PAR PÉRIODE	59
TABLEAU 4 : ÉCART-TYPE MÉDIAN PAR PLAGE DE MESURE	60
TABLEAU 5 : VALEURS MOYENNES DE CONCENTRATION DE REQ <sub>M</sub> PAR PÉRIODE	61
TABLEAU 6 : ERREUR DE PRÉDICTION DE LA REQ <sub>M</sub> , PÉRIODE 3	62
TABLEAU 7 : COEFFICIENT DE LA DROITE DE RÉGRESSION	63
TABLEAU 8 : INFORMATIONS PAR JEU DE DONNÉES	77
TABLEAU 9 : PERFORMANCE DE L'ALGORITHME SBDR SANS TRANSPOSITION	87
TABLEAU 10 : INJECTION DE DONNÉES D'UN INSTRUMENT DE RÉFÉRENCE	88
TABLEAU 11 : RÉSULTAT DE LA TRANSPOSITION DES PROFILS DE DISPERSION	90
TABLEAU 12 : APERÇU DES DEUX PÉRIODES D'ACQUISITION, VALEURS DES POLLUANTS EN $\mu\text{G}/\text{M}^3$	92
TABLEAU 13 : NIVEAU MOYEN DE POLLUTION ATMOSPHÉRIQUE À PARTIR DE DONNÉES ANNOTÉES, $\mu\text{G}/\text{M}^3$	96



# ACRONYMES ET ABRÉVIATIONS

A: Ampères, 51  
ABS: Acrylonitrile butadiène styrène, 48  
AC: Alternating current, courant alternatif, 48  
ANR: Agence Nationale de la Recherche, 15  
ASA: Acrylonitrile styrène acrylate, 49  
AWS: Amazon web services, services web Amazon, 23  
BC: Carbone suie "black carbon", 3  
BLE: Bluetooth low energy, bluetooth basse consommation, 26  
BOM: Bill of materials, liste des composants, 22  
BPCO: Bronchopneumopathie chronique obstructive, 27  
CAO: Conception assistée par ordinateur, 48  
CFD: Computational fluid dynamics, Mécanique des fluides numérique, 48  
CHU: Centre hospitalier Universitaire, 44  
CO: Monoxyde de carbone, 4  
COV: Composés organiques volatils, 3  
CVF: Capacité vitale forcée, 32  
DC: Direct current, courant continu, 51  
ETM: Écart-type moyen, 82  
FDM: Fused deposition modeling, dépôt de fil fondu, 49  
GDPR: General Data Protection Regulation, règlement général sur la protection des données, 23  
GPS: Global positioning system, système de positionnement global, 19  
GSM: Global System for Mobile Communications, système global pour les communications mobiles, 21  
Hz: Hertz, 32  
IAH: Indice d'apnée-hypopnée, 45  
IoT: Internet (of things) des objets connectés, 10  
kWh: Kilowattheure, 53  
Li-Po: Lithium polymer, lithium polymère, 21  
M2M: Machine to machine, 21; Machine to machine, de machine à machine, 21  
mAh: Milliampere hour, milliampère heure, 21  
MB: Megabytes, mégabits, 53  
MEMS: Microsystèmes électromécaniques, 21  
O<sub>3</sub>: Ozone, 4  
OEM: Original equipment manufacturer, fabricant d'équipement d'origine, 21  
OMS: Organisation mondiale de la Santé, 1  
OPC: Optical particle counter, compteur optique de particules, 20  
OTA: Over-the-air, à distance, 22  
PLA: Acide polylactique, 49  
PM: Particulate matter, particules fines, 2  
PSG: Polysomnographie, 45  
PWM: Pulse-width modulation, modulation de largeur d'impulsion, 30  
RAM: Random access memory, mémoire vive, 23  
REQM: Racine de l'erreur quadratique moyenne, 60  
RGPD: Le règlement général sur la protection des données, 17  
SBDR: Réduction de la dispersion intra-set, Sensor Batch Dispersion Reduction, 70  
SIM: Subscriber identification module, module d'identification, 21  
SOx: Oxydes de soufre, 4  
SpO<sub>2</sub>: Saturation pulsée en oxygène, 32  
TEOM: Microbalance oscillante à élément conique, " tapered element oscillating microbalance, 6  
V: Volts, 20  
VEMS: Volume expiratoire maximum par seconde, 32  
VGP: Grand Parc de Versailles, 25

## 1. INTRODUCTION À LA QUALITÉ DE L'AIR

Le développement de la société moderne repose largement sur la consommation d'énergies fossiles, relâchant ainsi davantage de gaz nocifs et de particules dans l'atmosphère et impactant négativement à la fois le climat et la santé des êtres vivants dont les êtres humains [1]. La pollution de l'air est une cause majeure de morbidités et mortalités prématurées, met à risque l'écosystème [2] et est le plus important risque environnemental existant en Europe [3].

L'organisation mondiale de la santé (OMS) classe la pollution atmosphérique comme un cancérigène certain [4] et a estimé qu'elle cause approximativement 7 millions de morts prématurées chaque année. Une estimation plus récente de cet impact à partir de données individuelles indique un total de 9 millions de décès seulement pour la pollution atmosphérique de l'extérieur des locaux [5]. L'impact sur la santé de la pollution de l'air se traduit dans des effets à court (aggravation de maladies pulmonaires obstructives chroniques...) et à long terme (aggravation de maladies cardiopulmonaires, du risque d'accident vasculaire cérébral, de cancers du poumon, d'infections respiratoires, de maladies métaboliques, de maladies neurodégénératives...). La pollution atmosphérique impacte aussi le développement des enfants ainsi que leurs capacités cognitives et constitue un facteur d'aggravation des inégalités. Par ailleurs, ce sont les classes sociales les moins favorisées qui sont les plus exposées à la pollution atmosphérique. En se basant sur des données relatives à l'implantation spatiale des sites Britanniques de distribution de charbon entre 1851 et 1860, Brian Beach a pu observer comment les quartiers les plus exposés, en fonction du vent et de la topologie, ont subi les effets sanitaires les plus importants [6], avec une augmentation de 8 à 10% la mortalité infantile. Il montre que quand bien même la variation de la taille des individus est principalement d'origine génétique, les personnes ayant vécu dans des quartiers très pollués étaient 2.5 cm plus petits que le reste de la population.

L'utilisation massive du charbon lors de la révolution industrielle à la fin du 18<sup>ème</sup> siècle a eu un impact considérable sur la qualité de l'air et sur les populations vivant en milieux urbanisés. La transition au gaz naturel (remplaçant le gaz de ville) et à l'électricité, grâce à l'avènement de l'énergie nucléaire, ont permis de démocratiser ces énergies plus propres pour l'usage domestique. Les émissions de fumées contenant du dioxyde de soufre relâchées lors de la combustion du charbon ont ainsi pu être éliminées du paysage urbain.

Aujourd'hui, la pollution de l'air a reculé dans les pays industrialisés, mais est toutefois toujours importante dans les pays en développement et les communautés pauvres. La législation contraint dans une moindre mesure les industriels et les individus à contrôler et à limiter leurs émissions. Les populations chauffent leur logement et cuisinent avec de la biomasse, du kérosène ou du charbon, sans système de filtration ou de ventilation adéquat. Les effets de la pollution dans les pays en développement sont ainsi comparables aux villes industrialisées à la fin du 19<sup>ème</sup> siècle en pleine révolution industrielle. Le coût sanitaire est absorbé par la collectivité au profit du développement de l'activité humaine.

Plusieurs sociétés savantes nationales de médecine ont déclaré considérer le droit de respirer de l'air propre comme un droit humain, insistant sur le fait que les mesures prises par

les gouvernements pour lutter contre la pollution de l'air ne sont pas suffisantes [7]. En effet, on observe également des effets sanitaires lors d'expositions prolongées à la pollution atmosphérique même à des doses faibles et ainsi aussi dans les pays industrialisés. Pour ces raisons, l'OMS a tout récemment revu à la baisse ses seuils recommandés de qualité de l'air [8]. L'Agence nationale de santé publique a récemment publié les résultats de l'évaluation de la mortalité associée à la pollution de l'air sur la période 2016-2019 et conclut que « chaque année, près de 40 000 décès seraient attribuables à une exposition des personnes âgées de 30 ans et plus aux particules fines » [9]. De plus, les enjeux climatiques actuels [10,11] font de la réduction de la pollution de l'air un défi mondial.

## 1.1. CHIMIE DES POLLUANTS ET RISQUES ASSOCIÉS

### 1.1.1. SOURCES DES POLLUANTS ATMOSPHÉRIQUES ET EFFETS SANITAIRES

On associe à la pollution de l'air de nombreuses sources de polluants atmosphériques et de nombreux risques. Ces dernières années, plusieurs études ont permis de confirmer les risques de développement de l'asthme dû à l'exposition aux particules, dioxyde d'azote et ozone issus du trafic routier. En revanche, bien que quelques études aient été menées pour corrélérer la pollution d'origine routière au développement de la bronchopneumopathie chronique obstructive, davantage de preuves seraient nécessaires pour établir une causalité [12]. Il existe également des sources de pollution découvertes relativement récemment, comme les microplastiques [13]. D'autres sources de pollution sont donc certainement à l'heure actuelle encore méconnues. Pour ces raisons, davantage d'études doivent être réalisées pour découvrir de nouvelles maladies associées à la pollution de l'air, mieux les comprendre et mieux évaluer les risques, pour permettre aux décideurs de légiférer dans le sens du bien commun.

#### PM : LES PARTICULES FINES

Les particules, en anglais « Particulate Matter » (PM), désigne l'ensemble des particules solides et liquides (gouttelettes) en suspension dans l'air. Certaines sont visibles car suffisamment grosses ou sombres pour être perçues à l'œil nu ; par exemple la poussière, la buée, le sable, ou la fumée. Les particules sont désignées en fonction de leur taille maximum en micromètre ( $\mu\text{m}$ ). Ainsi «  $\text{PM}_{10}$  » désigne les particules d'une taille de 10  $\mu\text{m}$  ou inférieur et «  $\text{PM}_{2.5}$  » désigne les particules qu'on appelle « fines » d'une taille de 2.5  $\mu\text{m}$  ou inférieur. Les  $\text{PM}_{2.5}$  et  $\text{PM}_{10}$  (les particules fines dont la taille est inférieure à 1 micromètre) sont reconnues comme étant les plus dangereuses puisqu'elles sont suffisamment fines pour agir mécaniquement sur le corps humain, en pénétrant et obstruant les bronches, le système sanguin, ainsi que le cerveau entre autres au travers de la cavité nasale [14]. Elles sont ainsi la cause d'arythmies cardiaques, de crises cardiaques, de crises d'asthme et bronchites dans le cas d'expositions à court-terme et de maladies cardio-métaboliques, neurodégénératives, cancers... dans le cas d'expositions à long-terme. Les  $\text{PM}_{2.5}$  et les  $\text{PM}_{10}$  sont monitorées. Les seuils recommandés sont fournis en microgramme par mètre cube ( $\mu\text{g}/\text{m}^3$ ). La composition chimique des particules a été rarement considérée, mais elle est importante pour la compréhension des mécanismes toxicologiques des PM.

### NO<sub>x</sub> : LES OXYDES D'AZOTE

Les oxydes d'azote sont un groupe de gaz réactifs qui peuvent se transformer en acides lorsqu'ils sont exposés à des températures élevées ; sous un rayonnement solaire important, ou dans un moteur à combustion. Dans ce groupe de gaz, le plus répandu est le NO<sub>2</sub>, puisqu'il est principalement émis par les véhicules. Dans les villes où le trafic est très important, les risques pour la santé liées à l'exposition au NO<sub>2</sub> sont observés de façon plus importante que les risques liés au SO<sub>2</sub> [15] ; le SO<sub>2</sub> est plus dangereux mais désormais présent en plus faibles quantités. S'il est présent en grande quantité dans l'air, ou dans le cas d'expositions répétées, il peut irriter le système respiratoire et provoquer des symptômes tels que de la toux, une difficulté à respirer, une respiration sifflante et aggraver les maladies respiratoires. Des données récentes mettent le NO<sub>2</sub> en relation avec le développement de l'asthme [16]. Il réagit avec l'oxygène (O<sub>2</sub>) et l'eau de pluie (H<sub>2</sub>O) formant ainsi de l'acide nitrique (HNO<sub>3</sub>). Il est donc responsable de pluies acides. Le NO<sub>2</sub> fait partie des polluants actuellement monitorés. Les NO<sub>x</sub> peuvent également être indiqués en ppb (partie par milliard, de l'anglais « parts per billion »).

### COVs : LES COMPOSÉS ORGANIQUES VOLATILS

Les Composés Organiques Volatils (COVs), en l'anglais « Volatile Organic Compounds » (VOCs), sont une variété de substances [17] pouvant être d'origine naturelle ou humaine. Les plus néfastes sur la santé sont généralement créés par l'homme. À l'échelle mondiale, la majorité des COVs sont d'origine naturelle et sont indispensables au bon fonctionnement de l'écosystème. Certains permettant par exemple aux plantes de communiquer entre elles, ou à attirer les insectes pollinisateurs [18]. A l'extérieur des locaux, on trouve les COVs produit par la combustion des moteurs. On peut trouver facilement des COVs dans l'air intérieur des habitations puisque de nombreux produits dégagent des COVs lorsqu'ils sont utilisés ou entreposés. Les COVs se trouvent dans les solvants, peintures, colles, conservateurs et produits d'entretien. L'exposition aux COVs peut causer des nausées, fatigues, réactions allergiques, maux de tête et pertes de mémoire. Certains procédés industriels produisent des COVs [19] ; les ouvriers peuvent ainsi être très exposés [20–23]. Tous les composés organiques volatils ne se valent pas en termes de risque ; certains sont connus comme dangereux (le perchloroéthylène [24] et le benzène [25]) et d'autres n'ont pas d'effet connu sur la santé.

### PUF : LES PARTICULES ULTRAFINES

Les particules ultrafines (PUF) ou UFPs de l'anglais « Ultrafine Particles » sont des particules de taille nanométrique (PM<sub>0.1</sub>). Les PUF ont un effet grave sur la santé puisqu'elles sont en mesure de pénétrer dans le système sanguin [26], jusque dans les cellules. Cependant, les particules ultrafines ne sont pas à ce jour réglementées.

### BC : LE NOIR DE CARBONE

Le noir de carbone, « Black Carbon » en anglais (BC), fait partie de la famille des particules et sa taille varie entre les particules fines (PM<sub>2.5</sub>) et les particules ultra fines (10 nanomètres). Il est principalement issu de combustions incomplètes d'énergies fossiles ou de biomasse. Le noir de carbone, comme les particules fines, est nocif pour la santé et est également associé à la présence de polluants, par exemple les hydrocarbures aromatiques polycycliques,

dont la toxicité est reconnue [27]. Le noir de carbone a également un impact significatif sur le changement climatique [28]. Il absorbe la lumière, impactant directement la flore et participant au réchauffement de l'air, ainsi qu'à la fonte des glaciers dont il recouvre en partie la surface.

#### CO : LE MONOXYDE DE CARBONE

Le monoxyde de carbone (CO) est un gaz extrêmement dangereux pour l'homme. Auparavant le gaz de ville contenait du CO ; cela a contribué à motiver son abandon. Le CO est inodore et provoque des vertiges ainsi que des pertes de conscience. Il peut entraîner la mort lorsqu'il est en grande concentration dans l'air. En effet, les molécules de CO s'associent beaucoup plus facilement à l'hémoglobine qu'à l'oxygène, causant donc un manque d'oxygénation du sang et donc des organes vitaux.

De forts niveaux de CO ne sont que rarement rencontrés en air extérieur, mais peuvent malgré tout être un danger pour les personnes avec des maladies cardiaques et dont le sang est naturellement faiblement saturé en oxygène. En air extérieur, le CO est principalement issu des moteurs à combustion. Il peut également être produit par la présence d'une fuite dans une cheminée, ou bien dans le cas de combustions incomplètes se produisant dans les systèmes de chauffage ou de cuisson au gaz mal entretenus. Il est systématiquement produit lors de la combustion de bois et de charbon.

#### SO<sub>2</sub> : LE DIOXYDE DE SOUFRE

Le dioxyde de soufre est un composé dangereux pour la santé faisant partie de la famille des oxydes de soufre (SO<sub>x</sub>). Il est le sous-produit de la consommation du charbon ou du fioul, qui relâchent le soufre qu'ils contiennent lors de leur combustion. Il est incolore, mais possède une odeur distincte et réagit avec les particules en suspension dans l'air. Le SO<sub>2</sub> est également nocif pour l'écosystème puisqu'il agit négativement sur la croissance des plantes et contribue à la formation de pluies d'acide sulfurique (H<sub>2</sub>SO<sub>4</sub>). Il est un irritant très soluble dans l'eau et est ainsi en partie absorbé par les muqueuses lorsque inhalé. Il peut également se fixer sur les particules fines et ainsi pénétrer profondément dans les poumons. Lors de cette réaction, l'air s'opacifie et la visibilité est ainsi réduite. Dans les pays en développement, le SO<sub>2</sub> reste un problème majeur de la qualité de l'air [29,30].

#### O<sub>3</sub> : L'OZONE

L'ozone (O<sub>3</sub>) présent dans l'atmosphère à 25 km d'altitude nous protège des rayonnements ultraviolets du soleil, mais est nocif lorsqu'il est présent dans notre environnement proche. L'ozone est issu de la transformation du NO<sub>2</sub> et VOCs sous l'action du rayonnement solaire. Il peut ainsi être présent dans l'air extérieur, particulièrement les jours ensoleillés où il peut atteindre des niveaux nocifs pour la santé. L'ozone peut fragiliser les poumons, causant l'essoufflement. Les personnes travaillant en extérieur, les enfants et les personnes asthmatiques sont les plus à risques [31].

#### 1.1.2. PERCEPTION DU RISQUE PAR LA POPULATION

Au sein de la communauté scientifique internationale, l'intérêt pour la pollution de l'air est croissant. Cet intérêt se traduit par une augmentation croissante du nombre de publications

scientifiques depuis plusieurs années [32]. En France, le baromètre de l'institut de radioprotection et de sûreté nucléaire [33] rapporte que la pollution de l'air était perçue par la population comme un sujet environnemental préoccupant. Cette préoccupation a reculé en 2020, avec seulement 8% des votants contre 22% en 2016 ; cette régression s'est faite au profit du dérèglement climatique (39 %). Elle reste malgré tout perçue comme un sujet important, puisque 67% des sondés indiquent que la pollution de l'air représente un niveau de risque élevé. À l'international on observe que la perception des risques liés à la pollution de l'air diffère en fonction des études [34], des populations ciblées [35,36] et de l'état de santé des individus ou de leur entourage [37]. Généralement, la perception de la pollution de l'air par la population ne coïncide pas avec l'exposition réelle des individus.

Depuis plusieurs dizaines d'années en Asie, les autorités de nombreux pays recommandent le port de masques pour limiter le risque de transmission des maladies respiratoires virales. Les autorités de certains pays recommandent désormais le port de masques dédiés à la pollution de l'air [36]. Dans les métropoles, les salles de sport et écoles aisées sont équipées de systèmes de filtration de l'air. Au niveau individuel, les habitants des zones les plus urbanisées investissent dans des systèmes de filtration permettant de garantir une qualité d'air convenable à l'intérieur des habitations durant les événements de pollution importants.

En Asie, les populations sont donc davantage conscientes des risques liés à la pollution de l'air, d'une part du fait des niveaux ambiants plus importants, mais également du fait de l'intégration de ces solutions techniques palliatives dans leur quotidien. Ces dispositifs sont moins répandus en Europe, où les niveaux de pollution de l'air sont plus faibles. Ils n'effacent pas le coût humain, financier et écologique lié à la pollution de l'air.

## 1.2. RÉGLEMENTATIONS DE LA POLLUTION DE L'AIR

Aujourd'hui, au sein de l'Union européenne, les législations relatives à la qualité de l'air sont issues des décisions et objectifs déterminés au niveau européen. Les normes européennes d'émission dans le secteur automobile [38] sont connues du grand public puisqu'elles impactent directement les décisions des consommateurs, au travers de taxes ou d'incitations financières [39]. Ces législations européennes ne portent cependant pas uniquement sur les émissions des véhicules routiers, mais visent également le transport maritime, le secteur de l'industrie, de l'agriculture et de l'énergie. Notons que l'OMS a récemment publié des seuils recommandés plus exigeants [8].

La commission européenne joue donc un rôle essentiel dans l'amélioration de la qualité de l'air sur l'ensemble du territoire européen [40,41]. Elle légifère et s'assure que les seuils qu'elle fixe soient respectés, sans quoi les états sont sanctionnés. Elle impose également aux États membres de rendre disponibles au public et aux organisations concernées, à intervalles réguliers, les informations et données recueillies relatives aux concentrations de polluants. Dans le cas d'événements où des seuils préalablement fixés seraient atteints, l'État est tenu d'informer la population et des mesures doivent être prises. L'objectif principal visé est ambitieux ; en 2050 l'Europe doit devenir le premier continent neutre en carbone.

La France a été condamnée par la Cour de Justice de l'Union européenne [42] car de nombreux manquements aux normes européennes ont été enregistrés dans une dizaine d'agglomérations. Par la suite, le Conseil d'État, saisi par plusieurs associations, a ordonné à l'État de mettre en place les mesures nécessaires. Similairement, de nombreux pays membres ont été condamnés ou menacés de sanctions.

### 1.3. OUTILS ET MÉTHODES DE MESURE DE LA QUALITÉ DE L'AIR

#### 1.3.1. *STANDARDS ET INSTRUMENTS DE RÉFÉRENCE*

Les outils et protocoles pour réaliser la mesure des différents polluants relèvent de la métrologie et sont précisément définis par des autorités compétentes ; l'Organisation internationale de normalisation (ISO), le Comité européen de normalisation (CEN), ou encore au niveau national, l'Association française de normalisation (AFNOR). La norme internationale ISO 16000-37:2019 [43] spécifie par exemple la méthode de mesure pour déterminer les concentrations de PM en air intérieur. Il existe également un équivalent européen de cette norme [44].

Dans le cas des particules aéroportées (PM), il est possible de mesurer et de recueillir les résultats en continu, en évaluant leur concentration massique dans l'air, par l'utilisation d'une microbalance oscillante à élément conique, en anglais « tapered element oscillating microbalance » (TEOM). Cet instrument de mesure possède un tube dont l'oscillation est contrôlée et surveillée électroniquement. Au travers de ce tube, l'air ambiant est aspiré par une pompe. Un filtre situé à l'extrémité du tube permet d'accumuler les particules présentes dans l'air. La fréquence d'oscillation du tube réduit en fonction de la masse des particules capturées. Cet instrument est difficilement transportable, fragile et coûteux, mais est considéré comme équivalent à la méthode de référence lorsque les conditions météorologiques sont favorables [45]. L'utilisation d'un TEOM est donc relativement commune lorsque l'on désire évaluer avec précision la concentration massique des particules dans l'air.

Parmi les outils fiables se rapprochant des méthodes de référence, il existe également des échantillonneurs passifs. Ce sont des pièges à polluants, dont l'aspect diffère en fonction du polluant ciblé. Comme leur nom l'indique, ils fonctionnent sans aucune source d'alimentation. L'échantillonneur est disposé à un endroit fixe, mais peut aussi être porté en mobilité. Il captera le polluant durant une période définie, généralement grâce à un procédé chimique, puis requerra une analyse en laboratoire pour déterminer précisément les niveaux auxquels l'échantillonneur a été exposé. Pour chaque polluant et en fonction du scénario, il existe donc un ou plusieurs protocoles et des outils permettant son évaluation.

### 1.3.2. STATIONS DE MESURE

Sur l'ensemble du territoire, les polluants réglementés sont surveillés par les associations agréées. Au niveau national, l'association ATMO fédère les associations de surveillance de la qualité de l'air, qui gère régionalement les réseaux de stations de mesure de la pollution atmosphérique. Ces associations ont à charge de réaliser de manière impartiale, des études sur la qualité de l'air, de prévoir, d'alerter, de publier des rapports [46,47] ainsi que les données recueillies. Historiquement, la qualité de l'air a été mesurée grâce à des stations fixes mesurant les polluants présents dans l'air, fournissant des données de haute qualité, utilisant des outils de mesure performants et coûteux, sur la coordonnée géographique où la station est implantée. Il existe également des stations de mesure remorquables.



FIGURE 1 : Station de mesure, boulevard périphérique de Paris

Ces stations de mesures peuvent également mesurer des polluants qui ne sont pas encore réglementés. Elles alimentent des algorithmes de modélisation qui permettent ainsi de cartographier la région environnante. Ces réseaux de stations de mesure permettent d'évaluer la pollution dite « de fond » ; loin des sources de pollution, mais aussi la pollution dite « de proximité » comme dans le cas de la Figure 1 ; grâce à des stations situées proches des sources d'émissions. Elles sont donc placées stratégiquement, dans le but d'obtenir la meilleure appréciation possible de l'exposition réelle des populations.

### 1.3.3. MODÉLISATION ATMOSPHÉRIQUE

Divers modèles mathématiques, basés sur la mécanique des fluides, sont en mesure de représenter le déplacement, la dispersion et la déposition des flux de polluants, dans l'espace et le temps, à l'échelle planétaire. Ces modèles exploitent les données topographiques, les interactions physiques des polluants avec des éléments naturels, les données connues sur les sources d'émission ainsi que les données météorologiques.

Ces outils puissants sont en mesure de fournir des informations sur les niveaux de polluants aux coordonnées géographiques souhaitées. Les applications sont donc

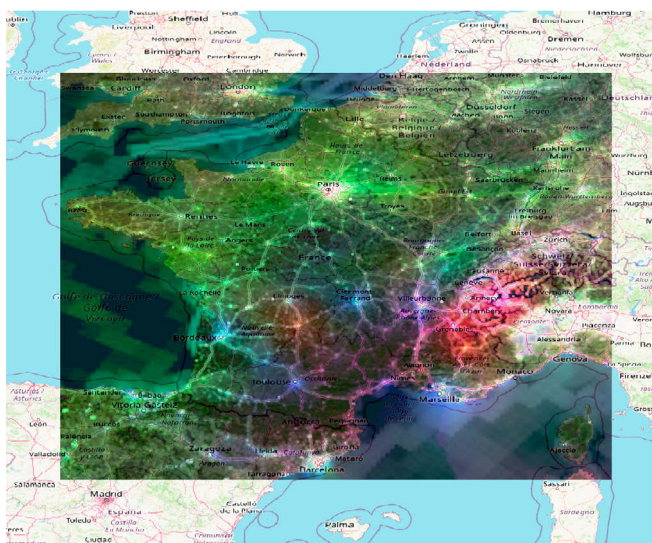


FIGURE 2 : Visualisation du modèle CHIMERE



multiples ; dans le cas d'études épidémiologiques [48], pour étudier les risques dans le cas de catastrophes industrielles, pour connaître les effets des actions gouvernementales sur les niveaux d'exposition des populations (en supprimant ou ajoutant fictivement des sources d'émissions de polluants), ou encore pour optimiser l'aménagement du territoire. En exploitant les données issues de modèles météorologiques de nouvelle génération, certains modèles permettent de prédire des événements de pollution et donc de prévenir les populations. L'obtention de données issues de ces modèles, en plus de requérir une certaine expertise, nécessite une puissance de calcul importante. Grâce aux avancées technologiques dans le domaine des processeurs, le coût de la puissance de calcul nécessaire tend à baisser.

Le modèle de chimie-transport CHIMERE est représenté sur la Figure 2, au travers d'une visualisation des valeurs de  $\text{NO}_2$  en France métropolitaine. C'est le modèle le plus utilisé par les associations françaises de surveillance de la qualité de l'air.

#### 1.3.4. TÉLÉDÉTECTION

La télédétection est un outil très performant pour évaluer la pollution de l'air là où les autres moyens de mesure sont indisponibles ou difficiles à déployer. Au niveau Européen, l'ESA (European Space Agency) produit et rend disponible en libre accès des données spatiales relatives à la qualité de l'air, via le programme Copernicus et au travers des missions Sentinel. Elles ont permis de déployer plusieurs satellites équipés d'appareils permettant la surveillance de certains polluants. En 2021, l'ESA prévoit de lancer Sentinel-4 et Sentinel-5, qui sont spécifiquement dédiés à la surveillance de la qualité de l'air sur le territoire européen. Ces nouveaux satellites seront en mesure de produire des données d'une résolution spatiale supérieure à l'existant. La Figure 3 présente l'antenne de réception satellite de la Maison de la Télédétection, à Montpellier.

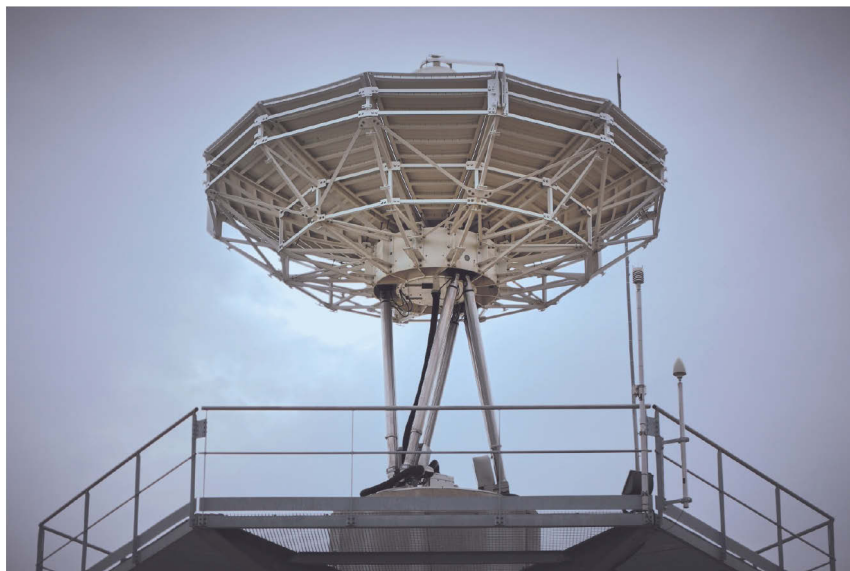


FIGURE 3 : Antenne de la Maison de la Télédétection

Il existe également un nombre de satellites de surveillance de la qualité de l'air couvrant eux le territoire américain [49] ; la mission TEMPO prévue pour 2022 [50] permettra de réaliser l'acquisition de plusieurs polluants, chaque heure, à un niveau de détail élevé. Dans la région

Pacifique-Asie, la Corée du sud a déployé en 2020 un satellite semblable ; le GEMS (spectromètre géostationnaire de surveillance de l'environnement monitoring) [51]. Ainsi, le GEMS, avec TEMPO et Sentinel-4 et 5, participeront à constituer une constellation de satellites de dernière génération permettant de surveiller la qualité de l'air à l'échelle mondiale.

Dans les années à venir, au travers du développement des outils de mesures et des satellites embarquant ces outils et considérant également l'avènement des satellites de petite taille, le grand public et la communauté scientifique va profiter d'une amélioration rapide des données disponibles et de leur qualité ; par l'augmentation de la résolution spatiale et temporelle, par l'amélioration de la pertinence des algorithmes permettant de traiter ces données, ainsi que par l'accessibilité de celles-ci.

#### 1.3.5. MICRO-CAPTEURS ET MODULES DE MESURE

Dans le langage scientifique et courant, le terme « capteur » peut désigner l'appareil de mesure tout comme les composants embarqués réalisant cette mesure. « capteur » décrit donc à la fois, la technologie qui détecte et communique les changements physiques à d'autres composants électroniques, mais aussi l'ensemble du dispositif intégrant cette technologie et le matériel nécessaire pour pouvoir être utilisé comme un appareil autonome. Cette absence de nuance existe en Français comme en Anglais. Nous désignerons donc les composants réalisant la mesure de la valeur physique comme étant des « micro-capteurs ».

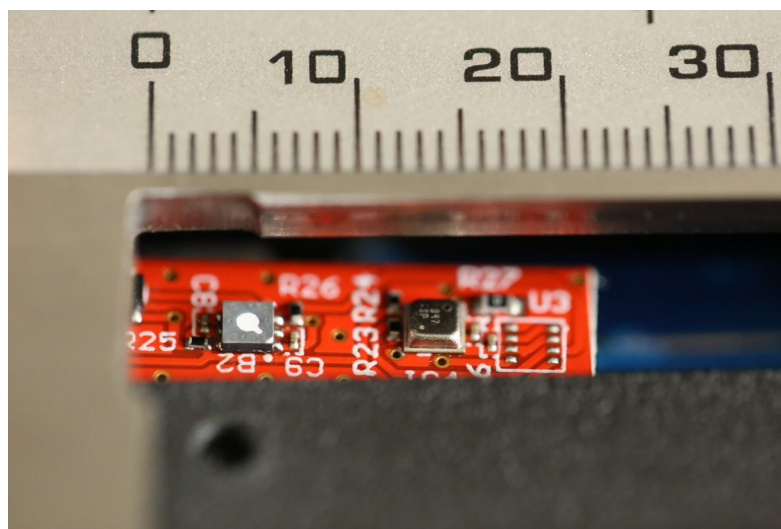


FIGURE 4 : Photomicrographie de deux micro-capteurs de la qualité de l'air

De nombreux micro-capteurs à faible coût sont aujourd'hui disponibles dans le commerce. Les fabricants proposent de nouvelles itérations dont les performances et capacités sont continuellement perfectionnées, à mesure de l'évolution des technologies permettant leur production. Ils sont ensuite intégrés sur une carte électronique équipée de composants qui permettent de communiquer les mesures acquises aux appareils environnants, à un serveur distant, ou directement à l'utilisateur.

En fonction de la technologie employée, le volume des micro-capteurs peut varier de quelques millimètres cubes, à plusieurs centimètres cubes (Figure 4). Ils sont disponibles pour un nombre grandissant de polluants, sont de plus en plus abordables et performants. Grâce aux récentes avancées technologiques, de nombreux micro-capteurs peuvent désormais être considérés comme des modules indépendants. Ces modules interfacent directement avec le microcontrôleur équipant le produit fini ; dans notre cas des capteurs portatifs. Ce nouveau type de micro-capteurs disposent de leurs propres circuits intégrés, circuits de régulation, capacités de calcul, mémoires embarquées, sont parfois programmables et sont en mesure de communiquer avec d'autres composants électroniques tout en requérant un nombre réduit de composants pour fonctionner. Pour ce faire, ces modules communiquent les mesures réalisées par des moyens d'échange de données désormais intégrés dans les microcontrôleurs tels que le protocole I<sup>2</sup>C (Inter-Integrated Circuit), SPI (Serial Peripheral Interface), ou UART (Universal Asynchronous Receiver-Transmitter). Ces capteurs ne peuvent être utilisés seuls par l'utilisateur final, mais sont désignés par les fabricants comme des « capteurs ». C'est pourquoi, dans le but de clarifier, nous désignerons cette catégorie de micro-capteurs par le terme de « module de mesure ».

#### 1.3.6. CAPTEURS PORTATIFS

Un capteur portatif moderne est ainsi composé d'un boîtier, de composants électroniques passifs, d'une batterie, d'un ou plusieurs micro-capteurs ou modules de mesure, de circuits intégrés et d'un microcontrôleur exécutant un logiciel embarqué. Dans le cas d'un capteur connecté, l'écosystème IoT (l'internet des objets, de l'anglais « Internet of Things ») intègre également un module de radio communication et un serveur distant hébergeant des services web dans le but de stocker, organiser et redistribuer les données acquises par les capteurs. La faible consommation électrique et la taille réduite de ces micro-capteurs et modules de mesure permettent aux produits finaux, les capteurs portatifs, de produire des niveaux de qualité et quantité de données jamais obtenues auparavant.

Les capteurs portatifs sont généralement dédiés à un unique polluant, les plus répandus étant les capteurs de PM, de NO<sub>x</sub>, de BC et récemment de COVs. Aujourd'hui, ces appareils ne sont ni aussi précis, ni aussi fiables que les instruments de référence, mais sont moins chers, ne nécessitent aucun entretien, peuvent être utilisés par le grand public, permettent de mesurer l'exposition individuelle lorsqu'ils sont embarqués, ne requièrent aucune connaissance technique pour être exploités, ils bénéficient des dernières avancées technologiques notamment dans le domaine des télécommunications, leur permettant par exemple de transmettre les données acquises en temps réel. Les appareils commerciaux abordables existants, ainsi que les modules de mesure des fabricants d'équipements d'origine (OEM) disponibles à la vente, ont majoritairement été évalués par la communauté scientifique [52–62]. De multiples méthodologies et outils ont été développés pour leur validation [63–66]. L'utilisation de capteurs de la qualité de l'air a été envisagée pour des applications de crowdsensing (de détection de foule) [67–73]. On observe cependant un nombre réduit de projets de recherche produisant de nouveaux capteurs portatifs [74–79] et étonnement, la quantité de capteurs portatifs commerciaux destinés au grand public, en dehors des applications industrielles, est encore plus limitée [80,81], de plus, l'accès aux données recueillies en temps réel n'est pas toujours possible.

Ces capteurs portatifs constituent donc de nouveaux outils permettant aux scientifiques de mieux comprendre l'exposition des individus à la pollution atmosphérique et les risques associés et plus particulièrement, représentent une source de données précieuses pour les épidémiologistes des maladies respiratoires. Le paradigme de la surveillance de la qualité de l'air a rapidement changé ; des réseaux de surveillance de la pollution de l'air à faible coût peuvent désormais fournir une résolution spatio-temporelle plus élevée que les stations de référence [82–88]. Les performances de ces appareils à bas coût augmentent rapidement. Ils pourraient à terme être suffisants pour être déployés à grande échelle et entièrement remplacer les outils actuellement utilisés dans les stations de mesure fixe.

### 1.3.7. EXPOSITION INDIVIDUELLE

Dans un environnement urbanisé, les individus sont exposés à un ensemble de différentes sources de pollution de l'air, dont les sources extérieures sont issues par exemple de l'industrie ou du trafic routier. Cependant, dans notre société moderne, les individus passent la plupart de leur vie en milieu intérieur, que ce soit à leur domicile, dans des locaux commerciaux et de divertissement, dans les transports, ou au travail. L'exposition des individus aux sources de pollution en milieu intérieur sont difficilement prédictibles. La ventilation, le type de nourriture consommée, les méthodes de cuisson, l'utilisation de produits chimiques pour le nettoyage de l'habitation, auront tous un impact considérable sur la qualité de l'air à laquelle les individus sont exposés. Par conséquent, chaque individu aura au cours de sa vie, une exposition unique à la pollution de l'air, à mesure qu'il se déplace dans différents environnements, au travers de ses habitudes, ses conditions de vie et de travail. Il est donc important, lorsque l'on étudie les effets de la pollution de l'air sur la santé et les mécanismes qui y sont associés, de prioriser les données d'exposition relevées au niveau individuel, plutôt que d'exploiter des données de concentrations extérieures ambiantes, puisqu'elles permettent de mieux décrire et donc de mieux comprendre l'impact de la pollution de l'air sur la santé humaine [88]. Cette compréhension du risque associé à l'exposition individuelle est d'autant plus nécessaire pour les individus susceptibles car atteints par des pathologies graves, telles que les maladies cardiorespiratoires. Leur qualité et leur espérance de vie pourraient ainsi être améliorées, dans la mesure du possible, en fournissant des conseils pour moduler leurs habitudes, ou adapter leur lieu de vie. Dans un futur proche, l'utilisation temporaire d'un capteur portatif pourra leur permettre de diagnostiquer si leurs conditions de vie, de travail, moyens de transports, sont un risque potentiel pour leur santé.

Précédemment, il n'était pas possible d'évaluer l'exposition individuelle à la pollution de l'air d'un groupe d'individus, en mobilité, de manière continue et sur une longue durée. Les capteurs portatifs étaient trop coûteux, trop encombrants, limités en termes d'autonomie, de capacité mémoire et de résolution temporelle, pour être utilisés comme outils d'évaluation de l'exposition individuelle. Tel que présenté ci-dessus, un certain nombre de capteurs utilisant des micro-capteurs de la qualité de l'air ont émergé. Ces capteurs peuvent être utilisés en mobilité dans un grand nombre de scénarios. Aujourd'hui, ces dispositifs sont donc le seul moyen d'évaluer l'exposition à la pollution de l'air au niveau individuel, sur de longues périodes, avec une haute résolution temporelle, en air intérieur comme extérieur.

## 1.4. APPROCHE ÉPIDÉMIOLOGIQUE ET PERSPECTIVES IOT

### 1.4.1. ÉTUDES ÉPIDÉMIOLOGIQUES

L'information est au centre de toute étude épidémiologique. Les méthodes de recueil de l'information diffèrent en fonction de la nature des études épidémiologiques et des objectifs visés. Au sein des études analytiques observationnelles, sur l'impact sanitaire de la pollution atmosphérique, on applique essentiellement les études écologiques, l'unité d'analyse est la population et non pas l'individu, pour lequel les données ne sont pas disponibles. L'objectif de ces études est de trouver une association entre une maladie et un polluant atmosphérique. Aucune information ne doit permettre d'identifier les individus issus de la population étudiée. Les études écologiques peuvent être facilement réalisées lorsque les données requises sont déjà disponibles, ou peuvent être extraites de données existantes.

En revanche, dans le cas d'une étude individuelle, on dispose généralement de données anonymisées par individu, que ce soit pour une étude transversale, une cohorte ou des cas-témoins. Si une partie des sujets d'une étude individuelle ne sont pas exposés, ils (les témoins) permettent alors de contrôler que les effets indésirables observés (sur les cas) sont bien attribuables à l'exposition au polluant ciblé. Pour les études écologiques, les données relatives à l'état de santé sont donc généralement issues de bases de données nationales telles que recueillies par le Système National des Données de Santé (SNDS) [89], tandis que dans le cas des études individuelles, elles résultent d'exams réalisés en hôpital, souvent accompagnés de questionnaires ou lors d'enquête épidémiologique ad hoc, nécessitant un examen clinique ou un test clinique. Concernant les données d'exposition, elles étaient auparavant extraites de modèles, ou de la télédétection. Il est dorénavant désirable, dans le cas d'études individuelles, d'utiliser des capteurs portatifs, puisqu'ils permettent d'obtenir des données d'exposition plus représentatives des expositions réelles.

Au sein des études observationnelles descriptives, il existe les études transversales et longitudinales. Une étude transversale, s'effectue à un moment donné (sur une durée limitée), dans une population dont on va établir des statistiques d'état de santé, la fréquence et la distribution des maladies en relation avec l'exposition à la pollution. Une étude longitudinale, en revanche, réalise le suivi dans la durée (possiblement sur plusieurs années) d'une population identifiée dont on sait qu'elle est exposée à des niveaux importants d'un polluant et permet de déterminer les effets à long terme.

### 1.4.2. MULTIPLICATION DES SOURCES DE DONNÉES

Nous introduisons dans les paragraphes 1.3.5 et 1.3.6 le rapide développement des capteurs portatifs et micro-capteurs. La miniaturisation et prolifération des micro-capteurs ne se limitent pas uniquement aux capteurs de la qualité de l'air. Les objets connectés les plus répandus sont actuellement les smartphones, les montres et plus rarement, les balances. Cependant, dans un avenir proche, l'ensemble de l'équipement électro-ménager équipant les habitations sera connecté, en plus du développement de la domotique ; le four, le frigo, la machine à laver, le chauffage, l'éclairage, les volets, les portes et fenêtres. Tous ces appareils embarqueront des micro-capteurs qui seront en mesure de fournir des informations possiblement exploitables d'un

point de vue épidémiologique, renseignant sur l'état de santé, la qualité de vie ainsi que les habitudes des individus. Ces données pourraient participer à compléter ou à simplifier la collecte de données nécessaires aux études épidémiologiques [90], qui requièrent à l'heure actuelle à minima un questionnaire et donc une contrainte pour le participant qui consent à partager anonymement ses données. À l'avenir, les études épidémiologiques pourraient ainsi avoir accès à des quantités importantes de données qui à l'heure actuelle restent inaccessibles dans le monde de la recherche médicale.

Dans un futur où la législation et la technologie le permettent, des bases de données portant sur l'exposition des populations à la pollution de l'air pourraient être constituées de manière autonome. Par exemple, en exploitant les données issues des capteurs embarqués de série sur certains véhicules. En effet, les constructeurs automobiles utilisent désormais des capteurs de la qualité de l'air pour moduler le fonctionnement du système de filtration, dans le but d'améliorer la qualité de l'air dans l'habitacle des véhicules. Ces données pourraient permettre d'obtenir des informations complémentaires sur la qualité de l'air à l'extérieur des habitations des individus, durant leurs trajets quotidiens, ainsi qu'à proximité de leur lieu de travail. Elles pourraient être exploitées dans le cadre d'une étude épidémiologique écologique. Les individus seraient sollicités à participer à cette étude directement depuis leurs smartphones, puisqu'étant en permanence connectés avec le véhicule. Avec leur consentement, les données d'exposition et de localisation pourraient être transmises, accompagnées des données produites par la multitude de capteurs dont les domiciles seront équipés.

#### 1.4.3. *LE QUANTIFIED SELF*

La multiplication des montres intelligentes (en anglais « smartwatches »), des appareils de suivi de la condition physique et de l'activité du sommeil, ont créé un nouveau phénomène culturel appelé "la mesure de soi" (en anglais « the quantified self »), dans lequel les membres de la population générale portent des appareils de suivi qui enregistrent en permanence leurs données, en échange d'améliorations potentielles de leur qualité de vie ou de leurs performances physiques.

Nous avons précédemment mentionné dans le paragraphe « 1.1.2 Perception du risque par la population » que la qualité de l'air est rapidement devenue une préoccupation majeure des populations vivant en zones urbanisées ou industrialisées. Malgré les législations en vigueur, ces populations peuvent être exposées à des niveaux supérieurs aux seuils recommandés. En réponse à cela, on observe ces dernières années une augmentation considérable de l'adoption par le grand public de capteurs de pollution de l'air, ou de produits en étant équipés, tels que des purificateurs d'air individuels ou portatifs [91]. À terme, ces capteurs seront démocratisés, car intégrés dans de nombreux appareils ménagers. Cette fonctionnalité de surveillance en temps réel des niveaux de pollution de l'air peut ainsi alerter les individus en cas de pics soudains ou d'augmentation progressive des polluants atmosphériques, ce qui leur permettra de modifier leur comportement (dans le cas de la pollution de l'air intérieur) ou d'éviter la source de pollution (dans le cas de la pollution de l'air extérieur). La promesse d'une amélioration de la qualité de vie et le désir du consommateur traduit par l'adoption généralisée de ces produits, met en évidence que cela relève du phénomène de quantified self.

Les projets de recherche sur l'exposition individuelle à la pollution de l'air n'exploitent généralement pas ce phénomène. Les participants équipés de capteurs de la qualité de l'air ne sont pas immédiatement informés des niveaux de polluants auxquels ils sont exposés. Ils ne reçoivent pas d'alertes en temps réel ni de compte rendu journalier. Le quantified self pourrait être exploité pour augmenter l'intérêt des participants à l'utilisation des capteurs et permettre ainsi de produire des données en plus grande quantité et de meilleure qualité. L'absence de son exploitation, spécifiquement dans le cas de la recherche sur l'exposition individuelle à la pollution de l'air, peut s'expliquer d'une part pour des raisons techniques ; les capteurs existants pouvant être utilisés pour la recherche médicale ne disposent actuellement pas de telles fonctionnalités et la mise en œuvre d'un service externe à l'écosystème des capteurs requerrait une expertise et des efforts de développement supplémentaires. D'autre part, pour des raisons scientifiques ; les participants pourraient changer leurs habitudes une fois informés des comportements les exposant à des niveaux élevés et donc en fonction de la question soulevée par l'étude, biaiser les résultats.

## 2. OBJECTIFS ET JUSTIFICATION DE LA THÈSE

### 2.1. OBJECTIFS

Dans le contexte présenté plus haut, cette thèse se propose de :

- Estimer l'exposition individuelle à la pollution atmosphérique mesurée en temps réel avec le capteur portatif CANARIN en tenant compte des variations spatio-temporelles
- Mettre en relation les mesures d'exposition à la pollution atmosphérique et à l'état de santé individuelles en temps réel dans le cas de la fonction respiratoire et de l'apnée du sommeil.
- Quantifier la dispersion des mesures des concentrations fournies par le CANARIN puis offrir une solution technique et algorithmique permettant sa réduction.
- Identifier les limitations des capteurs portatifs actuels et proposer une architecture de capteurs innovante.

### 2.2. PLAN DE THÈSE

À cet effet, la thèse a été organisée comme suit :

- La partie 3, « Populations et méthodes », introduira les différentes populations ainsi que les méthodes employées pour recueillir des données relatives à l'état de santé des individus, ainsi qu'à leur exposition à la pollution de l'air, tel que réalisé dans le cadre des projets POLLUSCOPE et POLLAR.
- La partie 4, « Projet CANARIN », retrace comment une collaboration internationale a donné naissance à plusieurs générations de capteurs portatifs permettant de mesurer la qualité de l'air en temps réel. Ces capteurs constituent un nouveau type d'outil pour la recherche épidémiologique, puisqu'ils permettent la mesure de l'exposition individuelle aux particules fines. Nous indiquerons les spécificités techniques des modèles que nous avons participé à développer.
- Dans la partie 5, « Projet POLLUSCOPE : observatoire participatif pour la surveillance de l'exposition individuelle à la pollution de l'air en lien avec la santé », nous présenterons ce projet de recherche financé par l'Agence Nationale de la Recherche (ANR), les outils employés et les spécificités liées au recrutement de volontaires. Nous expliquerons comment nous avons été en mesure de mettre en relation l'évolution de la fonction respiratoire des participants, avec leur niveau individuel d'exposition aux particules fines, dans un scénario où la durée d'acquisition et le nombre de sujets sont restreints.
- La partie 6, « Projet POLLAR : étude de la relation entre l'exposition à la pollution atmosphérique et l'apnée du sommeil » exposera comment des capteurs CANARIN ont permis d'explorer la relation entre l'exposition individuelle aux particules et l'apnée du sommeil, au travers d'un partenariat avec le Centre Hospitalier Universitaire de Grenoble.
- Dans la partie 7, « Banc d'essai mobile pour l'évaluation de la dispersion », nous énoncerons comment nous avons imaginé, simulé numériquement, puis construit et utilisé, un banc d'essai ayant pour but de permettre d'évaluer la dispersion présente dans un set de capteurs portatifs de la qualité de l'air.



- La partie 8, « Processus de réduction de la dispersion intra-set », proposera un algorithme, qui lorsqu'utilisé avec le banc d'essai présenté partie 7, permettra de qualifier et minimiser la dispersion.
- La partie 9 « Contextualisation des données », consistera à exposer les méthodes actuelles de contextualisation des données individuelles, dans le cadre du suivi détaillé d'un volontaire du projet POLLUSCOPE. Nous indiquerons également les limitations actuelles, notamment celles liées à la qualité de l'acquisition des données de géolocalisation.
- Dans la partie 10, « Évolution des architectures de capteurs », en tenant compte de la technologie disponible, nous discuterons d'un concept d'architecture de capteurs innovants, permettant de mieux répondre aux besoins actuels de la recherche épidémiologique.
- Enfin, la partie finale du manuscrit de thèse est dédiée à la discussion et aux perspectives.

### 3. POPULATIONS ET MÉTHODES

#### 3.1. POPULATIONS CONSIDÉRÉES

Dans cette thèse, trois types de populations sont considérées. Premièrement, des sujets issus de la population générale. Ils sont généralement sains, sans pathologies. Ils ont participé de manière bénévole et volontaire. Ils ont donc été particulièrement intéressés, motivés et ont participé activement à l'utilisation des divers outils nécessaires au bon déroulement des diverses campagnes de mesure. Ces individus sont donc particulièrement intéressants d'un point de vue technique ; les données disponibles sont complètes et variées.

Deuxièmement, nous avons étudié avec des populations issues de cohortes (cohorte RECORD [92]). Ces individus peuvent être sains, ou avoir certaines maladies. S'ils en ont, elles nous sont inconnues ; le protocole ainsi que le règlement général sur la protection des données (RGPD [93]) nous imposent de ne pas posséder ces informations.

Troisièmement, les individus ayant une maladie grave, recrutés par le biais de l'hôpital. Ces derniers ne sont pas mobiles, mais sont très sensibles aux variations des niveaux de pollution atmosphérique. Ce sont donc les participants les plus intéressants d'un point de vue épidémiologique.

##### 3.1.1. CRITÈRES D'EXCLUSION

Les participants ont été sélectionnés en fonction de leur lieu de résidence, leur lieu de travail ainsi que les moyens de transports qu'ils emploient. Les personnes de moins de 18 ans, les personnes n'étant pas suffisamment disponibles durant les périodes correspondant aux campagnes de mesure et les personnes effectuant des déplacements en dehors des lieux ciblés (la région Île-de-France pour le projet POLLUSCOPE), ont été exclues.

#### 3.2. TRAITEMENT DES DONNÉES DE POLLUTION ATMOSPHÉRIQUE

Les données issues des capteurs de pollution n'ont subi que très peu de post-traitement ; les valeurs négatives (possible dans le cas de dysfonctionnement du capteur AE51) ont été éliminées. Dans le cas de données manquantes, les données sont générées sur une durée maximale de 10 minutes. Les valeurs maximums ont été écrêtées afin de ne pas fausser les moyennes dans le cas d'exposition directe prolongée avec des sources de pollution. Dans le cadre du projet POLLUSCOPE, les valeurs de seuil ont été déterminées sur la base des valeurs seuil maximum recommandées et sur les limitations techniques des capteurs ;  $300\mu\text{g}/\text{m}^3$  pour le CANARIN (particules fines),  $250\mu\text{g}/\text{m}^3$  pour le Cairsens  $\text{NO}_2$  (obtenue depuis la valeur en ppb),  $50000\text{ ng}/\text{m}^3$  pour l'AE51 (BC). Nous avons utilisé le langage Python et les bibliothèques NumPy, Pandas [94], Matplotlib [95], Seaborn [96] et Sklearn [97], pour produire les résultats fournis dans les graphiques et tableaux. Les données furent téléchargées depuis la plateforme d'hébergement des données du projet.

### **3.3. ACQUISITION DE DONNÉES CONTEXTUELLES**

Les participants ont été sollicités pour fournir des données de contextualisation par annotation. Ces données doivent permettre d'indiquer si les événements extérieurs permettent d'expliquer l'évolution des niveaux de polluants auxquels les individus sont exposés. Les participants et sujets sont donc sollicités pour prendre note des environnements dans lesquels ils se situent, des modes de déplacement, ou bien des actions menées qui pourraient engendrer une baisse ou amélioration de la qualité de l'air ; cuisine, ménage, ouverture de fenêtres.

Cette tâche a dans un premier temps été réalisée sur papier dans le but d'établir de la pertinence des informations fournies. Cependant le papier est un support qui nécessite pour être numérisé de reporter l'information sur un tableur. Le manque d'uniformité et d'organisation des informations annotées rendait difficile leur intégration dans la base de données du projet. Pour simplifier cette tâche et répondre aux difficultés rencontrées, une application Android dédiée au projet a été développée. Cette application permettait à son utilisateur d'accéder à une liste de choix multiples. Pour chaque information fournie, l'heure était automatiquement enregistrée. Les données générées ont été stockées localement sur une tablette Android, puis transférées au serveur du projet par les techniciens en charge du rapatriement des différentes flottes de capteurs.

### **3.4. INDICATEURS DE SANTÉ**

Deux capteurs ont été utilisés pour évaluer les paramètres de santé ; un spiromètre permettant d'effectuer un test de contrôle de la fonction respiratoire (spirométrie), et un oxymètre permettant de quantifier la saturation en oxygène dans le sang (oxymétrie). Les participants furent sollicités pour effectuer trois séries de trois tests par jour. L'objectif était de réduire le risque qu'un test de spirométrie soit mal effectué. En effet, un test de spirométrie doit généralement être encouragé et nécessite un effort physique considérable. Les formateurs ont eu pour rôle de transmettre cette information et de former les participants à la bonne utilisation des oxymètres. Ainsi, pour chaque période de la journée (matin, midi, soir), seules les données du test fournissant le résultat le plus élevé ont été exploitées.

## 4. PROJET CANARIN

La recherche médicale requiert des outils dédiés pour évaluer et comprendre l'exposition individuelle à la pollution de l'air, puis étudier les effets néfastes qu'elle peut avoir sur la santé. Le projet CANARIN réunit une équipe internationale de chercheurs et d'étudiants avec un objectif commun : produire et utiliser des capteurs de pollution atmosphérique à faible coût.

Ce projet a produit plusieurs générations de capteurs portatifs, robustes et validés, permettant de recueillir des informations sur l'exposition individuelle aux particules fines. Les meilleures technologies disponibles ont été exploitées, permettant aux capteurs de produire des données de niveau granulaire, rendues immédiatement disponibles et pouvant être utilisées dans le cadre de la recherche en santé publique.

### 4.1. HISTORIQUE DU PROJET CANARIN

Le premier CANARIN a été créé en 2016, lorsque l'Institut Asiatique de Technologie, Sorbonne Université et l'Université de Bologne se sont associés dans le cadre du projet SEA-HAZEMON [98], visant à surveiller les niveaux de particules fines provenant d'épisodes récurrents de « brume sèche » dans la région de l'Asie du Sud-Est. Une série de capteurs CANARIN de première génération ont été déployés dans l'espace urbain, permettant de collecter des données sur les particules fines ( $PM_1$ ,  $PM_{2,5}$ ,  $PM_{10}$ ), comme le font les stations fixes de surveillance de la pollution atmosphérique, mais avec des coûts de déploiement et de fonctionnement restreints.

En 2018, une seconde version du CANARIN a vu le jour ; le CANARIN II. Cette nouvelle version possède de nouvelles options de connectivité sans fil ainsi qu'un GPS (système de positionnement global, ou en anglais, « Global Positioning System ») [99]. Des CANARINs II furent également montés sur des vélos électriques accessibles au public et ont ainsi été disséminés dans l'environnement urbain de Bologne. La possibilité de visualiser en temps réel des données géo-localisées sur la pollution de l'air a constitué une amélioration majeure pour le projet CANARIN.

L'utilisation de capteurs portatifs pour évaluer l'exposition humaine avait déjà été étudiée à l'Institut asiatique de technologie [100]. Plusieurs capteurs CANARIN II ont été déployés dans la Bibliothèque Joanina de Coimbra afin de surveiller l'impact des particules fines sur son écosystème. En effet, la population de chauves-souris y joue un rôle essentiel dans la préservation de la bibliothèque, en se nourrissant des vers de livres pendant la nuit [54].

En 2019, un CANARIN plus petit et plus léger a été créé : le CANARIN nano. Nous avons activement participé à son développement, à sa production et accompagné son déploiement. Il a depuis été utilisé dans le projet POLLAR, avec pour objectif principal d'explorer si l'apnée du sommeil est aggravée par l'exposition à des microparticules [101]. Les progrès technologiques réalisés dans le domaine des microcontrôleurs, des détecteurs de particules et de VOCs, ont permis de réduire la taille du capteur sans compromettre la qualité des données acquises.

#### 4.1. TECHNOLOGIES EMPLOYÉES

Les capteurs CANARINs utilisent tous des cartes de développement et des modules de mesure disponibles dans le commerce. Cela permet de réduire le temps de développement nécessaire. De plus, les modules de mesure utilisés ont déjà été mis en œuvre et testés [102]. Une fois le compteur optique de particules (OPC de l'anglais « Optical Particle Counter ») sélectionné, nous nous sommes efforcés d'intégrer les composants nécessaires dans un boîtier en suivant les recommandations du fabricant, garantissant ainsi que le module de mesure puisse fonctionner dans les meilleures conditions possibles. Les composants électroniques nécessaires ont été assemblés sur un circuit imprimé que nous avons fait fabriquer. Aucun traitement n'est appliqué aux données avant ou après leur transmission.

##### 4.1.1. CANARIN II

Le CANARIN II utilise une carte développement fonctionnant sous une distribution Linux, possédant une puissance de calcul qui offre la possibilité d'implémenter des fonctionnalités tel que le traitement d'images en temps réel. Le CANARIN II transmet les données qu'il recueille via Wi-Fi. Il ne dispose pas de connectivité cellulaire et requiert donc un point d'accès Wi-Fi lorsqu'on désire transmettre les données en temps réel. Dans le cas d'une utilisation en mobilité, les individus portant le capteur ont la possibilité de créer un point d'accès mobile. Enfin, le CANARIN II peut fonctionner sans point d'accès et stocker les données acquises pendant plusieurs mois, dans l'attente de la disponibilité d'un point d'accès.

La carte électronique de développement du CANARIN II consomme à elle seule environ 1,8 Watt-heure au repos, tel que présenté sur la Figure 5. Il est donc nécessaire d'équiper le capteur d'une batterie conséquente pour pouvoir maintenir cette carte en fonctionnement durant une journée normale d'utilisation. C'est pourquoi le CANARIN II possède une batterie d'une capacité totale de 6800 mAh (mAh) dont la tension de sortie est de 12 volts (V) nominaux. Elle est composée de deux éléments, comprenant chacun 3 accumulateurs de type 18650, dont la capacité individuelle est de 3400 mAh.

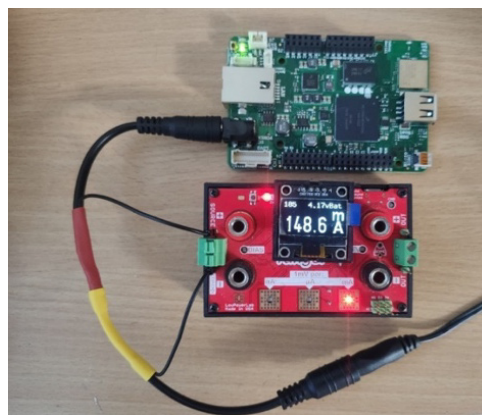


FIGURE 5 : Mesure de la consommation sur carte de développement CANARIN II

Le poids et les dimensions du CANARIN II (environ 1,2 kg, 15×12.5×7 cm) lui permettent d'être monté sur un véhicule, mais deviennent contraignantes dans le cas où c'est à des individus de le transporter, surtout si les individus en question souffrent de troubles respiratoires, comme c'était le cas pour certains sujets dans le projet POLLUSCOPE. De tels inconvénients avaient déjà été soulignés lors de l'évaluation comparative réalisée à l'étape de test et sélection des capteurs réalisée par le projet [63]. Plusieurs participants du projet POLLUSCOPE ont également commenté le manque de confort de port, tel rapporté par les formateurs et techniciens qui ont été en contact avec eux.

#### 4.1.2. CANARIN NANO

Le CANARIN nano, la dernière version du CANARIN, a été créé pour répondre aux retours des utilisateurs, pour proposer un capteur plus compact, plus léger et entièrement autonome. Il utilise un microcontrôleur basé sur le processeur ARM Cortex-M au lieu du microprocesseur UDOO NEO FULL [103] Cortex-A9 utilisé sur le CANARIN II, qui requiert davantage d'énergie. Les progrès réalisés par les fabricants de capteurs de particules leur ont permis de proposer à la vente des OPCs davantage compacts et moins énergivores. Les récents développements dans le domaine des capteurs électrochimiques ont permis, pour la première fois, d'accueillir un capteur de composés organiques volatils (COV). Le CANARIN nano embarque également un modem 3G, de la mémoire flash, un module GPS et son antenne active ainsi qu'un capteur de température, humidité et pression. Cet équipement permet de fournir le même niveau de performance que celui offert précédemment par le CANARIN II, tout en offrant une autonomie maximale de 14 heures.

Le capteur de COVs est monté sur un circuit imprimé secondaire qui peut être facilement ajouté ou retiré de la carte principale. Ce circuit imprimé secondaire offre donc la possibilité d'intégrer au CANARIN nano des capteurs électrochimiques, des microsystèmes électromécaniques (MEMs) ou des capteurs de gaz à oxyde métallique plus performants, sans requérir de produire à nouveau l'ensemble du circuit imprimé du capteur CANARIN. Une mémoire flash a été utilisée pour éviter la perte de données dans le cas où le modem 3G embarqué perdrait toute connectivité avec les tours cellulaires. Un fournisseur tiers de cartes SIM (Subscriber Identity Module) IoT ou M2M (Machine to machine) a été sélectionné plutôt et place des services proposés par défaut par Particles services [104], le fabricant de la carte de développement sélectionnée. Le Tableau 1 présente certains des composants utilisés dans le CANARIN nano.

TABLEAU 1 : Composants du CANARIN nano

Composant	Référence
Carte de development	Particle Electron 3G-U270
Microcontrôleur	STMicroelectronics STM32F205
Module GSM	U-BLOX SARA-U270
Antenne Cellulaire	Molex 2072350100
Fournisseur de carte SIM IoT/M2M	ThingsMobile
Module GPS	U-BLOX NEO-M8N
Antenne GPS	Taoglass AGGBP.25B.07.0060A
Batterie	GEB OEM 4000 mAh Li-Po
Chargeur	CUI Inc. SWI10-5-E-I38
Mémoire flash	Winbond Electronics W25Q64FWZPIG8MB
Capteurs de COVs	Sensirion SGP30, Bosch BME680

Après s'être assuré que les premiers prototypes étaient fonctionnels en utilisation réelle et après avoir obtenu les premiers retours d'expérience, il a été décidé de reconcevoir entièrement le CANARIN nano. Précédemment, la fabrication était presque entièrement locale, restreinte au sein des laboratoires universitaires et rendait relativement long la production des

CANARINs nano. Un circuit imprimé personnalisé et un nouveau boîtier ont été conçus et produits en plus grande quantité. La complexité de la liste des composants (BOM), de la logistique et les délais liés à l'approvisionnement dans un cadre universitaire, ont participé à augmenter le coût global par capteur. Le design du boîtier et du circuit imprimé ont permis en revanche de gagner en autonomie et de faciliter l'assemblage et la maintenance, tout en améliorant sa fiabilité. Le boîtier assure la protection de l'électronique et permet de rendre visible différentes diodes électroluminescentes et de donner l'accessibilité aux ports, boutons interrupteurs disponibles, visibles sur la FIGURE 6.



FIGURE 6 : Capteur CANARIN nano

Le CANARIN nano a été imaginé comme un capteur fonctionnant de manière continue. Les patients et les participants instrumentés ne doivent avoir aucune interaction avec l'appareil, en dehors de devoir le transporter avec eux et de le recharger quotidiennement. Les participants peuvent toutefois choisir de l'éteindre, désactivant ainsi le module GPS embarqué, lorsqu'ils désirent arrêter l'enregistrement de leur localisation dans le but de protéger leur vie privée. Le fait d'avoir un capteur connecté en permanence par le biais de la 3G procure plusieurs avantages ; des données échantillonnées sont envoyées chaque minute à notre infrastructure web et sont immédiatement disponibles pour les chercheurs. Cela permet de contrôler à distance le bon fonctionnement des capteurs, avant, après et pendant un déploiement. Les utilisateurs des CANARIN nano peuvent donc être suivis plusieurs fois par jour, limitant donc le risque de perte de données dans le cas d'un dysfonctionnement ou d'une erreur utilisateur. Les modifications de paramètres ou les mises à jour complètes du micrologiciel du CANARIN nano peuvent être effectuées à distance, autrement dit via une mise à jour OTA (over the air), sans que l'utilisateur final ne s'en aperçoive. La réduction de la taille et du poids du capteur a permis de créer un dispositif répondant mieux aux diverses exigences de la recherche médicale que d'autres capteurs commerciaux sous le tarif plancher de 1000 euros. Des améliorations seront apportées tant que la créativité des étudiants et des chercheurs continuera à alimenter le projet CANARIN.

#### 4.1.3. INFRASTRUCTURE CLOUD

Toutes les minutes, les données recueillies par les différents modèles de capteurs CANARINs déployés sont transmises au serveur de l'infrastructure cloud du projet CANARIN. Cette infrastructure est hébergée par Amazon Web Services (AWS) sur une instance EC2 T2 [105]. Un processeur virtuel (vCPU) et 1 Go de RAM sont suffisants pour assurer le service pour environ 100 capteurs CANARINs nano. Plusieurs conteneurs docker [106] exécutent des scripts Python pour ingérer les données reçues. La base de données est également hébergée par AWS à Francfort, conformément à la GDPR (General Data Protection Regulation). Le portail web [107] est sécurisé et permet aux chercheurs de télécharger ou de visualiser les données acquises par les capteurs en temps réel.

#### 4.1.4. FACE AUX CAPTEURS EXISTANTS

Durant ces dernières années, le projet CANARIN a donc cherché à résoudre des problèmes tels que la connectivité, le coût, la taille, la facilité d'utilisation, ainsi que la qualité des données acquises. Les capteurs disponibles dans le commerce ne sont souvent pas en mesure de fonctionner en totale autonomie. Les capteurs non-connectés nécessitent généralement de transférer manuellement les données produites vers un poste de travail. Les capteurs connectés requièrent, soit de donner l'accès à internet au capteur au travers d'un point d'accès Wi-Fi, soit de lui permettre de communiquer avec le smartphone du porteur du capteur via Bluetooth. Dans ce dernier cas, c'est le smartphone qui se chargera de la transmission des acquisitions, permettant alors la mise à disposition des données recueillies sur internet. Les données ne sont généralement partagées qu'avec l'utilisateur du capteur, sont prétraitées et ne permettent donc pas d'être immédiatement exploitées pour la recherche médicale.

Le CANARIN II a été évalué par le projet POLLUSCOPE, qui requerrait des solutions à la fois mobiles et abordables pour évaluer l'exposition individuelle à la pollution de l'air en Île-de-France. Lorsque le CANARIN II a été testé face aux différents capteurs de particules disponibles sur le marché, il a été en mesure de produire de meilleurs résultats, comme le rapportent les experts nationaux [63]. Le CANARIN II a donc été sélectionné puis utilisé par le projet POLLUSCOPE.

Les modules de mesure permettant de mesurer les concentrations de particules fines utilisées dans le projet CANARIN sont d'une qualité supérieure à ceux généralement embarqués dans les capteurs portatifs. Bien que ces modules soient abordables, leur intégration impose aux constructeurs des contraintes de coût, taille et consommation énergétique, qui sont à l'heure actuelle trop contraignantes pour un produit destiné à la population générale.



## **5. PROJET POLLUSCOPE : OBSERVATOIRE PARTICIPATIF POUR LA SURVEILLANCE DE L'EXPOSITION INDIVIDUELLE À LA POLLUTION DE L'AIR EN LIEN AVEC LA SANTÉ**

Nous avons présenté le projet CANARIN et ses deux capteurs, produits en milieu académique et spécifiquement conçus pour des projets de recherche en lien avec sur la santé et l'exposition individuelle à la pollution de l'air : le CANARIN II ainsi que le CANARIN nano. Dans le cadre du projet POLLUSCOPE, nous avons étudié les résultats de leur utilisation auprès de cohortes réelles de participants. Nous avons été amenés à réaliser plusieurs supports de communication pour le projet ; vidéos et flyers d'information et d'instruction. Un poster illustrant les différents acteurs au sein du projet est présenté en annexes, présenté durant le colloque national des capteurs et sciences participatives, page 126.

### **5.1. OBJECTIFS**

L'objectif du projet POLLUSCOPE fut de développer et tester une plateforme de collecte et d'analyse de données issues de capteurs portatifs mesurant des paramètres environnementaux et plus spécifiquement, les paramètres relatifs à la qualité de l'air, ainsi que des paramètres de santé. Autrement dit, le projet POLLUSCOPE a cherché à concevoir, développer puis mettre en œuvre un « Observatoire participatif pour la surveillance de l'exposition individuelle à la pollution de l'air en lien avec la santé ». Le projet POLLUSCOPE a été financé par l'ANR, subvention 556 ANR-15-CE22-0018 [108].

Le projet POLLUSCOPE a nécessité la participation et l'articulation d'un vaste arbre compétences ; des physico-chimistes de l'atmosphère ont évalué et sélectionné les capteurs puis validé leur bon fonctionnement durant le projet. Des informaticiens ont développé une plateforme permettant de recueillir et héberger, traiter, puis rendre disponibles au public et aux chercheurs les données recueillies. Des électroniciens ont réalisé la maintenance des capteurs et ont développé des outils de test et de validation. Des épidémiologistes ont corrélé les données environnementales avec les paramètres de santé. Des statisticiens ont permis d'identifier les différents profils d'exposition en fonction des paramètres environnementaux et contextuels. Des techniciens ont déployé les capteurs, transmis les informations utiles aux participants, puis manuellement recueilli les données lorsque stockées localement sur les capteurs.

Le projet utilise des technologies existantes, mais leur utilisation dans un tel contexte est relativement novatrice. Les données obtenues ont permis de produire des résultats scientifiques pertinents. L'expérience terrain, les succès et échecs fournissent un retour d'expérience nécessaire aux projets de plus grande envergure qui succéderont.

#### *5.1.1. SÉLECTION DES CAPTEURS*

Le projet POLLUSCOPE, pour créer un observatoire participatif de la pollution de l'air au niveau individuel, devait donc se doter de capteurs portatifs à faible coût. L'un des premiers

objectifs du projet fut donc de tester, de comparer et d'identifier quels sont les capteurs qui seraient les mieux adaptés à cette tâche. Les capteurs sélectionnés ont ensuite été évalués par un groupe d'experts, qui les ont testés entre eux, puis face à des appareils de référence [63], afin de s'assurer que leur performance était suffisamment satisfaisante pour être utilisés dans des scénarios de mobilité réels. Le CANARIN II a ainsi été soumis à des tests statiques et dynamiques. Le capteur CANARIN II a ainsi prouvé ses aptitudes en matière de qualité métrologique et a été sélectionné par le projet.

POLLUSCOPE est un projet qui s'est établi sur plusieurs années ; les capteurs peuvent donc rester inutilisés pendant des mois. Ainsi, pour limiter le risque de déployer un capteur défectueux, tous les capteurs de pollution atmosphérique sélectionnés ont été remis aux experts de la pollution atmosphérique du projet pour être testés vis-à-vis d'instruments de référence, pour être évalués avant de procéder à plusieurs semaines de mesures. Pour l'évaluation des CANARIN II, un Thermo Scientific™ 1405-F (TEOM) et un ThermoFisher 1400 (TEOM) ont été utilisés [63]. Si les capteurs se comportaient de manière inattendue, ils étaient alors mis de côté pour être révisés.

#### 5.1.2. INSTRUMENTATION DES PARTICIPANTS

Le deuxième objectif du projet fut d'évaluer l'exposition individuelle à la pollution atmosphérique en fonction de la mobilité et des activités. Trois groupes de population ont été considérés : des participants recrutés autour du Grand Parc de Versailles (campagne VGP), des participants à l'enquête de population RECORD [109] (campagne RECORD), ainsi que des participants atteints de maladies respiratoires. Chacun des participants a dû donner leur consentement écrit pour que leurs données soient exploitées, puis répondre à une enquête en ligne permettant de recueillir des données socio-économiques importantes avant le début de l'étude. Une partie des participants ont été équipés de capteurs de santé permettant de recueillir des données en mobilité ; un spiromètre et un oxymètre portatifs. Ces deux capteurs sont simples d'utilisation, ne requièrent pas au participant de les recharger et sont à la fois très compacts et légers. Les acquisitions sont transmises directement à un serveur respectant les restrictions de la RGPD pour les données de santé, aucune donnée n'est stockée localement.

Les participants ont ensuite porté un set de capteurs portatifs pour une période moyenne d'une semaine. Les capteurs portatifs de pollution de l'air ont été rechargés quotidiennement et il a été demandé aux participants de les brancher quand c'est possible afin d'éviter la perte accidentelle de données due au manque de batterie. Davantage de recommandations d'utilisations ont été communiquées aux participants lors de leur formation, notamment afin de veiller à ce que les capteurs ne soient pas placés à proximité des sources directes de pollution ou de chaleur (une plaque de cuisson, une cheminée par exemple), ni exposés à des niveaux anormaux d'humidité (dans une salle de bain, sous la pluie), qui pourraient à la fois impacter la pertinence des mesures durant cette période, mais aussi détériorer de manière permanente leur bon fonctionnement.

Les participants ont systématiquement été équipés d'un CANARIN II, présenté sur la Figure 7. Le CANARIN II embarque un capteur de particules ( $PM_{10}$ ,  $PM_{2.5}$ ,  $PM_1$ ), un GPS,

ainsi qu'un capteur de température, humidité et pression [99]. Le boîtier des CANARIN II est imprimé en 3D et a été conçu de façon à pouvoir y accrocher un second capteur. À cet emplacement est inséré un capteur de dioxyde d'azote (NO<sub>2</sub>) produit par la société Envea [110] ; le Cairsens NO<sub>2</sub> [111]. Un nombre limité de participants ont été équipés de capteurs de noir de carbone (BC). Ces capteurs étant plus coûteux n'étaient disponibles qu'en nombre limité. Chacun des capteurs de pollution atmosphérique a été configuré pour réaliser une mesure par minute, de manière continue sur toute la semaine de mesure.

Le kit de capteurs de pollution de l'air pouvait être complété par des capteurs mobiles mesurant des paramètres de santé dans le but d'analyser les effets sanitaires de l'exposition à la pollution atmosphérique. Les deux capteurs de santé sélectionnés sont vendus par la société MIR [112], qui propose une solution logicielle pour l'acquisition des données, leur d'hébergement et la mise à disposition des données. Les capteurs en question ; un spiromètre ainsi qu'un oxymètre portatif, tous deux communiquant via Bluetooth faible énergie (BLE) et ne nécessitant pas d'être rechargés quotidiennement. L'utilisation de ces deux capteurs requiert au participant quelques minutes d'effort pour effectuer plusieurs tests de spirométrie. Les participants ont été sollicités à effectuer trois tests de spirométrie, trois fois par jour, totalisant donc 9 tests de spirométrie par jour. Lors du traitement des données, seul le test de spirométrie fournissant les meilleurs résultats sera conservé.



FIGURE 7 : CANARIN II avec un capteur Cairsens NO<sub>2</sub>

Une tablette Android (Archos 101 Oxygen 4G [113]) accompagne systématiquement le kit de capteurs. Cette tablette est équipée d'un GPS et modem 4G et permet de remplir cinq rôles. Premièrement, la tablette sert d'outil d'annotation de données contextuelles, via une application dédiée développée par le projet. Deuxièmement, elle permet de fournir une source additionnelle de données GPS. Sa batterie étant plus conséquente qu'un smartphone, elle permet d'enregistrer plusieurs points GPS par minute, sur une durée de 16H environ, avec une seule charge. Troisièmement, elle fournit un point d'accès Wi-Fi auquel le CANARIN II se connecte pour transmettre les acquisitions. Ceci permet de monitorer en temps réel le bon fonctionnement du CANARIN et donc de contacter le participant ou d'intervenir en cas de problème technique. Quatrièmement, la tablette contient des vidéos éducatives relatives à l'utilisation des capteurs, qui sont ainsi facilement accessibles en cas d'oubli du participant. Cinquièmement, elle permet aux participants de l'étude épidémiologique de réaliser les tests de spirométrie et d'oxymétrie. En effet, les deux capteurs de paramètre de santé communiquent avec la tablette via Bluetooth basse consommation (BLE). Une application mobile réalise l'enregistrement puis la transmission des données par 4G.

### 5.1.3. DISSÉMINATION DES DONNÉES D'EXPOSITION

Le troisième objectif du projet fut de fournir aux participants et aux chercheurs l'information mesurée. Pour servir ce rôle, une plateforme web dédiée a été développée. Elle dispose de plusieurs types d'accès qui servent chacun un rôle différent pour un public distinct.

Le premier type d'accès, destiné aux techniciens, permettait l'envoi des données récoltées par les capteurs non-connectés ; le Cairsens NO<sub>2</sub>, l'AE51, ainsi que les données disponibles sur la tablette ; celles issues de l'annotation des activités et de l'enregistrement de la géolocalisation par GPS. Le technicien doit déclarer chaque participant, avec son kit de capteurs et fournir le nom de la campagne, permettant aux données d'être ingérés et traités avant d'être rendus disponibles sur la base de données du projet. La seconde, fut l'interface de développement web Jupyter [114], hébergée par le serveur du projet POLLUSCOPE, permettant aux chercheurs d'accéder directement à la base de données du projet et de traiter les données acquises, en développant des scripts en langage Python. Des algorithmes ont ainsi été mis en œuvre afin de compléter partiellement les données manquantes, en utilisant les données disponibles, qu'elles soient issues d'autres capteurs de pollution de l'air, des données GPS, ou des données de contextualisation fournies par l'utilisateur [115]. La troisième, se destinait aux participants, pour lesquels il est possible de visualiser l'ensemble de leurs données individuelles via une interface dynamique [116]. POLLUSCOPE s'est assuré de la conservation de l'anonymat des individus puisque la plateforme n'héberge aucune information identifiante. Seuls les recruteurs ont connaissance de l'identité et de l'identifiant unique de chaque participant.

Il fut admis d'un commun accord qu'il était nécessaire de faire davantage pour le public qui n'est pas désireux d'utiliser l'interface web suscitée. Tous les participants ont ainsi obtenu un rapport numérique au format PDF, facilement consultable sur un smartphone ou ordinateur. Le rapport fournissait des informations générales ; graphiques, cartes et statistiques permettant notamment d'identifier les endroits et activités où l'air est le plus pollué. En plus de ces informations, lorsque c'est nécessaire, les rapports sont complétés par des commentaires et annotations personnalisés. L'automatisation d'une telle tâche n'a été limitée que par notre capacité à comprendre les données disponibles. Des informations pertinentes peuvent être fournies si nous savons dans quel contexte se produit l'exposition à la pollution atmosphérique. Nous avons démontré que la multiplicité des données du capteur portatif, complétées par les annotations de l'utilisateur, peuvent être utilisées pour améliorer les connaissances sur l'exposition personnelle à la pollution de l'air [117].

## 5.2. RECRUTEMENT DE PARTICIPANTS

L'objectif du projet fut de couvrir une population totale de 200 participants. Au sein de cette population, 50 participants devaient être affectés par une maladie respiratoire chronique telle que la bronchopneumopathie chronique obstructive (BPCO). Étant donné que ces participants sont particulièrement sensibles aux variations de la qualité de l'air, ils permettront de mettre en exergue la corrélation entre l'évolution de la fonction respiratoire et leur exposition à la pollution atmosphérique. Ils participent à une étude épidémiologique avec un examen de santé avant et après la semaine de mesure et pourront ainsi être comparés à 50 participants

sains. La centaine de participants restants seront issus de la population générale. À ce jour, cette partie de l'étude est encore en cours. L'épidémie de la COVID 19 ayant rallongé les temps.

Chacun des participants ont été contactés par courrier électronique ou par téléphone afin de les informer des objectifs du projet. Des vidéos informatives ont permis aux potentiels participants de s'informer sur la nature des capteurs à porter ainsi que sur les opérations requises pendant la semaine d'acquisition. En fonction des habitudes de fonctionnement des partenaires et des populations visées, c'est deux processus de recrutement qui ont été utilisés pour mener à bien les campagnes de recrutement et de mesure.

### 5.2.1. RECRUTEMENT AVEC ACCOMPAGNEMENT

Le premier processus de recrutement choisi fut axé sur l'accompagnement des participants (Figure 8). Des participants ayant précédemment participé à des études sur la mobilité, donc susceptibles d'être intéressés de participer au projet, ont été identifiés puis contactés par courriers électroniques. S'ils répondent positivement, ils ont alors été conviés à participer à un entretien téléphonique, après lequel les capteurs furent apportés à leur domicile. Les participants ont bénéficié d'une formation individuelle sur le bon usage des capteurs et outils, qui fut dispensée par le technicien lors de la remise des capteurs.

À l'issue de la formation et après avoir pris pleine connaissance des contraintes, les volontaires sont questionnés sur leur motivation à participer. Ils ont donc eu la liberté de se rétracter et de refuser de participer à la campagne. L'objectif de cette demande fut de limiter le risque d'abandon rapide du participant, puisque très coûteux à la fois en temps et en logistique. À l'issue de la semaine de mesure, les capteurs furent récupérés au domicile des participants. Notons que durant cette formation, des tubes remplaçables en carton furent utilisés par le formateur pour éviter tout risque sanitaire. Les capteurs ont été nettoyés à l'aide d'une solution hydro-alcoolique désinfectante, avant déploiement des capteurs, ainsi qu'à leur récupération.

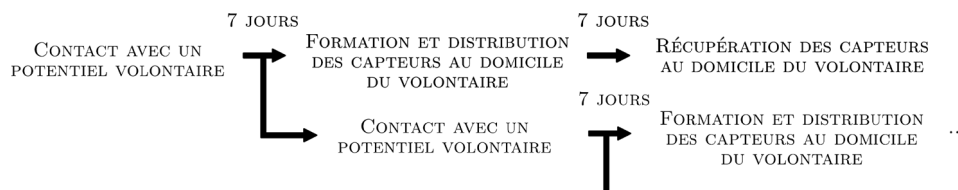


FIGURE 8 : Flux opérationnel du processus de recrutement avec accompagnement

Ce processus de recrutement est le plus confortable pour le participant, qui n'a pas à se déplacer pour obtenir les capteurs. Ils furent davantage accompagnés à leur utilisation et disposaient d'un contact direct avec le formateur en cas de difficulté, de doute, ou dans le cas d'un dysfonctionnement des capteurs. Du fait de la relation créée avec le formateur, nous avons observé que les participants ont davantage contacté leur formateur en utilisant ce processus. Étant donné qu'il est logistiquement et humainement complexe de distribuer rapidement plusieurs capteurs sur Paris et sa périphérie, les campagnes de mesure utilisant ce processus de recrutement se sont établies sur plusieurs semaines.

### 5.2.2. RECRUTEMENT PAR VOLONTARIAT

Dans le cas du recrutement par volontariat, les participants furent informés par des affiches, stratégiquement placées dans des lieux publics, informant l'existence du projet, de ses enjeux et des futures campagnes de recrutement. Les personnes intéressées furent invitées à contacter un recruteur qui en retour leur a transmis des vidéos et documents informatifs, afin de s'assurer de l'intérêt porté par chaque participant. Tous les participants contactés furent ensuite conviés à se rendre dans un lieu à proximité du lieu d'affichage afin de participer à une session commune de formation à l'usage des capteurs. Les sessions furent conviviales et animées par plusieurs formateurs. Les participants eurent l'opportunité de poser des questions aux formateurs durant la formation, tout en ayant la possibilité de les contacter au cours de la campagne de mesure. À l'issue de chaque campagne de mesure, similairement à la journée de déploiement, une demi-journée fut organisée pour la remise des capteurs.



FIGURE 9 : Flux opérationnel du processus de recrutement par volontariat

Ce processus de recrutement fut plus efficace que le premier, puisque la campagne se déroule sur une unique semaine et les capteurs furent déployés et récupérés très rapidement grâce à l'effort individuel des participants. Ces participants sont donc issus de la population générale.

### 5.2.3. DÉROULEMENT DE L'ENQUÊTE

Avant de se voir remettre un kit de capteurs, les participants ont dû remplir un questionnaire en ligne. Le questionnaire a permis de recueillir des données socioéconomiques sur la population participant à l'étude. Le questionnaire fut hébergé par l'équipe du Laboratoire DAVIDE de l'UVSQ, en utilisant le logiciel d'enquête en ligne open source LimeSurvey [63]. Les résultats produits ne peuvent être généralisés, du fait que l'échantillon de population ayant participé à ce jour soit trop réduit.

Le questionnaire est exhaustif puisqu'il totalise 72 questions, produisant un total de 393 variables. Nous avons exploré ces données en lien avec les mesures de qualité de l'air réalisées par les capteurs portatifs. Afin de faciliter l'exploration des données issues des questionnaires, nous avons effectué une compilation des questions visant à réduire le nombre de variables à explorer. Les questions relatives à la pratique sportive sont par exemple composées de 146 variables possibles, que nous avons décidé de réduire à deux variables ; le nombre d'heures par semaine et ratio du nombre d'heures effectuées en extérieur. Lorsque c'était possible, nous avons produit des variables conditionnelles. Quelques variables n'ont pas été conservées par manque de pertinence. Ainsi après compilation, 25 variables ont été obtenues. Deuxièmement, les participants devaient être disponibles durant deux weekends pour la remise et récupération des capteurs et devaient être intéressés par le projet. Dans le cas du premier processus de recrutement privilégiant l'accompagnement, les participants avaient tous une première expérience avec le port de capteurs et donc, avaient tous l'autorisation d'apporter les capteurs sur leur lieu de travail.

### 5.3. TECHNOLOGIE DES CAPTEURS ET INSTRUMENTS EMPLOYÉS

#### 5.3.1. COMPTEUR OPTIQUE DE PARTICULES

Le CANARIN sélectionné par le projet POLLUSCOPE embarque un OPC. Le principe de fonctionnement des OPCs est de mesurer la diffraction de la lumière causée par les particules traversant un faisceau lumineux émis à l'intérieur d'une chambre de détection, dont l'air est extrait ou forcé de manière constante. La source de lumière doit être suffisamment lumineuse ; pour cette raison, c'est un laser qui est systématiquement utilisé dans le cas des micro-capteurs. Pour mesurer la diffraction, une cellule photo détectrice ou un capteur d'image peut être utilisé. Dans ce dernier cas, il faut alors que l'OPC soit équipée d'un bloc optique. La taille des particules est déterminée par la longueur d'onde de la lumière incidente avec la particule [118].

Ces dernières années, plusieurs OPCs de petite taille, issus de différents fabricants, ont été commercialisés et évalués par la communauté scientifique [55,62,119,120]. Ci-dessous, la vue intérieure d'une OPC de la marque chinoise Plantower, modèle PMS7003. L'OPC mesure 47mm par 37mm pour 12mm d'épaisseur et pèse seulement 26 grammes. Le flux d'air continu est produit par un ventilateur contrôlé par modulation de largeur d'impulsion (PWM). Le capteur se compose d'une structure en plastique, enveloppée dans de fines feuilles d'acier de 0.25mm d'épaisseur, qui se clipsent sur la structure. Ces feuilles d'acier réalisent l'étanchéité avec l'air extérieur, la mise à la masse du boîtier et fournissent un ratio poids, résistance et épaisseur qu'il ne serait pas possible de réaliser avec un boîtier en plastique. La structure en plastique étant placée entre les deux feuilles d'acier, elles permettent de créer une cavité sur chacune des faces.

Sur la Figure 10, on peut voir le micro-ventilateur équipant le capteur OPC. Ce micro-ventilateur a accès à l'air extérieur. L'air extérieur est introduit dans la cavité est représenté (en bleu). L'air est ensuite amené, par la pression générée par le ventilateur, dans d'un conduit restreint au travers duquel il est exposé au laser (en rouge) et où une cellule optique (en vert) compte les particules. L'air est ensuite amené dans la seconde cavité au travers de trous dans la structure (en haut) jusqu'à être extrait à l'extérieur du capteur via une aération à proximité du ventilateur. Ce choix de design permet au constructeur de placer l'entrée d'air à proximité de la sortie d'air, facilitant donc l'intégration de ce module dans des capteurs portatifs, puisque transporter un capteur en veillant à ne pas obstruer les trous destinés à la ventilation peut être difficile pour son porteur.

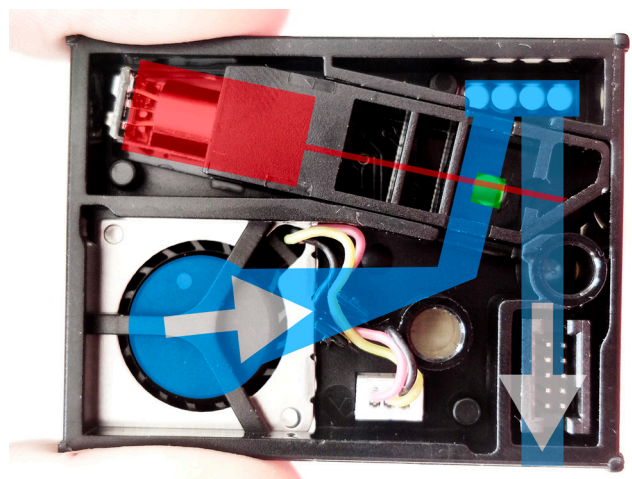


FIGURE 10 : Fonctionnement d'un compteur de particules optique

### 5.3.2. CAPTEUR ÉLECTROCHIMIQUE DE $\text{NO}_2$

Un capteur électrochimique permet de mesurer la concentration d'un gaz dans l'air, en utilisant des électrodes réagissant physiquement avec le gaz cible présent dans l'air. La réaction électrochimique d'oxydation ou de réduction génère un courant qui est mesuré par un circuit. Ces cellules sont sensibles aux variations de températures, pression et humidité et réagissent également aux autres gaz présents dans l'atmosphère et peuvent ainsi introduire une erreur de mesure positive comme négative.

L'une des principales difficultés dans la conception d'un capteur électrochimique est donc de mitiger cet effet pour améliorer la sélectivité du capteur. Dans le cas des capteurs électrochimique de  $\text{NO}_2$ , ils sont particulièrement sensibles à la présence de gaz oxydants tels que l' $\text{O}_3$ . Pour pallier ce phénomène, deux solutions techniques existent ; l'utilisation de filtres permettant de retirer l' $\text{O}_3$  de l'air, ou d'un second capteur d' $\text{O}_3$  hautement sélectif. Le Cairsens  $\text{NO}_2$  utilisé dans le projet POLLUSCOPE embarque une cellule électrochimique ainsi qu'un filtre à  $\text{O}_3$  qui permet donc d'améliorer la sélectivité de la cellule électrochimique et de limiter l'effet de l'humidité [111].

### 5.3.3. CAPTEUR DE BC

Pour mesurer l'exposition individuelle au BC, l'aethalomètre AE51 de la société Américaine Aethlabs a été sélectionné. Ce capteur n'est pas récent, a été donc déjà évalué et utilisé pour évaluer l'exposition individuelle [121–123]. Le principe de fonctionnement de l'aethalomètre consiste à mesurer l'atténuation de la lumière transmise au travers d'un filtre fibreux. L'atténuation est causée par l'accumulation des particules de BC sur ce filtre et est proportionnelle à son opacité. Connaissant la spectrométrie d'absorption du BC, en utilisant plusieurs longueurs d'ondes, il est possible ainsi de mieux discriminer le BC des autres particules présentes dans l'air et pouvant être capturées par le filtre.

C'est la force d'aspiration créée par une pompe embarquée et fonctionnant continuellement qui amène les particules de BC à se fixer sur le filtre. Le filtre en question doit être remplacé à intervalles réguliers. En fonction de la vitesse de fonctionnement de cette pompe, après quelques heures d'utilisation, l'accumulation des particules sur le filtre et l'augmentation de la densité, rend la mesure moins fiable ; elle peut être corrigée si l'on connaît les heures de changement des filtres [124].

### 5.3.4. SPIROMÈTRE

La spirométrie est le test standard d'évaluation de la fonction pulmonaire. Il permet de mesurer la fonction respiratoire ; la quantité d'air que les poumons peuvent emmagasiner et la vitesse maximale d'expiration ou d'inspiration. La spirométrie permet de diagnostiquer des maladies telles que la fibrose, la mucoviscidose et la bronchopneumopathie chronique obstructive. Parce que la taille des organes varie entre les individus, pour pouvoir juger des résultats d'un test de spirométrie, il est important de prendre en considération la taille, le poids, sexe et origine ethnique. La spirométrie peut être effectuée différemment en fonction de l'équipement utilisé. Les spiromètres portatifs Spirobank Smart utilisés par le projet



POLLUSCOPE mesure l'expiration seulement et requiert l'utilisation d'un smartphone ou tablette, ainsi aucune donnée n'est stockée localement. Ils ont déjà été utilisés dans le cadre de nombreuses études [125,126] et sont suffisamment performant pour être utilisés dans le cadre de la recherche médicale, à condition que le protocole approprié soit utilisé [127]. Voici une description des paramètres les plus importants mesurés durant la spirométrie.

- CVF : La capacité vitale forcée correspond au volume maximum en litres d'air qu'une personne peut expirer après avoir effectué une inhalation profonde.
- VEMS : Le volume expiratoire maximal par seconde correspond au volume expiré après une seconde, lors d'une expiration forcée.
- VEMS/CVF : Le ratio entre VEMS et CVF, obtenu par exploration fonctionnelle respiratoire, aussi appelé coefficient de Tiffeneau [128], permet d'évaluer le degré d'obstruction des bronches, notamment pour évaluer

Dans notre cas, nous avons exploité le paramètre VEMS/CVF. Le patient doit inspirer profondément et expirer le plus fort et le plus vite possible, dans le tube du spiromètre, tout en maintenant une expiration forcée le plus longtemps possible. Un élément mécanique est alors mis en rotation par le souffle du patient. Une cellule optique mesure la fréquence de rotation de cet élément en continu pendant la durée du test, les données sont recueillies par la tablette, qui s'assure du bon déroulement du test puis transmet les données à un serveur distant, sans conserver de données de santé localement.

#### 5.3.5. OXYMÈTRE

L'oxymètre permet de mesurer de manière non invasive la saturation en oxygène du sang, en mesurant le pouls. La valeur relevée est la saturation pulsée en oxygène ( $SpO_2$ ). Le sang transporte l'oxygène qui s'attache à l'hémoglobine et alimente les organes et muscles du corps humain. La valeur de  $SpO_2$  relevée est insuffisante lorsqu'elle est en dessous de 95%. La méthode la plus fiable pour mesurer l'oxymétrie est de réaliser une prise de sang. Cependant cette méthode est invasive, coûteuse et inconfortable pour le patient. L'hémoglobine non chargée en oxygène (Hb) n'a pas la même absorption spectrale que l'hémoglobine chargée en oxygène (oxyhémoglobine,  $HbO_2$ ). Il existe une région, à 800 nanomètres, où l'absorption est similaire. Les oxymètres portatifs exploitent donc cette propriété de l'hémoglobine pour mesurer sa présence dans le sang. Deux émetteurs et récepteurs de lumière, de longueur d'onde différentes permettent de mesurer la différence entre Hb et  $HbO_2$ . Dans notre oxymètre, les longueurs d'onde utilisées sont 660nm et 870nm. La lumière doit pouvoir traverser le corps humain pour être mesurée, ainsi l'oxymètre doit être placé sur le doigt, ou le lobe de l'oreille, permettant au capteur d'être compact, simple d'utilisation et relativement confortable. Afin que le capteur ne soit pas influencé par la lumière ambiante, l'oxymètre ne mesure pas en continu et utilise une lumière modulée à 200 Hz. Dans l'oxymètre, un microcontrôleur a pour rôle de réaliser l'acquisition du signal analogique, d'enregistrement de plusieurs dizaines d'échantillons et de calculer la valeur de  $SpO_2$ , en environ une seconde. Les résultats de mesure sont ensuite transmis à un périphérique via BLE.

## 5.4. RÉSULTATS

### 5.4.1. DESCRIPTION DES PARTICIPANTS

Les données recueillies par le projet à la date du 1er avril 2021 se composent de données issues de la population générale. 86 individus ont participé à l'étude, dont 64 ont complété le questionnaire ; 34 femmes et 30 hommes, dont l'âge varie entre 26 et 71 ans, avec un âge médian de 46 ans. 42% des participants au questionnaire ont un niveau d'éducation équivalent ou supérieur à un Master (Bac +5). Parmi ces participants, 25 ont été équipés de capteurs de paramètres de santé, afin d'être les témoins pour le groupe de participants ciblés par l'étude épidémiologique. 19 d'entre eux ont participé à deux occurrences, permettant de recueillir deux expositions individuelles espacées de plusieurs mois, auxquels s'ajoutent 5 participants ayant participé à deux occurrences, mais sans capteurs de paramètres de santé. En finalité, le serveur POLLUSCOPE a donc enregistré à cette date 110 enregistrements d'exposition individuelle, contenant un individu, un kit de capteur et une campagne de mesure correspondant à une période donnée.

### 5.4.2. DISPARITÉS LIÉES AU RECRUTEMENT

Bien que les participants soient tous issus de la population générale, deux biais sont introduits par ce processus de recrutement. Ce sont deux populations distinctes qui sont ciblées. Il existe un certain nombre de disparités directement introduites par le processus de recrutement. Premièrement, parce que la couverture géographique de la campagne va être limitée dans le cas du recrutement sur la base du volontariat ; la population recrutée va inévitablement résider autour du lieu de recrutement, tel que présenté.

Le recrutement par volontariat, dans le cas de la campagne VGP, comme son nom l'indique, a été effectuée par des affichages dans Versailles Grand Parc. Les données recueillies sont affichées sous forme de carte de chaleur sur la Figure 12, en fonction du type de recrutement. On peut ainsi observer que la majorité des données GPS recueillies sont situées proche du point de recrutement ; proche de la mairie de Versailles.

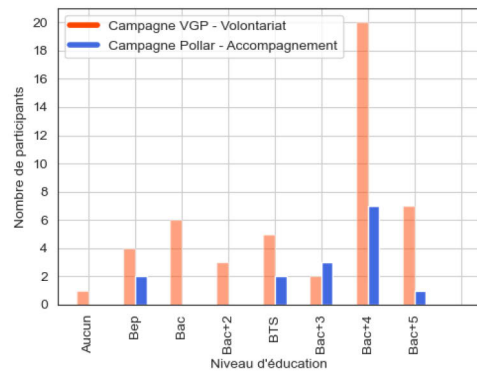


FIGURE 11 : Niveau d'éducation par processus de recrutement

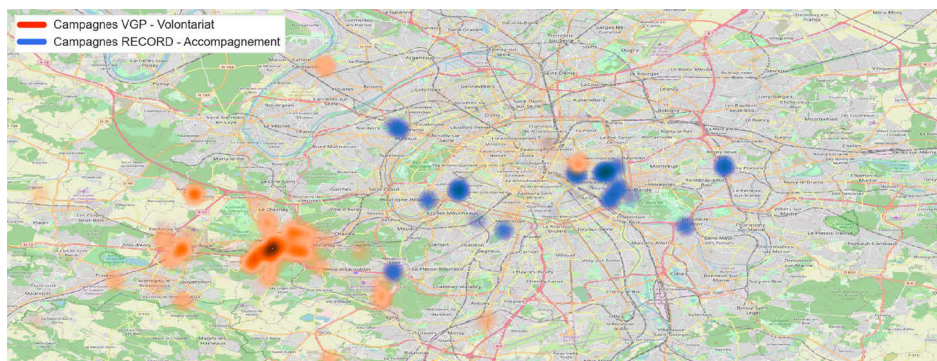


FIGURE 12 : Localisation des volontaires en fonction de la méthode de recrutement

## 5.5. EXPLOITATION DES DONNÉES DE SPIROMÉTRIE

La spirométrie nécessite un effort physique considérable, surtout lorsqu'elle est répétée plusieurs fois. Les tests de spirométrie sont normalement réalisés sous l'œil d'un professionnel de santé. Ce n'est cependant pas le cas avec des capteurs portatifs utilisés en conditions réelles. Au sein du projet POLLUSCOPE, le rôle des formateurs est donc fondamental ; ils assurent premièrement la transmission des informations permettant aux volontaires de réaliser le test spirométrie puis contrôlent que le volontaire les a bien assimilés, en lui requérant de réaliser une ou plusieurs spirométries sous sa surveillance.

Le manque de données ou leur faible qualité furent un handicap majeur pour l'exploitation des données de spirométries et sont majoritairement liées au fait que les volontaires réalisent les spirométries en parfaite autonomie. Il n'y a pas de contrôle immédiat de la qualité de la spirométrie, ni de rappel automatisé dans le cas où les volontaires n'effectuent pas le nombre de spirométries minimum requis. Pour répondre à ces difficultés, la société MIR [129], fabricant des spiromètres, propose deux mesures ; l'envoi de courriers électroniques après un délai que l'on peut modifier sur l'interface de gestion en ligne et une vérification de la bonne qualité du souffle effectuée localement sur l'application, qui sollicitera automatiquement le participant à répéter le test dans le cas où la spirométrie aurait été très mal effectuée. Dans le projet POLLUSCOPE, la fonctionnalité d'alertes par courriers électroniques a été exploitée mais sa fonction a été détournée. D'une part pour des questions organisationnelles ; la personne effectuant la distribution des kits n'était pas formée ni autorisée à l'utilisation de cette interface de gestion en ligne et n'était donc pas en mesure de renseigner les adresses e-mail des volontaires. D'autre part, pour des questions d'accessibilité ; l'âge moyen des volontaires équipés de spiromètres est de 57 ans, rendant donc incertaine l'utilité des alertes par courriers électroniques, puisque pour être effectives, elles nécessitent que le participant consulte ses courriers électroniques plusieurs fois par jour. Un contrôle partiel a donc premièrement été effectué manuellement durant les campagnes de mesure pour s'assurer que les volontaires équipés de spiromètre effectuent des séries de trois spirométries ou plus, au minimum trois fois par jour. Deuxièmement, les courriers électroniques d'alertes ont été dirigés vers des boîtes e-mail portant les identifiants des spiromètres, puis systématiquement redirigés vers une unique boîte mail, permettant au responsable de la campagne d'être notifié à la place des participants, tout en conservant leur anonymat. Les participants ont ainsi été contactés par téléphone en cas d'oublis successifs.

Malgré ces précautions, un nombre important de spirométries ne sont pas correctement effectuées. Une disparité importante existe en termes de qualité et quantité des données produites par les participants ayant utilisé les spiromètres portatifs. Il est donc nécessaire, dans un premier temps, d'identifier les volontaires n'ayant pas eu une utilisation rigoureuse du spiromètre, puis de filtrer les spirométries dans le but de rejeter celles qui ont été mal effectuées. Un processus de sélection des données de spirométries a donc été développé, dans le but final d'être en capacité observer si un lien existe entre l'exposition individuelle aux particules fines, qui est produite de manière autonome par les différents capteurs de la qualité de l'air, tandis que les données de santé sont issues des tests de spirométrie.

La consigne transmise par le formateur fut de réaliser au minimum trois sessions de spirométries par jour ; matin, midi et soir. Ces sessions de spirométrie doivent être composées d'un minimum de trois spirométries. Le nombre de trois spirométries par session nous permettent d'identifier à posteriori, pour chaque session, la spirométrie dont les résultats sont les plus satisfaisants. Chaque semaine d'acquisition contient donc environ trois sessions par jour. Au niveau de chaque session de spirométrie, nous effectuerons une sélection du test à exploiter. Au niveau des participants, nous identifierons et exclurons de notre analyse ceux ayant produit des données de qualité insuffisante, ou bien un nombre insatisfaisant de données. Le nombre de participants équipés de spiromètres est initialement réduit. Des efforts ont donc été effectués pour permettre à ces données d'être malgré-tout exploitables de façon objective.

Nous avons utilisé les module Python Pandas [94] pour la manipulation de données, le module Statsmodels [130] pour réaliser puis évaluer les modèles de régressions linéaire multiple et enfin le module Matplotlib [95] pour produire les différents graphiques.

#### 5.5.1. ÉVALUATION DU TAUX D'UTILISATION

Nous avons dans un premier temps cherché à déterminer quels sont les participants les plus engagés à l'utilisation du spiromètre dans l'objectif d'exclure les participants ayant produit un nombre insuffisant de données, ou ayant produit des données de manière trop stochastique pour qu'elles soient exploitables. Les participants ayant utilisé les spiromètres sont tous issus de la cohorte RECORD [109]. Deux campagnes de mesures composent ces données, désignées par RECORD\_1, de juin à septembre 2019 et RECORD\_2, de février à mars 2020. Sept des vingt individus ayant réalisé les spirométries, soit 35%, ont participé aux deux campagnes et donc comptabilisent deux semaines d'acquisition. A partir des données produites par ces participants, nous avons observé si le spiromètre a été utilisé de manière plus régulière durant la seconde campagne.

Ainsi, nous avons établi un indicateur ; le taux d'utilisation, calculé à partir de la formule suivante (en %) :

$$\text{Taux d'utilisation} = \frac{1}{3} \left( \frac{\text{Nombre de sessions}}{\text{Jours} \times 3} + \int_{n_{j+1}}^{n_j-1} \left( \frac{\text{Sessions}}{n \text{ Jour entier} \times 3} \right) + \frac{\text{Nombre d'acquisitions}}{\text{Jours} \times 9} \right)$$

Il tient compte du nombre de sessions de spirométrie effectuées relativement à la durée d'acquisition, de la régularité de l'utilisation (le respect des 3 sessions par jour) et du nombre total de spirométries effectuées. Il est possible que le volontaire effectue plus de trois sessions par jour, ou davantage que le maximum requis de 3 spirométries par session. C'est pourquoi il est nécessaire de tenir compte de la régularité d'utilisation et non pas seulement du total des sessions réalisées. La qualité des spirométries effectuées n'est en revanche pas une variable considérée pour le calcul du taux d'utilisation.

Le nombre de jours réel d'utilisation des capteurs par le volontaire est une variable centrale pour le calcul du taux d'utilisation. Cette information est fournie manuellement lors de la transmission des données recueillies au serveur du projet. Le format de la date fait que l'information transmise ne contienne pas d'heures. Cette information telle qu'elle nous a été fournie ne pouvait donc pas représenter précisément le nombre de jours où le volontaire a été

en capacité d'utiliser les capteurs. De plus, il est possible que ces dates aient été annotées de façon décalée (avec un jour supplémentaire par ex.) par simple convenance. Exploiter cette information aurait introduit un biais de plusieurs heures, vis-à-vis des dates réelles de port. Ce ne serait pas un biais significatif dans le cas où les durées d'acquisition seraient plus importantes, cependant la durée moyenne déclarée des semaines d'acquisition RECORD est d'environ 8 jours. Pour établir correctement le nombre de jours de port réel des capteurs, nous avons donc choisi d'observer la durée en jours comprise entre la première et la dernière acquisition de données de spirométries. Pour estimer le taux d'utilisation, nous avons réalisé la moyenne de trois paramètres s'exprimant en pourcentage

Dans notre population, nous avons observé un taux d'utilisation moyen par semaine d'acquisition de 85% et un minimum de 54%. Indépendamment des participants, 17% des semaines d'acquisition avaient un taux de participation de 100%. Les deux participants dont le taux d'utilisation est inférieur à 60% ont été exclus de l'analyse. Aucune relation n'a pu être établie entre le taux moyen d'utilisation, ou l'évolution du taux d'utilisation entre les deux campagnes, en fonction des données socioéconomiques issues des questionnaires.

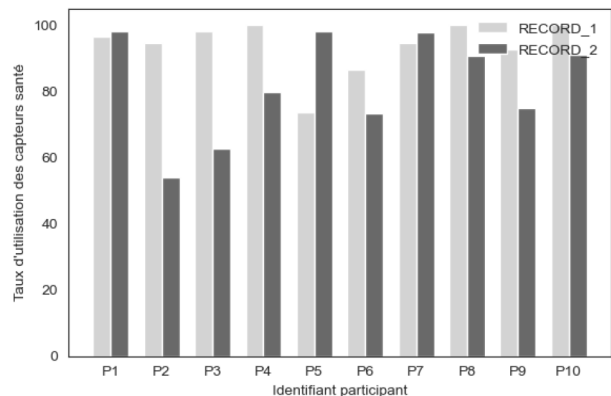


FIGURE 13 : Taux d'utilisation durant l'acquisition

L'évolution du taux d'utilisation pour les participants à deux campagnes est présentée sur la Figure 13. Le taux d'utilisation moyen est de 93.7% pour la première semaine, contre 82.1% durant la seconde. On relève donc, au sein de ce groupe de participants, une baisse de 11.6% du taux moyen d'utilisation entre les deux campagnes. On présume que cette baisse est le résultat du désintéressement progressif des participants à l'utilisation des capteurs.

### 5.5.2. SÉLECTION DES SPIROMÉTRIES

Il est bien connu que les données produites par le test de spirométrie peuvent être inexploitable dans le cas où le participant n'effectue pas le test correctement ; si le souffle du sujet est hésitant, s'il n'est pas immédiatement à son débit maximum, ou encore si l'expiration forcée n'est pas maintenue plusieurs secondes à la fin du test. Idéalement, les spirométries doivent être effectuées sous l'œil d'un spécialiste et les résultats doivent être contrôlés individuellement. Dans notre cas, il a été nécessaire d'automatiser cette tâche afin de faire face à la disparité de la qualité des acquisitions et à l'impossibilité de sélectionner manuellement les spirométries valides pour chaque session. La première session a été systématiquement rejetée puisque les volontaires apprennent généralement à souffler sous l'œil d'un formateur qui participe également à montrer l'utilisation du spiromètre.

Une session d'acquisition de spirométrie contient un ou plusieurs tests de spirométrie et possède une date de début et une fin. Considérant qu'il est possible qu'un participant effectue

des spirométries de manière stochastique, il est alors difficile de déterminer quand une session de spirométrie prend fin. Considérant que la fonction respiratoire n'évolue pas rapidement, nous avons choisi de regrouper ces acquisitions par heure. Ainsi, dans l'éventualité où plusieurs spirométries sont réalisées dans l'heure qui suit le premier test, elles seront alors incluses dans la même session et seule l'acquisition répondant au mieux à nos critères sera exploitée. Cette contrainte permet en revanche aux participants de réaliser davantage de sessions que les 3 requises par jour. Les sessions où le participant a effectué moins de trois spirométries n'ont pas été prises en compte et ont été identifiées comme ayant un nombre d'acquisitions insuffisantes.

Pour qu'une spirométrie soit valide, il faut qu'elle soit reproductible ; les autres spirométries acquises durant la même session doivent comporter des valeurs similaires. Si une session contient un nombre important de spirométries, il est alors plus facile de discriminer les spirométries divergeant de la tendance. Nous considérons reproductible une spirométrie dont les valeurs ne sont pas filtrées par un algorithme permettant d'exclure les valeurs aberrantes. Dans notre cas, nous avons choisi d'utiliser l'algorithme présenté dans le paragraphe 8.7.3. L'exploitation de tout autre algorithme excluant les valeurs aberrantes permettrait d'obtenir des résultats similaires. Cette discrimination des valeurs aberrantes a donc été appliquée aux valeurs de CVF et de VEMS, pour les données de chaque session, sans tenir compte des antécédents.

A partir des valeurs de CVF reproductibles, nous avons déterminé la CVF médiane et avons rejeté les acquisitions dont les valeurs de CVF sont en dehors d'une plage de plus ou moins 20% autour de cette médiane. Enfin, parmi les spirométries répondant aux quatre critères évoqués ci-dessus (plus de trois spirométries par session, CVF et VEMS reproductibles et plus ou moins 20% autour de la CVF médiane), seule la spirométrie portant la valeur de VEMS la plus élevée a été sélectionnée ; là où le participant aura donc réalisé l'effort maximum au moment de l'expiration. A l'exception d'un participant où 78% de ses sessions ne comportaient qu'une unique spirométrie, 94,1% des sessions effectuées par nos participants ont pu être exploitées car elles contenaient un minimum de 3 acquisitions.

La Figure 14 est une représentation graphique du résultat de la sélection des spirométries, effectuée dans le cas d'un participant pour une semaine d'acquisition. Les valeurs de la CVF, du VEMS, ainsi que le coefficient de Tiffeneau (VEMS/CVF) sont représentés individuellement. Pour mémoire, le coefficient de Tiffeneau [128] est le rapport VEMS sur CVF.

Chaque spirométrie est classifiée et colorisée selon 6 catégories possibles, résultantes du processus de sélection décrit ci-dessus. Les sessions sont numérotées en bas de graphique. Comme ce serait le cas pour une série temporelle, l'écart visible entre chaque session est représentatif du temps écoulé entre chaque session. On peut remarquer par exemple la proximité temporelle entre les sessions 1 et 2, puis 18 et 19 ; elles sont donc traitées indépendamment car distantes de plus de 60 mn.

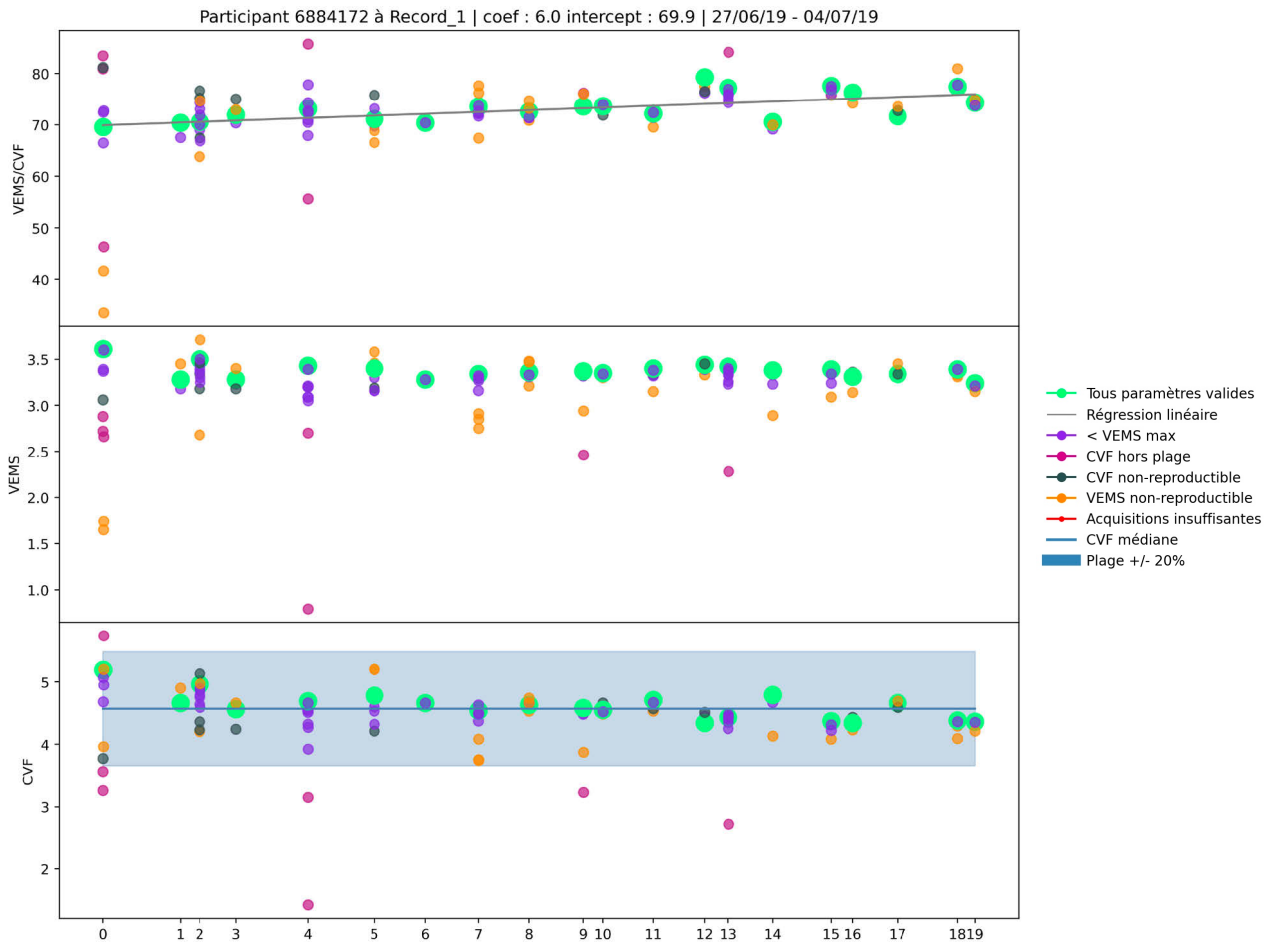


FIGURE 14 : Représentation graphique de la sélection des spirométries

L'utilisation de la visualisation graphique fut un outil indispensable lors du développement de l'automatisation de la sélection des spirométries. Cette représentation permet désormais de contrôler la cohérence de l'algorithme de sélection vis-à-vis des choix de qui seraient effectués dans le cas d'une sélection manuelle. Lorsque effectué dans une interface de développement (Spyder 5 [131], avec affichage via Qt 5 [132]), l'affichage est dynamique et permet d'observer le détail de chaque session. Notons que la session 2, sans agrandissement, est représentée par un amas de points au milieu duquel il est difficile de distinguer la classification colorimétrie des spirométries et donc de juger de la pertinence de la sélection.

Un agrandissement a été fait sur la session 2 (Figure 15). L'agrandissement permet d'observer le déroulement des acquisitions dans le temps et de lire plus aisément les niveaux relevés. Le nombre de spirométrie est élevé pour une session puisqu'elle comporte au total 13 spirométries, alors que 3 sont normalement suffisantes. Pour cette session, le nombre important d'acquisitions permet d'illustrer au mieux la classification automatique des spirométries. Les valeurs extrêmes de VEMS ont été classifiées comme non-reproductibles (en orange) et étaient effectivement éloignées du reste des autres acquisitions. Concernant la CVF, trois acquisitions

ont été classifiées comme non-reproductibles (en gris). Pour cette session, aucune acquisition ne dépasse le seuil de 20% (en bleu) autour de la CVF médiane. Enfin, 7 acquisitions de spirométrie sont reproductibles (en violet), mais seule une spirométrie sera sélectionnée ; celle portant la valeur de VEMS maximum (en vert).

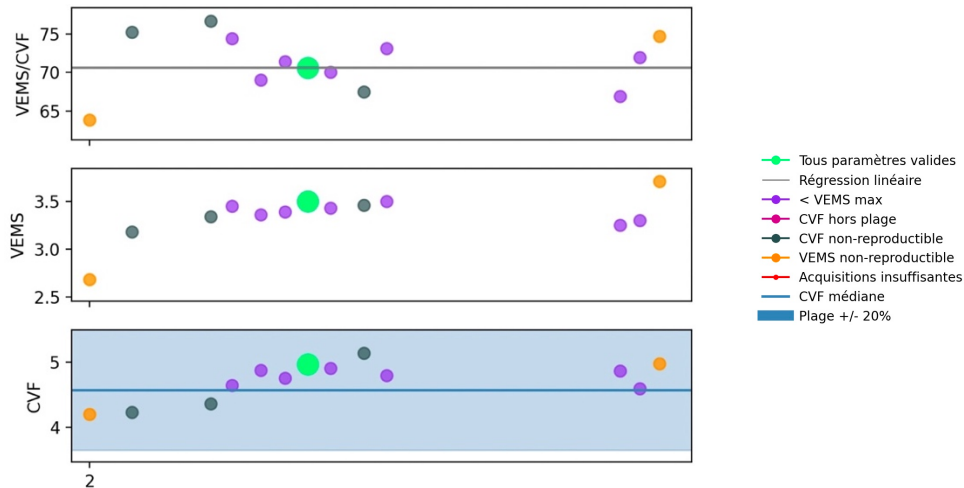


FIGURE 15 : Agrandissement sur une session d'acquisition

### 5.5.3. MISE EN RELATION DE LA FONCTION RESPIRATOIRE AVEC LES DONNÉES D'EXPOSITION À LA POLLUTION ATMOSPHÉRIQUE

Pour mettre en relation les valeurs de fonction respiratoire avec les données d'exposition à la pollution atmosphérique (PM, NO<sub>2</sub>, BC...), nous avons exploité la pente de la droite de régression linéaire du coefficient de Tiffeneau (en gris sur la Figure 14), produite à partir des spirométries sélectionnées, pour chaque participant. Pour ce faire, nous avons produit, pour chacune des sessions de spirométrie et pour chacun des polluants relevés, la moyenne de l'exposition individuelle précédant la session de spirométrie. Nous requérons arbitrairement que les capteurs relevant l'exposition individuelle à la pollution aient été en fonctionnement à minima durant 15% des périodes précédant chaque session. Les périodes ne respectant pas ce critère ne pourraient être représentatives de l'exposition réelle et resteront donc inexploitées. Similairement à ce qui a été précédemment réalisé pour les données de spirométrie, nous produirons une régression linéaire à partir des niveaux moyens d'exposition cumulés par session.

Pour montrer que la fonction respiratoire évolue conjointement avec les niveaux de pollution atmosphérique moyens, nous avons effectué une régression linéaire multiple permettant de décrire la pente de la droite de régression linéaire issu des tests de spirométrie (l'évolution de la fonction respiratoire), en fonction de l'ordonnée à l'origine (l'état initial de la fonction respiratoire) et de la régression linéaire réalisée sur les données d'exposition aux différents polluants (le niveau initial de polluants et son évolution durant la semaine d'acquisition).



Une semaine contenant un nombre limité de sessions ne doit pas permettre d'évaluer avec confiance le lien entre l'évolution de la fonction respiratoire et l'exposition individuelle aux polluants. Le nombre de sessions minimum requis durant chaque semaine d'acquisition est donc un paramètre supplémentaire qui va davantage restreindre les données exploitables.

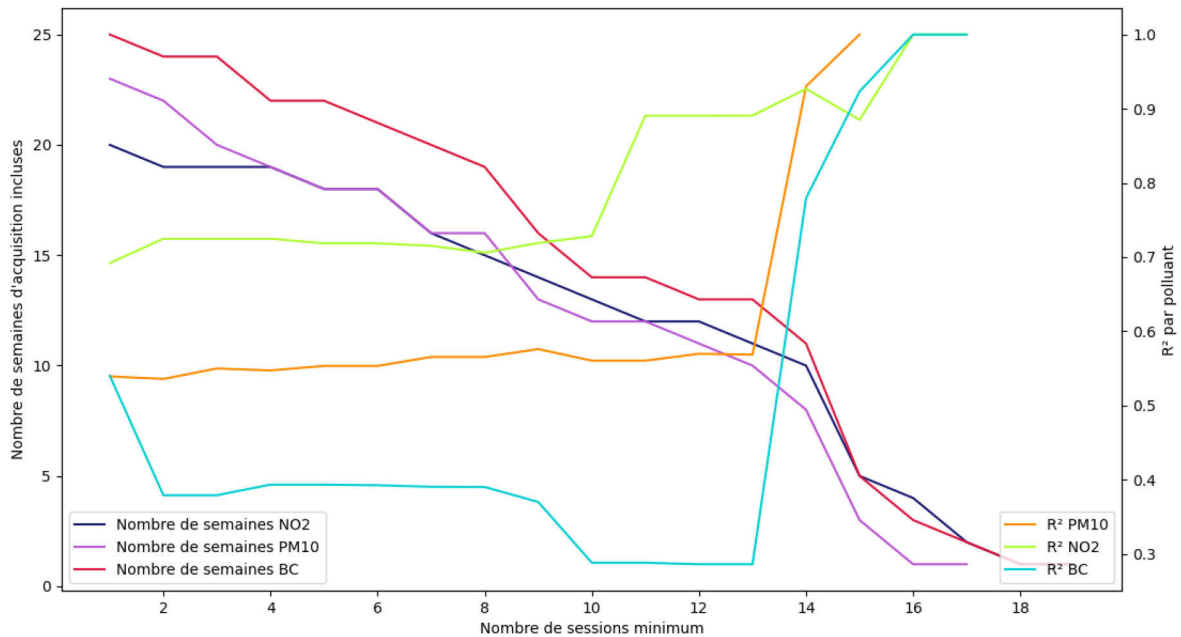


FIGURE 16 : Aide à la sélection du nombre de sessions requises

Sur la Figure 16 nous avons pour chaque type de polluant indiqué le nombre de semaines d'acquisition incluses en fonction du nombre de sessions minimum disponibles, ainsi que le coefficient de détermination ( $R^2$ ) issu de la régression linéaire multiple appliquée chaque polluant (en bleu, orange et vert) également en fonction du nombre de sessions minimum disponible. Naturellement, le nombre de semaines d'acquisition incluses décroît lorsque l'on requiert davantage de sessions d'acquisition. On observe que le coefficient de détermination augmente lorsque le nombre de semaines d'acquisition se rapproche de 0 ; l'augmentation du coefficient de détermination est désirable, mais il est nécessaire d'inclure un maximum de semaines dans le modèle pour qu'il soit représentatif de la réalité des acquisitions faites par l'ensemble des volontaires. Nous ne pouvons tenir compte des résultats lorsque le nombre de semaines d'acquisition incluses est proche de 0. Le nombre de participants ayant été équipés d'un capteur de BC est réduit du fait du coût important du capteur. Le nombre de semaines d'acquisition possiblement incluses est donc plus faible dans le cas du BC vis-à-vis des autres polluants. Nous estimons ainsi, dans le cas des données présentes, qu'un nombre de 7 sessions minimum requises est optimal ; nous exploiterons donc entre 16 et 20 semaines d'acquisition en fonction des polluants et conservons un coefficient de détermination relativement élevé ( $R^2 = 0.56$  pour les  $PM_{1, 2.5}$  et  $PM_{10}$ ,  $R^2 = 0.72$  pour le  $NO_2$  et  $R^2 = 0.39$  pour le BC).

#### 5.5.4. RÉSULTATS : RELATIONS ENTRE L'EXPOSITION À LA POLLUTION ATMOSPHÉRIQUE ET LA FONCTION RESPIRATOIRE

Nous avons modélisé l'évolution de la fonction respiratoire (pente de la droite de régression VEMS/CVF), vis-à-vis de leur fonction respiratoire initiale (ordonnée à l'origine VEMS/CVF), des niveaux de pollution initiaux (ordonnée à l'origine polluants) auxquels sont exposés les participants, au début des semaines d'acquisition exploitables et de leur évolution dans le temps (pente de la droite de régression polluants), au travers d'une régression linéaire multiple. Nous visualiserons uniquement le modèle dans le cas des données produites par le capteur de particules CANARIN et plus spécifiquement, dans le cas des PM<sub>10</sub>, dans le cas d'un minimum de 7 sessions requises par semaine d'acquisition.

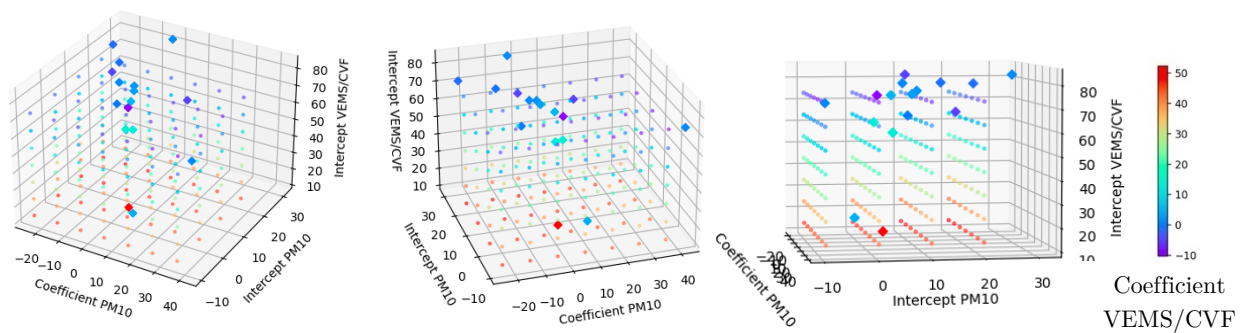


FIGURE 17 : Visualisation 3D du modèle

16 semaines d'acquisition, issues de 14 participants distincts, ont permis de créer le modèle représenté sur la Figure 17. Leurs valeurs individuelles sont représentées par un losange. Il est important de noter que dans le cas des PM<sub>10</sub> une corrélation négative significative existe entre la pente de la droite de régression et son ordonnée à l'origine ( $r = -0.68$ ,  $p = 0.003$ ) ; le niveau initial de polluants était généralement faible lorsque les niveaux de polluants (dans notre cas les PM<sub>10</sub>) ont augmenté durant la semaine d'acquisition. Les niveaux moyens de polluants, tout comme la fonction respiratoire, ne peuvent augmenter ou baisser indéfiniment.

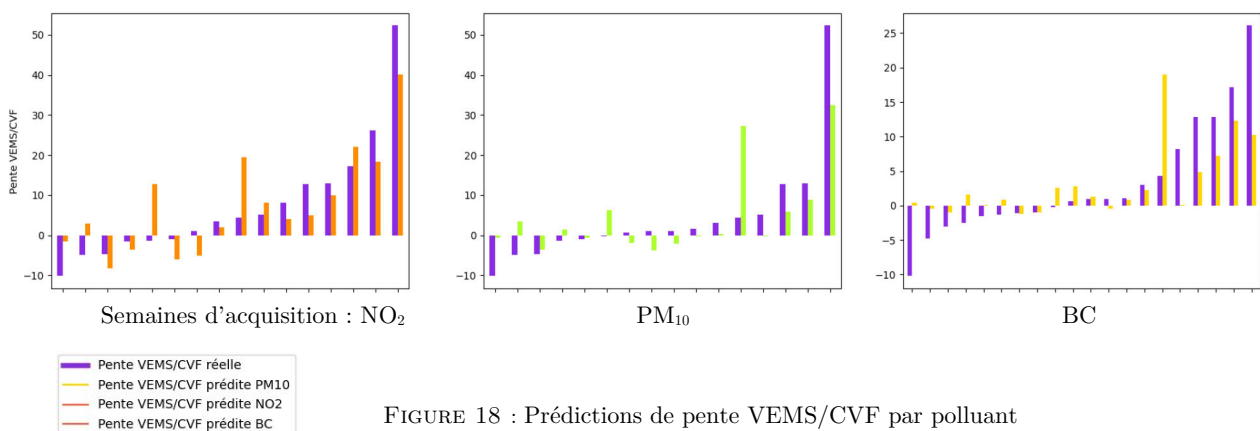
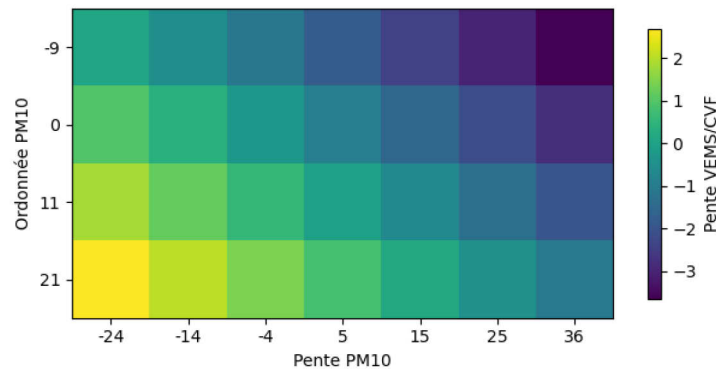


FIGURE 18 : Prédications de pente VEMS/CVF par polluant par semaine d'acquisition

Pour chacun des polluants, nous avons fourni au modèle les données ayant permis de le produire (Figure 18). On peut ainsi observer les valeurs réelles de pente VEMS/CVF, vis-à-vis des valeurs prédites à partir des valeurs réelles de pente et ordonnées à l'origine, pour chacun des trois polluants. Les valeurs de pentes VEMS/CVF sont ordonnées par ordre croissant, de gauche à droite. Les semaines d'acquisition exploitées diffèrent pour chacun des polluants.



Ordonnée à l'origine VEMS/CVF médiane (73,6)

FIGURE 19 : Vue en coupe du modèle PM<sub>10</sub>

La Figure 19 est une matrice 2D, représentant une vue en coupe du modèle, dans le cas des PM<sub>10</sub>, dans le cas de la valeur médiane de l'ordonnée à l'origine du coefficient de Tiffeneau (73,6). Seules trois variables sont donc présentées. On peut ainsi lire que la fonction respiratoire des volontaires tend à s'améliorer lorsque le volontaire est exposé initialement à des niveaux moyens élevés et que ces niveaux décroissent rapidement durant la semaine d'acquisition. Inversement, la fonction respiratoire se dégrade lorsque les niveaux moyens sont initialement faibles et augmentent rapidement durant la semaine d'acquisition.

## 5.6. DISCUSSION ET CONCLUSIONS

Nous avons préparé, traité, filtré, puis analysé des données issues de capteurs mixtes ; des capteurs de santé non-invasifs nécessitant une action de la part de l'utilisateur et des capteurs de la qualité de l'air fonctionnant en autonomie. La visualisation en nuage de points du modèle (FIGURE 17) montre que nous manquons de données pour pouvoir généraliser nos observations. Dans le cas où nous aurions à disposition les données de plusieurs centaines de participants et sur des durées plus longues, nous serions alors en mesure de produire des résultats plus pertinents. Nous pourrions alors évaluer la performance du modèle à partir de semaines d'acquisition n'ayant pas été exploitées pour construire ce même modèle, évitant ainsi le phénomène de sur-apprentissage. Deuxièmement, nous pourrions intégrer davantage de variables ; tel que l'âge des participants (issue du questionnaire), la mobilité des individus (issue des données GPS), ou les types d'environnements empruntés (issue des annotations).

Outre le faible nombre de participants équipés de spiromètres inclus dans le projet, le manque de données de spirométrie est majoritairement dû à facteur humain ; manque de formation à l'utilisation des spiromètres, surveillance de l'enregistrement des résultats

insuffisante, ou abandon de la part des volontaires. Il est également impossible d'exploiter les données de spirométries produites dans le cas d'un dysfonctionnement ou d'une mauvaise utilisation du ou des capteurs de la qualité de l'air. Le manque de données exploitable peut ainsi être tout autant dû au faible nombre de spirométrie exploitables qu'à l'absence de données relatives à l'exposition individuelle à la pollution de l'air. L'étape de filtrage des données, quand bien elle permet de traiter des données plus qualitatives (Figure 16), participe également à restreindre davantage le nombre d'acquisitions exploitables. En définitive, nous avons exploité entre 16 et 20 semaines d'acquisition, quand nous devrions théoriquement pouvoir traiter un total de 34 semaines d'acquisitions.

Dans le projet POLLUSCOPE, les durées d'acquisition sont d'environ une semaine et ne permettent pas d'observer une augmentation puis une baisse de la qualité de la fonction respiratoire des volontaires. Ainsi il n'a pas été possible d'intégrer une variable temporelle au modèle. Autrement dit, sans connaître les données d'exposition aux polluants antécédents au début de la semaine d'acquisition, il est difficile de mettre en évidence les délais et la vitesse de récupération ou de dégradation de la fonction respiratoire. La courte durée des périodes d'acquisition peut également introduire un biais du fait que les volontaires vont se familiariser à l'effort physique requis pour la spirométrie et donc produire un score VEMS/CVF plus élevé après quelques jours ; nous observons en effet une augmentation du coefficient de Tiffeneau pour en moyenne 55% des périodes d'acquisitions pour le modèle issu des données de BC et 63% dans le cas des PM et NO<sub>2</sub>.

Le processus que nous avons présenté a permis de décomplexifier l'information granulaire recueillie par ces capteurs hétérogènes et de produire des résultats, malgré le handicap qu'implique la faible quantité de données à disposition. Cette approche, que ce soit dans le filtrage ou la sélection des données, peut être adaptée ou transposée à des projets de recherche soulevant des problématiques similaires.

## 6. PROJET POLLAR : ÉTUDE DE LA RELATION ENTRE L'EXPOSITION À LA POLLUTION ATMOSPHÉRIQUE ET L'APNÉE DU SOMMEIL

Nous avons présenté comment évaluer les effets court-terme de la pollution de l'air au travers l'étude du syndrome d'apnées-hypopnées du sommeil. Les capteurs CANARIN II et CANARIN nano ont permis pour la première fois de comparer les résultats de la polysomnographie avec les données granulaires sur la pollution de l'air obtenues pendant et après le sommeil. Ces résultats donnent des perspectives sur le futur des capteurs de la pollution de l'air dédiés à la recherche médicale, en soulignant la nécessité d'obtenir davantage de données contextuelles afin de mieux déterminer l'exposition environnementale qui y est liée.

Le projet POLLAR a été financé par l'EIT (European Institute of Innovation and Technology, subvention 18354 - Impact of POLLution on Asthma and Rhinitis), Fondation partenaire FMC-VIA-LR. Les données produites par la polysomnographie et les données relatives à la planification du déploiement des CANARIN au sein de l'étude POLLAR ont été fournies par l'équipe du CHU de Grenoble, par le biais d'un fichier Excel. Aucun post-traitement n'a été appliqué pour compenser la perte de données ou les erreurs des utilisateurs ; toutes les figures et tous les tableaux ont été produits à partir de données brutes non traitées.

### 6.1. OBJECTIFS DU PROJET

Le but premier du projet POLLAR fut d'étudier les interactions possibles entre la pollution atmosphérique, le sommeil et les maladies allergiques. Pour répondre à cette problématique, l'axe principal du projet fut le développement d'une application mobile basé sur le crowdsensing et la modélisation des pollens présents dans l'air au niveau européen. Les utilisateurs de l'application, en échange des informations sur la présence de pollens dans l'air, étaient conviés à renseigner la gravité de leur réaction allergique et la qualité de leur sommeil. Grâce à ces données, en plus de pouvoir cartographier les résultats produits, il a été possible d'étudier à plus large échelle la causalité entre les vagues de pollens, les symptômes allergiques et les conséquences sur la qualité de vie.

### 6.2. POPULATION ET MÉTHODES

L'étude clinique EIT-HEALTH POLLAR [101] a débuté en mars 2019, avec pour objectif d'estimer l'impact de l'exposition à la pollution atmosphérique et au pollen sur la rhinite, l'asthme et le sommeil. Les données sur le pollen ont été obtenues par le biais du Réseau National de Surveillance Aérobiologique (R.N.S.A., réseau français d'aérobiologie [133]). L'impact sur le sommeil a été étudié au Centre hospitalier Universitaire (CHU) de Grenoble Alpes où des patients souffrant d'apnée du sommeil ont été recrutés pour effectuer, à deux périodes différentes, une



FIGURE 20 : Sujet instrumenté avant polysomnographie

polysomnographie (PSG), suivie par une période de plusieurs jours de port du CANARIN Nano ou du CANARIN II pour estimer l'exposition à la pollution atmosphérique. La PSG est une étude complète du sommeil qui exige que le sujet reste à l'hôpital pendant une nuit équipé d'appareils de surveillance, voir Figure 20. Le principal résultat de la PSG pris en compte dans l'étude est la gravité de l'apnée du sommeil, donnée par l'indice d'apnée-hypopnée (IAH). Les données produites par les CANARINs, dans le cas du projet POLLAR, ont été obtenues à partir du site web du projet CANARIN [107], via une page web protégée par un mot de passe.

### 6.3. EXPOSITION AUX PARTICULES FINES EN RELATION AVEC L'APNÉE DU SOMMEIL

LaM Figure 21 montre l'exposition horaire moyenne aux particules pour les sujets de l'étude POLLAR atteints d'apnée du sommeil (21 sujets avec un IAH > 5, en bleu) et les sujets avec un IAH normal (22 sujets avec un IAH moyen  $\leq 5$ , en orange). Dans les deux groupes, la valeur supérieure des diagrammes montre à quel point certains sujets ont été exposés à des valeurs anormalement élevées de PM. Dans cette figure, les valeurs aberrantes extrêmes ont été dissimulées, car un participant enregistrerait systématiquement des valeurs excessivement élevées de PM durant la nuit, avec des échantillons de  $PM_{10}$  supérieurs à  $1000 \mu\text{g}/\text{m}^3$ , excédant la plage de fonctionnement du module  $PM_{10}$ . De telles valeurs peuvent être enregistrées par les OPC lorsqu'ils sont exposés à la vapeur des e-cigarettes, mais aucune autre information n'est disponible pour contextualiser les données. Les valeurs médianes sont très proches entre les deux groupes. Entre 9 h et 21 h, les valeurs de  $PM_{10}$  du 75e percentile sont plus élevées dans le groupe touché par l'apnée du sommeil.

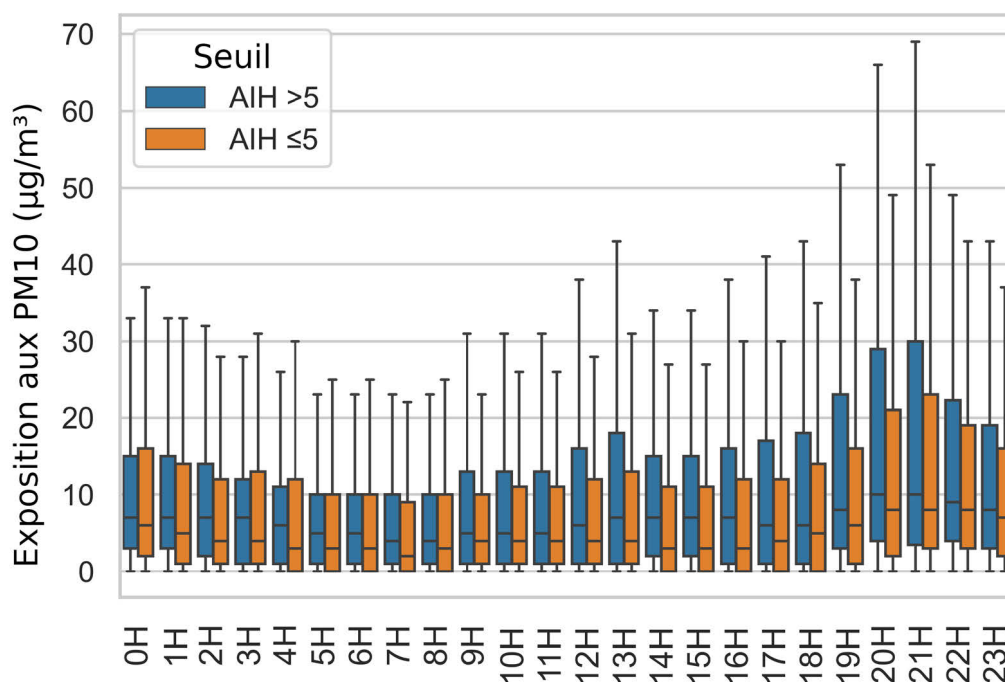


FIGURE 21 : Concentrations horaires de  $PM_{10}$  selon le groupe de l'indice d'apnée-hypopnée (IAH).

#### 6.4. DISCUSSION ET CONCLUSIONS

POLLAR a montré avec succès, dans le cadre d'une étude pilote, comment les capteurs personnels peuvent offrir de nouvelles possibilités pour la recherche en matière de santé publique. L'étude a rencontré des difficultés techniques au cours des premiers mois ; les prototypes du CANARIN nano n'étaient pas équipés de stockage flash, ce qui a entraîné une certaine perte de données lorsque aucune couverture 3G n'était disponible. L'option optimale aurait été de conserver une base de données stockée de manière décentralisée et constamment mise à jour, en plus du service en ligne, mais cela n'était pas techniquement possible à ce moment-là en raison des limites matérielles de la carte de prototypage choisie. L'autorisation légale d'enregistrer des données géo-localisées n'ayant pas été obtenue, le GPS a dû être désactivé à distance. La conception de l'étude visait à évaluer l'exposition personnelle à la pollution de l'air pendant les jours suivant le dépistage du PSG a rendu impossible de savoir si l'exposition à la pollution durant les jours précédant au dépistage entraînait une dégradation de la qualité du sommeil. La comparaison des deux périodes à l'aide d'une régression linéaire ne donne aucune corrélation entre l'exposition moyenne aux particules et l'AHI. Une telle analyse pourrait donner des résultats intéressants si les données d'exposition étaient enregistrées préalablement à la PSG.

## 7. BANC D'ESSAI MOBILE POUR L'ÉVALUATION DE LA DISPERSION

Les récents développements dans la miniaturisation des capteurs de la qualité de l'air offre aux épidémiologistes des maladies respiratoires de nouveaux outils pour évaluer l'exposition individuelle à la pollution atmosphérique et les risques sur la santé qui y sont associés ; les capteurs portatifs mesurant la qualité de l'air. Pour une flotte de capteurs visant à être déployée, il est essentiel d'être en mesure de pouvoir contrôler et valider individuellement le bon fonctionnement des capteurs, en amont comme postérieurement à leur utilisation. Dans le cas de capteurs neufs, n'ayant donc jamais été en fonctionnement, ce contrôle permet de s'assurer qu'ils répondent aux spécifications du constructeur. Dans le cas de capteurs ayant déjà été utilisés, ce contrôle permet alors de vérifier l'intégrité des modules ou micro-capteurs réalisant la mesure physique. En effet, les performances de ces derniers pourraient être dégradées, soit par un usage abusif antérieur qui aurait altéré de manière irrémédiable le capteur, soit dans le cadre du vieillissement normal des composants. Si un comportement anormal est observé, des mesures correctives peuvent alors être appliquées par post-traitement des données acquises ; soit en écartant les données sujettes à une dispersion importante, soit en intégrant une variable d'incertitude, ou une correction.

Il existe donc, au sein d'une flotte de capteur de la qualité de l'air exposés aux mêmes conditions, une dispersion entre les différentes mesures produites. Nous cherchons ainsi à observer, comprendre, quantifier et à idéalement être en mesure de caractériser cette dispersion. De par notre participation dans le développement et le déploiement de plusieurs flottes de capteurs portatifs, nous avons été d'une part, soumis aux obligations logistiques et prévisionnelles qu'un contrôle du bon fonctionnement des capteurs requiert et d'autre part, nous avons pu observer la présence de ce phénomène de dispersion ainsi que les incertitudes qui l'accompagnent. Les outils et infrastructures nécessaires à ce contrôle ne sont pas accessibles au public et ne peuvent être immédiatement disponibles. Souvent, elles ne sont pas mobiles et ne permettent donc pas d'exposer les capteurs en test à de multiples compositions de la pollution de l'air.

Pour répondre à ces problématiques, nous avons imaginé, évalué par simulation numérique, puis conçu et utilisé, un banc d'essai mobile à bas coût, permettant d'évaluer la dispersion présente dans un set de capteurs portatifs de la qualité de l'air, en les exposant à des environnements représentatifs des conditions réelles d'utilisation. Un certain nombre de paramètres ont été déterminés par des simulations numériques en amont de la fabrication du banc. Cependant, afin de permettre une meilleure interprétation des résultats de cette simulation, nous présenterons premièrement le principe de fonctionnement et le processus de fabrication. Ensuite, nous présenterons des résultats de la simulation numérique ayant permis d'établir la nécessité d'implémenter un système de diffusion. Dernièrement, nous analyserons les données produites, le comportement de la dispersion entre les différentes périodes de mesure, notamment vis-à-vis des conditions météorologiques.



## 7.1. OUTILS LOGICIELS

La modélisation 3D des composants du banc d'évaluation a été réalisée sur le logiciel de conception assistée par ordinateur (CAO) Fusion 360. La licence étudiante étant entièrement gratuite. La simulation des fluides fut réalisée sur le logiciel Ansys® Academic Research computational fluid dynamics (CFD) [134], dont la location de licence nous a été attribuée à titre gracieux. Cette simulation a permis de déterminer le placement des capteurs et l'espacement des trois sections, ainsi que de s'assurer de l'homogénéité de la vitesse du flux d'air à l'intérieur du banc.

## 7.2. NOMENCLATURE

Ci-dessous la liste indicative détaillée des composants, outillages et consommables ayant été nécessaires pour la réalisation du banc d'évaluation. Cette liste doit être modulée en cas de reproduction du banc, notamment concernant les dimensions de l'armoire électrique, des plaques et des moyens utilisés pour l'alimentation des capteurs.

### COMPOSANTS PRINCIPAUX :

- x1 Armoire électrique en ABS avec fermeture à clef (50 x 70 x 25 cm)
- x3 Plaques en polycarbonate ou acrylique (3 à 5mm d'épaisseur, 40 x 60 cm)
- x9 Tiges filetées taille M6
- x20 Ecrous en acier taille M6
- x18 Rondelles taille M6, diamètre 4 cm
- x2 Ventilateurs 12V Noctua NF-P12
- x6 Filtres anti-poussières pour ordinateur
- x3 Concentrateur USB alimenté en 12 V(16 ports)
- x48 Collier de serrage réutilisables en nylon
- x1 Transformateur 230V AC vers 12V DC 15 A
- x1 Extension d'alimentation 230V

### OUTILLAGE :

- Outil rotatif multifonction
- Perceuse avec variateur de vitesse
- Imprimante 3D à dépôt de filament (FDM)
- Pistolet à cartouche
- Fer à souder

### CONSOMMABLES :

- x3 Bobines de filament ABS ou ASA (2 à 3kg)
- x1 Cartouche de mastic imperméabilisant
- x4 Disques diamanté
- x2 Forêts taille M6 bois et métal
- x1 Tube de colle époxyde (25 ml)
- Étain et Gaine thermorétractable



FIGURE 22 : Environnement de test en air intérieur (1<sup>er</sup>)

### 7.3. PRINCIPE DE FONCTIONNEMENT ET FABRICATION

Lorsque le banc d'évaluation mobile a été imaginé et avant tout processus de simulation ou de fabrication, la contrainte technique majeure que nous avons considérée, fut que l'ensemble des matériaux et composants nécessaires doivent être abordables et aisément trouvables dans le commerce. Cette contrainte découle de la nécessité que ce banc d'évaluation soit peu coûteux et facilement reproductible. Similairement, sa fabrication doit être aisée ; avec un nombre d'opérations et de composants nécessaires réduits. L'usinage et l'assemblage ne doivent pas nécessiter d'outillage spécifique qui ne pourrait pas être facilement disponible (tel qu'une machine de découpeuse laser, par exemple). Dans notre cas, il nous a été malgré-tout nécessaire d'utiliser une imprimante 3D pour produire certaines pièces. Cet outil est néanmoins abordable et de plus en plus répandu.

Le composant central du banc est une armoire électrique en ABS résistante aux éclaboussures, avec fermeture à clef, dans laquelle seront disposés l'ensemble des capteurs à évaluer. Plusieurs perçages sont ensuite effectués et des bouches d'aération et diffuseurs sont ajoutés, permettant d'assurer le renouvellement continu de l'air tout en préservant l'intégrité des capteurs face aux risques extérieurs.

L'armoire électrique s'intégrant facilement dans l'environnement urbain, elle permet ainsi de déployer le banc d'évaluation en environnement extérieur, tout en limitant le risque d'interactions avec le public ; le contenu du banc n'est ni visible ni directement accessible. Sur les côtés de l'armoire électrique, plusieurs perçages doivent être réalisés, permettant l'entrée de l'air extérieur dans la cavité du banc. Nous recommandons d'utiliser un outil rotatif équipé d'un disque diamanté qui permettra de découper aisément le corps de l'armoire. Ces aérations compromettent la capacité du boîtier à protéger les capteurs de la pluie. Pour maintenir un certain niveau de protection, nous avons dessiné et produit par des bouches d'aération protégeant de la pénétration de la pluie et n'obstruant pas l'entrée d'air plus que nécessaire. Ces pièces ont été imprimées en 3D par procédé dépôt de matière fondue (FDM). L'une des contraintes techniques les plus importantes dans le cadre de l'impression FDM est la surface d'impression. Cette surface définit la taille maximale des pièces imprimées. Dans notre cas, l'impression bouche d'aération requiert la quasi-totalité de la surface d'impression disponible. Le modèle de l'imprimante utilisé est « Prusa Mini » ; il est très répandu en Europe, le volume maximal d'impression est de 18 cm<sup>3</sup>. Les bouches d'aération ont été produites en acrylonitrile butadiène styrène (ABS) ou acrylonitrile styrène acrylate (ASA) ; ces polymères thermoplastiques ont de meilleures propriétés vis-à-vis de l'acide polylactique (PLA) qui est très souvent utilisé pour l'impression 3D FDM. L'ABS ou l'ASA sont davantage résistants aux chocs, aux chaleurs importantes, ainsi qu'à la décoloration due à l'exposition aux rayons ultraviolets.



FIGURE 23 : Environnement de test en air extérieur  
(2<sup>ème</sup>)

Entre le corps de l'armoire et chaque bouche d'aération deux pièces supplémentaires sont introduites. Premièrement, un filtre à poussière destiné aux ordinateurs de bureau. Ces derniers sont facilement trouvables dans le commerce, perméables aux gaz et particules fines et permettront d'arrêter les insectes ou autres corps étrangers. Deuxièmement, un diffuseur permettant d'assurer l'homogénéité du renouvellement de l'air dans le banc en dirigeant et modérant la vitesse du flux d'air auxquels les capteurs sont exposés. Les filtres et diffuseurs sont placés sur la surface intérieure du corps de l'armoire et sont maintenus à l'aide de vis et écrous de diamètre 6 mm, qui permettent également de maintenir solidement les bouches d'aération à l'extérieur du banc. Une colle permettant d'imperméabiliser l'assemblage, de type mastic, résistante aux UV, a été appliquée aux points de jonction de chacune des pièces permettant de se prémunir contre les infiltrations d'eau dans le cas où le banc d'évaluation serait exposé à la pluie.

L'air dans la cavité du banc est continuellement renouvelé grâce à une ventilation forcée réalisée par deux ventilateurs fixés sur le fond du banc d'évaluation. L'ajout de 4 pieds est donc nécessaire pour surélever de quelques centimètres le banc d'évaluation et permettre ce fonctionnement des deux ventilateurs. Les pieds ont été également imprimés et contiennent un filetage. Ils sont ainsi vissés et collés à l'armoire en ABS, permettant de se prémunir de la possibilité qu'ils soient accidentellement arrachés et des risques qui y sont associés. Aucun filtre n'a été placé sur les ventilateurs réalisant l'extraction. La ventilation étant continue lorsque le banc d'évaluation est en fonctionnement, le ventilateur agit comme une barrière physique et empêche toute pénétration de corps étrangers ou insectes.

Sur le dos de l'armoire électrique se trouve une plaque en acier galvanisé originellement destinée au montage de l'équipement électrique. Cette plaque d'acier a été percée à l'aide de forêts à métaux, afin de réaliser le montage des tiges filetées. Leur longueur doit être inférieure à l'épaisseur de l'armoire. Au dos de la plaque d'acier galvanisé, des rondelles, écrous et de la colle époxyde ont été utilisés afin de limiter le débattement possible des tiges filetées. Des plaques en polycarbonate, dont les dimensions permettent d'être insérées dans le volume intérieur de l'armoire ont été préalablement percées afin de pouvoir être traversées par les tiges filetées qui sont désormais fixées au dos du banc grâce à la plaque d'acier galvanisé (Figure 24).



FIGURE 24 : Vue intérieure du banc d'évaluation

Ces plaques de polycarbonate ont pour rôle de diviser le volume intérieur du banc d'évaluation en plusieurs sections, de fournir un support où les capteurs seront maintenus à l'aide de colliers de serrage réutilisables en nylon et ainsi de maximiser l'exploitation du volume à notre disposition. La friction créée par le collier de serrage suffit dans notre cas à maintenir

le capteur en place. Naturellement, dans le cas de types ou de tailles de capteurs significativement différentes, d'autres solutions techniques pourraient être envisagées. Les plaques en polycarbonate conservent une distance prédéfinie entre elles, à chaque assemblage du banc, grâce à des tubes insérés sur chaque tige filetée (Figure 25).

La longueur de ces tubes détermine donc l'espacement des sections entre elles. Nous avons observé, en utilisant le même processus de simulation numérique que celle présentée chapitre 7.5 : Validation par simulation numérique par simulation numérique, qu'il était requis d'inverser le sens de montage d'une des trois sections afin de permettre à toutes les sections d'être exposées aux flux d'air pénétrant au travers des bouches d'aérations. Similairement, nous avons observé qu'il était préférable d'installer le maximum de bouches d'aération.

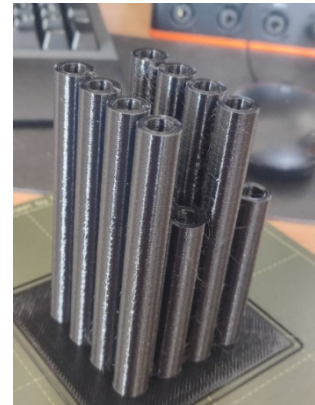


Figure 25 : Tubes séparant les sections

Un unique écrou en fin de chaque tige filetée permet ainsi de maintenir l'ensemble des sections. Ces tubes participent ainsi à accélérer considérablement l'assemblage et la mise en œuvre du banc d'évaluation. Pour chacune des trois sections, un concentrateur USB alimente jusqu'à 16 capteurs, permettant au banc d'évaluation d'accueillir un total de 48 capteurs dans des dimensions contenues. La plupart des capteurs portatifs peuvent fonctionner lorsqu'ils sont alimentés par USB. Un unique transformateur 12V et 15 ampères (A) a été modifié afin de fournir la connectique nécessaire ; il nous permet d'alimenter nos trois concentrateurs USB et nos deux ventilateurs. Le banc ne requiert ainsi qu'un câble d'alimentation 230 V pour fonctionner. Au dos du banc se trouve une extension de câble de 230V d'une dizaine de mètres, monté de façon permanente.

Nous avons sélectionné une armoire électrique de 50 cm de largeur, 70 cm de hauteur et 25 cm de profondeur ; ces dimensions permettent d'accueillir jusqu'à 45 capteurs CANARIN nano, tout en étant suffisamment légère pour être transportable par un individu seul. L'armoire électrique est originellement certifiée IP65 [135], donc résistante à la pénétration de la poussière et aux jets d'eau à la lance. L'épaisseur des plaques de polycarbonate sélectionnées est de 5 mm, permettant de supporter le poids de l'équipement. Le concentrateur USB sélectionné permet d'alimenter jusqu'à 16 capteurs par section. La norme USB nous permet de disposer de 500 mA par capteur, quand seuls 450 mA sont nécessaires (Figure 27). Chaque concentrateur possède son propre transformateur, mais il est cependant désirable de limiter les branchements nécessaires à l'intérieur banc d'évaluation. Pour cette raison, nous avons adapté un transformateur d'alimentation 230 V AC vers 12V DC 12A, en



FIGURE 26 : Pieds permettant l'extraction d'air

l'équipant de 4 sorties 12 V, permettant ainsi d'alimenter les 3 concentrateurs en plus des ventilateurs. Un unique transformateur permet ainsi d'alimenter l'ensemble des éléments présents dans le banc d'évaluation ; la mise sous tension du banc d'évaluation ne requiert donc qu'un point d'alimentation en 230V. Nous avons utilisé des ventilateurs de 120mm. Leur vitesse de rotation est de 1700 tours minute, pour un débit de 120 m<sup>3</sup>/heure. Le banc est surélevé du sol grâce à des pieds de 5 cm de haut (Figure 26), permettant à l'air d'être évacué par le dessous. Une fois posé au sol, l'extraction d'air est invisible pour le public. Aucune grille n'obstrue l'extraction d'air au sol, permettant de maximiser le volume d'air extrait. Même à pleine vitesse, les ventilateurs ne présentent aucun risque de blessure, aucune coupure ou choc important. Les risques associés à l'utilisation de ce banc sont donc extrêmement limités, pour le public, pour le technicien manipulant le banc, ainsi que l'infrastructure où le banc est installé.

#### 7.4. COÛTS ASSOCIÉS

La fabrication de ce banc d'évaluation mobile requiert uniquement quelques heures et son coût de fabrication, en dehors de l'outillage, est de moins de 350€. Le coût lié à la consommation électrique peut être approximativement évalué en mesurant la consommation individuelle des capteurs et ventilateurs. Nous ignorerons l'inefficacité des transformateurs 12V et 5V. Afin d'évaluer la consommation d'un ventilateur, nous nous sommes référés aux spécifications constructeur du ventilateur [136] qui indiquent une consommation de 200 mA par ventilateur. Pour l'évaluation de la consommation des capteurs CANARIN nano, nous avons utilisé l'outil de mesure CurrentRanger [137] ainsi qu'un oscilloscope permettant d'enregistrer la consommation en milliampères sous forme de série temporelle à une fréquence de 50 échantillons seconde.

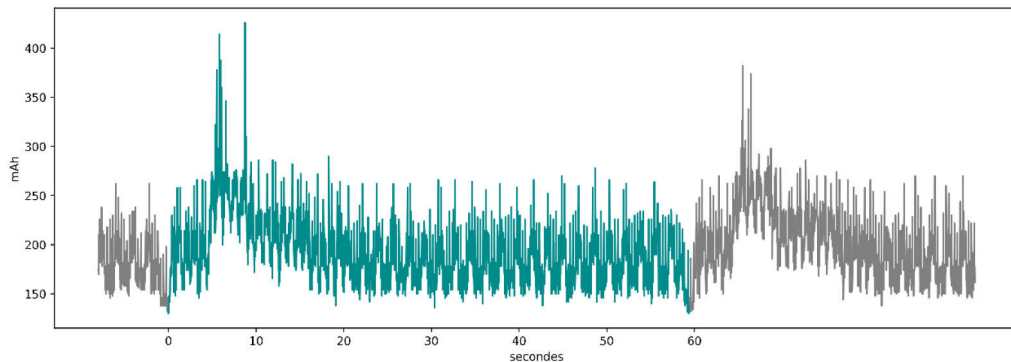


FIGURE 27 : Consommation énergétique du CANARIN nano en fonction du temps

Sur la Figure 27, nous avons représenté 100 secondes d'acquisition. La batterie du capteur CANARIN nano a été retirée afin de ne pas fausser les résultats en fournissant une source d'alimentation tierce. Seul un câble USB, où nous réalisons la mesure, permet d'alimenter le capteur avec une tension de 5V. Le micrologiciel du CANARIN nano réalise un cycle toutes les minutes. Nous avons donc représenté en vert ce cycle. On peut observer à 8 secondes une augmentation de la consommation qui coïncide avec la réalisation des mesures de la qualité de l'air et leur transmission par réseau cellulaire. En isolant un cycle d'acquisition, nous calculons donc une consommation moyenne de 192 mA. Le banc d'évaluation possède une capacité d'emport maximum de 48 capteurs. Entièrement équipé, il requiert donc un minimum 46 watts

permettant l'alimentation des 48 capteurs, auxquels s'ajoutent 7.2 watts pour l'alimentation des deux ventilateurs. Sur la base d'un tarif de 0,16 centimes d'euros le kilowattheure, le coût de la consommation électrique s'élève donc à seulement 6,12 € par mois.

Les CANARIN nano sont capteurs connectés permettant la disponibilité immédiate des données et requérant donc une carte SIM leur permettant de communiquer avec un serveur distant. Le serveur va permettre de rendre disponible pour l'utilisateur, l'ensemble des données recueillies au travers d'une page web. Chaque CANARIN nano consomme par jour environ 0.52 mégabits (MB). Le fournisseur de SIMs M2M sélectionné pratique un tarif de 0,1€ par MB. Le coût opérationnel mensuel serait donc d'environ 75 € pour 48 capteurs, auxquels s'ajoutent une cinquantaine d'euros de frais permettant de couvrir les coûts des services web AWS.

### 7.5. VALIDATION PAR SIMULATION NUMÉRIQUE

Les capteurs présents dans le banc d'évaluation mobile doivent être exposés aux mêmes conditions, or la ventilation forcée accélère la vitesse de l'air, autour des bouches d'aérations (entrées) et des ventilateurs (sorties). Plusieurs simulations numériques ont été nécessaires pour déterminer l'espacement entre les sections, le placement des capteurs, leur orientation, ainsi que la position et le nombre de bouches d'aérations nécessaires. Originellement, seules les dimensions internes de l'armoire en ABS et la présence de plaques permettant de maintenir les capteurs dans un le volume de l'armoire étaient connues.

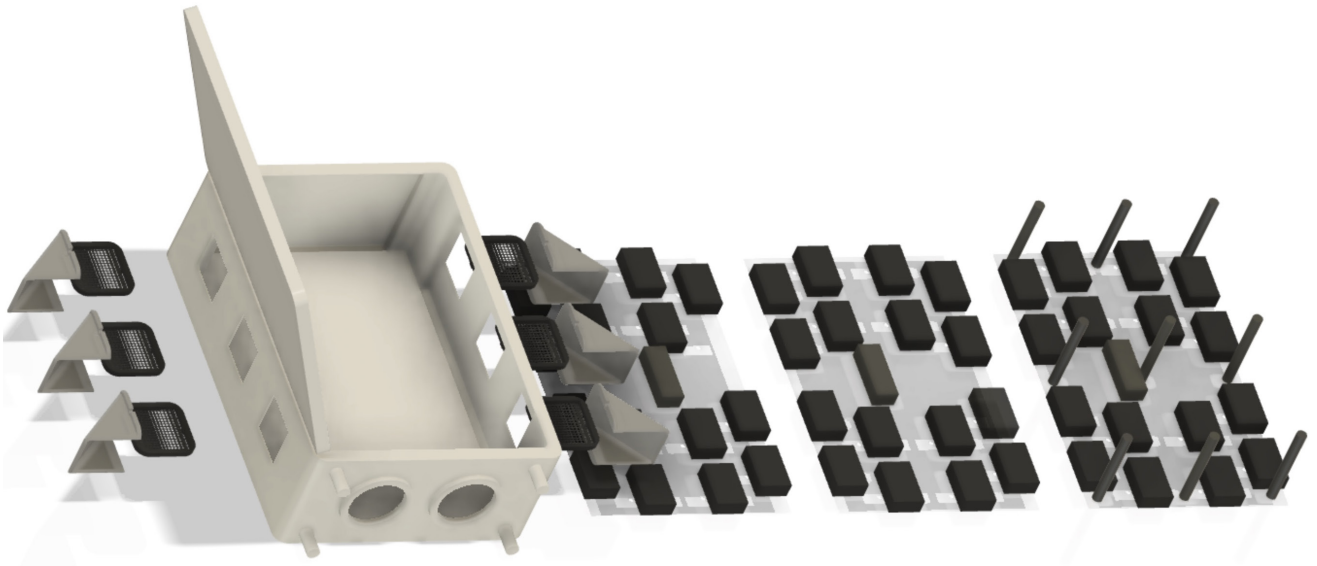


FIGURE 28 : Vue en éclaté du banc d'évaluation

Afin de pouvoir juger de l'amélioration de l'homogénéité de la vitesse des flux d'air en fonction de la modulation de ces paramètres, il a été nécessaire de réaliser un modèle 3D du banc. Nous avons utilisé le logiciel de modélisation Fusion 360 [138]. La cavité a été modélisée avec fidélité, tandis que les capteurs, concentrateurs USB, bouches d'aération et tiges filetées ont été modélisées grossièrement, dans le but de réduire les temps de calcul nécessaire lors des

simulations numériques. La Figure 28 présente le modèle 3D final du banc d'évaluation ; où tous les paramètres ont été déterminés. Une animation en référence permet de mieux visualiser la vue en éclaté et l'assemblage des sections [139]. À partir de ce modèle 3D, la licence Ansys [134] mise à notre disposition à titre gracieux nous a permis de réaliser plusieurs simulations. Nous avons ainsi utilisé les logiciels Ansys Workbench, DesignModeler, Fluent et enfin, CFD Post.

### 7.5.1. *DIFFUSEUR*

Les résultats initiaux nous ont permis d'observer que la vitesse à laquelle les capteurs placés face aux entrées d'air sont exposés est significativement plus importante vis-à-vis des autres capteurs. Nous présenterons ci-dessous les résultats qui ont mis en évidence la nécessité de l'implémentation d'un élément permettant de préserver les capteurs des flux d'air entrants dont la vitesse est importante. La seule modulation des paramètres suscités n'était pas suffisante pour assurer à l'ensemble des capteurs d'être exposés à des vitesses de flux d'air homogènes. Nous avons donc conçu un diffuseur ayant pour rôle de permettre aux capteurs de ne pas être directement exposés aux flux d'air pénétrant dans la cavité du banc d'évaluation. Plusieurs itérations de design et simulations nous ont amené à conserver le diffuseur présenté sur la Figure 29.



FIGURE 30 : Assemblage diffuseur et filtre



FIGURE 29 : Diffuseur du banc d'évaluation

L'air capté par le centre du diffuseur est dirigé vers le haut, tandis que l'air pénétrant sur les côtés est lui dirigé sur les côtés et le haut. La présence du diffuseur participe à apporter une sécurité supplémentaire contre l'intrusion de corps étrangers ; le filtre à poussière (en noir sur la Figure 30) ne suffirait pas à empêcher une main de pénétrer.

La Figure 31 est le résultat d'une simulation ayant pour but de montrer le principe de fonctionnement du diffuseur seul. La vue en coupe à gauche, comme les vecteurs à droite, permettent de visualiser la trajectoire et les variations de vitesse du flux d'air. L'échelle de mesure est omise puisqu'elle n'est pas pertinente pour cette démonstration. On observe que le diffuseur permet de projeter l'air sur le haut et les côtés du conduit carré, en évitant toute

projection vers le bas ou directement en face du diffuseur. Une animation en référence permet de visualiser la convergence du modèle vers un état stable [140].

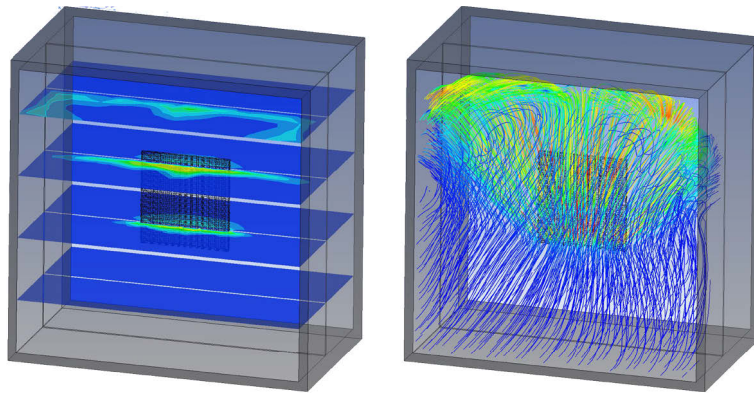


FIGURE 31 : Principe de fonctionnement du diffuseur

### 7.5.2. EFFET DES DIFFUSEURS

Les spécifications constructeur des ventilateurs sélectionnés [136] indiquent un débit d'air de 133.7 m<sup>3</sup>/h. Le diamètre des ventilateurs est de 110.5 mm. L'extraction d'air est réalisée par deux ventilateurs, modélisés par deux cercles de 95 cm<sup>2</sup>, représentant les ventilateurs. La vitesse de l'air au niveau de la bouche d'extraction est estimée à 3.87 m/s.

Sur la Figure 31, nous avons observé le comportement d'un diffuseur seul. Le banc d'évaluation est équipé de 6 bouches d'aérations et donc de 6 diffuseurs. Dans le banc d'évaluation, l'air est dirigé dans le haut du banc par les diffuseurs, mais également en direction de la plaque d'acier galvanisée et en direction de la porte. Nous sommes donc désormais en mesure d'interpréter la Figure 32, qui illustre la vitesse du flux d'air en fonction de la présence des bouches d'aération. Les vues en coupe présentées sont issues de simulations réalisées avec des conditions initiales identiques, où l'échelle est identique, à 20 et 60 cm de hauteur relativement au sol, soit au niveau du haut des bouches d'aération les plus basses et les plus hautes.

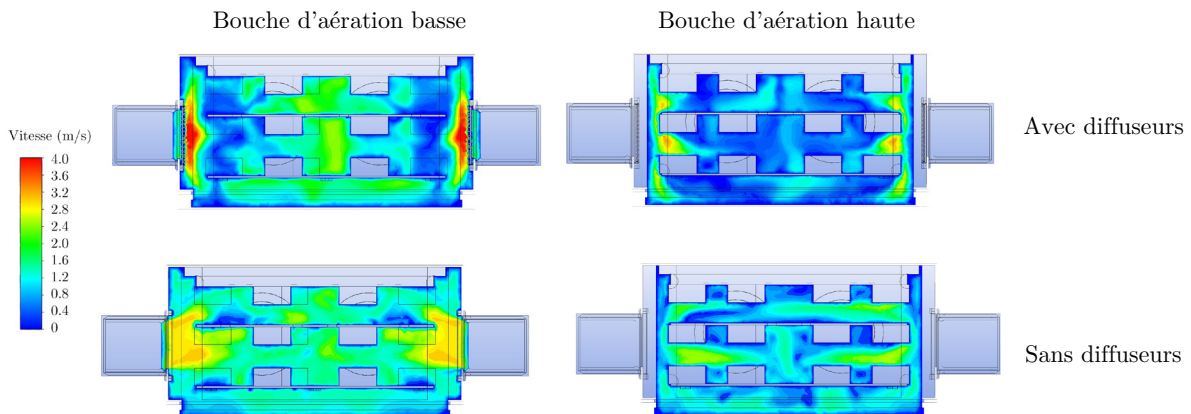


FIGURE 32 : Vitesse de l'air en fonction de la présence des diffuseurs, vue de dessus



Lorsque le banc est équipé de diffuseurs, on observe que l'air capté en bas du banc est projeté avec significativement plus de vélocité que l'air capté dans le haut du banc ; l'aspiration est plus forte dans le bas du banc du fait de la proximité avec les ventilateurs d'extraction. En l'absence de diffuseurs, des flux dont la vélocité est importante pénètrent entre les sections. Ce phénomène est cependant absent lorsque les diffuseurs sont présents. La Figure 33, similairement à la Figure 32, illustre la vélocité en vue de coupe, mais cette fois en vue de face. Premièrement, nous observons que la vélocité à proximité des ventilateurs (dans la partie extrême basse et au centre) reste inchangée. Tel qu'observé précédemment sur la Figure 33, la présence des diffuseurs préserve les capteurs placés directement en face des bouches d'aération d'être exposés à une vélocité importante, notamment pour le centre de la section 2. On note que la présence des diffuseurs participe à la formation, au centre du banc, d'un flux d'une vélocité d'environ 2 mètres seconde (en vert). Ce flux est présent jusqu'au niveau de la bouche d'aération la plus haute et sur l'ensemble des trois sections. Lorsque le banc n'est pas équipé de capteurs, ce flux est absent. Tout comme pour la Figure 31, une animation de la convergence du modèle est disponible en référence [141].

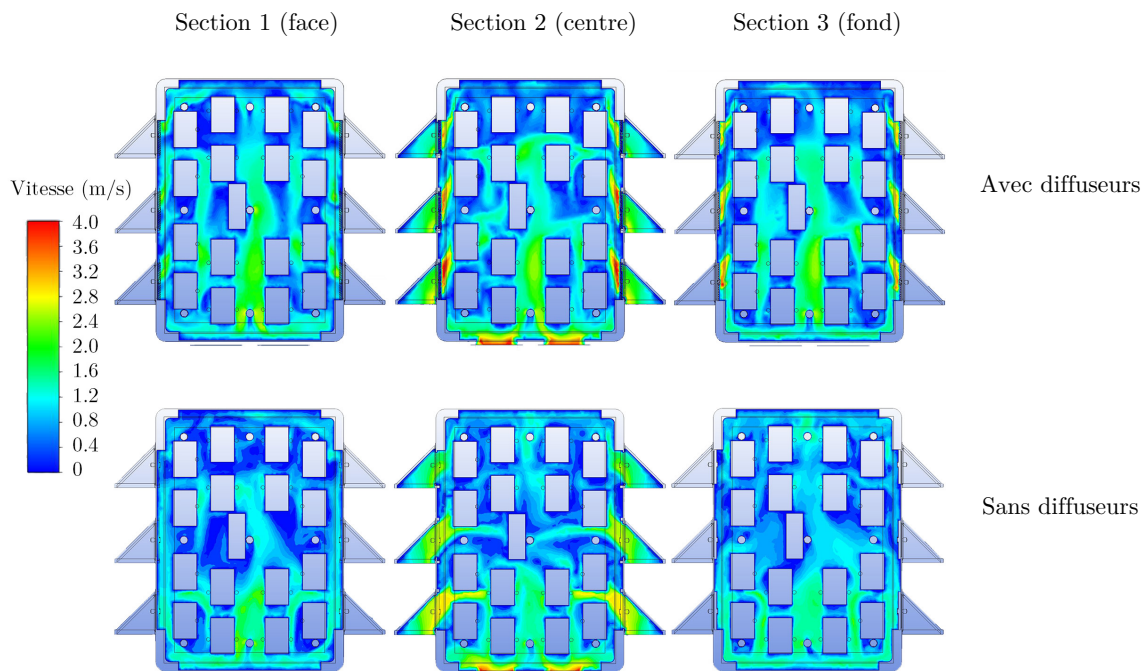


FIGURE 33 : Vélocité de l'air en fonction de la présence des diffuseurs, vue de face

Nous avons précédemment observé et décrit le comportement des flux d'air pénétrant dans le banc lorsque diffuseurs sont présents. Afin de quantifier l'effet produit par le diffuseur, nous avons relevé les valeurs de vélocité aux niveaux des modules de mesure pour chacun des capteurs, avec et sans diffuseur. La Figure 34 représente les trois vues en coupe utilisées pour relever les données et détaille les points permettant d'évaluer la variance de la vélocité ; là où été relevée dans le cas de la première section. Notons que la coupe est réalisée à hauteur des capteurs. Pour chacune des 3 sections, 16 valeurs sont relevées, comptabilisant donc un total de 48 points, pour les deux configurations possibles.

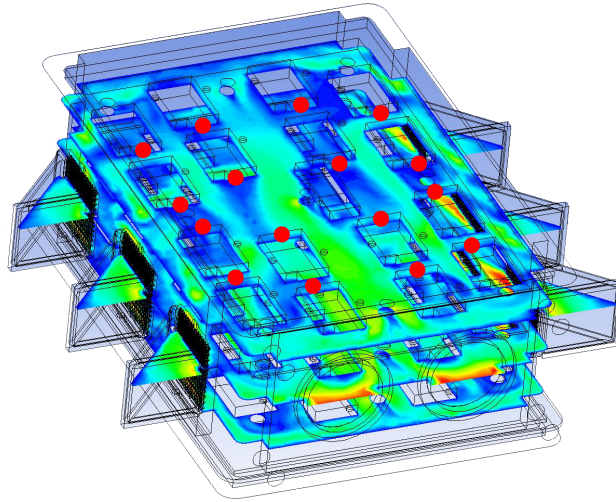


FIGURE 34 : Points d'évaluation de la vitesse

Nous obtenons une variance de 0.265 en l'absence de diffuseurs, contre 0.116 lorsqu'ils sont présents. En observant la variance par section, on peut observer que le gain le majeur est réalisé au niveau de la section centrale ; une variance de 0.482 sans diffuseur contre 0.144 lorsqu'ils sont présents. Le Tableau 2 nous informe de la vitesse médiane en fonction de la hauteur des capteurs dans le banc, relativement à la valeur médiane maximale observée. Les diffuseurs participent donc bien à augmenter la vitesse médiane de l'air auxquels les capteurs sont exposés dans le haut du banc. En revanche, on observe que les capteurs placés à environ 40 cm, sont eux exposés à une vitesse presque identique relativement aux valeurs maximales observées.

TABLEAU 2 : Vitesse médiane en fonction de la position des capteurs

Hauteur (cm)	Avec diffuseur (m/s)	Avec diffuseur (%)	Sans diffuseur (m/s)	Sans diffuseur (%)
$\approx 50$	0,385	60.2	0.310	33.5
$\approx 40$	0.270	42.2	0.375	40.5
$\approx 30$	0.640	100.0	0.655	70.8
$\approx 20$	0.635	99.2	0.925	100.0

### 7.5.3. CONCLUSIONS MODÉLISATION ET PROTOTYPAGE

Nous avons montré comment l'homogénéité de la vitesse de l'air au niveau des modules de mesures a été améliorée. Les performances initiales du banc ont donc été améliorées par l'ajout de diffuseurs. Notre expérience montre qu'en modulant les contraintes techniques et budgétaires, la réalisation d'un banc d'évaluation encore plus performant est possible.

La conception et production du banc d'évaluation auraient été compromis sans les outils logiciels de modélisation et simulation numérique. L'impression 3D des diffuseurs a également joué un rôle important. C'est un témoignage de la nécessité de rendre disponibles ces outils modernes, mais également de former les étudiants et chercheurs à leur utilisation.

## 7.6. ÉTUDE DE LA DISPERSION

Le banc d'évaluation mobile a été déployé en conditions réelles, équipé d'un set de capteurs, durant trois périodes, dans trois environnements différents. Ces acquisitions ont été motivées par un objectif principal et deux objectifs secondaires.

L'objectif principal fut de produire des jeux de données dans de multiples environnements, permettant d'évaluer les performances du processus présenté dans la partie 8, « Processus de réduction de la dispersion intra-set », page 69. Nous exploiterons les différents jeux de données pour caractériser la dispersion individuelle des capteurs en test vis-à-vis des autres capteurs. Par la suite, une fois les capteurs utilisés en scénarios réels (par les sujets d'un projet de recherche), nous serons en mesure de minorer cette dispersion par post-traitement.

Les objectifs secondaires furent premièrement de montrer la linéarité entre les variations de la dispersion et l'évolution des niveaux moyens du polluants ciblé, puis deuxièmement de s'assurer que ni les conditions météorologiques, ni le placement des capteurs à l'intérieur du banc ne jouent un rôle significatif sur les niveaux de dispersion relevés.

### 7.6.1. DESCRIPTION DES ENVIRONNEMENTS

Le premier environnement fut en air intérieur, dans l'enceinte de la faculté de médecine de Saint-Antoine (Figure 22), dans une pièce sans individus au 8<sup>ème</sup> étage, durant environ 6 jours (143 heures), du 18 au 24 février 2021 et dont la fenêtre est restée entrouverte.

La seconde période de mesure a été effectuée (Figure 23) en air extérieur, mais cette fois au troisième étage de la faculté de médecine, durant 7 jours (166 heures), du 5 au 29 mars 2021, faisant face à l'enceinte de l'hôpital de Saint-Antoine. Enfin, la troisième période de mesure fut réalisée en extérieur, au 7<sup>ème</sup> étage d'un appartement résidentiel, sur une terrasse, à proximité de l'église de Saint-Ambroise, durant environ 10 jours (232 heures).



FIGURE 35 : Environnement de test en extérieur (3<sup>ème</sup>)

La durée totale d'acquisition pour cette démonstration fut de 22 jours et 12 heures. Ce sont 32 capteurs neufs, n'ayant jamais fonctionné auparavant, qui ont été placés dans le banc. A l'issue des trois acquisitions, un total de 6 capteurs furent exclus du fait de divers dysfonctionnements. Ce sont les données de 26 capteurs qui ont finalement été exploitées. Ces capteurs ont mesuré la pollution de l'air environ chaque minute sur l'ensemble des trois périodes. Le Tableau 3 présente les pourcentages d'acquisitions relevé par capteur et par période. Considérant qu'une mesure est possible chaque minute sur l'ensemble de la durée des trois périodes, nos capteurs ont été en capacité de réaliser la mesure et transmettre les données acquises dans 98,2% des cas.

TABLEAU 3 : Pourcentage de mesures réalisées par période

Période	Moyenne	Médiane	Minimum	Maximum
Période 1	97.27	99.34	75.04	99.77
Période 2	98.71	99.41	91.86	99.83
Période 3	98.61	99.54	78.90	99.88

Chacun des capteurs a conservé sa position dans le banc d'évaluation à l'exception de la période 3, où une partie des capteurs ont été déplacés en vue de tester si la position des capteurs à l'intérieur du banc affecte les mesures réalisées. Afin d'observer si les conditions extérieures peuvent influencer le comportement des capteurs, les données météorologiques correspondant aux trois périodes ont été obtenues via une station météo amateur située à moins de 2 kilomètres des lieux où les trois essais ont été effectués, dont les données sont publiquement disponibles sur internet [142].

### 7.6.2. VARIATIONS DE LA DISPERSION

Pour chacune des 32 462 minutes d'acquisition et à partir des mesures de  $PM_{10}$  réalisées toutes les minutes par les différents capteurs, nous avons calculé la valeur médiane et l'écart-type. La Figure 36 permet de visualiser la relation entre l'évolution de l'écart-type en fonction du niveau de  $PM_{10}$ , au travers de la médiane. Sur cette figure, les trois périodes sont représentées et chaque point correspond donc à une minute d'acquisition. Nous observons une relation entre la médiane et l'écart-type, entre 10 et 40  $\mu\text{g}/\text{m}^3$ . Cette observation correspond au comportement attendu d'un OPC ; plus les valeurs mesurées sont importantes, plus l'incertitude de mesure augmente. Lorsque les concentrations médianes relevées sont supérieures à 10  $\mu\text{g}/\text{m}^3$ , nous avons colorés en vert et en orange les minutes

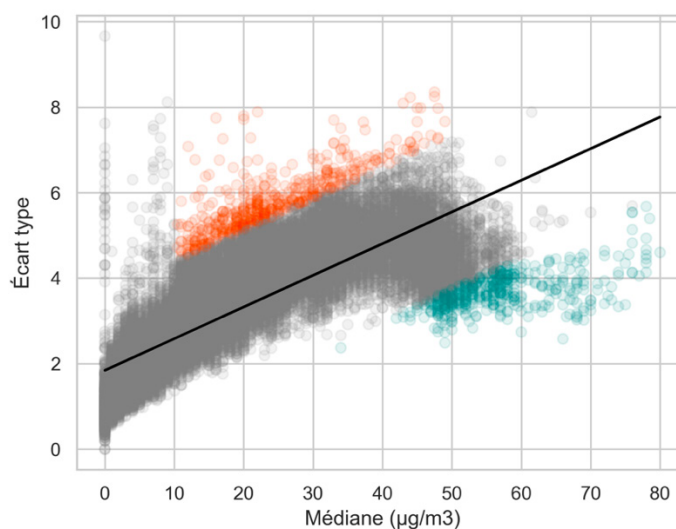


FIGURE 36 : Visualisation de la relation entre écart-type et médiane

où l'écart-type observé est élevé ou faible, vis-à-vis de la droite de régression linéaire (plus ou moins 2 unités d'écart-type). Les points en orange représentent donc les minutes où les capteurs ont davantage dispersé.

Le Tableau 4 détaille l'évolution de l'écart-type médian sur des plages de mesure de 10  $\mu\text{g}$ , en fonction des trois périodes. Aucune valeur n'a été renseignée pour les plages les plus élevées durant les périodes 1 et 2 car les niveaux observés ont été plus faibles. On observe que la variation de l'écart-type au sein du set de capteur est avant tout déterminée par le niveau de pollution atmosphérique ; l'écart-type médian est consistant sur chacune des plages.

TABLEAU 4 : Écart-type médian par plage de mesure

Plage (PM <sub>10</sub> $\mu\text{g}/\text{m}^3$ ) :	$\geq 0 < 10$	$\geq 10 < 20$	$\geq 20 < 30$	$\geq 30 < 40$	$\geq 40 < 50$	$\geq 50 < 60$	$\geq 60 < 70$	$\geq 80 < 90$
Période 1	1.81	3.15	4.41	NaN	NaN	NaN	NaN	NaN
Période 2	1.57	3.14	4.21	4.95	4.9	4.53	NaN	NaN
Période 3	1.89	3.23	4.05	4.63	4.38	4.07	3.9	4.14
Moyenne	1.76	3.17	4.22	4.79	4.64	4.3	3.9	4.14

### 7.6.3. POSITIONNEMENT ET VARIATIONS DE LA DISPERSION

L'un des objectifs principaux du banc d'évaluation est de permettre de s'assurer que la dispersion mesurée correspond à la dispersion propre aux capteurs et n'est pas due à des paramètres extérieurs. Les caractéristiques de la dispersion doivent donc rester similaires malgré le changement d'environnement. La racine de l'erreur quadratique moyenne (REQM) permet de quantifier l'erreur absolue vis-à-vis de la médiane, pour chacune des trois périodes.  $\hat{\mathcal{Y}}$  correspond à l'ensemble des échantillons relevés chaque minute par nos capteurs.  $\mathcal{Y}$  est la médiane des échantillons produits à chaque minute. Nous obtenons ainsi ;

$$REQM = \sqrt{\frac{\sum_{t=1}^T (\hat{\mathcal{Y}}_t - \mathcal{Y}_t)^2}{T}}$$

Nous avons permuté le placement de 12 des 26 capteurs, entre les Périodes 2 et 3, sur les sections 1 et 2, dans le bus d'évaluer si le placement des capteurs dans le banc d'évaluation engendre une évolution significative du niveau de dispersion et donc des valeurs de REQM. La Figure 37 illustre ce déplacement. Le déplacement est identique entre les deux sections ; un déplacement symétrique sur les axes horizontaux et verticaux.

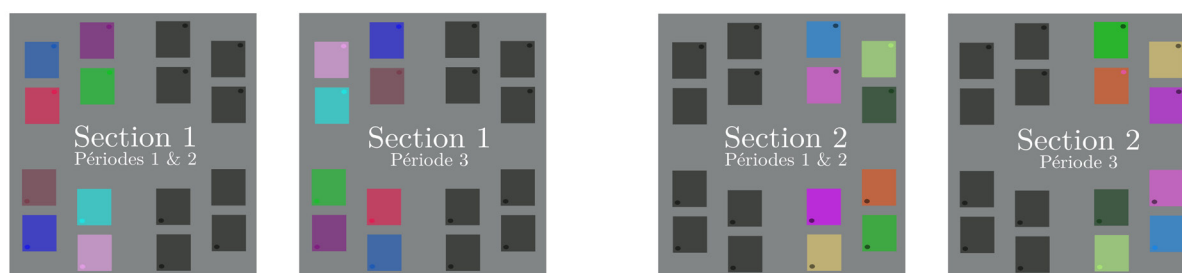


Figure 37 : Positionnement des capteurs

La Figure 38 représente la REQM pour chacun des capteurs et pour chacune des périodes. Les capteurs sont organisés en fonction de leur valeur moyenne de REQM. La teinte permet d'identifier les capteurs ayant été déplacés entre la période 2 et 3.

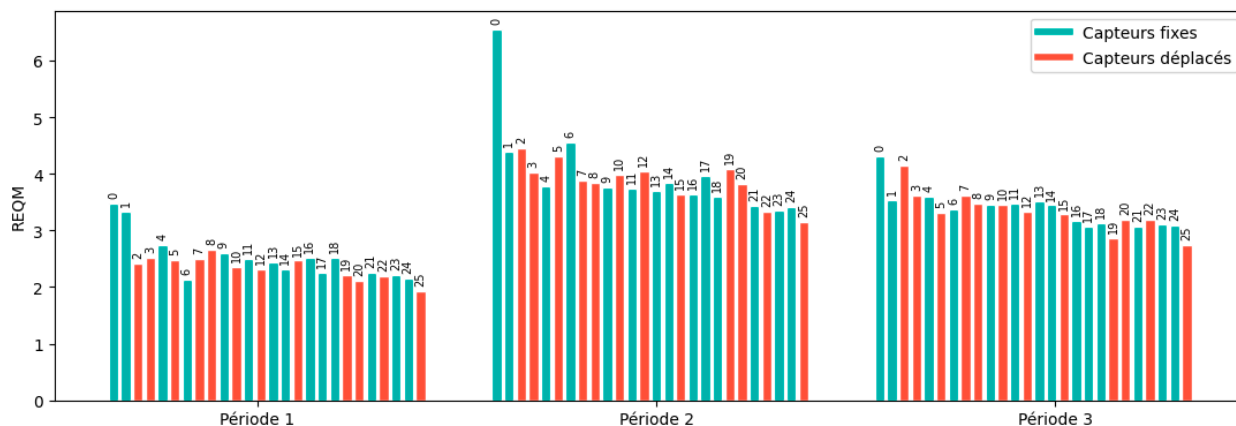


FIGURE 38 : REQM par capteur et par période

Comme le montre la Figure 36, la Figure 38 ainsi que le Tableau 5, la REQM corrèle avec l'évolution des niveaux de concentration de  $PM_{10}$ . Pour un certain nombre de capteurs, on observe que les disparités entre les niveaux de REQM sont conservées entre les trois périodes. Notons que ces disparités sont équitablement distribuées entre les capteurs fixes et les capteurs ayant été déplacés.

TABLEAU 5 : Valeurs moyennes de concentration de REQM par période

Périodes :	Période 1	Période 2	Période 3
Concentration moyenne $PM_{10}$ ( $\mu g/m^3$ )	8	26	17
REQM Moyenne	2.4	3.9	3.4

Les niveaux de REQM semblent évoluer linéairement avec les niveaux de concentration moyens relevés. Nous avons donc effectué une régression linéaire multiple, afin de pouvoir prédire les niveaux de REQM attendus pour la période 3, à partir des valeurs de REQM produites par les capteurs n'ayant pas été déplacés pendant les périodes 1 et 2. Nous obtenons ainsi l'équation suivante ;

$$REQM \text{ Période } 3 = 1.73 + (0.125 \times REQM \text{ Période } 1) + (0.336 REQM \text{ Période } 2)$$

Notons que la régression linéaire multiple étant réalisée à partir des données issues des capteurs fixes, par effet de sur-ajustement, l'erreur vis-à-vis des données réelles est naturellement plus faible. A partir des valeurs de REQM individuelles estimées pour la période 3, dans le Tableau 6 nous avons reporté l'erreur moyenne, ainsi que le pourcentage d'erreur entre les valeurs de REQM prédites et réelle ;

$$Erreur \text{ de } \text{prédiction} \text{ moyenne} = \frac{1}{n} \sum_{i=0}^n |REQM - REQM \text{ prédite}|$$

On observe que le changement de la REQM introduite par le déplacement des capteurs à l'intérieur du banc d'évaluation est négligeable. Ces observations ne tiennent pas compte de la possible évolution des caractéristiques de la dispersion des capteurs en fonction du niveau ou de la nature des polluants, mais suffisent à montrer que l'évolution de la REQM, dans notre cas, évolue en fonction des niveaux de concentrations de particules fines, indépendamment de la position des capteurs dans le banc d'évaluation. On peut ainsi conclure que la position des capteurs dans le banc ne suffit pas à expliquer les différences de niveaux de REQM, observées entre les différents capteurs sur la Figure 38.

TABLEAU 6 : Erreur de prédiction de la REQM, Période 3

	REQM réelle moyenne	Erreur de prédiction moyenne absolue	Pourcentage d'erreur vis-à-vis de la REQM réelle
Capteurs fixes	3.42	0.147	4.48 %
Capteurs déplacés	3.29	0.195	5.91 %

#### 7.6.4. CONDITIONS MÉTÉOROLOGIQUES ET VARIATIONS DE LA DISPERSION

Les CANARINs nano sont individuellement équipés de capteurs de température et d'humidité. Les données produites par ces modules de mesure produisent des informations potentiellement plus pertinentes que la station météo qui elle est distante et donc n'est pas exposée à des conditions parfaitement identiques. La Figure 39 montre que les données de température et humidité produites dans le premier environnement, en air intérieur, ne corrèlent pas avec les données issues de la station météo, malgré que le banc d'évaluation ait été placé dans une pièce ventilée, ni chauffée ou climatisée. Les périodes 2 et 3 quant à elles suivent une fonction linéaire vis-à-vis de la station météorologique. Les données des périodes 2 et 3 confirment la capacité des capteurs à produire des données de qualité.

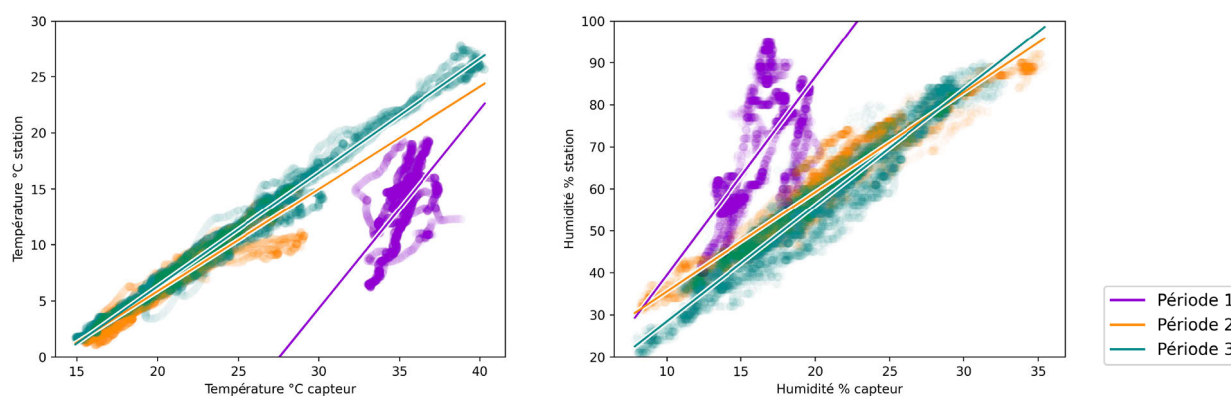


FIGURE 39 : Corrélation entre les sources de données météorologique

Nous avons observé une augmentation de la corrélation entre l'évolution des niveaux de dispersion et les variations d'humidité et températures à l'intérieur du banc d'évaluation, lorsque nous exploitons les données de température et humidité produites par nos capteurs plutôt que celles issues de la station météo. Les capteurs intégrés dans le CANARIN nano sont

eux-mêmes proches de composants qui émettent leur propre chaleur, tels que les micro-contrôleurs, régulateurs, modules de communication radiofréquence. Il est nécessaire d'appliquer une correction permettant d'obtenir des valeurs se rapprochant au mieux de la réalité. Afin de trouver l'équation permettant de corriger les données relevées à l'intérieur du banc, nous avons exploité uniquement les données produites en extérieur, durant les périodes 2 et 3. À partir de régressions linéaires, nous obtenons les équations suivantes ;

$$\text{Température corrigée} = (\text{Température brute} * 1.013) - 14,021$$

$$\text{Humidité corrigée} = (\text{Humidité brute} * 2.60) + 5.44$$

La Figure 40 présente le comportement de l'écart-type en fonction des niveaux de température et humidité. L'écart-type tend à diminuer lorsque les capteurs sont sujets à une augmentation de la température et une réduction du niveau d'humidité, ce qui correspond à nouveau au comportement attendu compte-tenu de la technologie employée. La colorisation sur la Figure 40 correspond aux classes introduites sur la Figure 36. On peut noter que les échantillons dont l'écart-type est faible sont situés dans une plage d'humidité de 45 à 80%. On ne remarque cependant pas de clustérisations qui pourraient indiquer une corrélation étroite avec ces variables.

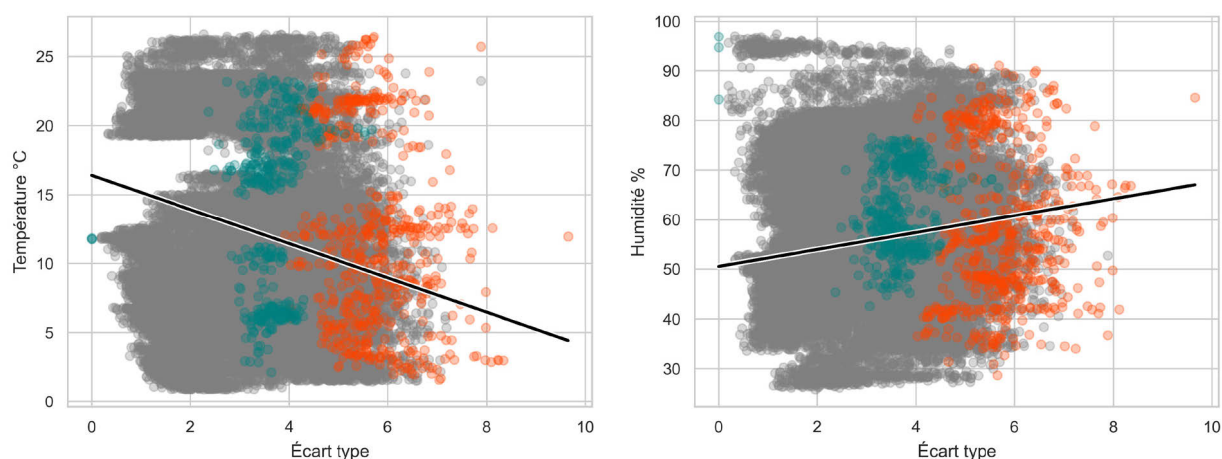


FIGURE 40 : Relation entre écart-type, température et humidité

Afin d'appuyer le choix de favoriser l'exploitation des données issues des capteurs, tel que précédemment évoqué, le Tableau 7 fournit les coefficients de la droite de régression, dans le cas où les données de la Figure 40 auraient été issues de la station météo. Notons que sans les données relevées au niveau des capteurs, l'observation du rôle de l'humidité et la température à augmenter la dispersion n'aurait alors pas été possible.

TABLEAU 7 : Coefficient de la droite de régression

Source de la donnée	Température	Humidité
Capteur	-1.23	1.71
Station météo	-0.21	-1.05



Les diagrammes en boîte représentés Figure 41 décrivent les trois classes d'écart-type introduites sur la Figure 36. Les valeurs aberrantes ont été cachées afin de conserver la lisibilité de la figure. La classe correspondant à l'écart-type faible possède une température médiane de 16.2°C contre 10 à 11° pour les classes dont l'écart-type est moyen ou élevé. D'autres variables issues de la station météo ont été évaluées mais sans montrer de différences significatives.

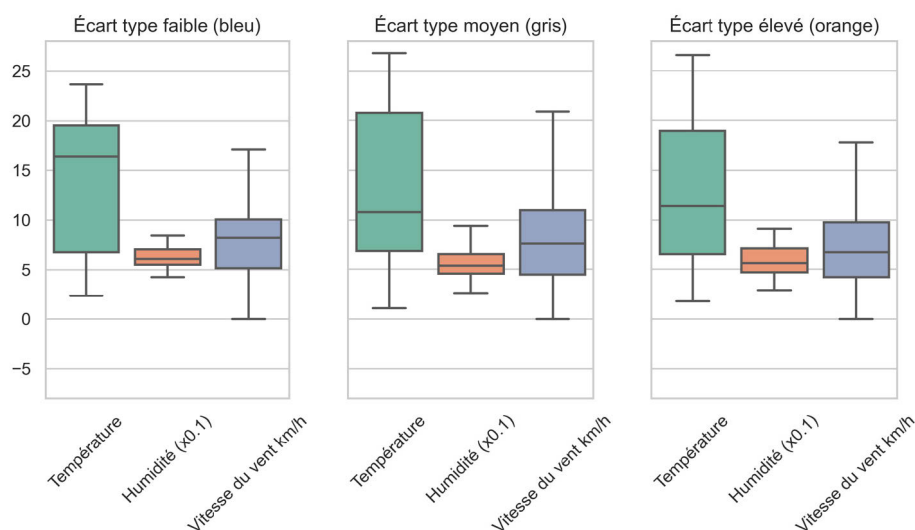


FIGURE 41 : Profil des classifications d'écart-type

Les données météorologiques produites par la station nous informent également sur la direction du vent. La médiane et l'écart-type, ainsi que la distribution du nombre d'acquisitions sont représentées en fonction de la direction du vent sur la Figure 42. Pour chaque variable et pour chacun des 16 points cardinaux, seule la valeur médiane est représentée. Nous observons ainsi que l'écart-type médian semble assez cohérent peu importe la direction du vent. Notons que la période 1, parce qu'elle fut réalisée en intérieur, n'est pas sujet à une possible influence de la direction du vent.

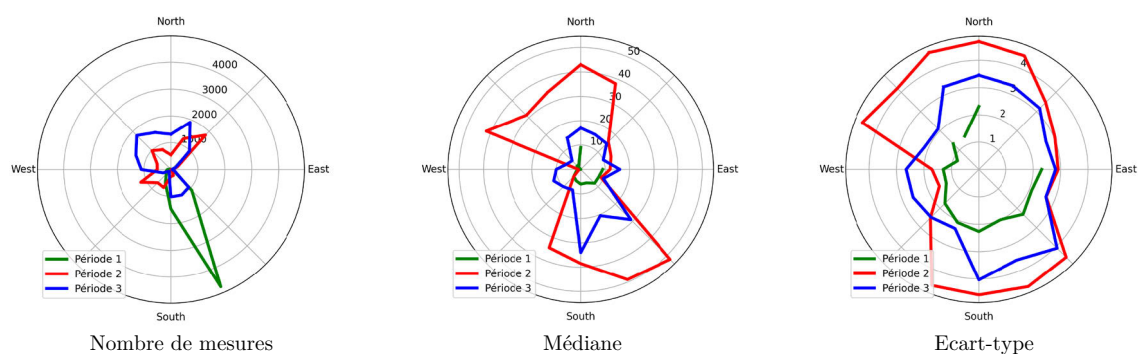


FIGURE 42 : Fonction de la direction du vent

Nos observations montrent le lien connu entre la température et l'humidité et l'augmentation de la dispersion ; la condensation d'eau étant perçue comme une particule fine par la technologie employée par nos capteurs. Nous soulignons la nécessité d'équiper le banc d'évaluation de capteurs environnementaux si les capteurs en test en sont dépourvus et

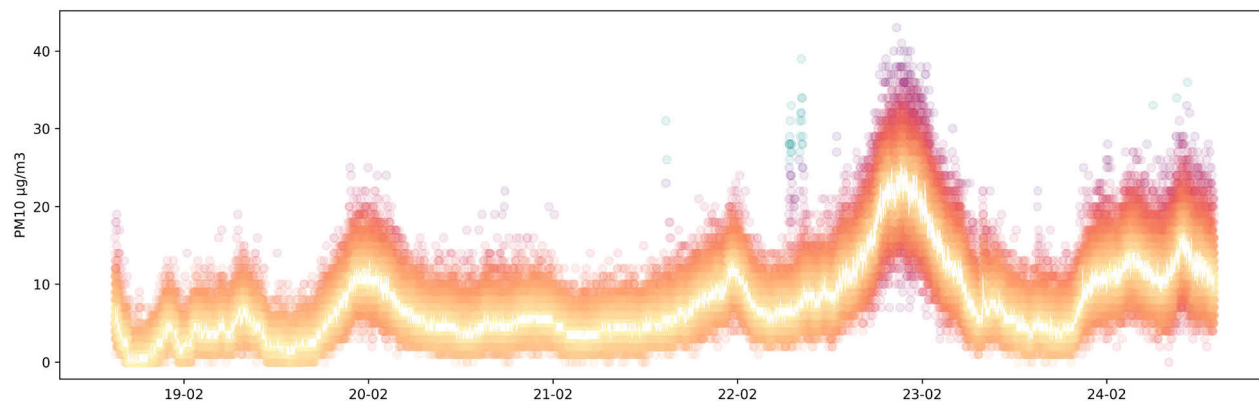
également de capteurs météorologiques tels qu'un anémomètre et d'une girouette, car l'acquisition de ces paramètres, tel que présenté dans le Tableau 7, doit être suffisamment représentative de l'exposition réelle du banc d'évaluation pour être exploités. Les technologies pour ces capteurs sont rendues de plus en plus abordables.

#### 7.6.5. *ÉVÉNEMENTS DE DISPERSION INDIVIDUELS*

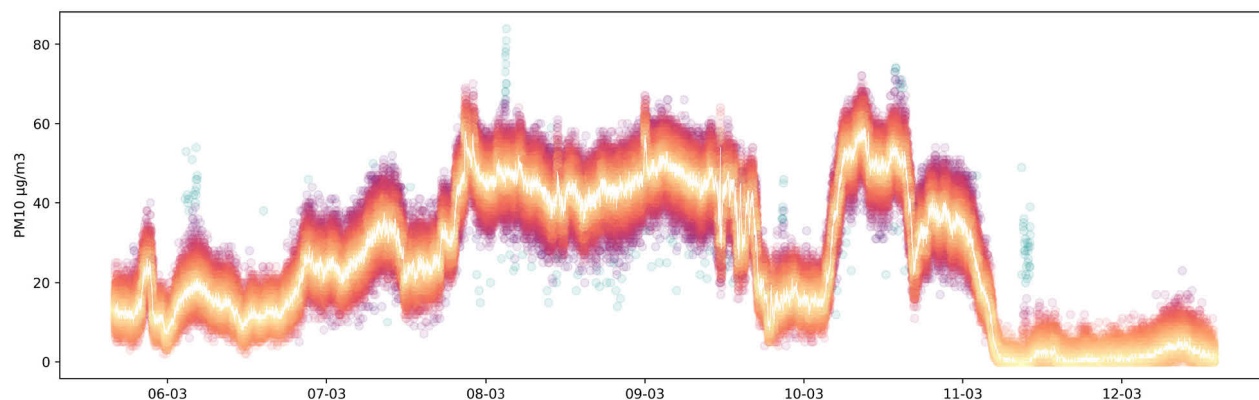
Nous avons jusqu'à présent traité la dispersion comme étant le fait de l'ensemble des capteurs et utilisé l'écart-type de chaque minute d'acquisition comme étant le critère permettant de déterminer lorsque la dispersion est présente. Nous chercherons désormais à identifier les événements de dispersion individuels ; en différenciant les capteurs qui dispersent vis-à-vis de ceux qui ne dispersent pas. Nous exploiterons la différence entre les valeurs relevées par chaque capteur vis-à-vis de la valeur médiane.

Sur la Figure 43, les niveaux de  $PM_{10}$  sont représentés pour chacune des trois périodes par une série temporelle. Chacun des échantillons individuellement acquis par les capteurs lors des périodes d'acquisitions est représenté. Cette figure permet premièrement de visualiser la variation du niveau de pollution en  $PM_{10}$  pour chaque période. Deuxièmement, elle permet de mettre en évidence certains événements ; la couleur attribuée à chaque échantillon est déterminée par l'absolu de la différence, entre la mesure acquise par le capteur et la valeur médiane produite par l'ensemble des capteurs à la minute d'acquisition donnée. La période 1, où les niveaux de dispersion et niveaux moyens de  $PM_{10}$  ont été plus faibles que durant la période 2 et 3, est donc représentée dans une teinte plus claire. La valeur médiane est tracée au centre de chaque période, en blanc. Il est ainsi possible d'identifier les événements où la dispersion a augmenté anormalement. Les événements où la dispersion est importante ont été filtrés en définissant un seuil d'écart absolu médian de 20 et sont représentés en bleu cyan afin d'améliorer la lisibilité du graphique.

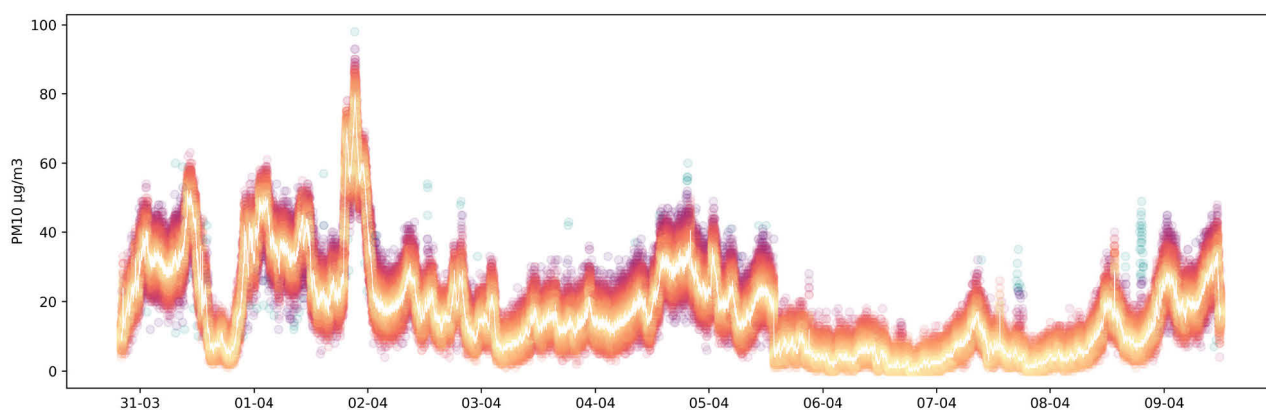
FIGURE 43 : Isolation des événements de dispersion



Période 1



Période 2



Période 3

Écart absolu médian



Le nombre d'échantillons dépassant notre seuil est cohérent avec les tendances précédemment observées ; 28 échantillons pour la période 1, 133 pour la période 2 et 88 pour la période 3. Cependant on note une grande disparité de la distribution de ces échantillons hors seuil au sein des différents capteurs tel que présenté Figure 44. La majorité des capteurs comptabilisent moins de 15 échantillons hors seuil. Seuls trois capteurs, soit 11.5% du set, représentent à eux seuls 38% des échantillons hors seuil.

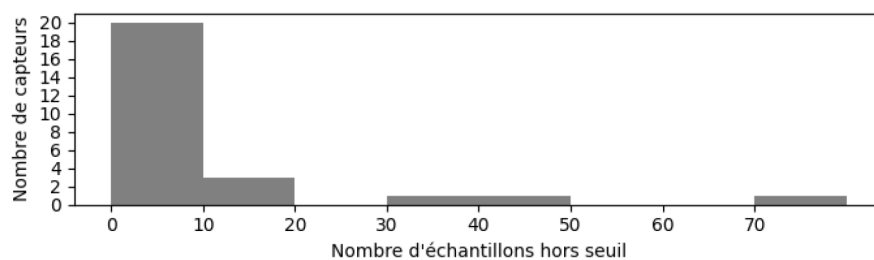


FIGURE 44 : Distribution des échantillons hors seuil

Sur la Figure 43, nous avons observé des clusters d'échantillons dépassant notre seuil, que nous pouvons qualifier d'événements de dispersion ; ils se caractérisent par leur courte durée dans le temps et par leur variance. La Figure 45 nous informe que ces événements sont le fait d'un nombre réduit de capteurs, mais n'est pas suffisante pour expliquer ces événements. Premièrement, les capteurs en question pourraient disperser aux mêmes instants ; nous pourrions alors théoriser que la dispersion est le fait d'un événement extérieur au banc d'évaluation. Deuxièmement, il est possible que les capteurs reproduisent ce comportement lorsqu'ils sont exposés aux niveaux auxquels la dispersion a été observée. Pour répondre à ces questions, les caractéristiques de la dispersion des trois capteurs en cause (désignés par capteurs

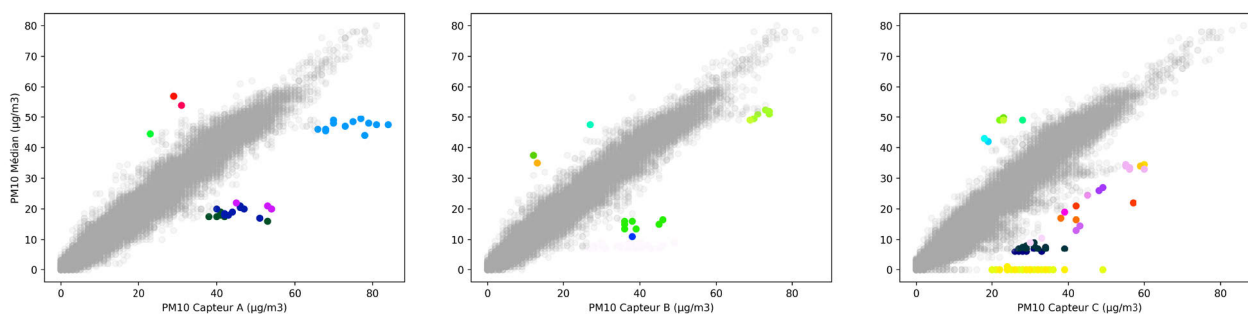


FIGURE 45 : Événements de dispersion individuels par capteur

A, B et C) sont présentées sur la Figure 45.

Seuls les échantillons dépassant le seuil sont représentés en couleur et ont été groupés lorsqu'ils subviennent à moins d'une heure d'intervalle. Chaque groupe correspond ainsi à un événement de dispersion, peut être composé d'échantillons issus de plusieurs capteurs et a la possibilité de durer plusieurs heures. Ces trois capteurs couvrent la quasi-totalité des échantillons dépassant le seuil déterminé précédemment (Figure 43) et forment un total de 34 événements de dispersion, pour 151 échantillons. La durée moyenne d'un événement est de 11 minutes et la durée maximale observée est de 58 minutes. Les événements sont donc

majoritairement courts dans le temps. Seul un unique événement a été observé sur deux capteurs au même instant et ne contient qu'un unique échantillon ; il est donc négligeable. La totalité des événements observés sont donc le fait d'un capteur, tel que le montre la colorisation des événements sur la Figure 45. On note également que les événements sont à la fois répartis sur la plage de mesure des capteurs, uniquement supérieurs à la médiane lorsque le fait de plusieurs échantillons et sont minoritaires vis-à-vis du reste des échantillons. Certains événements apparaissent à des niveaux proches (Capteur A), mais nous notons que les événements semblent disséminés sur l'ensemble de la plage de mesure des trois capteurs.

## 7.7. DISCUSSION ET CONCLUSIONS

Nous avons premièrement présenté un concept innovant de banc d'évaluation mobile à faible coût, spécifiquement dédié au contrôle de capteurs miniatures de la qualité de l'air en conditions réelles. Son processus de fabrication a été détaillé et avons fait la démonstration de son utilisation. L'importance de l'utilisation d'outils modernes de modélisation, de simulation numérique et de prototypage a été démontrée. Grâce à ces derniers, nous avons été en mesure d'appliquer une solution de modération des flux d'air à l'intérieur du banc.

Davantage d'efforts doivent être faits pour permettre d'augmenter l'homogénéité de l'air à l'intérieur du banc, tout en conservant un design facilitant sa reproduction. Nous avons par exemple envisagé la possibilité de capter l'air extérieur par toute la surface du haut du banc, en s'inspirant de la forme des conduits de ventilations utilisés par les abris souterrains. Il serait alors nécessaire de disposer d'une surface d'impression considérablement plus importante, ou d'assembler plusieurs pièces, puis de contrôler l'étanchéité de cet assemblage.

Au travers de l'analyse des données produites, nous avons pu mettre en évidence que les événements de dispersion les plus importants ne sont pas le fait d'éléments extérieurs. Nous sommes désormais en mesure d'affirmer que ces événements relèvent donc d'un dysfonctionnement de capteurs neuf. Précédemment, du fait que de nombreux éléments extérieurs pouvaient interagir avec les capteurs, il n'était pas possible de tirer de telles conclusions. Deux solutions sont envisageables : filtrer et corriger ces événements lorsqu'il est possible de les identifier, ou appliquer une correction matérielle, en remplaçant les capteurs défectueux.

Des efforts doivent également être réalisés pour mieux expliquer ces dysfonctionnements. Ils pourraient être dus soit à l'accumulation de particules ne pouvant être extraites par la ventilation équipant ce modèle d'OPC, soit à un défaut de conception interne aux capteurs. Une évaluation de la dispersion produite par des OPCs ayant été utilisés de manière continue et pendant plusieurs mois pourrait permettre de fournir des informations importantes relatives à la durée de vie réelle de ces capteurs.

## 8. PROCESSUS DE RÉDUCTION DE LA DISPERSION INTRA-SET

Sensor Batch Dispersion Reduction

### 8.1. INTRODUCTION

#### 8.1.1. DÉFINITION DE LA PROBLÉMATIQUE

Ces dernières années, le nombre et la qualité des solutions miniatures permettant de mesurer les niveaux de pollution atmosphérique ont considérablement augmenté et ont permis l'essor d'un certain nombre de capteurs commerciaux. Ces derniers visent à répondre à la demande, d'une part du public qui désire connaître son exposition individuelle à la pollution de l'air et d'autre part aux institutions, qui visent à étendre les connaissances scientifiques, sensibiliser la population et légiférer en conséquence. La prolifération de modules de mesure et leur faible coût ont également amené citoyens et chercheurs, à expérimenter, développer et produire leurs propres capteurs, intégrant ces technologies, souvent dans le but de répondre à des contraintes financières ou techniques.

Ces solutions de mesures de la qualité de l'air, portatives ou fixes, commerciales ou expérimentales, sont désormais fréquemment utilisées dans le cadre d'études épidémiologiques, permettant ainsi de caractériser l'évolution de l'état de santé de populations et cohortes, en fonction des données individuelles d'exposition à la pollution de l'air. L'utilisation de tels outils dans le cadre de la recherche médicale et épidémiologique requiert donc un contrôle rigoureux des performances des capteurs. Généralement, une validation périodique du bon fonctionnement des capteurs est effectuée.

Cette validation requiert de réaliser plusieurs jours d'acquisition où la flotte de capteurs en test est placée à un endroit sans accès au public, ainsi qu'un instrument de référence (une TEOM dans le cas des OPCs) [63]. Elle permet donc d'évaluer les performances de chaque capteur vis-à-vis d'instruments de référence et de s'assurer qu'ils répondent aux critères de performance attendus. De tels équipements ont un coût d'achat, d'utilisation et d'entretien importants. Les outils et infrastructures requis pour cette validation ne sont donc pas toujours accessibles aux entités de recherche médicale qui font l'acquisition de sets de capteurs portatifs. La récente démocratisation de l'utilisation de flottes de capteurs portatifs pour la recherche épidémiologique a donc rendu indispensable l'implémentation de procédures et d'outils alternatifs qui permettent d'effectuer la validation périodique des capteurs sans requérir à d'importants moyens techniques.

### 8.2. CHANGEMENT DE PARADIGME

Historiquement, c'est la capacité des capteurs à produire des résultats proches des instruments de référence qui fut valorisée ; la dispersion propre. Dans le cadre de la recherche épidémiologique, on cherche à mettre en relation des données d'exposition individuelles à la pollution de l'air issues de capteurs portatifs et l'évolution de l'état de santé d'une population réduite. Les études limitent leur observation à la population équipée de capteurs.

En réalité, ce sont les variations entre les sujets qui sont observées. Ainsi, le référentiel des données recueillies n'est pas l'instrument de référence, mais l'étalon autour duquel chacun des capteurs déployés disperse. Il est donc pertinent de chercher à évaluer et réduire cette dispersion quand c'est possible. L'évaluation de la dispersion, visera à contrôler le comportement individuel des capteurs, autrement dit, leur homogénéité vis-à-vis de la tendance du reste des capteurs et non pas vis-à-vis d'un instrument de référence. Si cette dispersion est répétable et correctement caractérisée, les statisticiens en charge de traiter et d'analyser ces données peuvent alors minimiser cette imprécision de mesure par post-traitement.

### 8.3. OBJECTIFS

C'est donc dans ce contexte que nous avons développé un processus réduction de la dispersion « intra-set », autrement dit, une réduction de la dispersion interne à une flotte de capteurs, qui s'apparente à une forme étalonnage puisqu'elle permet d'améliorer les qualités métrologiques des capteurs dont les données sont traitées.

Nous mettons donc à disposition un outil (un banc d'évaluation mobile présenté chapitre 7.3), une méthode (l'acquisition de données dans plusieurs environnements, telle que réalisée chapitre 7.6) et un algorithme en capacité de produire une correction transposable aux données obtenues en utilisation réelle (présenté ci-dessous). Ce processus, nous le désignerons par le nom de « SBDR » (de l'anglais « Sensor Batch Dispersion Reduction »).

La SBDR a été spécifiquement créée pour permettre aux entités de recherche d'effectuer une validation de flottes de capteurs, sans pour autant disposer d'instruments de référence ni d'infrastructure dédiée. Ces travaux visent donc à rendre plus accessible et permettre de généraliser la validation des capteurs portatifs mesurant la qualité de l'air. Dans le cas où les instruments de référence sont disponibles, alors l'algorithme présenté dans la SBDR peut également être intégré dans un processus de validation préexistant.

Dans un premier temps, nous introduirons les éléments et la terminologie nécessaire à la compréhension du processus de la dispersion intra-set. Nous présenterons ensuite le principe de fonctionnement de l'algorithme, puis l'appliquerons à divers jeux de données. Nous ferons ainsi la démonstration du caractère transposable de la caractérisation de la dispersion, dans le cas des OPCs. L'algorithme de réduction de la dispersion intra-set (SBDR) est délivré sous la forme d'un script python [143].

### 8.4. MÉTHODES

#### 8.4.1. ÉTALONNAGE DE CAPTEURS

Dans le cas d'un capteur professionnel, le fabricant propose généralement de recalibrer le capteur moyennant un coût, après une période d'utilisation déterminée, sous laquelle il est en mesure de garantir les spécifications de son produit. Les capteurs professionnels peuvent être calibrés dans le respect de procédures standardisées et définies par des normes européennes ou internationales. Cependant, la complexité d'un tel étalonnage est souvent trop importante pour être envisagée par l'acheteur du capteur professionnel, que ce soit un laboratoire public, ou une

entreprise privée. Dans le cas des micro-capteurs, un contrôle de leur fonctionnement est effectué durant leur fabrication, conformément aux normes et standards définis dans leurs fiches techniques respectives. En fonction de la technologie employée, les fabricants peuvent appliquer une correction dans le but de corriger la déviation d'un batch de capteurs. Pour un capteur commercial se destinant au grand public, des algorithmes peuvent être développés par le fabricant du capteur pour compenser la déviation du module de mesure dans le temps.

Lorsqu'un fabricant développe un capteur intégrant des modules de mesure, l'éventuelle incapacité à évaluer la qualité des données de manière approfondie, les contraintes temporelles et le besoin de profit, peuvent conduire à un mauvais choix de conception qui dégradera la capacité de fonctionnement du module de mesure. Dans le cas de prototype de capteur, il est difficile de s'assurer du bon fonctionnement du module de mesure embarqué. L'évaluation du prototype dans un laboratoire spécialisé face à un instrument de mesure de référence validé n'est pas toujours possible. Le développement d'un algorithme d'étalonnage n'est donc pas envisageable sans connaissances approfondies du comportement du module de mesure embarqué. Pour ces raisons, les capteurs prototypes n'embarquent pas d'algorithmes de traitement permettant d'en améliorer la qualité métrologique.

Par conséquent, l'utilisateur final d'un capteur portatif à bas coût n'a généralement aucun moyen d'accéder aux informations qui permettraient d'effectuer un étalonnage ; elles peuvent être inexistantes, ou relever strictement de la propriété intellectuelle du fabricant du capteur ou des micro-capteurs. En définitif, les contraintes de temps et de coût liées à l'étalonnage ou au contrôle approfondi des capteurs, conduisent les entreprises et les chercheurs à collecter des données dont la qualité est potentiellement compromise. Les micro-capteurs et modules ne sont donc pas exempts du besoin d'une réétalonnage, mais souvent aucune solution n'est proposée pour les capteurs à bas coût.

#### 8.4.2. VALIDATIONS DE CAPTEURS

Au sein du projet POLLUSCOPE, nous avons eu l'opportunité d'observer la nécessité de pouvoir évaluer l'ensemble des capteurs et valider leur bon fonctionnement ; à plusieurs reprises, des capteurs ont été identifiés comme dysfonctionnant et ont dû être remplacés et envoyés en maintenance. Lorsqu'un capteur est déployé sans validation préalable, son bon fonctionnement est incertain et la qualité des données recueillies est potentiellement compromise. Les études épidémiologiques utilisant des capteurs portatifs se doivent donc d'effectuer une validation systématique des capteurs portatifs utilisés.

La validation est donc l'action de réaliser des jeux de données où tous les capteurs sont exposés à des conditions identiques. Ensuite, les données produites sont analysées et des indicateurs statistiques permettent de déterminer si les différents capteurs répondent à des critères statistiques, sans quoi ils sont exclus de la flotte et ne seront pas déployés.

La SBDR a donc pour but de permettre cette étape de validation en l'absence d'instruments de référence, contrairement à ce qui est traditionnellement réalisé. En revanche, il est nécessaire d'utiliser un banc d'évaluation mobile pour collecter plusieurs jeux de données, dans plusieurs environnements, tout en limitant l'impact d'éléments extérieurs autres que la



pollution de l'air, tel que réalisé dans le chapitre 7.6, « Étude de la dispersion », page 58. Cette étape est nécessaire pour permettre la caractérisation de la dispersion.

Dans le cas d'un projet de recherche où les capteurs seraient déployés individuellement et de manière stochastique, l'évaluation périodique n'est alors pas une possibilité. Il est alors nécessaire d'évaluer les capteurs avant et après chaque déploiement, face à un sous-groupe de capteurs, qui ne seront eux jamais déployés et face auxquels les capteurs individuels seront évalués. On veillera également à surveiller l'évolution des paramètres de dispersion à l'intérieur de ce sous-groupe.

#### 8.4.3. LA DISPERSION PROPRE

La dispersion propre désigne les caractéristiques de la dispersion de chaque capteur vis-à-vis de la valeur physique réelle mesurée et donc à défaut d'être en capacité d'obtenir de meilleures informations, des données produites par instruments de référence. Lorsqu'elle est inconnue, dans le cas de la recherche en épidémiologie, elle réduit la possibilité de mettre en perspective les résultats produits entre les études ; la réponse caractéristique du module de mesure divergera en fonction de la technologie employée, de la composition chimique des polluants atmosphériques et des paramètres environnementaux s'ils sont significativement différents. Afin de minimiser les incertitudes liées à cette dispersion, il est nécessaire de réaliser une évaluation approfondie des modules de mesure employés par le capteur. Cette évaluation requiert des chambres de test et outils de mesure de référence permettant d'exposer le module de mesure à plusieurs polluants en environnement contrôlé, dans le but de caractériser le comportement du module de mesure et donc la dispersion vis-à-vis d'autres technologies existantes [57]. Dans une certaine mesure, cette caractérisation de la dispersion propre pourrait se transposer aux capteurs de marque et modèles identiques ; il subsisterait alors uniquement la dispersion intra-set. Les outils et infrastructures pour la caractérisation de la dispersion propre sont particulièrement coûteux, nécessitent d'être entretenus, sont non mobiles, inaccessibles au grand public, requièrent une expertise pour opérer et représentent donc une contrainte importante, que ce soit sur l'aspect logistique, technique, ou financier.

#### 8.4.4. LA DISPERSION INTRA-SET

La dispersion intra-set est matérialisée lorsque des capteurs de la pollution atmosphérique déployés dans le même environnement, donc exposés aux mêmes conditions, ne sont pas en mesure de fournir des résultats de mesures identiques. La dispersion intra-set est donc issue de l'accumulation des imperfections du micro-capteur ou module de mesure, des tolérances de l'intégration de ce dernier dans le capteur et de la position des multiples capteurs dans le micro-environnement où est effectuée la mesure.

Parce qu'elle se caractérise par la dispersion de chaque capteur vis-à-vis du set, elle ne requiert pas d'instrument de référence pour être évaluée. La Figure 46 représente l'ensemble des données du set « Banc2 », rééchantillonnées à une fréquence de 60 minutes. Les séries temporelles de chaque capteur sont individuellement représentées (en noir) et permettent d'observer les tendances de certains capteurs à surévaluer ou sous-évaluer, sur l'ensemble du jeu de données. L'opportunité d'appliquer une correction logicielle permettant de minimiser cette erreur de mesure est ainsi mise en évidence.

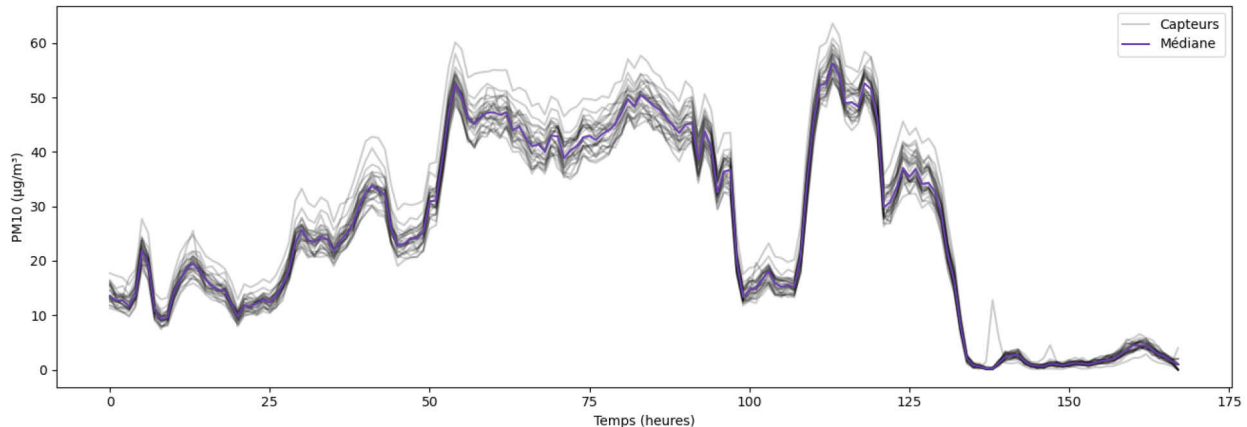


FIGURE 46 : Jeu de données Banc2 ré-échantillonné

Dans le cas où la caractérisation de la dispersion intra-set est produite dans différents environnements et que les profils de correction produits sont similaires, alors on considère qu'il est possible d'appliquer une correction par post-traitement qui sera en capacité de réduire la dispersion pour des données qui n'ont pas été encore acquises. La correction entraînera une réduction de la variance au sein des données, indépendamment de l'environnement. Cette correction permettra d'améliorer la cohérence des capteurs entre eux et donc la qualité des données produites en situation réelle ; lorsque le capteur est porté par un sujet.

#### 8.4.5. CARACTÉRISATION ET PARAMÈTRES DE DISPERSION

Le résultat de la caractérisation de la dispersion intra-set peut être obtenu après avoir produit un jeu de données — similairement à la validation d'une flotte de capteurs — dans des conditions permettant de minimiser l'impact des éléments extérieurs sur les capteurs. Dans notre cas, nous avons utilisé le banc d'évaluation précédemment présenté.

La caractérisation de la dispersion consiste à déterminer quel est le comportement des capteurs vis-à-vis de la tendance de ces derniers. Cette tendance est représentée par la médiane sur la Figure 46. La caractérisation permet de connaître quels sont les capteurs surévaluant ou sous-évaluant, dans quelle mesure et à quels niveaux de leur plage de mesure.

La caractérisation de cette dispersion se matérialisera par un tableau 2D, qui fournira les valeurs de correction à appliquer pour compenser la dispersion intra-set, pour chacun des capteurs en test, sur la plage de mesure couverte durant l'acquisition des données.

#### 8.4.6. *TRANSPOSABILITÉ DE LA CARACTÉRISATION*

La caractérisation de la dispersion réalisée dans la SDBR est transposable lorsque la caractérisation effectuée sur un jeu de données permet de réduire la dispersion sur la grande majorité des capteurs d'un second jeu de données. La transposabilité de la caractérisation est une condition nécessaire pour pouvoir appliquer une correction sur les données produites en conditions réelles.

Pour établir cette transposabilité, il est donc nécessaire de réaliser plusieurs jeux de données de validation, d'une durée de plusieurs jours et dans différents environnements, dont les variations de niveaux de polluants auront couvert plusieurs fois la plage de mesure du capteur. Ensuite, le caractère transposable est démontré en caractérisant la dispersion à partir d'un jeu de données, puis en appliquant la correction produite sur les autres jeux de données disponibles. La combinaison des caractérisations/corrections doit pouvoir produire une réduction systématique de la dispersion, sur la majorité des capteurs ; la transposabilité sera alors avérée. Le processus de la SBDR est finalisé par une dernière caractérisation, qui doit être produite à partir de l'ensemble les jeux de données disponibles. Si la quantité de données présentes par type d'environnements est déséquilibrée, les jeux de données doivent être ré-échantillonnés. Finalement, la caractérisation obtenue permettra alors de corriger les données acquises en scénarios réels, antérieurement ou postérieurement à la création de jeux de données de validation.

### 8.5. **SOURCES DE DISPERSION**

La dispersion est multifactorielle. Il est donc particulièrement difficile de correctement évaluer les sources de dispersion et en quelle mesure elles sont responsables sur des niveaux de dispersion finaux. La dispersion, qu'elle soit intra-set ou propre, impacte de manière permanente des mesures, mais n'est visible que lors de la validation des capteurs, ou lors de la caractérisation de la dispersion dans le cadre de la SBDR.

De nombreux éléments peuvent perturber le bon déroulement d'une validation de capteurs et rester inaperçus. À minima, il faut s'assurer que l'horloge interne des capteurs en test est synchronisée du mieux possible ; chaque mesure doit être réalisée au même instant dans le temps. Des erreurs humaines commises lors de la mise en place des capteurs ; des variations dans leur orientation, dans leur espacement. L'obstruction accidentelle d'entrées d'air peut rester inaperçue et aura des conséquences importantes sur les résultats.

#### 8.5.1. *CONDITIONS EXTÉRIEURES*

Dans le cas de validations réalisées en air extérieur et s'étendant sur plusieurs jours, sans surveillance humaine ; l'intervention de personnes étrangères à la procédure de test, la présence d'animaux, d'insectes, d'éléments solides autre que de la fine poussière en suspension dans l'air, des conditions météorologiques incertaines. Lorsque les capteurs sont exposés aux conditions extérieures, l'absence d'outils adaptés à l'évaluation de la dispersion intra-set contribue donc à augmenter les incertitudes quant aux sources réelles de la dispersion. Sans ces outils, la

caractérisation de la dispersion sera incertaine et les résultats relatifs à transposabilité des corrections seront dégradés.

Pour répondre à cette nécessité, nous avons donc présenté partie 7 un banc d'évaluation mobile permettant de faciliter l'évaluation de la dispersion et réduisant les risques susmentionnés. Cet outil permet au groupe de remplir les conditions nécessaires à la caractérisation de la dispersion intra-set. Les capteurs sont agencés de manière à les exposer aux mêmes conditions.

#### 8.5.2. INTÉGRATION DES MICRO-CAPTEURS

Au sein d'un set de capteurs de la qualité de l'air, la dispersion peut être causée par les limitations propres aux micro-capteurs réalisant la mesure. En effet, chaque micro-capteur possède une dispersion définie par la plage d'erreur de mesure permise par le constructeur. Cette plage d'erreur tient compte des limitations physiques de la technologie employée et des tolérances permises lors du processus de fabrication. L'intégration du micro-capteur peut également être en cause. Les composants électroniques en fonctionnement à proximité des micro-capteurs peuvent être source d'interférences (électriques, magnétiques, lumineuses). Un processus durant l'assemblage des composants, des conditions dégradées de transport, de stockage, ou une exposition accidentelle à des polluants durant le conditionnement pourraient nuire de manière permanente au bon fonctionnement du module de mesure.

La capacité du micro-capteur à correctement évaluer le paramètre physique ciblé, sera impactée par le placement et l'orientation du dit micro-capteur dans le boîtier composant le capteur, ainsi que par les choix esthétiques et ergonomiques réalisés par le constructeur du capteur. Dans le cas des capteurs de la qualité de l'air, le placement des aérations joue un rôle critique dans la capacité des capteurs à effectuer des mesures en mobilité. Il existe donc dans ce cas, lors de l'utilisation des capteurs, une dispersion induite par les préférences de port des individus, qui peut être évaluée, mais qui est imprédictible et qui ne peut être corrigée.

#### 8.5.3. DURÉE DE VIE EN UTILISATION RÉELLE

Lorsque ces capteurs sont déployés et utilisés en situation réelle par la population (dans notre cas, des participants participant au projet POLLUSCOPE) ils subissent le vieillissement naturel du module de mesure et le dérèglement de leur correction interne et donc à terme, une dégradation substantielle de ses performances. De plus, en situation réelle, les micro-capteurs sont sujets à des conditions d'utilisation inconnues. Les utilisateurs des capteurs peuvent possiblement exposer les modules de mesure à des conditions environnementales (humidité, chaleur, poussière) dépassant des seuils recommandés. Cette exposition, si elle est répétée, pourrait entraîner une détérioration permanente du micro-capteur. Cette détérioration se traduira par une dégradation de la qualité de mesure, ou par des dysfonctionnements imprévisibles, autrement dit, par une augmentation de la dispersion.

## 8.6. JEUX DE DONNÉES EXPLOITÉS

Nous exploiterons au total 6 jeux de données. Trois d'entre eux ont été produits dans le cadre de l'utilisation du banc d'évaluation (voir chapitre 7, « Banc d'essai mobile pour l'évaluation de la dispersion »). Ces jeux de données sont désignés par Banc1 (en intérieur), Banc2 (face à un hôpital) et Banc3 (sur la terrasse d'une résidence). Ils ont donc été spécifiquement produits pour permettre de démontrer la caractérisation de la dispersion intra-set. Ci-dessous, nous fournissons le contexte relatif aux trois jeux de données supplémentaires.

### 8.6.1. JEUX DE DONNÉES ISSUS DE VALIDATIONS :

Le jeu de données désigné par « Aiparif » a été produit entre le 27 avril et le 9 mai 2019, à l'aide de 15 capteurs Cairsens NO<sub>2</sub>. Il nous a été fourni par l'association agréée de surveillance de la qualité de l'air Airparif au travers de Christophe Debert. Ces capteurs ont été placés dans l'enceinte d'une station de surveillance de la qualité de l'air, en air extérieur, à proximité du boulevard périphérique de Paris. Le jeu de données « Lsce » a été produit à l'aide d'un set de 9 CANARINs II, en air extérieur, dans une zone non industrialisée, par le Laboratoire des sciences du climat et de l'environnement, entre le 6 et le 7 mai 2019. Ces deux jeux de données furent produit dans le cadre de validation de capteurs pour le projet POLLUSCOPE [108]. Notons que les niveaux relevés sont particulièrement faibles du fait que la zone environnante n'est pas urbanisée.

### 8.6.2. JEUX DE DONNÉES ISSUS DE CAPTEURS FIXES :

Le jeu de données désigné par « Coimbra » nous a été fourni par Davide Aguiari et se compose de 57 jours d'acquisitions, réalisées entre le 11 décembre 2018 et le 12 février 2019, pour 5 capteurs CANARIN II. Les capteurs en question ont été disposés dans la Bibliothèque Joanina à Coimbra, au Portugal. Dans cette configuration, les capteurs ont été fixés et ont permis de suivre l'évolution des niveaux ambiants de particules à l'intérieur de la bibliothèque [144]. Une légère variation des niveaux de PM relevés peut en partie être expliquée de par leur disposition dans la bibliothèque. Néanmoins, ces capteurs ont relevé des données dans le même micro-environnement et une quantité importante de données est disponibles. Le public n'a normalement aucune interaction avec les capteurs et de par la fragilité de l'écosystème où ils résident, ils ne sont pas exposés à des risques de dispersion importants. Ce jeu de données est donc une opportunité formidable pour appliquer l'algorithme SBDR.

### 8.6.3. RÉCAPITULATIF

Nous disposons donc de plusieurs jeux de données, réalisés dans des environnements différents, contenant un nombre d'échantillons et de capteurs qui vont nous permettre d'exposer le fonctionnement de l'algorithme SDBR. Notons que tous les capteurs ont été configurés pour échantillonner la pollution de l'air une fois par minute. Le Tableau 8 nous informe pour chaque jeu de données, du nombre d'échantillons collectés par capteur (en moyenne, maximum et minimum) et à titre de rappel, réunit l'ensemble informations présentées ci-dessus ;

TABLEAU 8 : Informations par jeu de données

Nom	Moyenne	Maximum	Minimum	Polluant	Capteurs	Environnement	Contexte
Banc1	8378	8594	6096	PM <sub>10</sub>	26	Intérieur	Caractérisation
Banc2	9876	9989	9159	PM <sub>10</sub>	26	Hôpital	Caractérisation
Banc3	14192	14390	11595	PM <sub>10</sub>	26	Résidence	Caractérisation
Airparif	14485	16940	11640	NO <sub>2</sub>	15	Boulevard	Validation
Lsce	1513	1795	1025	PM <sub>2.5</sub>	9	Plein air	Validation
Coimbra	81838	84556	79922	PM <sub>2.5</sub>	5	Bibliothèque	Fixe

## 8.7. ALGORITHME D'ESTIMATION DES CENTRES

La capacité à estimer la valeur centrale au sein d'une série de valeurs dont certaines dispersent est une fonction centrale du processus SBDR. C'est pourquoi nous présenterons avant toute chose l'algorithme utilisé pour réaliser cette « estimation des centres », que l'on peut apparenter à un filtrage des valeurs aberrantes.

Le calcul de la médiane est une solution simple, rapide et efficace. Dans l'algorithme SBDR, nous désignerons son utilisation par « l'estimation des centres médians ». Cependant, la médiane n'est pas systématiquement adaptée. Dans le but de maximiser les résultats de l'algorithme SBDR, nous avons cherché à produire un algorithme d'estimation des centres dont la sélectivité peut être modulée.

### 8.7.1. DESCRIPTION DE L'ALGORITHME D'ESTIMATION DES CENTRES

Dans un premier temps, nous produirons pour chacun des points d'acquisition dans le temps, l'absolu de la différence des échantillons entre eux. Dans le cas d'une série d'échantillons ( $x_1, x_2, x_3$ ) issus donc de trois capteurs différents, nous obtenons donc une matrice symétrique que nous décomposerons ainsi ;

$$x_{c1}, x_{c2}, x_{c3} = \begin{bmatrix} |x_1 - x_1| & |x_2 - x_1| & |x_3 - x_1| \\ |x_1 - x_2| & |x_2 - x_2| & |x_3 - x_2| \\ |x_1 - x_3| & |x_2 - x_3| & |x_3 - x_3| \end{bmatrix}$$

À partir de cette matrice, nous produirons pour chaque échantillon, un score ( $x_s$ ) résultant du rapport entre la somme et le maximum des différences précédemment produites ( $x_c$ ). Pour chaque échantillon, ce score sera image de sa distance moyenne avec les autres échantillons ;

$$x_{s1}, x_{s2}, x_{s3} = \frac{[\sum x_{c1} \quad \sum x_{c2} \quad \sum x_{c3}]}{[Max(x_{c1}) \quad Max(x_{c1}) \quad Max(x_{c1})]}$$

Nous ordonnerons les échantillons en fonction du score obtenu et sélectionnerons une partie des échantillons dont la distance est la plus faible. Le nombre d'échantillons sélectionné sera déterminé par le critère de sélectivité choisi par l'utilisateur et dont la variation module la sélectivité de l'algorithme ;

$$N \text{ } x_{i \text{ sélectionnés}} = \text{Sélectivité} (x_i \text{ où } x_{si} \text{ minimum})$$

Enfin, le centre sera obtenu en réalisant la moyenne des échantillons sélectionnés ;

$$\text{Centre} = \sum_{i=0}^N x_{i \text{ sélectionnés}}$$

Nous serons ainsi en capacité de moduler la sélectivité de cet algorithme grâce à la variation du critère de sélectivité (*Sélectivité*). Il sera donné en pourcentage ; 100 % produira un résultat strictement équivalent à la moyenne, tandis qu'une sélectivité plus faible favorisera la sélection d'un centre dont les niveaux relevés sont plus proches entre eux. Nous serons ainsi en mesure de rejeter les échantillons dispersants, même dans le cas où ils seraient majoritaires.

### 8.7.2. ESTIMATION DES CENTRES SUR DONNÉES SIMULÉES

Nous ferons la démonstration de l'algorithme de sélection des centres au travers de données simulées. Nous illustrerons l'effet du critère de sélectivité sur l'estimation des centres, dans plusieurs scénarios et sur une unique figure.

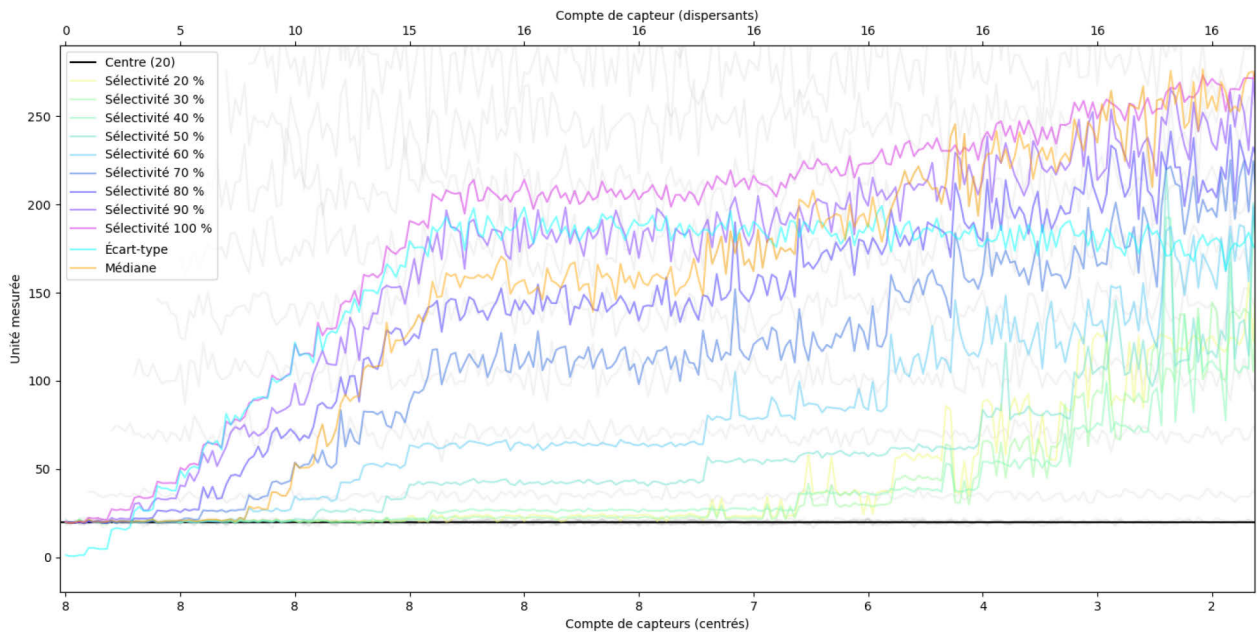


FIGURE 47 : Évolution de la sélection des centres en fonction de la sélection

Deux types de capteurs ont été simulés sur la FIGURE 47 ; ils sont individuellement représentés en gris (axe des ordonnées à gauche). Premièrement, nous avons simulé des capteurs centrés sur la valeur 20 (repère en noir), possédant une dispersion faible et dont le compte est renseigné sur l'axe des abscisses en bas de graphique. Deuxièmement, nous avons simulé des capteurs dispersants (visibles en partie en fond de figure, en gris), dont le compte est renseigné sur le second axe des abscisses, en haut de graphique. L'écart-type est également fourni (en cyan). Les niveaux et la variance de ces derniers augmente à mesure qu'ils sont introduits afin d'exacerber les niveaux de dispersion.

La Figure 47 permet donc d'illustrer la capacité de l'algorithme d'estimation des centres à moduler en fonction du critère de sélectivité. On peut ainsi produire une estimation des centres se rapprochant de la moyenne (en rose), de la médiane (en orange), ou d'être très sélectif, afin de favoriser les capteurs dont les niveaux sont proches (en bleu ou vert clair).

## 8.7.3. ESTIMATION DES CENTRES SUR DONNÉES RÉELLES

Nous réaliserons à présent l'estimation des centres sur des données réelles. Nous exploiterons le jeu de données « Banc2 » (introduit sur la Figure 46 et chapitre 8.6.). La totalité des données disponibles sont représentées. Les séries temporelles de chaque capteur sont individuellement représentées en noir et les résultats de l'estimation des centres en violet.

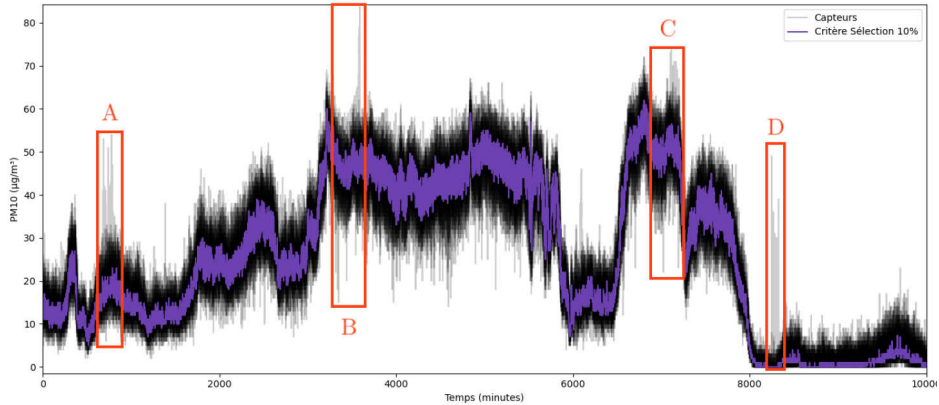


FIGURE 48 : Identification de valeurs aberrantes

Quatre évènements où la dispersion observée est plus importante sont fenêtrés (en orange) et identifiés (A, B, C, D). Chacun de ses évènements sont représentés en détail FIGURE 49. La dispersion relevée dans ce cas-là est le fait d'un nombre réduit de capteurs. Ces évènements de dispersion seront donc correctement discriminés par notre formule d'estimation des centres et leur présence sera ignorée.

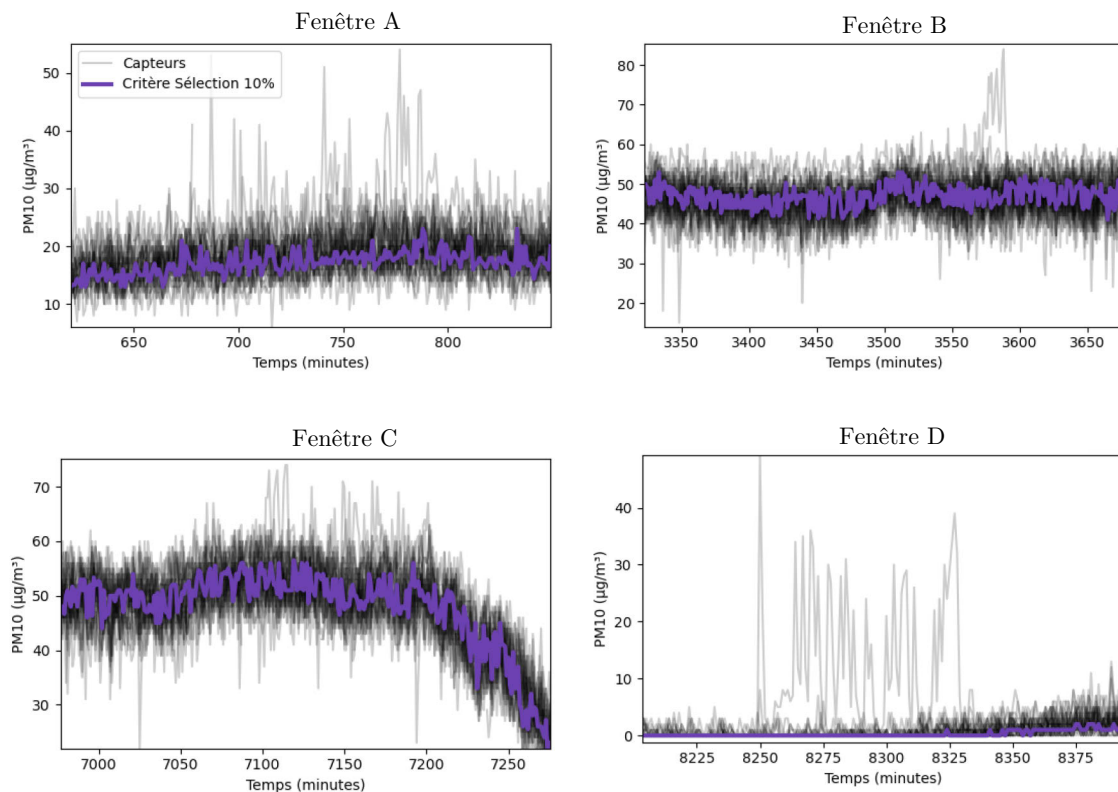


FIGURE 49 : Démonstration du rejet de valeurs aberrantes



## 8.8. DESCRIPTION DU PROCESSUS SDBR

Sensor Batch Dispersion Reduction algorithm

À présent, nous présenterons le fonctionnement de l'algorithme SDBR. Il permet d'une part, d'établir les niveaux de dispersion intra-set individuels et d'autre part, génère une table nous informant sur les corrections à appliquer pour chaque valeur de la plage d'acquisition des différents capteurs, dans le but de réduire cette dispersion.

De par son fonctionnement, l'algorithme ne prend pas en compte de l'aspect temporel ; chaque échantillon recueilli est traité individuellement. Le jeu de données peut être traité de manière totalement aléatoire sans que ça n'affecte les résultats. Trois des jeux de données ne permettent pas d'appliquer l'algorithme SDBR dans plusieurs environnements (Airparif, Lsce, Coimbra). Nous diviserons donc ces jeux de données de manière aléatoire et répèterons cette opération plusieurs fois lors de leur évaluation, pour n'observer que la moyenne des résultats, dans le but de limiter les biais liés au sur-apprentissage.

### 8.8.1. ILLUSTRATION DU PARADIGME SDBR :

Les qualités métrologiques d'un capteur, quel qu'il soit, sont limitées. Les données produites ne représenteront jamais à la perfection la valeur physique réelle du polluant mesuré. Cette imprécision de mesure, aussi faible soit-elle, fait que la fonction des niveaux réels face à la plage de mesure d'un instrument de mesure peut être représentée par l'illustration en vert, Figure 50.

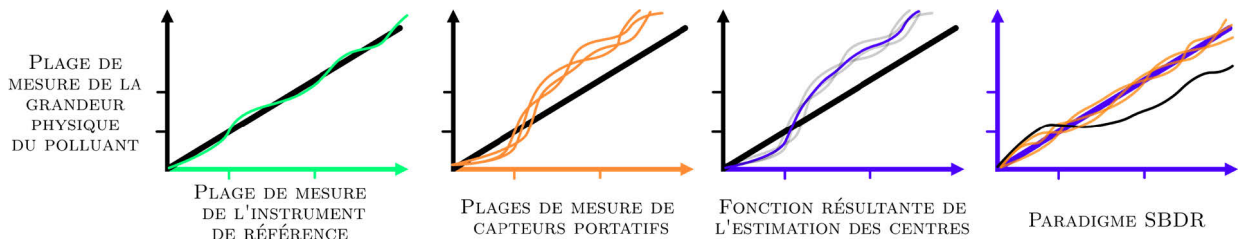


FIGURE 50 : Introduction au référentiel SDBR

La fonction produite par un groupe de capteurs portatif vis-à-vis des niveaux réels dispersera davantage donc que l'instrument de référence, tel qu'illustré par la seconde illustration en orange. Nous émettons la théorie que cette dispersion sera caractéristique au modèle de capteur et à la technologie employée ; les profils de dispersion tracés sont donc similaires. La dispersion observée, pour chacune des valeurs possibles sur la plage de mesures des capteurs, possède une valeur centrale représentative de cette caractéristique ; c'est la fonction résultante de l'estimation des centres, illustrée en bleu.

La SDBR ignore la référence réelle des données et privilégie la réduction de la dispersion intra-set. On cherche donc à réduire l'erreur entre les données produites par les capteurs portatifs (en orange), afin qu'ils se rapprochent au mieux de la fonction produite par l'estimation des centres ; c'est le « paradigme SDBR », tel que présenté sur la dernière illustration.

### 8.8.2. DÉCOMPOSITION DU PROCESSUS SBDR

Nous ferons la description des 7 étapes composant le processus SBDR, complétées par l'illustration du processus, ci-dessous (Figure 51).

#### I. CRÉATION D'UN JEU DE DONNÉES

La première étape du processus SBDR consiste à produire un jeu de données à l'aide d'un set de capteurs réalisant des acquisitions de manière synchronisée. Ils produiront à la fréquence d'échantillonnage paramétrée, autant d'échantillons que de capteurs présents dans le set. Le processus SBDR requiert plusieurs jeux de données, produits dans des environnements différents. Dans notre cas nous disposons des trois jeux de données précédemment décrits.

#### II. PRÉTRAITEMENT

Il est possible durant la création du jeu de données qu'une partie des capteurs dysfonctionnent. Certains points d'acquisition dans le temps ne disposeront donc pas du nombre d'échantillons attendus. Il est donc nécessaire d'exclure ces événements, puisqu'ils ne pourraient permettre d'évaluer avec précision la tendance du set de capteurs. Le nombre de capteurs en fonctionnement minimum est à l'appréciation de l'utilisateur.

Si la fréquence d'échantillonnage employée par les capteurs est hétérogène, alors le jeu de données devra être ré-échantillonné afin de maximiser le nombre d'échantillons disponible par point d'acquisition dans le temps.

#### III. ESTIMATION DES CENTRES

Pour chacun des points d'acquisition dans le temps, nous réaliserons une estimation des centres utilisant un algorithme dédié, tel que définit 8.7 « Algorithme d'estimation des centres ». Cette première estimation des centres décrira donc la tendance du set de capteur dans le temps.

#### IV. CRÉATION DES PROFILS DE CORRECTION

Nous regrouperons pour chaque capteur et pour chaque valeur de centre, les valeurs d'échantillons relevées. Par exemple, si l'estimation des centres, sur tout le jeu de données, estime que la valeur « 4 » a été rencontrée à 10 reprises, alors nous aurons, pour chaque capteur, 10 échantillons associés au centre de valeur 4.

Nous réaliserons pour chacune des valeurs de centre et pour chacun des capteurs, une seconde estimation des centres. Cette estimation consistera cette fois à déterminer, sur l'ensemble de la plage de mesure couverte durant la création du jeu de données, quelles sont les tendances individuelles des capteurs vis-à-vis du reste du set.

Nous obtiendrons ainsi, pour chacun des capteurs, des profils de correction, indiquant pour toute la plage de mesure des capteurs, la correction à appliquer pour se rapprocher au mieux de l'estimation des centrées réalisée à l'étape III.

## V. APPLICATION DE LA CORRECTION

Les profils de correction précédemment créés peuvent alors être appliqués à un sous-ensemble du jeu de données, n'ayant pas été exploité pour la création de ces derniers. Idéalement, on aura divisé le jeu de données de manière aléatoire, en deux sets ; l'un pour générer le profil de correction et un pour l'évaluer. Aucune correction ne pourra être appliquée sur les parties de la plage de mesure n'ayant pas été couvertes durant la création du jeu de données, ou dont les données sont manquantes.

## VI : ÉVALUATION

Nous chercherons premièrement à produire la réduction de la dispersion intra-set la plus importante possible en modulant divers paramètres de l'algorithme SBDR. Ensuite, nous contrôlerons et évaluerons la transposabilité des profils de correction produits, en s'assurant qu'ils permettent une réduction de la dispersion intra-set entre les différents jeux de données. Enfin, nous produirons des profils de correction à partir de l'ensemble des jeux de données et non plus sur des sous-ensembles.

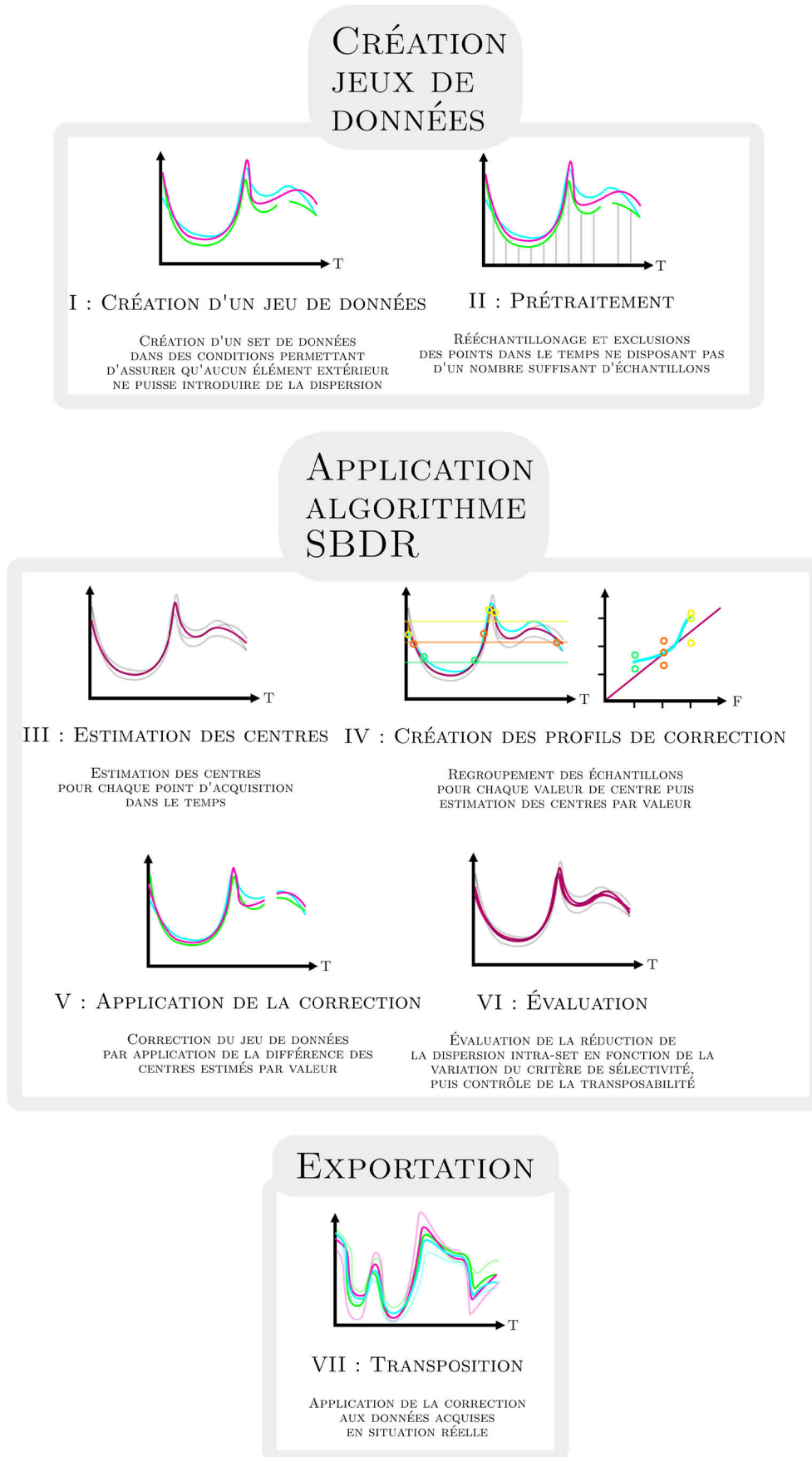
Pour évaluer la réduction des niveaux de dispersion intra-set observée dans les jeux de données corrigés avec l'algorithme SBDR, nous utiliserons le niveau d'écart-type moyen (ETM), correspondant à la moyenne des écarts type entre les capteurs ( $\mathcal{C}$ ), pour chaque acquisition dans le temps ( $\mathcal{E}$ ) ;

$$\text{Écart - type moyen} = \frac{1}{\mathcal{E}} \sum_{i=1}^{\mathcal{E}} \sqrt{\frac{1}{n} \sum_{i=1}^n (C_i - \bar{C})^2}$$

## VII : TRANSPOSITION

Si la transposabilité est avérée, alors nous archiverons les données produites ainsi que les profils de correction individuels des capteurs. Grâce à ces profils de correction, nous serons en mesure de minimiser par post-traitement la dispersion intra-set dans les données acquises en situation réelle.

FIGURE 51 : Illustration du processus SBDR



## 8.9. PARAMÈTRES DE L'ALGORITHME SBDR

### 8.9.1. POIDS DE LA CORRECTION

Les utilisateurs de l'algorithme SBDR ont à leur disposition les moyens de moduler le niveau de correction appliqué en fonction de la quantité de données disponibles par valeur. Il est possible que certains niveaux de la plage de mesure des capteurs ne disposent que de très peu d'occurrences d'échantillons ; typiquement dans les valeurs les plus hautes, où seuls quelques événements auront été relevés durant la création des jeux de données. Dans le cas de la technologie employée (OPCs), les valeurs les plus élevées possèdent l'erreur de mesure la plus importante. Un faible nombre d'échantillons ne permettront pas d'estimer correctement la correction à appliquer. La pertinence de la correction proposée pour des valeurs possédant un nombre limité d'occurrences est donc incertaine. C'est pourquoi il est nécessaire de pondérer l'application des profils de correction en fonction du nombre d'occurrences, pour chacune des valeurs de la plage de mesure des capteurs.

Nous introduirons deux paramètres permettant de moduler le pourcentage de correction appliqué, relativement au nombre d'échantillons disponibles. Grâce à ces deux critères, nous serons en mesure de déterminer les limites hautes et basses, là où la correction sera entièrement appliquée et là où elle n'aura plus d'effet.

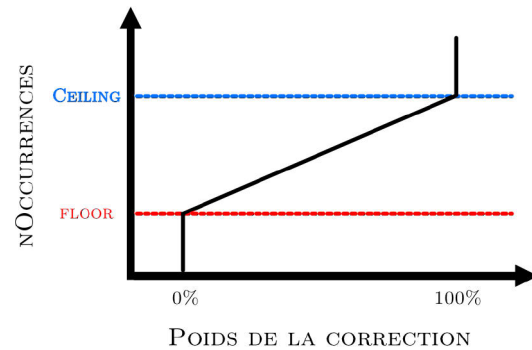


FIGURE 52 : Modulation du poids de la correction

Le premier paramètre (*CeilingRatio*) permet de déterminer le nombre minimum d'occurrences requis pour que la totalité de la correction estimée soit appliquée (*Ceiling*). Le second paramètre (*FloorRatio*) permet de déterminer à partir de quel nombre d'occurrences la correction n'aura plus d'effet (*Floor*). Dans l'algorithme de correction, les profils de correction générés contiennent le compte d'occurrences pour chaque valeur (*nOccurrences*). Le « plafond » (*Ceiling*) à partir duquel la correction sera entièrement appliquée, est déterminé par la multiplication du nombre d'occurrences médian par le paramètre utilisateur ;

$$Ceiling = CeilingRatio \times M(nOccurrences)$$

Le « plancher » (*Floor*) à partir duquel la correction n'aura plus aucun effet, est déterminé par la multiplication du *Ceiling* avec le paramètre utilisateur *FloorRatio* ;

$$Floor = Ceiling \times FloorRatio$$

Le niveau de correction appliqué sera déterminé pour chacune des valeurs de la plage de mesure, en fonction des deux paramètres du nombre d'occurrences disponibles ;

$$Correction\ appliquée = Correction\ totale - Correction\ totale \times \left( 1 - \frac{nOccurrences - Floor}{Ceiling - Floor} \right)$$

Le nombre d'occurrences (*nOccurrences*), lorsque situé entre les seuils *Ceiling* et *Floor*, déterminera donc linéairement le niveau de correction à appliquer, tel qu'illustré Figure 52.

### 8.9.2. MODULATION DES CRITÈRES DE SÉLECTIVITÉ

Nous avons précédemment présenté l'algorithme réalisant l'estimation des centres section 8.7.1 puis présenté les résultats de sa modulation section 8.7.2. Lors de l'application du processus SBDR, l'utilisateur peut choisir de moduler ce critère tout en évaluant la réduction de la dispersion intra-set, permettant de sélectionner les valeurs de critère de sélectivité les mieux adaptées.

Au sein de l'algorithme SBDR, l'algorithme d'estimation des centres est appliqué à trois reprises (désignés par  $C_1$ ,  $C_2$  et  $C_3$ ). Sa première application ( $C_1$ ) permet d'estimer la tendance du set de capteurs, pour chaque point d'acquisition dans le temps. En d'autres termes ; à partir des séries temporelles issues des capteurs, nous produirons une série temporelle image de la tendance du set. À partir de cette nouvelle série temporelle, nous regrouperons les échantillons relevés par les différents capteurs, pour chacune occurrences de chaque valeur. Nous disposerons donc, pour chaque capteur, d'autant de groupes d'échantillons que de valeurs présentes sur la série temporelle image de la tendance du set.

La seconde application de l'algorithme d'estimation des centres ( $C_2$ ) permettra d'établir la valeur centrale, dans chacun de ces groupes. La différence de cette valeur centrale vis-à-vis de la valeur auxquels ces échantillons sont associés représente l'offset à appliquer pour corriger la déviation des capteurs vis-à-vis de la tendance estimée du set. Nous pourrons ainsi produire des profils de correction, qui indiquerons pour chaque valeur et chacun des capteurs, l'offset à appliquer.

Enfin, la dernière application de l'algorithme d'estimation des centres ( $C_3$ ) sera réalisée lorsque nous appliquerons ces profils de correction à un nouveau jeu de données ; il sera alors nécessaire de produire la série temporelle image de la tendance du set de capteurs, permettant de déterminer l'offset pour chacun des points d'acquisition dans le temps.

### 8.9.3. PARAMÈTRES SECONDAIRES

Il existe cinq paramètres secondaires, qu'il est nécessaire de renseigner pour pouvoir exploiter l'algorithme SBDR. Dans certains cas les profils de correction générés ne couvrent pas toute la plage de mesure des capteurs. Le premier paramètre correspond à la possibilité de réaliser une interpolation des valeurs manquantes.

Le deuxième et le troisième paramètre correspondent au nombre de décimales lors du calcul de l'estimation des centres et des profils de correction. Ces deux paramètres sont indépendamment renseignés. Un profil de correction contenant plusieurs décimales peut induire une forme de sur-apprentissage. Le quatrième paramètre correspond à la sélection de la fonction permettant l'estimation des centres. C'est avec ce paramètre que nous pourrons choisir d'utiliser les centres médians plutôt que l'algorithme d'estimation des centres.

Enfin, le dernier paramètre correspond à la possibilité de retourner à l'utilisateur, lors de l'application des profils de correction, une alerte indiquant les valeurs de centre ne pouvant être corrigées sur le jeu de données cible.

Dans notre démonstration, nous avons utilisé les paramètres suivants ; sans interpolation des valeurs manquantes, 1 décimale pour les profils de correction et 0 pour l'évaluation de la tendance du set. Nous fournissons les paramètres  $C_1$ ,  $C_2$  et  $C_3$  ; nous n'avons donc pas exploité les centres médians.

### 8.10. APPLICATION DE L'ALGORITHME SBDR

Pour démontrer l'application de l'algorithme SBDR, nous modulerons les 5 paramètres de l'algorithme SBDR et relèverons, pour chaque scénario, lesquels ont permis d'obtenir la réduction de la dispersion intra-set la plus importante

. Voici les paramètres explorés ;

- Les valeurs de *CeilingRatio* explorées : 0.1, 0.25, 0.5, 1, 2 et 3.
- Les valeurs de *FloorRatio* explorées : 0.05, 0.1, 0.25 et 0.5.

Les valeurs de  $C_1$ ,  $C_2$ ,  $C_3$  explorées : 0.1, 0.25, 0.50 et 0.75.

Les résultats ont été archivés dans une matrice possédant 5 dimensions ; chacune des dimensions représentant un des 5 paramètres. C'est donc 1536 calculs de la dispersion intra-set qui ont été nécessaires pour évaluer chacun des scénarios. Le temps de calcul requis n'est donc pas négligeable malgré l'utilisation d'un processeur de dernière génération [145].

Nous explorerons divers scénarios et évaluerons la réduction de la dispersion intra-set résultante de l'application de l'algorithme SBDR. Les différents scénarios seront par la suite décrits de la sorte ;

*Jeu de données cible*  $\leftarrow$  (*Jeu de données exploitées pour la création des profils de correction*)

La combinaison de plusieurs jeux de données peut également être exploitée :

*Jeu de données cible*  $\leftarrow$  (*Jeu de données A + Jeu de données B*)

#### 8.10.1. PERFORMANCE SANS TRANSPOSITION

Nous évaluerons les performances de l'algorithme SBDR, appliquant les profils de correction sur les jeux de données ayant permis de créer ces derniers ; sans transposition des profils de correction. Nous évaluons ainsi la capacité de l'algorithme SBDR à estimer les tendances individuelles des capteurs pour un set de données. L'ETM résultant de l'application de la correction en utilisant la totalité des données est renseigné dans le Tableau 9 ; « ETM post jeu entier ». Dans ce scénario, les performances seront optimales du fait du phénomène de sur-apprentissage. À partir de ces résultats, nous avons relevé les meilleures combinaisons de paramètres.

Les jeux de données seront ensuite divisés en sous-ensembles de manière aléatoire dans l'objectif de limiter le sur-apprentissage. 40 % des échantillons permettront la création des profils de correction et 60 % permettront d'évaluer l'ETM après application des profils de correction. Nous avons utilisé les paramètres précédemment identifiés. Les résultats fluctueront du fait du caractère aléatoire de la division des données. En conséquence, nous réaliserons la moyenne de l'ETM résultant de 10 corrections successives. Elle sera désignée dans le Tableau 9 par « ETM post moyen (x10) ».

À titre informatif, les temps de calcul furent ;

- 1 heures et 58 minutes pour le scénario Banc1  $\leftarrow$  Banc1.
- 2 heures et 18 minutes pour le scénario Banc2  $\leftarrow$  Banc2.
- 3 heures et 19 minutes pour le scénario Banc3  $\leftarrow$  Banc3.
- 0 heures et 9 minutes pour le scénario Lsce  $\leftarrow$  Lsce.
- 1 heures et 31 minutes pour le scénario Airparif  $\leftarrow$  Airparif.
- 6 heures et 45 minutes pour le scénario Coimbra  $\leftarrow$  Coimbra.

TABLEAU 9 : Performance de l'algorithme SBDR sans transposition

Scénarios	ETM pré-correction	ETM post jeu entier	ETM post moyen x10	% Correction jeu entier	% Correction division x10	Ceiling Ratio	Floor Ratio	$C_1, C_2$ et $C_3$
Banc1 $\leftarrow$ Banc1	2.24	2.08	2.09	7.06 %	6.75 %	0.1	0.05	0.75, 0.75, 0.75
Banc2 $\leftarrow$ Banc2	3.52	3.19	3.22	9.57 %	8.63 %	0.1	0.05	0.75, 0.75, 0.75
Banc3 $\leftarrow$ Banc3	3.03	2.86	2.89	5.71 %	4.74 %	0.1	0.05	0.75, 0.75, 0.75
Lsce $\leftarrow$ Lsce	1.74	1.12	1.17	35.82 %	33.03 %	0.1	0.05	0.75, 0.75, 0.75
Airparif $\leftarrow$ Airparif	4.63	1.99	2.02	56.86 %	56.22 %	0.1	0.05	0.75, 0.75, 0.75
Coimbra $\leftarrow$ Coimbra	2.97	2.49	2.52	16.25 %	15.18 %	0.5	0.05	0.1, 0.75, 0.5

Les paramètres optimums sont identiques pour l'ensemble des jeux de données, à l'exception de Coimbra dont le nombre d'échantillons et le nombre de capteurs est considérablement différent du reste de jeux de données. En sélectionnant des paramètres allant au-delà des minimas (pour *CeilingRatio* et *FloorRatio*) et des maximums (pour  $C_1, C_2$  et  $C_3$ ), nous n'avons pas observé un gain significatif qui justifierait de réaliser à nouveau l'ensemble des calculs. Il est cependant nécessaire d'affiner le choix des paramètres explorés.

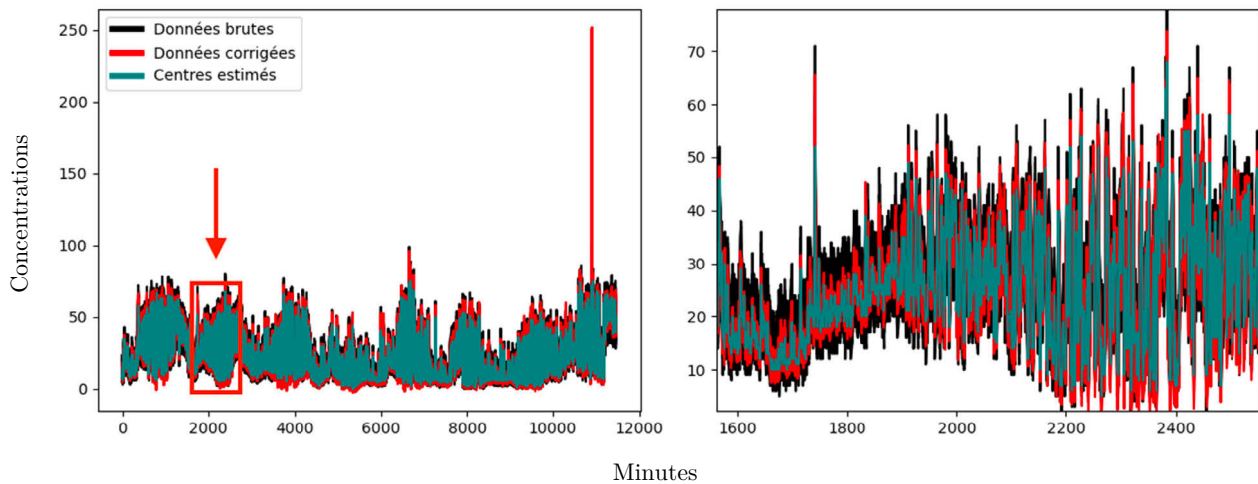


FIGURE 53 : Visualisation de la correction de la dispersion intra-set (Airparif)

Nous observons que dans le cas des jeux de données Banc1, 2 et 3, le niveau de réduction de la dispersion intra-set est faible comparativement aux jeux de données Lsce, Airparif et Coimbra. Les modules OPC utilisés pour la production de ces jeux de données sont neuf et présentent certainement une dispersion moins importante que dans le cas d'OPC ayant fonctionné durant plusieurs mois. Enfin, nous visualiserons (Figure 53) le résultat de



l'application des profils de correction, dans le cas du jeu de données Airparif, où nous avons observé la réduction de la dispersion intra-set la plus importante (56 %).

#### 8.10.2. INJECTION DE DONNÉES ISSUES D'UN INSTRUMENT DE RÉFÉRENCE

Dans le cas du jeu de données « Airparif », nous disposons, en plus des données des capteurs Cairsens NO<sub>2</sub>, des données recueillies par un instrument de référence. Nous chercherons à démontrer que ces données ne permettent pas de réduire davantage la dispersion intra-set, vis-à-vis de l'estimation des centres produite par l'algorithme SBDR.

Les données de cet instrument de référence possèdent une fréquence d'échantillonnage de 1 seconde. Elles ont donc été ré-échantillonnées et moyennées à la minute, afin de pouvoir être mises en relation avec les données recueillies par les capteurs portatifs CANARIN II. Au total c'est 11460 minutes qui ont pu être mises en relation avec les acquisitions produites par les capteurs portatifs. Sur ces données ci, l'ETM sans correction est de 4.63. Nous avons utilisé les paramètres déterminés précédemment.

Les données issues de l'instrument de référence seront injectées à l'algorithme de façon à être exploitées comme étant le résultat de l'estimation des centres (là où est appliqué). Ainsi, l'algorithme effectuera ce que l'on pourrait apparenter à une calibration standard ; il cherchera à rapprocher les valeurs produites par chaque capteur à l'instrument de référence. Le résultat de ce scénario est renseigné dans le Tableau 10 « Airparif  $\leftarrow$  ( Airparif )  $\sim$  Ref ». Ensuite, nous avons procédé à la même expérimentation, mais cette fois en retirant certains capteurs CANARIN du set (Ext0, 1 et 2), puis en injectant leurs données afin qu'elles soient interprétées comme étant le résultat de l'estimation des centres.

TABLEAU 10 : Injection de données d'un instrument de référence

Scénarios	ETM pré-corrrection	ETM post jeu entier	% Correction jeu entier
Airparif $\leftarrow$ ( Airparif )	4.63	1.99	56.86 %
Airparif $\leftarrow$ ( Airparif ) $\sim$ Ref	4.63	2.12	54.16 %
Airparif $\leftarrow$ ( Airparif ) $\sim$ Ext0	4.75	2.03	57.18 %
Airparif $\leftarrow$ ( Airparif ) $\sim$ Ext1	4.53	2.15	52.49 %
Airparif $\leftarrow$ ( Airparif ) $\sim$ Ext2	4.63	2.15	53.50 %

Notons que dans le scénario « Airparif  $\leftarrow$  ( Airparif )  $\sim$  Ext0 », la réduction de la dispersion intra-set est plus importante (57.18 % contre 56.86 % pour le scénario original), car un capteur a été retiré du set. On observe que les données issues de l'instrument de référence ne permettent pas de produire une réduction de la dispersion intra-set supérieure à ce qui est réalisable à partir de l'algorithme d'estimation des centres. Nous démontrons ainsi que lorsqu'il s'agit de réduire la dispersion intra-set, la présence d'un instrument de référence n'est pas nécessaire.

### 8.10.3. SURVEILLANCE PRÉVENTIVE

Nous pouvons utiliser l'algorithme SBDR dans le but d'évaluer la dispersion intra-set, sans appliquer de correction. En effet, les profils de correction générés permettent de connaître le profil de dispersion des capteurs.

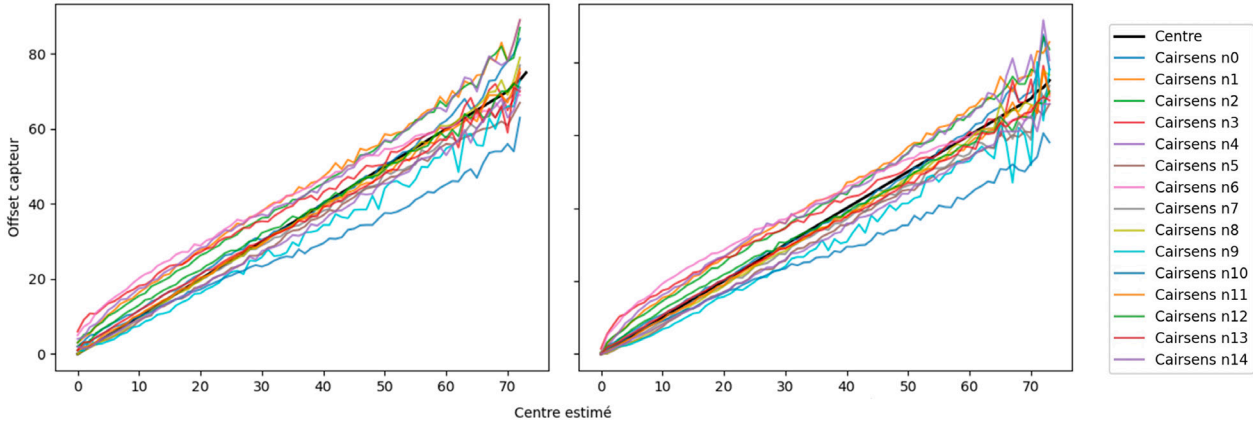


FIGURE 54 : Profils de correction (Airparif)

La FIGURE 54 présente deux profils de correction, sur la plage de mesure couverte (de 0 à 80 ppb). Chacun de ces profils de correction est produit à partir d'une moitié du jeu de données Airparif. On observe que ces profils sont visuellement très proches. Le même type de visualisations peuvent ainsi être utilisées pour évaluer l'évolution du comportement des capteurs dans le temps.

Lorsqu'un capteur présente un comportement semblant anormal, il est alors possible de se référer aux profils précédemment produits pour établir si le dit capteur doit être davantage surveillé, ou exclu du set. L'évaluation intra-set et la production de profils de correction peuvent ainsi constituer une forme de surveillance préventive des capteurs portatifs de la qualité de l'air.

### 8.10.4. TRANSPOSABILITÉ DES PROFILS DE CORRECTION

Nous chercherons à démontrer la transposabilité de la dispersion intra-set à partir des jeux de données Banc1, Banc2, Banc3. Pour chaque jeu de données, la réduction de la dispersion intra-set a été évaluée en appliquant les profils de correction obtenus en combinant les deux autres jeux de données restants. La réduction intra-set du jeu de données Banc1 a donc été évaluée en appliquant les profils de correction produits à partir du jeu de données combinant les jeux de données Banc2 et Banc3.

À titre informatif, les temps de calcul furent ;

- 2 heures et 6 minutes pour la combinaison Banc1  $\leftarrow$  (Banc2 + Banc3).
- 2 heures et 19 minutes pour la combinaison Banc2  $\leftarrow$  (Banc1 + Banc3).
- 2 heures et 39 minutes pour la combinaison Banc3  $\leftarrow$  (Banc1 + Banc2).

Nous établirons la performance du processus SBDR en simulant les conditions réelles. Pour ce faire, nous appliquerons les profils de correction produits avec deux des jeux de données, sur un troisième (exemple : Banc1  $\leftarrow$  (Banc2 + Banc3)). Nous évaluerons la réduction de la dispersion intra-set dans les trois cas de figure. Ainsi, nous considérons que les jeux de données cible représentent les données acquises en situation réelle et devant être corrigées. Le Tableau 11 fournit le détail des résultats optimaux obtenus en modulant les paramètres de l'algorithme SBDR.

Cependant, en conditions réelles, ces paramètres optimaux nous sont inconnus puisque la dispersion des données acquises, lorsque les capteurs sont portés, est impossible à évaluer. Nous avons donc également renseigné les résultats obtenus en utilisant les paramètres optimaux déterminés précédemment, Tableau 10.

TABLEAU 11 : Résultat de la transposition des profils de dispersion

Scénarios	ETM pré-corrrection	ETM post-corrrection	Pourcentage de correction	Ceiling Ratio	Floor Ratio	C <sub>1</sub> , C <sub>2</sub> , C <sub>4</sub>
Banc1 $\leftarrow$ (Banc2 + Banc3)	2.24	2.17	3.21 %	3	0.05	0.1, 0.75, 0.25
Banc2 $\leftarrow$ (Banc1+ Banc3)	3.52	3.35	5.02 %	0.2	0.05	0.75, 0.75, 0.75
Banc3 $\leftarrow$ (Banc1 + Banc2)	3.03	2.96	2.32 %	1.0	0.05	0.75, 0.75, 0.25
Banc1 $\leftarrow$ (Banc2 + Banc3)	2.24	2.17	3.17 %	0.1	0.05	0.75, 0.75, 0.75
Banc2 $\leftarrow$ (Banc1+ Banc3)	3.52	3.35	4.99 %	0.1	0.05	0.75, 0.75, 0.75
Banc3 $\leftarrow$ (Banc1 + Banc2)	3.03	2.98	1.84 %	0.1	0.05	0.75, 0.75, 0.75

## 8.11. DISCUSSION ET CONCLUSIONS

Nous avons introduit un processus permettant de minimiser l'erreur de mesure au sein de flottes de capteurs portatifs, dans le but d'améliorer la qualité des données acquises en situations réelles.

Des efforts importants ont été réalisés, d'une part pour mettre en évidence la dispersion intra-set et d'autre part, pour développer les outils permettant sa réduction ; un banc d'essai et un algorithme dédié.

Les résultats relatifs à la transposabilité des profils de correction sont positifs et mettent en évidence la pertinence du processus SBDR.

### 8.11.1. MULTI-DIMENSIONNALITÉ

L'algorithme tel que présenté, intègre uniquement une dimension ; les données issues des capteurs dont on désire réduire la dispersion. Or, les données d'humidité et de température peuvent impacter les niveaux de dispersion des capteurs de type OPC tel que montré dans l'article 0 «

Conditions météorologiques et variations de la dispersion ». Il semble donc nécessaire d'améliorer l'algorithme afin de produire une correction plus fine, multidimensionnelle, qui intégrerait les données issues des micro-capteurs embarqués — comme c'est le cas avec les capteurs CANARINS —, ou issus de capteurs dédiés — dans notre cas, un hygromètre et un

thermomètre de précision — qui accompagneraient la flotte de capteurs, tel que suggéré page 62. Dans ce scénario, pour être en mesure de tenir compte de ces paramètres supplémentaires (dimensions), une variabilité élevée des niveaux de polluants, d'humidité et de température seraient alors nécessaires pour produire un jeu de données permettant de correctement caractériser la dispersion. Seul un laboratoire équipé d'une enceinte climatique et d'un outil permettant l'émission de polluants de manière contrôlée serait en mesure de produire ces données. Nous nous écarterions donc de l'objectif originel ; fournir un outil accessible et ne nécessitant pas d'infrastructure dédiée. Par ailleurs, les jeux de données exploités dans cette partie sont pour la plupart tout juste suffisants pour caractériser les données sur une dimension — dans notre cas les particules —. L'ajout de paramètres (de dimensions supplémentaires) multiplierait donc le nombre de données nécessaires et donc allongerait inévitablement les durées d'acquisition requises.

#### 8.11.2. *PERSPECTIVES*

Des améliorations substantielles pourraient découler de l'optimisation de l'algorithme, notamment en améliorant l'automatisation de l'exploration des paramètres de l'algorithme SBDR.

L'évaluation de jeux de données produit à plusieurs mois d'intervalles permettrait d'observer l'évolution dans le temps des tendances présentes au sein d'un set de capteurs. Ils permettraient aussi d'observer potentiellement une réduction de la dispersion intra-set plus importante, vis-à-vis des OPC neufs employés pour les jeux de données Banc1, 2 et 3.

Il est nécessaire d'évaluer si les profils de correction peuvent être produits sans l'utilisation du banc d'essai mobile d'évaluation de la dispersion. Une future validation des capteurs du projet POLLUSCOPE pourrait permettre de répondre à ces questions.

La majorité des jeux de données exploités sont le produit de capteurs homogènes ; issus du même fabricant et de modèle identique. La SBDR peut cependant être utilisée avec un set de capteurs hétérogènes, lorsque l'on désire réduire la dispersion d'une flotte de capteurs mesurant les mêmes paramètres mais de marque ou de conception différentes. Plus d'efforts sont donc nécessaires pour connaître les performances du processus SBDR lorsque appliqué à d'autres technologies de mesure, pour d'autres polluants, ou encore pour d'autres types de capteurs environnementaux portatifs.

## 9. CONTEXTUALISATION DES DONNÉES

Grâce à notre implication dans le projet POLLUSCOPE, nous avons déterminé que, dans le cadre d'études épidémiologiques visant à analyser l'exposition individuelle d'un nombre réduit d'individus, des efforts supplémentaires importants doivent être réalisés pour permettre de mieux contextualiser les données acquises. Nous présenterons les résultats de l'analyse de données produites dans le cadre de ce projet, ainsi que les difficultés rencontrées. Les observations que nous réaliserons permettront d'introduire (Partie 10) une approche spécifique pour l'acquisitions de données contextuelles ; au travers de la conception d'une architecture innovante des capteurs portatifs, dont l'objectif final sera d'améliorer l'interprétation possible des données d'exposition individuelle à la pollution atmosphérique

Dans cette partie, nos résultats démontreront la pertinence de l'utilisation de capteurs portatifs pour la mesure de l'exposition individuelle à la pollution de l'air. Nous observerons la variabilité des niveaux d'exposition en fonction des micro-environnements traversés par les individus, en exploitant des données contextuelles annotées manuellement par les participants.

### 9.1. EXPOSITION INDIVIDUELLE EN FONCTION DE LA MOBILITÉ

Nous présentons le traçage spatial des données de  $PM_{10}$  acquises par un participant, collaborateur travaillant au CEREMA, membre du consortium POLLUSCOPE, totalisant six jours et demi de données (156 heures) provenant de deux sessions d'échantillonnage en novembre 2019. L'exposition individuelle aux PM a été exprimée en concentration ( $\mu\text{g}/\text{m}^3$ ). Les données sur la mobilité ont été acquises grâce au CANARIN II et au GPS embarqué d'une tablette Android. Ce participant a été sélectionné car il a accepté de publier ses données publiquement. Veuillez noter que les noms des rues ont été masqués sur les figures qui suivent.

Les données se composent de deux périodes d'échantillonnage de 8 jours chacune (384 heures au total). Le Tableau 12 indique, pour chacune des deux périodes d'échantillonnage, les niveaux moyens de pollution atmosphérique en  $\mu\text{g}/\text{m}^3$  et le décompte du nombre total d'échantillons enregistrés. Tous les capteurs utilisés ont été paramétrés pour échantillonner la pollution de l'air une fois par minute. Au total, 6040 acquisitions ont été effectuées pour la période A et 4227 acquisitions pour la période B. Cela représente environ la moitié du nombre maximal d'acquisitions qu'il est théoriquement possible de produire ; 8 jours étant équivalents à 11520 minutes.

TABLEAU 12 : Aperçu des deux périodes d'acquisition, valeurs des polluants en  $\mu\text{g}/\text{m}^3$

Période	Moyenne $PM_1$	Moyenne $PM_{2.5}$	Moyenne $PM_{10}$	Moyenne $NO_2$	Moyenne BC	Compte PM	Compte $NO_2$	Compte BC
A : 10/19/19 –10/27/19	8.1	11.6	19	5.8	0	5325	6040	0
B : 11/14/19 –11/22/19	8.6	14.5	15	7.9	1226.0	4027	4227	2628

Le participant a été exposé à des niveaux de pollution atmosphérique plus élevés au cours de la deuxième période (Tableau 12). Nous avons réalisé l’affichage de la distribution des données d’activité (Figure 55). Nous ne possédons pas d’information contextuelle relative à l’activité pour 55 % des acquisitions (« Période A + B »). Ces données, parce qu’elles sont annotées manuellement, peuvent contenir des erreurs et nécessiter un post-traitement. On observe qu’un pourcentage plus élevé d’acquisitions sont associées aux activités « Voiture » et « Bureau » lors de la Période B. Cependant, compte-tenu du pourcentage d’acquisitions où le contexte est « Non déclaré », toute différence observée entre ces deux périodes pourrait s’expliquer par la surreprésentation d’activités spécifiques. Cela indique premièrement que les capteurs ne devraient jamais être arrêtés durant la période d’acquisition. Deuxièmement, la présence de la totalité des annotations qui contextualisent ces données sont nécessaires pour permettre une analyse pertinente.

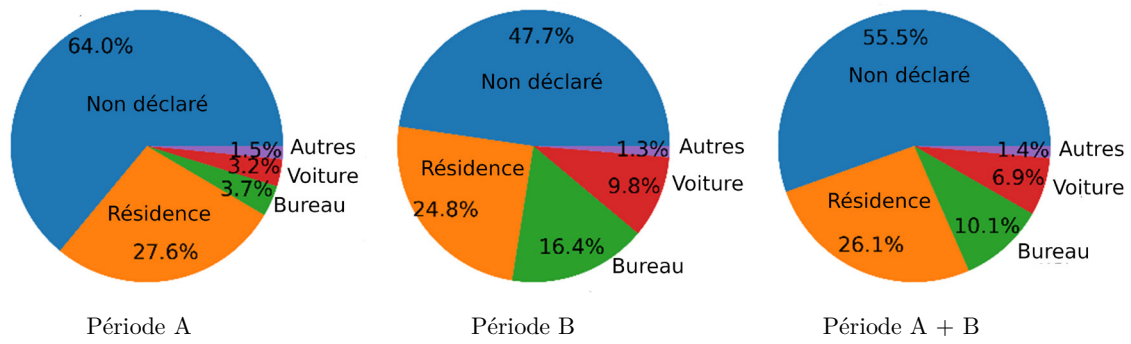


FIGURE 55 : Répartition des activités par période.

La Figure 56 introduit deux points d’intérêt (« Point Of Interest » ou P.O.I.), encerclés en noir, auxquels nous allons nous intéresser. Le reste de la carte montre l’ensemble du mouvement spatial enregistré par le GPS sur les deux périodes. Au nord-ouest, on relève (en rouge pale) quelques points d’acquisition désignant un unique trajet à l’écart d’une zone urbaine, se traduisant par une faible exposition aux  $PM_{10}$ . On peut observer que les niveaux de  $PM_{10}$  sont plus élevés le long du boulevard périphérique de Paris (nord-est de la carte) et diminuent lorsque l’on s’éloigne de l’axe routier principal (ouest).

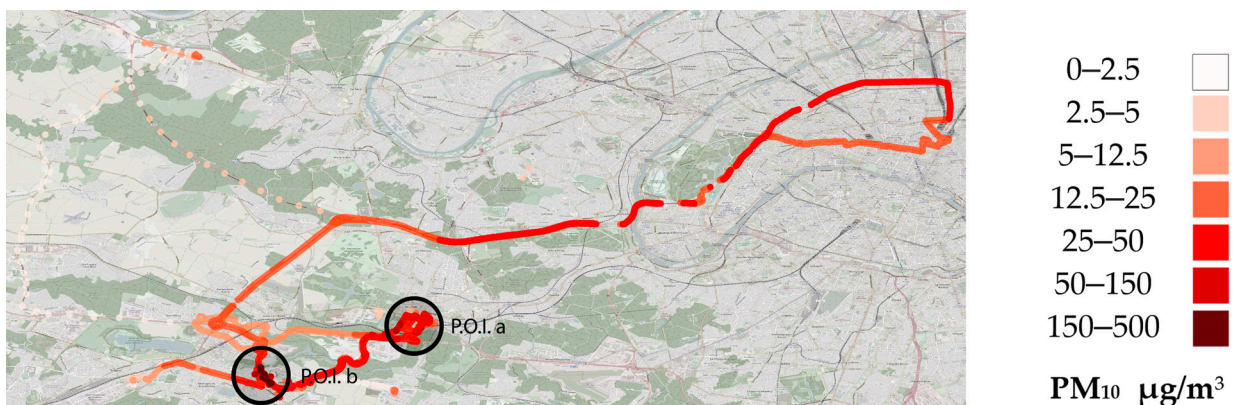


FIGURE 56 : Traçage spatial des  $PM_{10}$  ( $\mu g/m^3$ ) pour les deux périodes A et B

Sur la Figure 57, au sud, un trajet vers un « Magasin » est marqué par une terminaison rose. L'exposition aux  $PM_{10}$  y est sensiblement plus élevée (voir Figure 56). Les deux P.O.I. sur la Figure 57 contiennent donc des acquisitions réalisées autour d'un lieu de travail ou de vie.

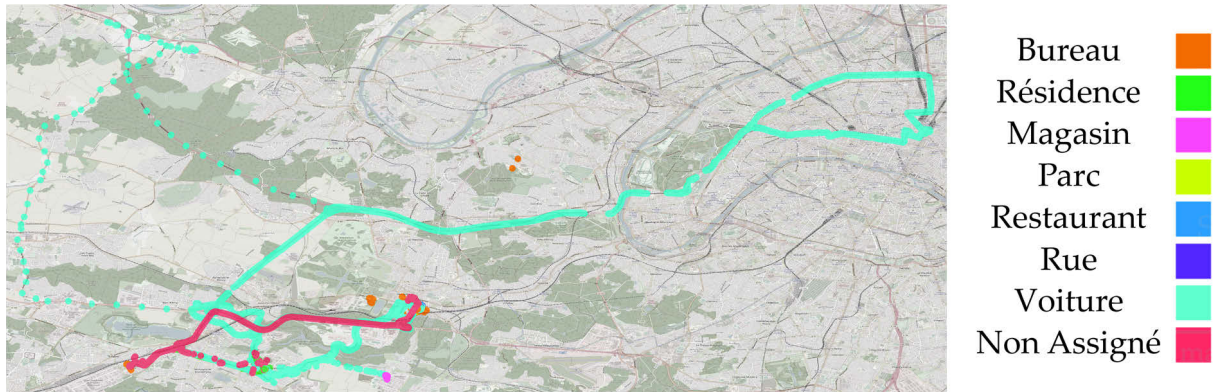


FIGURE 57 : Traçage spatial des annotations d'activité pour les deux périodes A et B

Un zoom sur le P.O.I. A (Figure 58), permet de confirmer que nous examinons le lieu de travail, tel qu'annoté par le volontaire. Le tracé rouge (visible également sur la Figure 57) correspond à un trajet en voiture ou en transports en commun dont ni les informations contextuelles, ni les acquisitions des niveaux de concentrations de  $PM_{10}$  ne sont disponibles ; seules les données GPS ont été enregistrées.

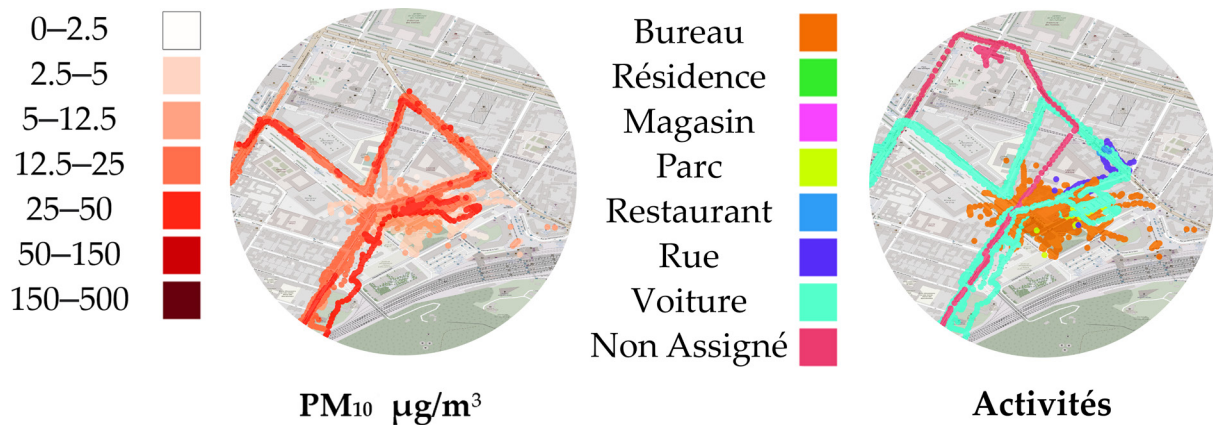


FIGURE 58 : Point d'intérêt géographique A

Le cluster « Bureau », montre un cas très commun de bruit de données GPS brutes causé par une mauvaise réception du signal GPS ; la précision de la localisation chute à l'intérieur des bâtiments. Les niveaux de  $PM_{10}$  enregistrés en air intérieur sont plus faibles que dans la zone urbaine environnante. Le cluster à l'ouest dans la Figure 57 et la Figure 56 correspond à un lieu de travail secondaire : dans une zone moins urbanisée, où les niveaux de  $PM_{10}$  plus faibles.

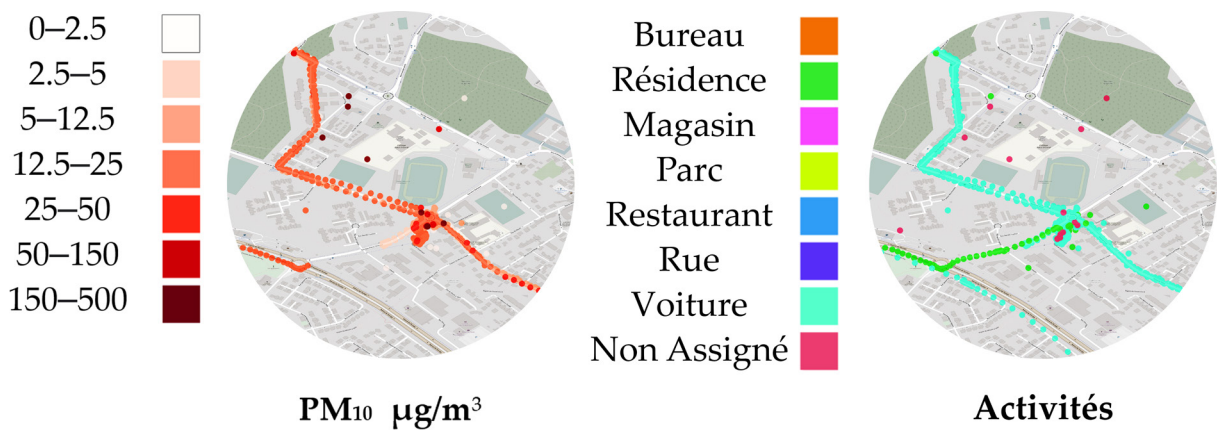


FIGURE 59 : Point d'intérêt géographique B

Le P.O.I. B (FIGURE 59) contient quelques acquisitions localisées de manière aléatoire et possédant des niveaux de PM<sub>10</sub> élevés. Ces quelques points autour de notre P.O.I. représentent en réalité 6905 échantillons dont les acquisitions les plus élevées sont affichées en premier. Cela révèle un emplacement intérieur où la réception GPS n'était pas possible. Les échantillons sont donc tous affichés sur le dernier emplacement identifié. Remarquez comment de multiples trajets en voiture mènent à ce groupe. L'un de ces trajets est accidentellement identifié comme étant le domicile (« Résidence ») au lieu d'indiquer un transport en véhicule (« Voiture » en vert). Cette erreur, si elle n'est pas traitée, aura un impact négatif sur l'estimation de l'exposition individuelle à la pollution de l'air dans la résidence de notre participant ; les niveaux de concentrations de PM<sub>10</sub> relevés dans l'environnement « Voiture » étant plus élevés.

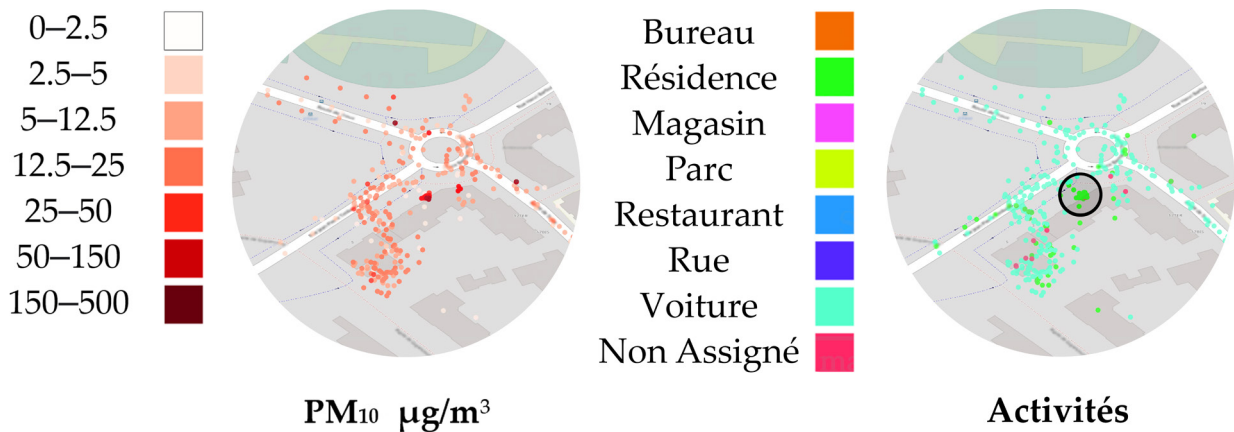


FIGURE 60 : Point d'intérêt B, gros plan sur « Résidence »

En zoomant davantage sur le cluster (Figure 60), on découvre sans surprises la résidence du participant. En dessous du cluster, parmi le bruit, on peut deviner où le véhicule utilisé est supposément garé. Les 321 échantillons de PM<sub>10</sub> encerclés sont fortement concentrés sur des coordonnées précises, indiquent ce qui pourrait être l'emplacement réel du lieu de vie du participant. L'acquisition de ces coordonnées pourrait avoir été réalisé à proximité d'une fenêtre. Cependant, nous avons confirmé avec le participant qu'aucun des tracés GPS présenté ne correspond à l'emplacement réel du lieu de « Résidence ». Sans plus d'informations



contextuelles, à cause de l'absence de réception GPS en intérieur, ces acquisitions spatialement concentrées pourraient induire en erreur les algorithmes visant à remplir ou à corriger les données d'activité. Les niveaux estimés d'exposition individuelle aux  $PM_{10}$  au domicile (« Résidence »), résultante du regroupement des données GPS, seraient alors considérablement biaisés.

TABLEAU 13 : Niveau moyen de pollution atmosphérique à partir de données annotées,  $\mu\text{g}/\text{m}^3$

Activité	$PM_{1.0}$	$PM_{2.5}$	$PM_{10}$	$NO_2$	BC
Bureau	2.3	3.55	3.9	5.7	586.8
Résidence	10.5	15.1	16.4	6.2	610.0
Magasin	2.7	4.2	4.9	10.4	NA
Parc	3.3	4.9	5.3	16.5	409.9
Restaurant	3.6	5.6	6.2	11.2	NA
Rue	3.0	4.6	5.1	18.5	NA
Voiture	9.2	13.1	14.3	20.6	2937.8

Le Tableau 13 indique les niveaux moyens de pollution atmosphérique par activité, en fonction des annotations de l'utilisateur. Les niveaux de  $PM_{10}$  dans « Bureau » sont faibles, comme observé précédemment sur la Figure 58. « Résidence » présente des moyennes d'exposition aux PM les plus élevées. Les niveaux de  $NO_2$  et de BC pendant les activités « Rue » (déplacements piétons dans la rue) et « Voiture » (déplacements en véhicule motorisé) sont les plus importants. « Voiture » est l'activité qui présente l'exposition globale à la pollution atmosphérique la plus élevée. La valeur moyenne des  $PM_{10}$  dans la région encerclée sur la Figure 60, est de  $34,1 \mu\text{g}/\text{m}^3$ , soit le double des niveaux de concentrations obtenus en exploitant uniquement données de contextualisation (Tableau 13, « Résidence »).

## 9.2. OBSERVATIONS

Nous avons montré comment les données des capteurs portatifs de pollution atmosphérique peuvent être interprétées en fonction des données de géolocalisation et des informations contextuelles fournies volontairement par l'utilisateur des capteurs. Nous avons également vu comment un examen scrupuleux de ces données pouvait en partie corriger les cas où l'utilisateur négligerait leur annotation. Des algorithmes de complétion partielle et correction ont été développés dans le cadre du projet POLLUSCOPE. Leur utilisation permet de pallier partiellement à ces manquements, mais ces algorithmes ne seraient en mesure de corriger les données recueillies dans le cas d'importants manquements.

Dans le cadre de l'évaluation de l'exposition individuelle, les participants sont tenus d'être constamment conscients de leur obligation de maintenir les appareils chargés, de se souvenir de les porter et de réaliser l'annotation du suivi de leurs activités quotidiennes sur une tablette. Ils doivent donc effectuer de multiples opérations manuelles pendant les moments les plus chargés de leur journée. Dans ce contexte, une application permettant de réaliser l'annotation des activités a été développée sous notre supervision, puis disséminée par le projet POLLUSCOPE. Elle a facilité cette étape d'annotation qui se faisait traditionnellement sur des carnets de notes.

En plus des activités présentées ci-dessus, davantage d'informations contextuelles pouvaient être annotées par les participants. L'application listant les causes possibles de changements de niveaux de pollution de l'air ; la proximité de toute source de pollution, la soudaine ventilation via l'ouverture d'une fenêtre, la proximité avec un fumeur, ou la cuisson d'aliments ; chacun de ces événements pouvant provoquer un changement de micro-environnement. Cependant nous avons observé un intérêt moindre pour le recueil de ce type d'information.

L'application mobile est fournie par défaut sur la tablette qui accompagne le kit de capteurs. Pour les participants équipés de smartphones Android, elle a été rendue optionnellement disponible à l'installation. Dans ce cas de figure, les participants ont alors été en mesure de réaliser l'annotation de leurs activités sans requérir d'accéder à la tablette, qui est généralement transportée à l'aide d'un sac à dos.

### 9.3. DISCUSSION ET CONCLUSIONS

Nous avons démontré, avec un participant de POLLUSCOPE l'avantage que présentent ces données contextuelles lorsque mises en relation avec les niveaux de concentration des polluants. Elles permettent de mieux comprendre les sources majeures de polluants, au niveau individuel et donc les risques associés pour la santé de chaque volontaire.

Nous avons mis en évidence les limitations technologiques existantes ; en observant que les volontaires, lorsque sollicités pour produire des données contextuelles, pouvaient facilement être une source d'erreur. Pire encore, ils peuvent choisir délibérément d'abandonner leurs obligations à mi-parcours de la période de mesure. Les capteurs portatifs de pollution atmosphérique produisent des données granulaires, mais des informations clés sur le contexte de leur acquisition sont nécessaires pour mieux comprendre en détail les causes sous-jacentes des changements de niveaux de polluants.

## 10. ÉVOLUTION DES ARCHITECTURES DE CAPTEURS

### 10.1. INTRODUCTION

Nous avons vu (Partie 9) comment des informations pertinentes peuvent être obtenues directement auprès de l'utilisateur des capteurs ; dans le cadre du projet POLLUSCOPE, nous avons sollicité nos participants, avant de leur remettre un ensemble de capteurs, pour qu'ils enregistrent activement leur activité ou les événements pouvant causer des évolutions importantes sur les niveaux de polluants relevés. Nous nous sommes toutefois appuyés uniquement sur leur vigilance et leur bonne volonté. Il nous manque des informations cruciales lors des événements de pollution ; les participants n'ont pas possibilité de savoir lorsque les capteurs qu'ils portent détectent une baisse ou une augmentation des valeurs de concentration de polluants. En conséquence, ils ne sont donc pas en mesure d'annoter des informations contextuelles pouvant expliquer la majorité des événements observés.

### 10.2. ILLUSTRATION DE L'ARCHITECTURE

Nous avons illustré ci-dessous l'architecture décrite dans cette partie. Au centre se trouve l'appareil de gestion communiquant avec l'utilisateur et un serveur distant. Autour de l'appareil de gestion, plusieurs capteurs portatifs forment un réseau local. À proximité de certains éléments se trouvent des lettres (A à I) auxquelles nous ferons référence pour décrire le fonctionnement de cette architecture.

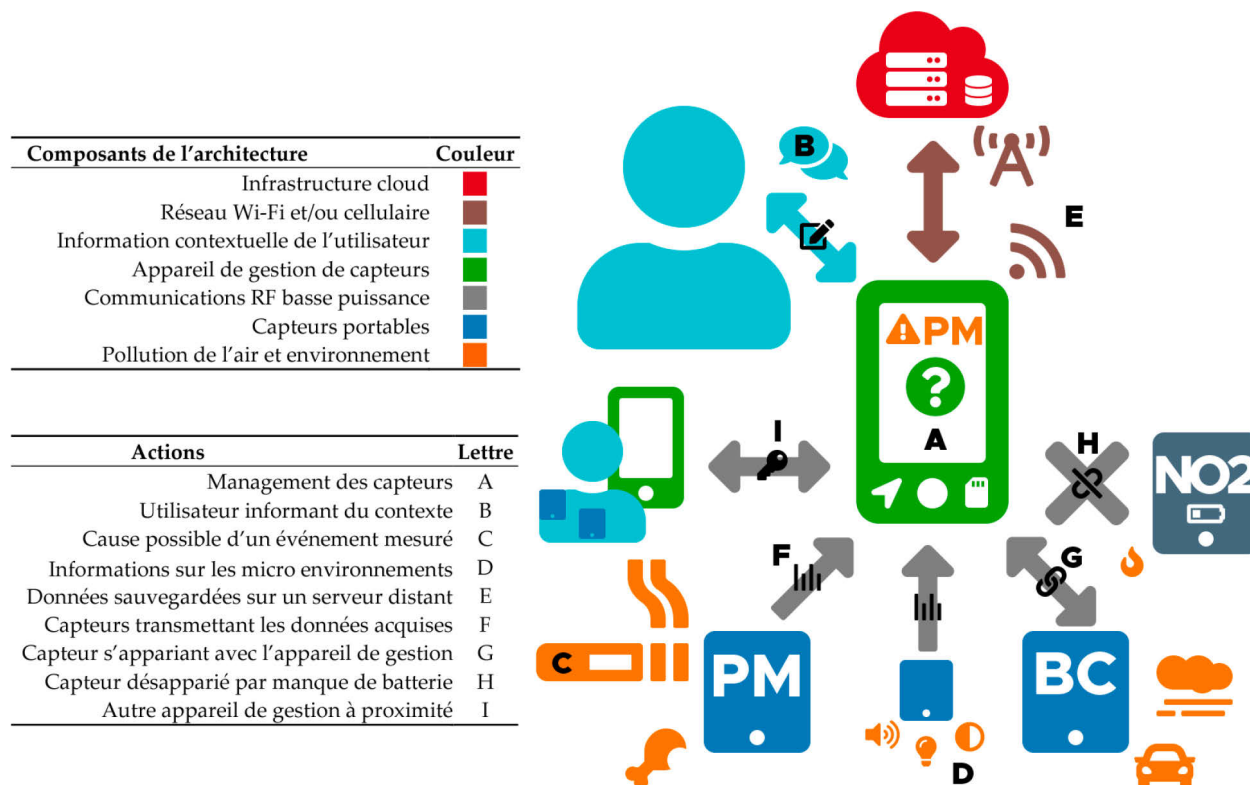


FIGURE 61 : Illustration de l'architecture du réseau local de capteurs

### 10.3. OBJECTIFS

Nous détaillerons les technicités d'une architecture de capteurs visant à automatiser la détection de ces événements ; en cessant d'exiger de l'utilisateur qu'il réalise les annotations de manière autonome. Il serait à la place, sollicité uniquement lorsque nécessaire, permettant ainsi de produire des annotations basées sur les variations observées par les capteurs en temps réel. Une telle architecture pourrait étendre les connaissances scientifiques relatives l'exposition individuelle à la pollution atmosphérique, puisqu'elle permettrait d'analyser plus précisément les données granulaires produites par les capteurs portatifs et permettrait donc d'augmenter la pertinence des analyses des données issues de la mesure de l'exposition individuelle à la pollution de l'air.

Nous suggérons une nouvelle approche lors du développement des capteurs de pollution de l'air ; par la création d'un appareil de gestion des capteurs qui se situera au centre d'un réseau local de capteurs portatifs (A). L'objectif de cet appareil de gestion sera de collecter des données et informations ; les données produites par les capteurs et les informations fournies par l'utilisateur. L'appareil de gestion sera la principale interface homme-machine de l'architecture. Elle sera en charge du recueil des données produites par les autres capteurs. Elle pourra indiquer à l'utilisateur, les données acquises en temps réel, la durée de vie restante des batteries équipant le ou les capteurs, ainsi que les informations nécessaires à l'utilisation du réseau de capteur dont le volontaire sera équipé.

L'appareil de gestion doit donc être suffisamment intelligent pour avertir l'utilisateur si les acquisitions réalisées dans un micro-environnement sont anormalement élevés ou faibles. Si un événement de pollution atmosphérique est détecté, si c'est estimé nécessaire, l'appareil de gestion en informera l'utilisateur et l'invitera à fournir les informations permettant de contextualiser l'événement (B). Cette sollicitation pourrait provoquer un changement de comportement de la part de l'utilisateur ; l'ouverture d'une fenêtre. Dans le cas d'une baisse subite des niveaux de polluants est observée, un nouvel avertissement permettra de contextualiser l'événement.

L'appareil de gestion permettra d'éliminer tout besoin de post-traitement des données contextuelles annotées par l'utilisateur. Il doit pouvoir réagir en quelques secondes, en informant l'utilisateur tandis qu'il est encore conscient des causes des événements de pollution (C). Le filtrage des signaux en temps réel doit être adapté à la caractéristique de chaque capteur, afin d'éviter l'envoi de fausses alertes, ou de notifications successives.

Les utilisateurs ne devraient pas avoir à procéder à des manipulations sur les capteurs autres que leur mise en charge. Les capteurs devraient tous posséder le même port de charge. Toutes les interactions avec l'appareil de gestion devraient être suffisamment faciles pour être effectuées par une personne n'ayant aucune connaissance de ces dispositifs. Optionnellement, les smartphones des utilisateurs pourront interfacer avec l'appareil de gestion et seront en mesure de fournir une interface utilisateur, similaire à celle présente sur l'appareil de gestion, mais sera doté naturellement d'une meilleure fluidité à l'usage.

## 10.4. SÉLECTION DU MATÉRIEL

L'appareil de gestion doit être construit autour d'un microcontrôleur suffisamment puissant, tel que les Teensy de PJRC [146] ou les cartes de développement basées sur la série STM32H7 de STMicroelectronics [147]. Ces deux modèles ont obtenu un score supérieur à 2000 dans CoreMark [148], un test d'évaluation de puissance de calcul de référence dédiée aux microcontrôleurs. Un bon candidat pour notre application serait le Portenta H7 [149] d'Arduino. Il est équipé d'une interface MIPI [150], permettant l'intégration d'écrans tactiles de haute qualité. La connectivité sans fil est assurée par les modules Wi-Fi et Bluetooth dont il est équipé. De plus, un circuit intégré permet de réaliser le contrôle de la charge de batteries lithium-polymère (Li-Po). Ces cartes de développement fourniraient une puissance de traitement suffisante pour afficher une interface utilisateur graphique (telle que LVGL [151]), tout en assurant un traitement continu des données acquises en temps réel, ainsi que l'accès aux enregistrements précédents, permettant d'adapter son comportement en permanence. Les modules permettant la communication au travers du réseau cellulaire, les modules RF (radiofréquence) de faible puissance (tel que le nRF24L01 [152]) et les autres modules et composants souhaités pourraient être interfacés par l'intermédiaire d'un circuit imprimé dédié.

Les capteurs de pollution existants sont des appareils spécialisés, qui surveillent rarement les paramètres qui ne sont pas directement liés à la pollution de l'air. Naturellement, nous utiliserions les données GPS pour détecter un changement d'environnement. Cependant, nous avons observé que dans certaines situations, ces données sont absentes ou insuffisantes. Nous pourrions en outre, utiliser la température de la couleur ambiante (avec un spectromètre) pour détecter si la lumière naturelle est présente (donc en extérieur), ou si au contraire, des ampoules fluorescentes sont présentes (donc en intérieur). Nous suggérons de tirer parti des capacités d'analyse temps réel de l'appareil de gestion. Des capteurs éprouvés et à bas coût pourraient permettre de mesurer l'intensité lumineuse, le spectre sonore et les mouvements. Ces informations peuvent permettre de détecter les changements de micro-environnements [77] (D). Ce type de traitements avancés des données est déjà exploité dans le domaine des smartphones. Ces capteurs pourraient être intégrés à l'appareil de gestion, ou alors disponibles en option, comme des membres du réseau de capteur, pouvant donc être couplés avec l'appareil de gestion

## 10.5. AUTONOMIE DE L'ARCHITECTURE

En cas de perte de connectivité avec le serveur distant (E), toutes les données seraient conservées et accessibles localement. Les algorithmes nécessaires au traitement temps réel des données ne seraient pas impactés ; l'architecture de capteur continuerait à fonctionner. La possibilité de perte de connectivité justifie à elle seule la nécessité d'exécuter la détection d'événements localement sur l'appareil de gestion. Les capteurs et les appareils de gestion doivent pouvoir résister à une utilisation quotidienne et continuer à fonctionner indépendamment de leur orientation, des conditions météorologiques et de l'activité de l'utilisateur — hormis la natation —. La surveillance à distance des capteurs doit être possible à partir d'un service web. Des alertes pourraient être envoyées aux utilisateurs ou aux techniciens si un dysfonctionnement est détecté. Les données recueillies seraient également transférées sur un serveur distant. Elles seraient donc consultables localement à partir de

l'appareil de gestion, mais aussi au travers d'un service web (D). Une réinitialisation permettant d'effacer la base de données interne de l'appareil de gestion devra être effectuée par un technicien pour des raisons de confidentialité avant de la remettre l'appareil de gestion à un nouvel utilisateur.

## 10.6. OPTIMISATION DE LA CONSOMMATION ÉNERGÉTIQUE

La consommation d'énergie est l'un des principaux facteurs à prendre en compte lors de la conception de nouveaux appareils portatifs [153] (p. 297). Dans le cas des capteurs mobiles, l'augmentation de l'autonomie se fait souvent au détriment du poids et de la taille de la batterie. Les ampèremètres pour les circuits de faible puissance ont récemment été démocratisés [137,154]. Associés à un oscilloscope numérique, ces outils permettent de lire des courants dont la valeur est inférieure au nanoampère (nA). Durant l'intégration des composants électroniques et le développement des logiciels embarqués, la consommation électrique des circuits intégrés et des modules de mesure doit être contrôlée et optimisée. Les tâches lourdes, telle que la géolocalisation, la connectivité internet et le stockage ne devraient être effectuées que par l'appareil de gestion, qui lui seul disposera d'une batterie conséquente, contrairement aux capteurs portatifs, qui doivent eux être assez légers pour être portés. Les capteurs qui sont appairés à l'appareil de gestion ne doivent donc pas être conçus pour fonctionner de manière autonome ; ils doivent au contraire préserver la durée de vie de la batterie et ne communiquer qu'avec l'appareil de gestion.

## 10.7. INTERFAÇAGE AVEC LES SMARTPHONES

En utilisant un appareil dédié plutôt qu'un smartphone ou une tablette, nous répondons aux éventuelles limitations matérielles (par exemple, pour permettre les communications RF avec les capteurs), au manque d'uniformité des équipements (capteurs de mouvement et de son intégrés, qualité de la réception du signal GPS), aux limitations logicielles (versions du système d'exploitation, protection de la vie privée et optimisation de la batterie limitant les possibilités de surveillance constante des capteurs). L'appareil de gestion pourra cependant être connecté au smartphone des participants via Bluetooth Low Energy (BLE) et utilisera donc une unique application propriétaire permettant d'accéder aux données des divers capteurs et de paramétrer ces derniers. L'utilisateur pourra également consulter l'historique des données hébergées localement, ainsi que visualiser les données acquises par les différents capteurs en temps réel.

## 10.8. CONTRAINTES DE DÉPLOIEMENT

Le déploiement de capteurs hétérogènes en situation réelle peut être d'une complexité inattendue. D'une part, chaque capteur possède ses propres exigences techniques. D'autre part, des informations cruciales doivent être correctement archivées pour permettre la mise en relation des données entre elles. Les efforts pour anonymiser les participants et l'existence de différentes sources de données, multiplie la possibilité d'erreurs d'annotation de la part des recruteurs. Dans le cadre du projet POLLUSCOPE, les recruteurs avaient à leur charge de créer et transmettre des invitations anonymes aux participants, de composer et distribuer un set de capteurs, puis de le distribuer. A la fin de la semaine d'acquisition, pour certains

capteurs, ils devaient alors récolter et archiver les données stockées localement, avant de les uploader sur la plateforme du projet. Sur cette plateforme, ils devaient reporter un à un les kits de capteurs de qualité de l'air tels qu'ils avaient été précédemment composés. Les données de santé étaient en revanche hébergées sur un second serveur, dans le respect de la GDPR. Les données étant anonymisées, il était essentiel de pouvoir s'assurer de l'association des données du questionnaire, ainsi que de la composition du kit de capteurs. Ayant collaboré activement avec les différents enquêteurs, nous avons pu observer la confusion et les erreurs que cette série de tâches peuvent engendrer.

Une première source d'erreur humaine fut induite par l'utilisation de capteurs non-connectés ; les données de ces derniers devant nécessairement être mises en ligne par un tiers. Les fichiers stockés localement sur ces capteurs portent soit dans le nom du fichier, soit dans l'entête du fichier, le numéro de série du capteur. Le contrôle doit donc être systématique avant ingestion des données. Dans le cas des données transmises par les capteurs connectés tels que les capteurs CANARINs, une connexion directe entre le serveur des deux projets a permis de dupliquer les données ciblées par les informations données par les enquêteurs sur le serveur dédié projet POLLUSCOPE. Une seconde source d'erreur humaine résidait dans la nécessité d'archiver manuellement plusieurs informations cruciales ; les numéros de capteurs, identifiants des participants, numéros de questionnaires, dates d'utilisation.

Dans le projet POLLUSCOPE, afin de pallier à toute erreur d'inattention de la part des recruteurs, nous avons implémenté une étape de contrôle supplémentaire avant le traitement des données, permettant de s'assurer de manière automatique de la conformité des trois sources d'informations ; le questionnaire, la composition des kits, les dates d'utilisation, ainsi que les tableaux réalisés par les enquêteurs, contenant ces deux informations. Lorsqu'une incohérence est identifiée, les possibles solutions sont alors automatiquement proposées, simplifiant le travail de correction qui lui doit être validé manuellement. Plusieurs erreurs mineures ont été faites, mais aucune erreur majeure nécessitant une correction sur la base de données n'a été relevée. Ces sources d'erreur devraient quand c'est possible être entièrement éliminées de manière automatique.

Pour lever ce frein, l'appareil de gestion rationalisera l'ensemble du processus de déploiement des capteurs. Les capteurs individuels seront immédiatement appairés et commenceront immédiatement à produire des données. Ils ne requerront donc pas de paramétrage de leur fréquence d'échantillonnage, de leur horloge interne, ou de leur mémoire ; ce seront des fonctionnalités réservées à l'appareil de gestion. En centralisant l'accès aux fonctionnalités sur un unique appareil, nous cherchons également à éviter aux volontaires les difficultés liées à l'utilisation d'une multitude d'applications mobiles. La création des différents kits de capteur et l'archivage des informations relatives au déploiement des capteurs doit être simplifié ; en limitant le nombre d'interventions humaines et en limitant les étapes d'annotation manuelles nécessaires. L'état des paires de capteurs devrait être visualisable sous forme de graphique pour les individus en charge du déploiement des capteurs et devrait être réalisé, soit au travers d'un appareil de gestion interfaçant avec les capteurs et les bases de données distantes, soit via l'interfaçage intelligent des capteurs entre eux [155].

Ces capteurs doivent pouvoir s'appairer ne pressant qu'un seul bouton, puis immédiatement commencer à transmettre des données à l'appareil de gestion. Tous les capteurs portatifs communiqueront avec le l'appareil de gestion par le même protocole de communication radio à faible puissance et courte portée (F), ce qui permettra un échange de données ultra rapide et économe en énergie entre les capteurs et l'appareil de gestion. Entre chaque échange de données, les microcontrôleurs qui pilotent les modules et micro-capteurs peuvent être mis en veille, ce qui permettra d'économiser en énergie. Comme ces communications ne sont pas sécurisées, aucune information sensible ne sera transmise par les capteurs ; seul l'identifiant unique du capteur, son type de modèle et les données brutes résultantes de la mesure. La transmission avec les capteurs est bidirectionnelle uniquement lors de l'appairage (G). Si un capteur appairé est perdu (H), le capteur sera désactivé et l'utilisateur sera notifié. Les appareils de gestion doivent pouvoir indiquer à d'autres appareils de gestion avec quels capteurs ils sont appairés, en veillant à ce qu'aucun autre appareil de gestion à portée ne puisse être accidentellement appairé aux capteurs en utilisation (I). Pour permettre l'essai et l'étalonnage des capteurs portatifs, on proposera, en remplacement de l'appareil de gestion, un appareil de gestion alternatif, spécifiquement imaginé pour permettre l'appairage et la communication avec un nombre important de capteurs portatif, sans nécessairement posséder des capacités de traitement ou de connectivité avancées.

## 10.9. DISCUSSION ET CONCLUSIONS

Le développement d'un tel écosystème nécessiterait des moyens considérables, tant pour le développement logiciel (algorithmes d'analyse des signaux en temps réel, interface utilisateur, micrologiciel, services cloud) que matériel (intégration électronique, boîtier). L'analyse des données en temps réel effectuée sur l'appareil de gestion doit être adaptée pour répondre aux besoins du chercheur en sciences des données, tout en s'assurant que les sollicitations ne sont pas inconfortables pour l'utilisateur.

Une telle architecture permettrait de réduire, au niveau des capteurs individuels, le matériel nécessaire (coût, poids, taille), le temps de développement logiciel et électronique, ainsi que les coûts de fabrication et d'entretien. Les modules de prototypage (émetteurs RF de faible puissance, chargeurs de batterie Li-Po, régulateurs de tension) sont très abordables et déjà disponibles. Les circuits imprimés sur mesure peuvent être rapidement conçus pour interfacier les modules et autres composants essentiels. Dans l'ensemble, cette approche permettrait une intégration rapide et la réalisation de tests en conditions réelles des plus récents modules de détection de la pollution de l'air dès leur mise sur le marché [156]. Le concept que nous venons de décrire va au-delà de la résolution du problème du manque de données contextuelles. Les utilisateurs seraient informés en permanence de la qualité de l'air qu'ils respirent et des effets de leur comportement. Nous espérons donc réduire considérablement les pertes accidentelles de données qui se produisent habituellement, en augmentant le degré d'implication et de sensibilisation de l'utilisateur.

Tous ces efforts mettent en évidence le fait que le manque d'informations contextuelles est l'une des lacunes rencontrées lors de l'acquisition de données granulaires. Nous avons abordé cette question en proposant le concept d'une architecture de capteurs de pollution de l'air qui



permettrait d'améliorer les connaissances sur l'exposition individuelle à la pollution de l'air en sollicitant l'utilisateur pour qu'il fournisse dynamiquement des informations basées sur les données relevées par les capteurs qu'il porte.

Parce que les connaissances et l'expérience acquises sont inestimables, nous préconisons aux chercheurs et aux étudiants de domaines hétérogènes de s'associer et de créer leurs propres appareils, comme cela a été fait pour le projet CANARIN. Enfin, la communauté scientifique devrait saisir l'opportunité de créer un standard pour les capteurs portatifs utilisés dans le cadre de la recherche médicale ; en employant les meilleurs composants et logiciels possibles pour produire une architecture axée sur la facilité d'utilisation, l'engagement du sujet et évidemment, la contextualisation des données.

## 11. CONCLUSION ET PERSPECTIVES

Historiquement, il a toujours été difficile de recueillir des données d'exposition à la pollution de l'air dans le contexte de la vie normale des individus. Cependant, ces dernières années, des innovations technologiques ont permis l'apparition d'une nouvelle gamme de capteurs ; les capteurs portatifs de la qualité de l'air. L'utilisation de ces derniers s'est démocratisée dans le domaine de la recherche épidémiologique ; ils sont plus abordables, plus légers, moins encombrants, fonctionnent en autonomie et peuvent ainsi être portés quotidiennement par des individus et sur de longues périodes. Ils offrent donc de nouvelles perspectives, notamment pour la recherche médicale, puisqu'ils permettent désormais d'étudier plus précisément l'impact de la pollution de l'air sur la santé.

### 11.1. CONCLUSION

C'est dans ce contexte que nous avons participé à lever ce verrou technologique, en participant au projet académique CANARIN (partie 4). Dans ce projet, nous avons accompagné le développement et l'amélioration de deux capteurs portatifs de la qualité de l'air ; le CANARIN II et le CANARIN nano. Ces derniers ont été utilisés par deux projets de recherche ; le projet POLLUSCOPE (partie 5) et le projet POLLAR (partie 6). Nous avons mis en relation des données issues de sources hétérogènes, en évaluant l'exposition des individus à la pollution de l'air en fonction de l'évolution de paramètres de santé ; la fonction respiratoire (chapitre 5.5) pour le projet POLLUSCOPE et l'apnée du sommeil (chapitre 6.3) pour le projet POLLAR.

Les capteurs portatifs de la qualité de l'air produisent des données de qualité moindre vis-à-vis des instruments de référence. Nous avons montré qu'ils sont sujets à la dispersion. Lorsqu'ils sont utilisés par des individus dans le contexte de la recherche épidémiologique, ils sont exposés à des conditions inconnues qui peuvent participer à accélérer le vieillissement des micro-capteurs et modules de mesures dont ils sont équipés. L'incertitude de mesure de ces capteurs portatifs est donc importante et sans retour d'expérience suffisant sur ces technologies, une surveillance périodique de leur bon fonctionnement doit nécessairement être mise en œuvre. Dans le cadre du projet POLLUSCOPE, nous avons observé que cette surveillance constitue une contrainte technique majeure. À l'heure actuelle, elle requiert l'utilisation d'instruments de référence et d'infrastructures dédiées.

Nous avons donc proposé un processus permettant de lever entièrement ces contraintes ; en substituant l'infrastructure et l'outil de référence par l'utilisation d'un banc d'essai mobile (partie 7). L'utilisation d'un algorithme dédié (partie 8) permet non seulement d'effectuer une surveillance préventive des capteurs, mais également de minimiser l'erreur de mesure de ces derniers. Dans ce manuscrit, nous avons donc présenté les résultats obtenus avec ce banc d'essai (chapitres 7.5, 7.6), puis quantifié la réduction de la dispersion possible grâce à cet algorithme, dans le cas d'un set de compteurs optique de particules (section 8.10.4).

Au sein du projet POLLUSCOPE, nous avons amené le développement et le déploiement d'une application mobile permettant de simplifier le recueil de données contextuelles par les participants. Nous avons montré comment ces multiples sources de données peuvent être traitées ; en associant les événements de pollution aux données spatio-temporelles et aux activités déclarées par les participants (partie 9). C'est ainsi que nous avons identifié la nécessité d'augmenter significativement la quantité et la pertinence de ces informations contextuelles. Un modèle d'architecture de capteurs, permettant d'automatiser l'acquisition et l'interprétation des données contextuelles a donc été proposé (partie 10).

Au sein des projets auxquels nous avons participé, l'encombrement des capteurs, leur taille, leur poids, leur apparence, furent les raisons principales motivant l'abandon d'une partie des participants. Le port d'un capteur portatif de la qualité de l'air implique à lui seul un certain nombre de contraintes. Des efforts doivent être réalisés, lors du développement de ces capteurs portatifs, afin de minimiser les potentielles frustrations que pourraient rencontrer les porteurs du capteur et donc le risque d'abandon. Ces efforts doivent s'inscrire au-delà de la seule miniaturisation des capteurs. Le confort de port doit être optimisé ; en développant des accessoires adaptés à plusieurs types d'utilisation. Le nombre de câbles et chargeurs nécessaires doit être limité ; en utilisant les standards existants. L'aspect visuel des capteurs doit être modulable et les porteurs doivent pouvoir accéder aux mesures en temps réel ; exploitant ainsi le phénomène de quantified self (section 1.4.3).

## 11.2. PERSPECTIVES

L'ensemble des technologies du domaine de la mesure de la qualité de l'air évoluent rapidement ; les modules de mesure et micro-capteurs disponibles sont de plus en plus performants et abordables. À terme, la qualité métrologique des capteurs portatifs utilisés dans le cadre de la recherche médicale se rapprocheront des instruments de référence actuels.

Les technologies liées à l'IoT ont rapidement été adoptées par le grand public. Dans un futur proche, chaque habitation et chaque véhicule sera équipé de capteurs de la qualité de l'air. La réduction du coût de ces technologies permettra aux associations en charge de la surveillance de la qualité de l'air de réaliser un maillage plus fin du territoire national en déployant des réseaux de capteurs à bas coût. De plus, nous avons relevé l'augmentation de la performance des outils de mesure embarqués par les satellites de surveillance de la qualité de l'air. En définitif, la pertinence et la quantité des données publiques relatives à la qualité de l'air ne pourront que s'accroître dans les prochaines années. Similairement, les technologies permettant de relever des paramètres de santé de façon non-invasive seront démocratisées. Les individus auront à leur disposition des informations relatives à leur exposition individuelle aux divers polluants, en air extérieur comme intérieur, dans les lieux publics, comme à leur domicile. Ces données seront automatiquement recueillies au travers de leur smartphone. Dans un tel scénario et si la possibilité est offerte aux individus de rendre leurs données accessibles à la demande, il est alors légitime de se questionner sur la pertinence des capteurs portatifs de la qualité de l'air tels qu'ils existent aujourd'hui.

## 12. ANNEXES

12.1 ARTICLE — ACADEMICALLY PRODUCED AIR POLLUTION SENSORS FOR PERSONAL EXPOSURE ASSESSMENT: THE CANARIN PROJECT, PAGE 108

12.2 POSTER ET FLYER — PROJET POLLUSCOPE, PAGE 126

## 12.1. ARTICLE — ACADEMICALLY PRODUCED AIR POLLUTION SENSORS FOR PERSONAL EXPOSURE ASSESSMENT: THE CANARIN PROJECT



Article

### Academically Produced Air Pollution Sensors for Personal Exposure Assessment: The Canarin Project

Boris Dessimond <sup>1,2,\*</sup>, Isabella Annesi-Maesano <sup>2</sup>, Jean-Louis Pepin <sup>3</sup>, Salim Srairi <sup>4</sup> and Giovanni Pau <sup>1</sup>

<sup>1</sup> Computer Science Laboratory, University of Pierre et Marie Curie, LIP6, NPA Team, 4 Place Jussieu, 75005 Paris, France; giovanni.pau@unibo.it

<sup>2</sup> EPAR Team, Medecine Faculty of Saint-Antoine, IPLESP, INSERM & Sorbonne University, 27 rue Chaligny, 75571 Paris, France; isabella.annesi-maesano@inserm.fr

<sup>3</sup> Laboratoire HP2, Université Grenoble Alpes, Grenoble, INSERM, U1042 and CHU de 24 Grenoble, France; JPepin@chu-grenoble.fr

<sup>4</sup> Centre for Expertise and Engineering on Risks, Urban and Country Planning, Environment and Mobility—Ile de France Territorial Division, Mobility Department, 12, rue Teisserenc de Bort, 78190 Trappes-en-Yvelines, France; salim.srairi@cerema.fr

\* Correspondence: boris.dessimond@etu.upmc.fr

**Abstract:** The World Health Organization has estimated that air pollution is a major threat to health, causing approximately nine million premature deaths every year. Each individual has, over their lifetime, a unique exposure to air pollution through their habits, working and living conditions. Medical research requires dedicated tools to assess and understand individual exposure to air pollution in view of investigating its health effects. This paper presents portable sensors produced by the Canarin Project that provides accessible, real time personal exposure data to particulate matter. Our primary results demonstrate the use of portable sensors for the assessment of personal exposure to the different micro-environments attended by individuals, and for inspecting the short-term effects of air pollution through the example of sleep apnea. These findings underscore the necessity of obtaining contextual data in determining environmental exposure and give perspectives for the future of air pollution sensors dedicated to medical research.

**Keywords:** Internet of things; sensor mesh architecture; particulate matter sensor; indoor air pollution; exposure assessment; health impact; mobility; sleep apnea



**Citation:** Dessimond, B.; Annesi-Maesano, I.; Pepin, J.-L.; Srairi, S.; Pau, G. Academically Produced Air Pollution Sensors for Personal Exposure Assessment: The Canarin Project. *Sensors* **2021**, *21*, 1876. <https://doi.org/10.3390/s21051876>

Academic Editor: Matjaž Finšgar

Received: 31 December 2020

Accepted: 4 March 2021

Published: 8 March 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

#### 1. Introduction

Modern societal development relies heavily on the consumption of fossil fuels that continually release more harmful gases and particles in the atmosphere, negatively impacting both the climate and human health [1]. Air pollution is a major cause of disease and premature death, and is the single largest environmental health risk in Europe [2].

In an urbanized environment, individuals are exposed to outdoor air pollution created by industry, traffic and domestic heating. However, individuals spend most of their lifetime indoors in settings where they are exposed to a different set of sources of air pollution, including combustion for heating and cooking, building materials, cleaning products, or cigarette smoke. As a consequence, exposure at the personal level comprises a unique composition of air pollutants as the individual moves across different environments. It is, thus, important when investigating the health impacts of air pollution and involved mechanisms to prioritize personal exposures over ambient concentration values.

Air pollution has historically been measured using fixed air pollution stations, providing high quality data at one geographical coordinate. The paradigm of air quality monitoring has changed as low cost networks of air pollution monitors are able to provide a higher spatio-temporal resolution than the more expensive reference stations [3–9]. Additionally, satellite imagery, although limited by its resolution, allows for the assessment of air pollution where other means are not available. Collectively, these measurements are

largely used in assessing health effects of air pollution at the population level. However, they cannot provide individual level data, nor can they provide assessments of indoor air pollution, of living and working conditions, behavior and habits, ventilation, cooking habits or the use of chemicals for cleaning, which will considerably impact the quality of the air for a given person. Thus, individual level data enable medical researchers to more accurately depict and understand how air pollution impacts human health [9].

The rise of wearable smartwatches, with fitness and sleep activity trackers, has created a cultural phenomenon called the quantified self, whereby members of the general population voluntarily wear tracking devices that continuously log their data, in exchange of potential improvements in their quality of life or physical performance. Real-time monitoring of air pollution levels can alert subjects about sudden peaks or slow rise of air pollutants, enabling them to change their behavior (in case of indoor air pollution) or to avoid the pollution source (in case of outdoor air pollution). Public adoption of wearable air pollution sensors and health monitoring devices has considerably increased over the last few years.

Sensor is a broad term that describes both the technology sensing and communicating physical change to other electronics, as well as the whole device embedding that technology and the necessary hardware to be used as a standalone device. A wearable sensor is composed of an enclosure, embedded battery, one or multiple sensors, integrated circuits, passive electronics and a microcontroller unit (MCU) running a dedicated software and possibly remote server hosting web services to store and subsequently access the data. Technology in sensing modules is evolving at a fast pace, as many companies are developing and producing small, efficient, affordable and reliable sensor modules, enabling the end products, the wearable sensors, to produce a level of quantity and quality of data that has never been reached before.

In the field of air quality sensors, commercial devices and original equipment manufacturer (OEM) low-cost sensors have been benchmarked by the scientific community [10–20] and multiple methodologies and tools have been developed for validation [21–24]. Such sensors have been considered for crowdsensing applications [25–31]. However, there are fewer wearable sensors produced by research projects [32–37], and the amount of commercial wearable sensors aimed at the general public, rather than use in industrial settings, is surprisingly even more limited [38,39]. While wearable and/or portable air pollution sensors do exist, accessing real-time geolocalized data obtained by these sensors is not necessarily available to third parties, such as a research team. Thus, Canarins were created with this purpose in mind—to make available a robust, validated, portable sensor for public health research requiring continual individual exposure information.

The primary purpose of this paper is to present the application of the Canarin portable sensors for medical research. Specific objectives include the description of the Canarin as a tool for assessing individual exposure to particulate matter (PM) in real time in relation to mobility (as in the POLLUSCOPE [40] project) and to associate it with sleep apnea (as in the POLLAR project). The Canarin project has already leveraged today's best available technologies to produce granular level data. Through our work on these projects, we have determined that further efforts in contextualizing the acquired data must be made. The final goal of this paper is to describe an innovative approach in sensor architecture design, specifically aimed at improving the interpretation of personal exposure to air pollution.

## 2. Materials and Methods

### 2.1. Overview of the Canarin Project

The Canarin project unites an international team of researchers and students with a common goal: to produce and utilize low-cost air pollution sensors to assess individual exposure to air pollution and its health effects. While alternative sensors are commercially available, they often are required to connect to an external device for data transmission, such as a mobile phone, and data are often shared only with the user, rather than a third

party. As such, this project aims to solve issues such as connectivity, cost, size, data acquisition and ease of use for medical and health study participants.

Canarin sensors use commercially available development boards and modules, which have the advantage of having been already implemented and tested [41]. We selected an optical particle counter (OPC), and made efforts to integrate modules in an enclosure following the manufacturer recommendations, ensuring the best expectable performance. The rest necessary components are mounted on printed circuit boards (PCBs) that we manufactured externally. No post-processing is applied to the data before transmission.

The first Canarin was created in 2016 by the Asian Institute of Technology, Sorbonne University, and the University of Bologna, who partnered in the frame of the SEA-HAZEMON project [42]. A series of Canarins were deployed, acting as low-cost fixed air pollution monitoring stations, collecting data on Particulate Matter (PM), namely particles of 1, 2.5 and 10  $\mu\text{m}$  of diameter ( $\text{PM}_1$ ,  $\text{PM}_{2.5}$ ,  $\text{PM}_{10}$ ), from the recurring haze events in the Southeast Asia region. Using portable sensors such as these to probe human exposure had been previously investigated at the Asian Institute of Technology [43]. In 2018, the Canarin II portable sensors were upgraded to include multiple wireless connectivity options and a GPS (global positioning system). [44] They were then dispersed over the urban environment of Bologna, while mounted on publicly available e-bikes. The ability to visualize geolocalized air pollution data in real time was a major improvement. Concurrently, other Canarin II devices were benchmarked by the POLLUSCOPE project, which required both mobile and affordable solutions to assess individual exposure to air pollution in the Île-de-France region. When tested against particulate matter sensors commercially available at the time, the Canarin II performed the best, as reported by national experts [21]. It was, therefore, selected and used extensively by the POLLUSCOPE project. Several Canarin II have also been deployed in the Johannine Library of Coimbra to monitor the impact of PM on its ecosystem, where the bat population has an essential role in the preservation of the library by eating the bookworms during the night [12]. These Canarin II are still acquiring data continuously to this date. In 2019, a smaller and lighter Canarin was created; the Canarin Nano. It has been used in the POLLAR project, with the primary goal of exploring whether sleep apnea is aggravated by exposure to microparticles [45].

## 2.2. Canarin Nano

The hardware of the Canarin II runs on Linux, and therefore, provides a good amount of processing power and software development possibilities. The main board consumes approximately 3 watts, requiring a battery able to sustain a normal day of use. The size and weight of the device (around 1.2 kg) is not ideal for individuals who have to carry the device all day long, especially if they are affected by respiratory diseases, as was the case in the POLLUSCOPE project. Such drawbacks had been previously outlined during the performed benchmarking [21], and volunteers of the POLLUSCOPE project commented on these aspects. Technological advancements in microcontrollers, PM and gas sensing technologies have made size reduction possible without compromising the quality of the acquired data. The Canarin nano uses a microcontroller based on the ARM Cortex-M processor instead of the more power hungry UDOO NEO FULL [46] Cortex-A9 processor used on the Canarin II. Onboard, a 3G modem, flash memory, 4000 milliamper hour (mAh) battery, a GPS module, a particulate matter sensor ( $\text{PM}_1$ ,  $\text{PM}_{2.5}$ ,  $\text{PM}_{10}$ ), a temperature humidity and pressure sensor all manage to provide the same level of performance previously offered by the Canarin II, while staying on for up to 14 h on a single charge. Improvements in particle sensor design allowed for a slimmer device, and recent development in electrochemical sensors have allowed, for the first time, to accommodate a volatile organic compound (VOC) sensor. The VOC sensor is mounted on a secondary printed circuit board (PCB) header board that can be easily plugged in and out of the main board. This secondary PCB header board opens the possibility to upgrade to newer electrochemical or metal-oxide gas sensors, without needing to remanufacture the entire device. Flash memory has been used to prevent data loss if the on-board 3G modem loses connectivity

with cell towers. A third party IoT (Internet of things) or M2M (Machine to machine) SIM (subscriber identity module) card provider has been selected instead of the default Particle cloud services [47]. Table 1 provides some of the components used in the Canarin Nano.

**Table 1.** Canarin Nano components.

Components	Reference
Development board	Particle Electron 3G-U270
Microcontroller	STMicroelectronics STM32F205
GSM <sup>1</sup> module	U-BLOX SARA-U270
Cellular antenna	Molex Part Number 2072350100
IoT <sup>6</sup> /M2M <sup>7</sup> SIM <sup>8</sup> card provider	ThingsMobile
GPS <sup>2</sup> module	U-BLOX NEO-M8N
GPS <sup>2</sup> antenna	Taoglass AGGBP25B.07.0060A
Battery	GEB battery OEM <sup>3</sup> 4000 mAh <sup>4</sup> Li-Po <sup>5</sup> battery
Charger	CUI Inc. SW110-5-E-I38
Memory flash storage	Winbond Electronics W25Q64FWZPIG8MB
VOC <sup>9</sup> sensors	Sensirion SGP30, Bosch BME680

<sup>1</sup> GSM: Global system for mobile, <sup>2</sup> GPS: Global positioning system, <sup>3</sup> OEM: original equipment manufacturer, <sup>4</sup> mAh: Milliampere hour, <sup>5</sup> Li-Po: Lithium polymer, <sup>6</sup> IoT: Internet of things, <sup>7</sup> M2M: Machine to machine, <sup>8</sup> SIM: Subscriber identity module, <sup>9</sup> VOC: Volatile organic compounds.

The cloud infrastructure server is hosted by Amazon Web Services (AWS) through an EC2 T2 instance [48]. One vCPU (virtual central processing unit) and 1 GB of RAM (random access memory) are enough to provide service for approximately 100 Canarin Nanos. Multiple docker containers [49] run Python scripts to ingest data. The database is also hosted by AWS in Frankfurt, following the GDPR (General Data Protection Regulation). The front end [50] allows users to download or visualize data in real time.

Once the first prototypes provided satisfying performance in a real-world scenario, a redesign of the device occurred, to switch from locally soldered and printed materials, to a custom PCB and a fused deposition modeling (FDM) 3D printed enclosure, produced externally in larger quantities. Complexity to source the bill of materials, logistics and delays increased the overall cost per unit. Any possibility to streamline the production chain was abandoned due to lack of manpower. Redesigning the entire device enclosure and PCB successfully eased the assembly and maintenance of the sensor, while greatly improving reliability. The enclosure provides protection for the electronics, visibility for multiple light-emitting diodes (LEDs) and accessibility to the available ports, buttons and switches, see Figure 1. The Canarin Nano is designed to be always on, as the instrumented patients and volunteers should have no interactions with the device other than to carry it with them. The participants can, however, choose switch it off, disabling the on-board GPS module, when privacy concerns are involved.

Having a sensor permanently connected through 3G provides several advantages; sampled data is sent each minute to our cloud service, and is immediately available for the researcher to visualize or download. The patients and volunteers can, therefore, be monitored multiples times per day, limiting the risk of data loss. Setting changes or full firmware updates can be triggered over-the-air, without the end user even noticing it happening. Reducing the sensor size and weight helped create a device able to fit more of the medical researchers' requirements than other commercial sensors in the sub-thousand euros price range. Improvements will occur as student and researcher creativity will continue to drive the Canarin project.





**Figure 1.** A Canarin Nano. The hook in the foreground is for carrying the device.

### 2.3. POLLUSCOPE Project

The primary goal of the POLLUSCOPE project is to create a participatory observatory for measuring air pollution at the individual level with low-cost portable sensors. One of the first tasks of the project was to test, compare and identify which sensors will be fit to the task. Selected sensors have been evaluated by a panel of experts, who compared these sensors against reference instruments [21] to ensure that the performance of the selected modules was satisfactory enough to be used in real-life mobility scenarios. The Canarin II went through both static and dynamic tests and was selected as the best particulate matter (PM) sensor. The project has since been using the Canarin II. POLLUSCOPE is a long-term project, and sensors can spend months unused. To moderate the risk of having a faulty sensor before going through several weeks of measurements, all selected air pollution sensors were regularly handed over to the project's air pollution experts to be tested against reference instruments. For the particulate matter, a Thermo Scientific™ 1405-F tapered element oscillating microbalance (TEOM), for PM<sub>10</sub>, and a ThermoFisher TEOM 1400 were used [21]. If sensors displayed unexpected behaviors, they were set aside for servicing.

The second goal of the project was to assess individual exposure to air pollution as a function of mobility. Three population groups were considered: volunteers from Versailles Grand Parc (VGP), participants of the population-based survey RECORD [51] and volunteers affected by respiratory diseases. VGP volunteers had to fill an online survey, and provide written consent for their data to be used. Volunteers were required to wear POLLUSCOPE's sensors, including a Canarin II (Figure 2), for one week. The volunteers were asked to keep the sensors turned on at all times, and to let them charge when possible. The minimum sensor set is composed of a Canarin II (laser diffraction particulate matter sensor), a portable electrochemical NO<sub>2</sub> (nitrogen dioxide) sensor attached to it (Cairsens NO<sub>2</sub> [52]) and a tablet (Archos 101 Oxygen 4G) to manually log activity through a dedicated app and provide a Wi-Fi access point for the Canarin II.

Each air pollution sensor used by the project samples air pollution every minute. Depending on the cohort, the sensor set can be completed with an AE51 (optical black carbon (BC) sensor, aethalometer) and a Bluetooth low energy (BLE) connected portable spirometer, as well as an oximeter to log health parameters, three time a day. Volunteers were invited to fill an online survey at the beginning of the study in order to collect some important socio-economic data.

The third goal of the project was to provide the participants with valuable feedback on their exposure to air pollution. Automating such a task was only limited by our ability to understand the available data. Relevant information can be provided if we know in which context air pollution exposure happens. We demonstrated that the multiplicity of portable

sensor's data, completed by user's annotations, can be used to improve the knowledge on personal exposure to air pollution.



**Figure 2.** Canarin II in grey. The Cairnsens NO<sub>2</sub> is conveniently attached to it.

#### 2.4. POLLAR Study

The EIT-HEALTH POLLAR clinical study [45] began in March 2019, with the objective of estimating the impact of air pollution and pollen exposure on rhinitis, asthma and sleep. Pollen data were obtained through the Réseau National de Surveillance Aérobiologique (R.N.S.A., French aerobiology network [53]). The impact on sleep was studied at the Grenoble Alpes University Hospital where sleep apnea patients were recruited to perform, at two different periods, a polysomnography (PSG), followed by three days of air pollution exposure using a Canarin Nano or a Canarin II. A PSG is a comprehensive sleep study that requires the subject to stay at the hospital for one night equipped with monitoring devices (Figure 3). The main PSG outcome considered in the present study is the severity of the sleep apnea, given by the apnea-hypopnea index (AHI). The Canarin II and Canarin Nano sensors allowed for the first time the comparison of PSG results with granular air pollution data obtained both during and after sleep.

#### 2.5. Quantifying Individual PM Exposure Relative to Mobility and Activities

Individual exposure to PM was expressed as a concentration in  $\mu\text{g}/\text{m}^3$ . Mobility data were acquired through the Canarin II and tablet onboard GPS. In the results are displayed the spatial tracing of the PM<sub>10</sub> data acquired by one volunteer, a member of the POLLUSCOPE consortium, totaling six and a half days of data (156 h) from two sampling sessions in November 2019. This volunteer was selected as he is a collaborator working at the CEREMA Île-de-France and have agreed to publish his data. Both raw GPS and Canarin II data were used.

### 2.6. Statistical Methods

We used Python’s NumPy and Pandas libraries to produce the results provided in the tables. The Canarin data were downloaded from the POLLUSCOPE projects data hosting server. POLLAR’s Canarin data were obtained from the Canarin project website [50], through a password-protected web page. Data produced by the polysomnography and Canarin deployment planning of the POLLAR study were provided through an Excel file. No post processing was applied to compensate for data loss or user errors; all figures and tables were produced using raw unprocessed data. Matplotlib and Seaborn were used to create the figures. Mobility data figures were produced using QGIS and an Open Street Map layer for understandability. In the study of mobility, the notion of points of interest (P.O.I.) introduces two contrasting clusters with regard to individual exposure and device behavior.



Figure 3. Subject instrumented before polysomnography (PSG) screening at the hospital.

## 3. Results

### 3.1. Individual PM Exposure as a Function of Mobility (The POLLUSCOPE Project)

The data presented below were acquired by the collaborator working at the CEREMA<sup>3</sup>. The data are composed of two sampling periods of 8 days each (384 h in total). Table 2 shows, for each of the sampling periods, the average air pollution levels in  $\mu\text{g}/\text{m}^3$  and the count of the total number of recorded samples. All sensors sampled air pollution once per minute. Over the course of each 8-day period, measurements were taken 6040 (period A) and 4227 (period B) times, representing approximately half and just under half of the potential maximum of 11,520 measurements, respectively (Supplementary S1).

Table 2. Overview of the two periods of acquisition; pollutants values in  $\mu\text{g}/\text{m}^3$ .

Period	Avg PM <sub>1</sub> <sup>1</sup>	Avg PM <sub>2.5</sub> <sup>1</sup>	Avg PM <sub>10</sub> <sup>1</sup>	Avg NO <sub>2</sub> <sup>2</sup>	Avg BC <sup>3</sup>	Count PM <sup>1</sup>	Count NO <sub>2</sub> <sup>2</sup>	Count BC <sup>3</sup>
A: 10/19/19 to 10/27/19	8.1	11.6	12.9	5.8	0	5325	6040	0
B: 11/14/19 to 11/22/19	8.6	14.5	16.5	7.9	1226.0	4027	4227	2628

<sup>1</sup> PM: particulate matter, <sup>2</sup> NO<sub>2</sub>: nitrogen dioxide, <sup>3</sup> BC: black carbon.

The volunteer was exposed to higher concentrations of air pollution during the second period (Table 2 Period . Activity data distribution in Figure 4 shows we lack activity for 55% of the acquired samples (Figure 4 “Period A + B”). Available activity data may contain mistakes and require post-processing. A higher percentage of activity data has been obtained from the user for the activities “Car” and “Office” (Figure 4 “Period B”). Any difference observed in terms of exposure to air pollution between the two periods could be explained by the lack of data, or by the over representation of specific activities. This underscore that sensors should ideally be never stopped during data logging, and how annotations that contextualize the sensor’s data are mandatory to provide a trustworthy analysis.

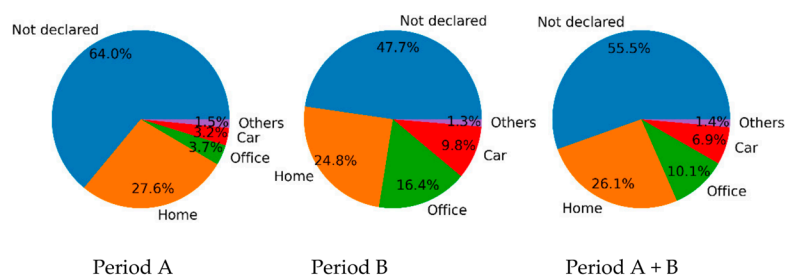


Figure 4. Distribution of the activities per period.

Figure 5 shows the two points of interest, circled in black, that we are going to focus on. The rest of the map displays the entirety of the spatial movement recorded by the GPS over the two periods. In the north-west, a single ride away from an urban area is visible, translated into lightly colored low PM<sub>10</sub> exposure. PM<sub>10</sub> exposure is higher along the Paris’s ring road (“Boulevard Périphérique,” in the north-east corner), and decreases away from the main road.



Figure 5. Spatial tracing of PM<sub>10</sub> (µg/m<sup>3</sup>) for both periods A and B.

In Figure 6, in the south, a commute to a store is marked by a bright violet end and a noticeably higher PM<sub>10</sub> exposure (see Figure 5). Both P.O.I. in Figure 5 contain a cluster of samples, which generally indicate a working or living place. A pink trace in Figure 6 shows a commute by car or public transports that have not been manually logged by the volunteer; therefore, the recorded PM data are not traced in Figure 5.

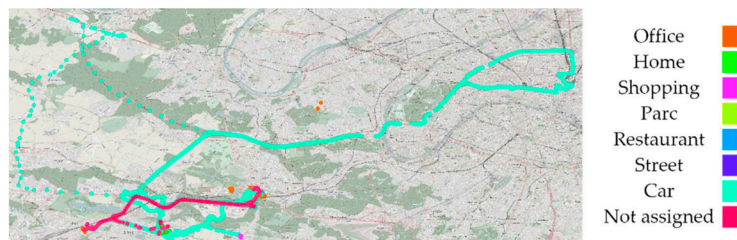


Figure 6. Spatial tracing of activity annotations for both periods A and B.

Zooming in on P.O.I. A confirms that we are observing the volunteer’s annotated “Workplace” (Figure 7). The missing tracking activity information also lacks  $PM_{10}$  samples. The “Workplace” cluster of points shows a very common case of raw GPS data noise caused by less-than-ideal GPS signal reception. The accuracy of localization drops indoors. From this figure, we can observe how lower levels of  $PM_{10}$  have been recorded indoors compared to the surrounding urban areas. The cluster in the west in Figures 5 and 6 correspond to the volunteer’s secondary workplace, in a less urbanized area, with lower  $PM_{10}$  levels.

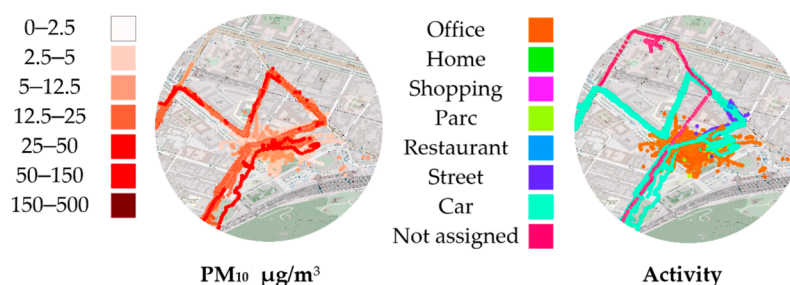


Figure 7. Geographical point of interest A for particulate matter ( $PM_{10}$ ) measurements “Workplace” cluster surrounded of other activities.

P.O.I. B (Figure 8) shows some randomly localized samples with high  $PM_{10}$  values. These few points around our P.O.I. comprise 6905  $PM_{10}$  samples, the higher samples being displayed first. This indicates an indoor location where GPS reception was not possible, and samples are all displayed on the last found location. Notice how multiple commutes by car leads to this cluster, with one accidentally marked as “Home” instead of “Car,” in green. This mistake will negatively impact the estimated exposure to air pollution at “Home.”

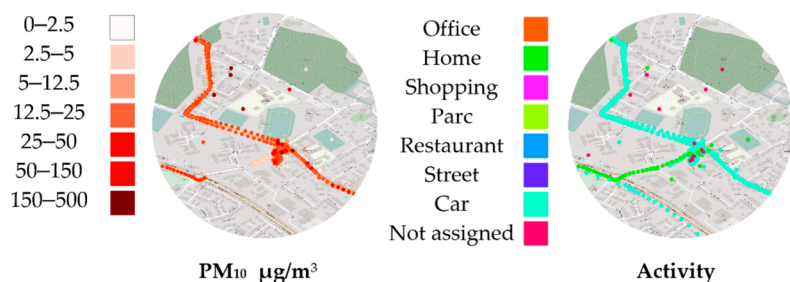


Figure 8. Point of interest for particulate matter ( $PM_{10}$ ) B “Home” cluster with lack of GPS reception.

Zooming further into the cluster (Figure 9) reveals the volunteer’s “Home.” Below the center, among the noise, we can guess where the car is supposedly parked. Circled are a total of 321 PM<sub>10</sub> samples, which are heavily concentrated on pinpointed coordinates, indicating what could be the real “Home” location, possibly next to a window. However, we confirmed with the volunteer that no GPS samples are located precisely on the “Home” location. Without more contextual information, the lack of indoor GPS reception and the low count of heavily concentrated samples could mislead algorithms aimed at filling or correcting activity data, or bias the estimated PM<sub>10</sub> value at “Home,” obtained from GPS data clustering.

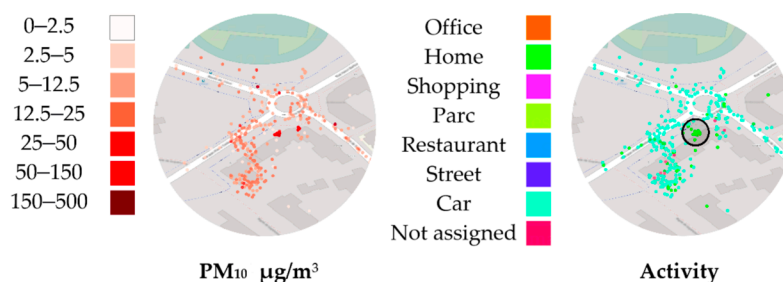


Figure 9. Point of interest for particulate matter (PM<sub>10</sub>) B, close up of “Home.”

Table 3 provides the average air pollution levels per activity, according to user’s annotations. PM<sub>10</sub> levels in “Office” are low, as observed previously in Figure 7. “Home” has the highest PM exposure averages, and “Car” is the activity with the highest overall air pollution exposure, specifically regarding NO<sub>2</sub> and BC levels. The average PM<sub>10</sub> value in the circled region in Figure 9 is 34.1 µg/m<sup>3</sup>, double what was obtained using the annotated data. NO<sub>2</sub> and BC levels during the activities “Street” and “Car” are predictably higher than any other activity.

Table 3. Average air pollution level from annotated data, µg/m<sup>3</sup>.

Activity	PM <sub>1.0</sub> <sup>1</sup>	PM <sub>2.5</sub> <sup>1</sup>	PM <sub>10</sub> <sup>1</sup>	NO <sub>2</sub> <sup>2</sup>	BC <sup>3</sup>
Office	2.3	3.55	3.9	5.7	586.8
Home	10.5	15.1	16.4	6.2	610.0
Shopping	2.7	4.2	4.9	10.4	NA
Park	3.3	4.9	5.3	16.5	409.9
Restaurant	3.6	5.6	6.2	11.2	NA
Street	3.0	4.6	5.1	18.5	NA
Car	9.2	13.1	14.3	20.6	2937.8

<sup>1</sup> PM: particulate matter, <sup>2</sup> NO<sub>2</sub>: nitrogen dioxide, <sup>3</sup> BC: black carbon.

### 3.2. Individual PM Exposure in Relation to Sleep Apnea (The POLLAR Study)

Figure 10 displays the hourly average PM exposure for POLLAR study subjects affected by sleep apnea (21 subjects with average AHI > 5, in blue), and subjects with normal AHI (22 subjects with average AHI ≤ 5, in orange). In each group, the upper value of the box plots indicates how much some subjects were exposed to unusually high values of PM<sub>10</sub>. In this figure, the extreme outliers were hidden, as one volunteer was consistently recording excessively high values of PM overnight, with PM<sub>10</sub> samples above 1000 µg/m, which exceeds the functioning range of the PM<sub>10</sub> module. Such values can be recorded by OPCs when exposed to e-cigarettes vapor, but no other information is available to contextualize the data. The median values are very close between the two groups. Between 9 a.m. and 9 p.m., the 75th percentile PM<sub>10</sub> values were higher in the group affected by sleep apnea.

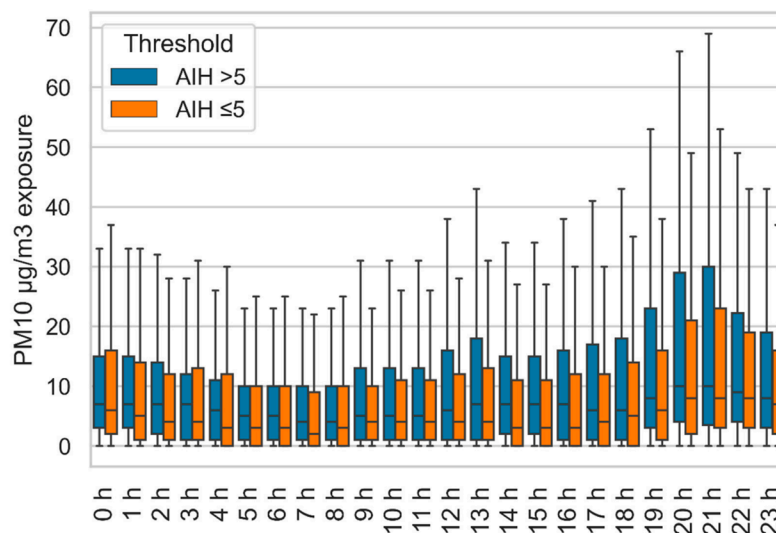


Figure 10. Hourly PM<sub>10</sub> concentrations according to the apnea-hypopnea index (AHI) group.

#### 4. Discussion

##### 4.1. Quantification of Individual PM Exposure

We have shown how portable air pollution sensor data can be interpreted based on user's information about their activity. We also saw how a scrupulous data review could partly correct activity data in case the user neglected to log their activity data. Data completion and activity correction algorithms are under development within the POLLUSCOPE project.

When assessing personal exposure, volunteers are required to be constantly aware of their obligation to keep the devices charged, to remember to carry the devices and possibly also track their daily activities. This requires them to perform multiple manual operations during the busiest moments of their day. A mobile app is delivered on a tablet, and is available on Android phones for the user's convenience. The app has been developed by the POLLUSCOPE project to ease this annotation step, which was historically done on paper notebooks. Volunteers are also asked to log in possible causes of air pollution level changes: close proximity to any pollution sources, ventilation from opening a window, proximity to cooking or smoking or other micro-environment changes. We observed how the end user could easily be a source of error and oversight or worse, abandon their duties mid-way through the measurement period. Portable air pollution sensors produce detailed data, but key information about the context is necessary to understand in detail the underlying causes of pollution change within the environments and the effects they may have on health outcomes.

##### 4.2. Individual PM Exposure in Relation to Sleep Apnea

POLLAR successfully showed, as a pilot study, how personal sensors can open new opportunities for public health research. The study faced technical difficulties during the first months; prototypes of the Canarin Nano were not equipped with flash storage, leading to some data loss when no 3G coverage was available. Keeping a locally stored, continually updated database in addition to the cloud service would be the optimal option, but was not technically possible at that time due to the hardware limitations of the prototyping board selected. Legal authorization to record geolocalized data was not obtained; therefore, GPS had to be remotely disabled. The study design to assess the personal exposition to air

pollution during the days following the PSG screening made it impossible to investigate if pollution exposure days prior to the screening leads to a degradation of the sleep quality. Comparing the two periods using a linear regression gives no correlation between the averaged PM exposure and the AHI. Such analysis might produce interesting results if exposure data were recorded anterior to the PSG screening.

#### 4.3. Suggestion of Sensor Architecture Design for Academic Research

We saw how relevant information could be obtained directly from the user; inside the POLLUSCOPE project, we asked our volunteers, prior to delivering a set of sensors, to actively log their activity or pollution events during the following week of data acquisition. For this, however, we relied only on the users' awareness and good will. We are still missing crucial information, as the POLLUSCOPE participants were never aware when the sensors detected a drop or increase in pollutant concentration values. Switching from requiring the user to manually annotate events to automatically prompt the user for annotations, based on the sensor's live measurements, would improve considerably the quality and quantity of the available annotated data. In the paragraphs bellow, we will be detailing how a sensor mesh aimed to automate the detection of these events could be the key to expand the scientific knowledge on personal exposure to air pollution, as it would dramatically increase our capacity to analyze more precisely the granular data produced by these modern portable sensors.

To address this issue, we are suggesting a new approach in the development of air pollution sensors; through the creation of a sensor managing device at the center of a sensor mesh network (A). The goal of this managing device is to obtain live information from both the sensors and the user. The managing device would act as the primary human-machine interface for all the other sensors, displaying the acquired sensor data, remaining battery life and other necessary information in real time. If an air pollution event is detected, the managing device will notify the user and prompt them to provide necessary information to contextualize for the event (B). Users' smartphones could be used to interface with the managing device. However, the managing device should keep its full ability to communicate with the user and register its inputs even if the user's smartphone is turned off (Figure 11).

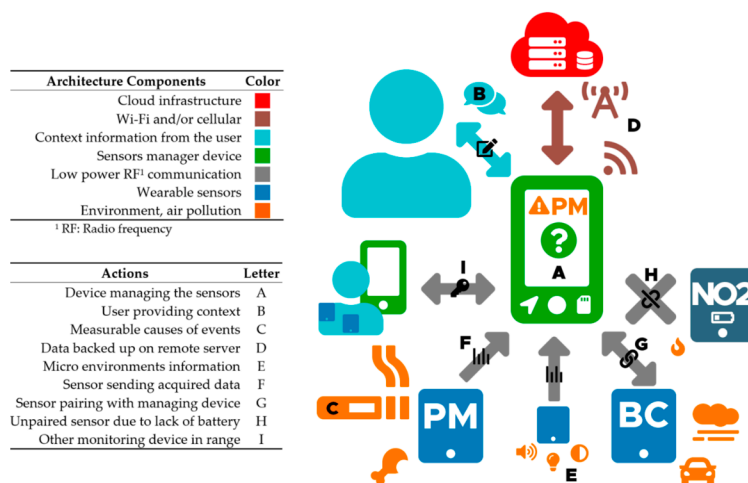


Figure 11. Illustration of the air pollution sensors mesh architecture.



The managing device should be able to react within seconds, reaching user for feedback while it can still be aware of the causes of the event (C). Filtering has to be scaled for each sensor's characteristic to avoid sending false or successive notifications. The managing device should be built around a powerful enough MCU such as PJRC's Teensy boards [54] or development boards based off STMicroelectronics STM32H7 series [55]. Both score above 2000 in CoreMark [56], a microcontroller-dedicated benchmark. A good candidate for our application would be the Portenta H7 [57] from Arduino. It is equipped with a MIPI [58] interface, allowing the integration of high quality touch screens, has wireless connectivity with on-board Wi-Fi and Bluetooth modules, as well as an integrated battery charger for lithium polymer (Li-Po) batteries. These developing boards would provide enough processing power to display a graphic user interface (such as LVGL [59]), while performing continuous processing of both the live acquired data and previous user records. Cellular network modules, low power RF (radio frequency) modules (such as nRF24L01 [60]) and other desired modules and components would need to be interfaced through a custom PCB.

The monitoring device should aim to eliminate any need of post-processing the annotated context data. We were able to experiment within these projects the ways in which deploying sensors in the field can be a burden, each sensor having its own requirements. To solve this issue, the managing device would streamline the entire process, as individual sensors would be paired and start displaying data and battery levels within seconds, on one unique device.

By using a dedicated device instead of a smartphone or tablet, we solve any concerns of hardware limitation (e.g., RF communications with sensors), uniformity of hardware (embedded movement and sound sensors, GPS signal consistency) and software limitations (operating system versions, privacy safeguards and battery optimization limiting constant monitoring of the sensors). Commercial portable sensors, in the best scenarios, are connected to the subject's smartphone via Bluetooth Low Energy (BLE), and use a proprietary app. This reserves the battery life necessary to communicate with a server and still allows the user to visualize sensor data in real time on their smartphones. It becomes cumbersome, however, for studies requiring subjects to wear multiple sensors; the subjects then have to go through multiple apps. Unless a standard emerges from the industry, we do think that a dedicated device would iron out these issues.

In case of connectivity loss with the remote server, all data would be preserved locally, and the architecture would continue functioning as expected. The possibility of connectivity loss by itself justifies the need of running event detection locally on the managing device. Both the sensors and monitoring devices should be able to withstand daily use and keep functioning independently of their orientation, weather conditions and the user's activity (aside from swimming). Remote monitoring of individual sensors should be possible from a web service. Alerts could be sent to users or technicians if a dysfunction is detected. All logged data would be backed up locally and stored permanently on the remote server. Data should be both accessible directly from the monitoring device as well as on the web service (D). A reinitialization clearing the monitoring device's internal database should be performed by a technician for privacy concerns prior to delivering it to another user.

Existing pollution sensors are dedicated devices, rarely monitoring parameters non-related to air pollution. As we would naturally use GPS to sense an environment change, we could, in addition, use ambient color temperature to sense indoor fluorescent bulbs or outdoor natural light. We are, therefore, suggesting to leverage the live analysis done locally by the managing device, by sensing other parameters using affordable and already available hardware. Light intensity, color temperature, sound spectrum and motion sensors would provide useful information to identify micro environments [35] (E). These sensors could be integrated with the monitoring device, or delivered as an optional sensor to pair with the monitoring device.

Power consumption is one of the main factors when considering designing new hardware [61] (p. 297). In the case of mobile sensors, increasing the battery life is always at the cost of the weight and size of the battery. Current meters for low-powered circuits have recently been made affordable [62,63]. Paired with a digital oscilloscope, such tools can read currents with values as low as nanoamperes. During the hardware and software development, the power consumption of integrated circuits and module should be monitored and optimized. Tasks such as geolocation, internet connectivity and storage should only be done by the monitoring device. Sensors that are paired to the monitoring device should not be designed to work as a standalone device. They should instead be designed to conserve battery life and communicate only with the managing device.

They should pair with the push of a button and start sending data to the monitoring device immediately. All sensors communicate to the managing device through the same low-power short-range radio communication (F), enabling ultra-fast and energy efficient data exchange between the sensors and the managing device. Between any data exchange, the microcontrollers running the sensors can be sleeping, saving on battery consumption. Because these communications are not secure, no sensitive information is transmitted by the sensors through low power RF; only the sensor's unique identifier, its model type and raw data. Transmission with the sensors is bidirectional only during pairing (G). If a sensor paired is lost (H), the sensor will be turned off and the user will be notified. Monitoring devices should be able to advertise to other monitoring devices which sensors they are paired with, ensuring that no other monitoring devices in range could accidentally pair to these sensors (I). For sensor testing and calibration, to alternatively use sensors without a monitoring device, a much simpler receiving device without advanced processing or connectivity capabilities could be proposed.

Users should not have to proceed to any kind of manipulation of the sensors other than charging them. Sensors should have a uniform charging port. All interactions with the managing device should be easy enough to be performed by someone with no knowledge of the devices. The managing device should be smart enough to notify the user if a location's samples are unusually high or low, prompting them to provide information accordingly, possibly triggering user behavior, and then prompting them again to document the event if another change in air quality is observed.

We acknowledge that the monitoring device requires considerable efforts in software and hardware development. The developers should ensure that the live data analysis done on the monitoring device is relevant for both the user and the data scientist. On the other hand, we would eliminate the need for standalone sensors, reducing for each unit of sensor; required hardware (cost, weight, size), software and electronic development time, manufacturing and servicing costs. Prototyping modules (low power RF transmitters, Li-Po battery chargers, voltage regulators) are very affordable and already available. Custom PCBs can be rapidly designed to interface modules, sensor modules and other essential hardware together. All in all, this approach would enable rapid integration and real-world testing of the latest air pollution sensing modules as they hit the market [12]. The concept we just described goes beyond solving the burden of the lack of contextual data. Users would be continuously informed on the quality of the air they breathe, and the effects of their behavior. We, therefore, expect to also drastically reduce the accidental data loss that usually occurs, by increasing the user's dedication and awareness.

## 5. Conclusions

In this paper, we presented the Canarin project, and two academically produced sensors: the Canarin II and the Canarin Nano. These sensors were tested against other similar sensors and compared to reference monitors, proving their performance capabilities. As these sensors were designed to aid in health studies, they were trialed in two projects, POLLUSCOPE and POLLAR, where we investigated their utility and ease of use with real cohorts of study participants. With a POLLUSCOPE volunteer, we demonstrated the advantage of applying contextual activity data to the pollutant concentration level

data in order to understand personal exposure and what that means on a daily basis. Collectively, these efforts highlight how the lack of contextual information is one of the shortcomings faced when acquiring granular data. We addressed this issue by offering the concept of an air pollution sensor mesh that would considerably improve the knowledge on individual exposure to air pollution by soliciting the user to provide information based on a sensor's input.

Because the acquired knowledge and experience is invaluable, we are advocating for researchers and students from heterogeneous fields to partner together and create their own devices, as has been done for the Canarin project. Finally, the scientific community should seize the opportunity to create its own open standard for portable sensors, using the best hardware for the task, with an architecture focused on ease of use, subject's engagement and, indeed, data contextualization.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/1424-8220/21/5/1876/s1>, Table S1: Individual PM Exposure of the total number of recorded samples.

**Author Contributions:** Conceptualization, B.D. and I.A.-M.; data curation and formal analysis, B.D.; methodology, B.D. and I.A.-M., supervision, I.A.-M., G.P. and S.S.; writing—original draft preparation, B.D.; writing—review and editing, I.A.-M., J.-L.P. and G.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** The POLLUSCOPE project was funded by the Agence nationale de la recherche (ANR) (Grant ANR-15-CE22-0018). The POLLAR project was funded by the EIT (European Institute of Innovation and Technology) Health (Grant ref: 18354—Impact of POLLution on Asthma and Rhinitis), Foundation 96 FMC-VIA-LR.

**Institutional Review Board Statement:** The POLLAR study was authorized by the CPP and ANSM on 6 December 2018. Following several modifications (Amendments), this study was finally authorized by the CPP and ANSM on 4 of April 2019 and the first patient inclusion was done on 8 April 2019.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The POLLUSCOPE volunteer's data is made available, however the GPS data has been removed for obvious confidentiality concerns. POLLAR data is not made public as obtaining each patients' approval isn't possible.

**Acknowledgments:** We thank the POLLUSCOPE (PI: K. Zeitouni, UVSQ University Paris-Saclay) and POLLAR (PI: S. Zeng, ATOS) consortia. We acknowledge patients and the medical staff who produced the data in Grenoble. Our thanks to the CEREMA Île-de-France staff, who made VGP measurement campaign. Thanks to Zeitouni K., El Hafyani H. and Ktaish A. from DAVID Lab—UVSQ University Paris-Saclay for providing access to POLLUSCOPE project server. Our thanks to the QGIS Development Team (2020). We are especially indebted to Aguiari D.<sup>1</sup>, who developed the Canarin cloud infrastructure and Canarin firmware, Naude J.M.<sup>4</sup> for conducting the VGP survey and authorizing the publication of his personal data and GPS traces and Terrail R. from Grenoble's Medical School, who ensured proper deployment of our sensors. We thank Maesano C. for the extensive review. Finally, we thank the reviewers for their careful reading and comments towards improving the article.

**Conflicts of Interest:** The authors declare no conflict of interest on submitted work.

## References

1. Kampa, M.; Castanas, E. Human Health Effects of Air Pollution. *Environ. Pollut. Barking Essex 1987* **2008**, *151*, 362–367. [[CrossRef](#)] [[PubMed](#)]
2. Thurston, G.D.; Kipen, H.; Annesi-Maesano, I.; Balmes, J.; Brook, R.D.; Cromar, K.; De Matteis, S.; Forastiere, F.; Forsberg, B.; Frampton, M.W.; et al. A Joint ERS/ATS Policy Statement: What Constitutes an Adverse Health Effect of Air Pollution? An Analytical Framework. *Eur. Respir. J.* **2017**, *49*, 1600419. [[CrossRef](#)]
3. Snyder, E.G.; Watkins, T.H.; Solomon, P.A.; Thoma, E.D.; Williams, R.W.; Hagler, G.S.W.; Shelow, D.; Hindin, D.A.; Kilaru, V.J.; Preuss, P.W. The Changing Paradigm of Air Pollution Monitoring. *Environ. Sci. Technol.* **2013**, *47*, 11369–11377. [[CrossRef](#)]
4. Park, J.; Jo, W.; Cho, M.; Lee, J.; Lee, H.; Seo, S.; Lee, C.; Yang, W. Spatial and Temporal Exposure Assessment to PM<sub>2.5</sub> in a Community Using Sensor-Based Air Monitoring Instruments and Dynamic Population Distributions. *Atmosphere* **2020**, *11*, 1284. [[CrossRef](#)]

5. Schneider, P.; Castell, N.; Vogt, M.; Dauge, F.R.; Lahoz, W.A.; Bartonova, A. Mapping Urban Air Quality in near Real-Time Using Observations from Low-Cost Sensors and Model Information. *Environ. Int.* **2017**, *106*, 234–247. [[CrossRef](#)]
6. Mueller, M.; Meyer, J.; Hueglin, C. Design of an Ozone and Nitrogen Dioxide Sensor Unit and Its Long-Term Operation within a Sensor Network in the City of Zurich. *Atmos. Meas. Tech.* **2017**, *10*, 3783–3799. [[CrossRef](#)]
7. Yuval; Magen Molho, H.; Zivan, O.; Broday, D.M.; Raz, R. Application of a Sensor Network of Low Cost Optical Particle Counters for Assessing the Impact of Quarry Emissions on Its Vicinity. *Atmos. Environ.* **2019**, *211*, 29–37. [[CrossRef](#)]
8. Tse, R.; Monti, L.; Prandi, C.; Aguiari, D.; Pau, G.; Salomoni, P. On Assessing the Accuracy of Air Pollution Models Exploiting a Strategic Sensors Deployment. In Proceedings of the 4th EAI International Conference on Smart Objects and Technologies for Social Good-Goodtechs '18, Bologna, Italy, 28–30 November 2018; ACM Press: Bologna, Italy, 2018; pp. 55–58.
9. Kumar, P.; Morawska, L.; Martani, C.; Biskos, G.; Neophytou, M.; Di Sabatino, S.; Bell, M.; Norford, L.; Britter, R. The Rise of Low-Cost Sensing for Managing Air Pollution in Cities. *Environ. Int.* **2015**, *75*, 199–205. [[CrossRef](#)] [[PubMed](#)]
10. Feenstra, B.; Papapostolou, V.; Hasheminassab, S.; Zhang, H.; Boghossian, B.D.; Cocker, D.; Polidori, A. Performance Evaluation of Twelve Low-Cost PM<sub>2.5</sub> Sensors at an Ambient Air Monitoring Site. *Atmos. Environ.* **2019**, *216*, 116946. [[CrossRef](#)]
11. Liu, H.-Y.; Schneider, P.; Haugen, R.; Vogt, M. Performance Assessment of a Low-Cost PM<sub>2.5</sub> Sensor for a near Four-Month Period in Oslo, Norway. *Atmosphere* **2019**, *10*, 41. [[CrossRef](#)]
12. Bauerová, P.; Šindelářová, A.; Rychlík, Š.; Novák, Z.; Keder, J. Low-Cost Air Quality Sensors: One-Year Field Comparative Measurement of Different Gas Sensors and Particle Counters with Reference Monitors at Tušimice Observatory. *Atmosphere* **2020**, *11*, 492. [[CrossRef](#)]
13. Sayahi, T.; Butterfield, A.; Kelly, K.E. Long-Term Field Evaluation of the Plantower PMS Low-Cost Particulate Matter Sensors. *Environ. Pollut.* **2019**, *245*, 932–940. [[CrossRef](#)] [[PubMed](#)]
14. Feinberg, S.; Williams, R.; Hagler, G.S.W.; Rickard, J.; Brown, R.; Garver, D.; Harshfield, G.; Stauffer, P.; Mattson, E.; Judge, R.; et al. Long-Term Evaluation of Air Sensor Technology under Ambient Conditions in Denver, Colorado. *Atmos. Meas. Tech.* **2018**, *11*, 4605–4615. [[CrossRef](#)]
15. Tryner, J.; L'Orange, C.; Mehaffy, J.; Miller-Lionberg, D.; Hofstetter, J.C.; Wilson, A.; Volckens, J. Laboratory Evaluation of Low-Cost PurpleAir PM Monitors and in-Field Correction Using Co-Located Portable Filter Samplers. *Atmos. Environ.* **2020**, *220*, 117067. [[CrossRef](#)]
16. Zheng, T.; Bergin, M.H.; Johnson, K.K.; Tripathi, S.N.; Shirodkar, S.; Landis, M.S.; Sutaria, R.; Carlson, D.E. Field Evaluation of Low-Cost Particulate Matter Sensors in High- and Low-Concentration Environments. *Atmos. Meas. Tech.* **2018**, *11*, 4823–4846. [[CrossRef](#)]
17. Moreno-Rangel, A.; Sharpe, T.; Musau, F.; McGill, G. Field Evaluation of a Low-Cost Indoor Air Quality Monitor to Quantify Exposure to Pollutants in Residential Environments. *J. Sens. Sens. Syst.* **2018**, *7*, 373–388. [[CrossRef](#)]
18. Zikova, N.; Hopke, P.K.; Ferro, A.R. Evaluation of New Low-Cost Particle Monitors for PM<sub>2.5</sub> Concentrations Measurements. *J. Aerosol Sci.* **2017**, *105*, 24–34. [[CrossRef](#)]
19. Badura, M.; Batog, P.; Drzeniecka-Osiadacz, A.; Modzel, P. Evaluation of Low-Cost Sensors for Ambient PM<sub>2.5</sub> Monitoring. *J. Sens.* **2018**, *2018*, 5096540. [[CrossRef](#)]
20. Crilley, L.R.; Shaw, M.; Pound, R.; Kramer, L.J.; Price, R.; Young, S.; Lewis, A.C.; Pope, F.D. Evaluation of a Low-Cost Optical Particle Counter (Alphasense OPC-N2) for Ambient Air Monitoring. *Atmos. Meas. Tech.* **2018**, *11*, 709–720. [[CrossRef](#)]
21. Languille, B.; Gros, V.; Bonnaire, N.; Pommier, C.; Honoré, C.; Debert, C.; Gauvin, L.; Srairi, S.; Annesi-Maesano, I.; Chaix, B.; et al. A Methodology for the Characterization of Portable Sensors for Air Quality Measure with the Goal of Deployment in Citizen Science. *Sci. Total Environ.* **2020**, *708*, 134698. [[CrossRef](#)] [[PubMed](#)]
22. Samad, A.; Melchor Mimiaga, F.E.; Laquai, B.; Vogt, U. Investigating a Low-Cost Dryer Designed for Low-Cost PM Sensors Measuring Ambient Air Quality. *Sensors* **2021**, *21*, 804. [[CrossRef](#)]
23. Papapostolou, V.; Zhang, H.; Feenstra, B.J.; Polidori, A. Development of an Environmental Chamber for Evaluating the Performance of Low-Cost Air Quality Sensors under Controlled Conditions. *Atmos. Environ.* **2017**, *171*, 82–90. [[CrossRef](#)]
24. Lesser, D.; Katra, I.; Dorman, M.; Harari, H.; Kloog, I. Validating and Comparing Highly Resolved Commercial “Off the Shelf” PM Monitoring Sensors with Satellite Based Hybrid Models, for Improved Environmental Exposure Assessment. *Sensors* **2020**, *21*, 63. [[CrossRef](#)]
25. Hasenfratz, D.; Saukh, O.; Sturzenegger, S.; Thiele, L. Participatory Air Pollution Monitoring Using Smartphones. In Proceedings of the 2nd International Workshop on Mobile Sensing, Beijing, China, 16–20 April 2012.
26. Dutta, J.; Chowdhury, C.; Roy, S.; Middya, A.L.; Gazi, F. Towards Smart City: Sensing Air Quality in City Based on Opportunistic Crowd-Sensing. In Proceedings of the 18th International Conference on Distributed Computing and Networking, Hyderabad, India, 5–7 January 2017; ACM: Hyderabad, India, 2017; pp. 1–6.
27. Sivaraman, V.; Carrapetta, J.; Hu, K.; Luxan, B.G. HazeWatch: A Participatory Sensor System for Monitoring Air Pollution in Sydney. In Proceedings of the 38th Annual IEEE Conference on Local Computer Networks-Workshops, Sydney, Australia, 21–24 October 2013; IEEE: Sydney, Australia, 2013; pp. 56–64.
28. Guo, B.; Yu, Z.; Zhou, X.; Zhang, D. From Participatory Sensing to Mobile Crowd Sensing. In Proceedings of the 2014 IEEE International Conference on Pervasive Computing and Communication Workshops (PERCOM WORKSHOPS), Budapest, Hungary, 24–28 March 2014.
29. Zeng, Y.; Xiang, K. Adaptive Sampling for Urban Air Quality through Participatory Sensing. *Sensors* **2017**, *17*, 2531. [[CrossRef](#)]

30. Galinina, O.; Mikhaylov, K.; Huang, K.; Andreev, S.; Koucheryavy, Y. Wirelessly Powered Urban Crowd Sensing over Wearables: Trading Energy for Data. *IEEE Wirel. Commun.* **2018**, *25*, 140–149. [[CrossRef](#)]
31. Capponi, A.; Fiandrino, C.; Kantarci, B.; Foschini, L.; Kliazovich, D.; Bouvry, P. A Survey on Mobile Crowdsensing Systems: Challenges, Solutions, and Opportunities. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 2419–2465. [[CrossRef](#)]
32. Maag, B.; Zhou, Z.; Thiele, L. W-Air: Enabling Personal Air Pollution Monitoring on Wearables. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2018**, *2*, 1–25. [[CrossRef](#)]
33. Jerrett, M.; Donaire-Gonzalez, D.; Popoola, O.; Jones, R.; Cohen, R.C.; Almanza, E.; de Nazelle, A.; Mead, I.; Carrasco-Turigas, G.; Cole-Hunter, T.; et al. Validating Novel Air Pollution Sensors to Improve Exposure Estimates for Epidemiological Analyses and Citizen Science. *Environ. Res.* **2017**, *158*, 286–294. [[CrossRef](#)] [[PubMed](#)]
34. Piedrahita, R.; Xiang, Y.; Masson, N.; Ortega, J.; Collier, A.; Jiang, Y.; Li, K.; Dick, R.P.; Lv, Q.; Hannigan, M.; et al. The next Generation of Low-Cost Personal Air Quality Sensors for Quantitative Exposure Monitoring. *Atmos. Meas. Tech.* **2014**, *7*, 3325–3336. [[CrossRef](#)]
35. Wong, M.; Yip, T.; Mok, E. Development of a Personal Integrated Environmental Monitoring System. *Sensors* **2014**, *14*, 22065–22081. [[CrossRef](#)]
36. Zhuang, Y.; Lin, F.; Yoo, E.-H.; Xu, W. AirSense: A Portable Context-Sensing Device for Personal Air Quality Monitoring. In Proceedings of the 2015 Workshop on Pervasive Wireless Healthcare, Hangzhou, China, 22 June 2015; ACM: Hangzhou, China, 2015; pp. 17–22.
37. Dutta, P.; Aoki, P.M.; Kumar, N.; Mainwaring, A.; Myers, C.; Willett, W.; Woodruff, A. Common Sense: Participatory Urban Sensing Using a Network of Handheld Air Quality Monitors. In Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems—SenSys '09, Berkeley, CA, USA, 4–6 November 2009; ACM Press: Berkeley, CA, USA, 2009; p. 349.
38. Flow, by Plume Labs: The First Smart Air Quality Tracker. Available online: <https://plumelabs.com/en/flow/> (accessed on 3 February 2021).
39. Atmotube—A Series of Portable Air Quality Monitors. Available online: <https://notanotherone.com/en/atmotube#> (accessed on 3 February 2021).
40. Observatoire Participatif Pour La Surveillance de L'exposition Individuelle à La Pollution de L'air En Lien Avec La Santé | ANR. Available online: <https://anr.fr/Projet-ANR-15-CE22-0018> (accessed on 9 February 2021).
41. Sensing the Air Quality: Research on Air Quality Sensors. Available online: <https://aqicn.org/sensor/> (accessed on 28 January 2021).
42. STIC-ASIA: SEA HAZEMON. Available online: <https://interlab.ait.ac.th/HAZEMON/> (accessed on 2 February 2021).
43. Tse, R.; Pau, G. Enabling Street-Level Pollution and Exposure Measures: A Human-Centric Approach. In Proceedings of the 6th ACM International Workshop on Pervasive Wireless Healthcare—MobiHealth '16, Paderborn, Germany, 5 July 2016; ACM Press: Paderborn, Germany, 2016; pp. 1–4.
44. Aguiari, D.; Delnevo, G.; Monti, L.; Ghini, V.; Mirri, S.; Salomoni, P.; Pau, G.; Im, M.; Tse, R.; Ekpanyapong, M.; et al. Canarin II: Designing a Smart e-Bike Eco-System. In Proceedings of the 2018 15th IEEE Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 12–15 January 2018; IEEE: Las Vegas, NV, USA, 2018; pp. 1–6.
45. Bédard, A.; Sofiev, M.; Arnavielhe, S.; Antó, J.M.; Garcia-Aymerich, J.; Thibaudon, M.; Bergmann, K.C.; Dubakiene, R.; Bedbrook, A.; Onorato, G.L.; et al. Interactions Between Air Pollution and Pollen Season for Rhinitis Using Mobile Technology: A MASK-POLLAR Study. *J. Allergy Clin. Immunol. Pract.* **2020**, *8*, 1063–1073. [[CrossRef](#)] [[PubMed](#)]
46. Introduction-UDOO Neo Docs. Available online: <https://www.udoo.org/docs-neo/Introduction/Introduction.html> (accessed on 1 February 2021).
47. Global IoT SIM Card & M2M Connectivity | Things Mobile. Available online: <https://www.thingsmobile.com/business> (accessed on 8 February 2021).
48. Instances T2 Amazon EC2—Amazon Web Services (AWS). Available online: <https://aws.amazon.com/fr/ec2/instance-types/t2/> (accessed on 27 January 2021).
49. What Is a Container? | App Containerization | Docker. Available online: <https://www.docker.com/resources/what-container> (accessed on 27 January 2021).
50. Canarin.Net. Available online: <https://canarin.net/> (accessed on 27 January 2021).
51. Accueil-Cohorte RECORD. Available online: <https://www.hal.inserm.fr/RECORD> (accessed on 1 February 2021).
52. Cairsens@NO2 | Micro-Sensors | Ambient | ENVEA. Available online: <https://www.envea.global/s/ambient-en/micro-sensors-a/cairsens-no2/> (accessed on 8 February 2021).
53. Accueil—Le Réseau National de Surveillance Aérobiologique—RNSA. Available online: <https://www.pollens.fr/> (accessed on 9 February 2021).
54. Teensy 4.1. Available online: <https://www.pjrc.com/store/teensy41.html> (accessed on 27 January 2021).
55. STM32H7-Arm Cortex-M7 and Cortex-M4 MCUs (480 MHz)—STMicroelectronics. Available online: <https://www.st.com/en/microcontrollers-microprocessors/stm32h7-series.html> (accessed on 27 January 2021).
56. Benchmark Product List-EEMBC-Embedded Microprocessor Benchmark Consortium. Available online: <https://www.eembc.org/products/> (accessed on 1 February 2021).
57. Portenta H7-Arduino Boards & Modules-Arduino. Available online: <https://store.arduino.cc/portenta-h7> (accessed on 27 January 2021).
58. MIPI Display Serial Interface (MIPI DSI). Available online: <https://mipi.org/specifications/dsi> (accessed on 1 February 2021).

59. Limited Liability Company, LVGL-Light and Versatile Embedded Graphics Library. Available online: <https://lvgl.io/> (accessed on 1 February 2021).
60. NRF24 Series-Nordic Semiconductor. Available online: <https://www.nordicsemi.com/Products/Low-power-short-range-wireless/nRF24-series> (accessed on 1 February 2021).
61. Morawska, L.; Thai, P.K.; Liu, X.; Asumadu-Sakyi, A.; Ayoko, G.; Bartonova, A.; Bedini, A.; Chai, F.; Christensen, B.; Dunbabin, M.; et al. Applications of Low-Cost Sensing Technologies for Air Quality Monitoring and Exposure Assessment: How Far Have They Gone? *Environ. Int.* **2018**, *116*, 286–299. [CrossRef]
62. Current Ranger | LowPowerLab. Available online: <https://lowpowerlab.com/guide/currentranger/> (accessed on 2 February 2021).
63. EEVblog UCurrent-Precision NA Current Measurement Assistant [v3] ID: 882-\$59.95: Adafruit Industries, Unique & Fun DIY Electronics and Kits. Available online: <https://www.adafruit.com/product/882> (accessed on 2 February 2021).

12.2. POSTER ET FLYER — PROJET POLLUSCOPE

**POLLUSCOPE**

Participatory observatory for the monitoring of individual exposure to air pollution in relation with health status

**Cohort of volunteers, wearing individual sensors**

- humidity, pressure, temperature
- PM1, PM2.5, PM10
- NO2
- Black Carbon
- Spirometry et oxymetry
- GPS tracking

**Data stored locally** in each sensor's internal memory

**OR**

Connected sensors with a **secured cloud service**

**Ability to monitor in real time and act fast** in case of a malfunction or a volunteer not properly using the sensors

The server will process the data to aggregate the different data sources. This will ease the life of the statisticians and the public. It's called **data fusion**.

Locally stored data are uploaded through a **web platform** dedicated to the project. They are then **totally anonymised** and hosted on a **secured server**.

**Anonymised and processed data**

At the end of the project, some of the data will be available to the public. The platform will enable volunteers to **visualise their personal data** as time trends, but also to plot it on a map. Advanced user will be able to **download their data**.

**Partners:** ANR, Inserm, AIRPARIF, david, IRENAV, EIVP, CNRS, cap-digital, LSCE, Sorbonne Université, Cerema, Université de Versailles Saint-Quentin-en-Yvelines.

## LE PETIT GUIDE DU PORTEUR DE CAPTEURS

# POLLUSCOPE

Observatoire participatif pour la surveillance de l'exposition individuelle à la pollution de l'air en lien avec la santé



## QUELQUES INFORMATIONS GENERALES

L'autonomie des capteurs est d'environ de 24H. Il faut donc veiller à charger les capteurs chaque soir à l'aide de leurs chargeurs respectifs. Il est possible de les laisser branchés.

### Lorsque transportés :

Les capteurs doivent être portés de la façon qui vous paraît la plus convenable. Veillez à ce qu'ils soient exposés à l'air ambiant pour la bonne validité des mesures. Dans le cas d'un port par une sangle, ils doivent être sécurisés afin d'éviter le risque de chute du capteur.

### Lorsque entreposés :

Dans le cas des capteurs qui mesurent en continu, veillez à ce que la prise d'air soit dégagée. Ils doivent être à l'écart de sources directes de pollution (cheminée, four, cuisson...), de source de chaleur (radiateur, réchaud, TV), ainsi que des endroits très ventilés ou humides. Les capteurs ne sont pas étanches. En cas de pluie, il faut les mettre à l'abri. Entreposez les à l'écart des enfants et des animaux, afin d'éviter le risque de chute du capteur.

POLLUSCOPE

## LE GUIDE DU PORTEUR DE CANARIN

Dans le cas où vous choisissez de porter le Canarin dans son sac de transport respirant :

- Il doit être posé à plat afin qu'il puisse se géolocaliser
- L'inscription «Canarin II» vers le ciel

Dans le cas où vous choisissez de le porter avec une sangle, portez le e la façon qui vous est le plus confortable.

Le **Canarin** est un capteur mesurant chaque minute les PM1, PM2.5 et PM10 par mesure optique. Il fait également l'acquisition de la température l'humidité de la pression, et la géolocalisation par GPS.

Le **Cairsens** est un capteur miniature indépendant qui permet de mesurer le dioxyde d'azote. Il est accroché au Canarin par souci de praticité.



POLLUSCOPE

## LE GUIDE DU PORTEUR DE L'AE51 PAGE 1

Le MicroAeth AE51 est un aethalomètre analysant la concentration en carbone élémentaire par la méthode classique d'absorption d'un rayonnement infrarouge.



### Recommandations :

- Veillez à ce que le tuyau soit non obstrué
- Chargez le capteur chaque soir à l'aide de son chargeur
- Changez le filtre chaque soir en veillant à respecter le sens la partie métallisée est du côté du bouton poussoir
- Arrêtez le capteur durant le changement du filtre
- Conservez le filtre usagé dans le sachet prévu à cet effet
- Ne jamais éteindre le capteur sauf si dysfonctionnement.



POLLUSCOPE

## LE GUIDE DU PORTEUR DE L'AE51 PAGE 2

Le capteur AE51 dispose d'un embout flexible qui doit être fixé autant que possible à hauteur de respiration (col de veste, sangle de sac) ou à défaut, à proximité de la prise d'air du Canarin (sur la sangle de celui-ci par exemple).

### Lorsque transporté :

Le capteur AE51 peut être disposé dans la poche du sac à dos fourni, ou dans un logement équivalent.

### Lorsque entreposé :

Le capteur AE51 doit être posé verticalement, en fixant l'embout du tuyau à hauteur d'homme.



POLLUSCOPE

## GUIDE DE L'UTILISATEUR DE LA TABLETTE 4G

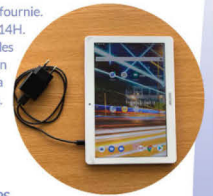
Une tablette 4G vous est fournie. Son autonomie est d'environ 14H.

Toute autre application que celles destinées aux mesures de santé ont un accès internet restreint afin d'éviter la surconsommation de données mobiles.

Merci de ne pas l'utiliser pour un usage autre que celui prévu.

### Rôle de la tablette :

- Enregistrer vos déplacements via GPS
- Fournir un point d'accès internet au Canarin
- Transmettre vos résultats de tests de spirométrie et oxymétrie



La **spirométrie** est un examen indolore qui permet de mesurer la quantité d'air dans vos poumons. Votre Spirobank doit être utilisé avec l'application PneumotelUploader. Les instructions d'appariage se trouvent à l'intérieur de la boîte du Spirobank.

L'**oxymétrie** permet d'évaluer votre état d'oxygénation au repos. Le test s'effectue à la suite de chaque test de spirométrie. L'application vous fournira des indications à l'écran.



POLLUSCOPE

## LE GUIDE DE L'UTILISATEUR DE SPIROBANK

L'expiration forcée peut entraîner de la fatigue, mais cela est passager.

### Instructions complémentaires :

- Activez le bluetooth de votre smartphone.
- Soyez debout en gardant le dos droit
- Respirez calmement quelques secondes
- Bouchez votre nez
- Saisissez l'appareil avec les deux mains
- Inspirez le maximum d'air possible
- Appuyez sur «Lancer un test de soufflé»
- Tenez le spiromètre à deux mains
- Soufflez par la bouche dans l'embout du spiromètre
- Vous devez expirer violemment, le plus fort possible, et essayer de continuer à expirer durant les 6 secondes nécessaire au test
- Si vous tousssez ou si vous inspirez, merci de recommencer
- Un message vous indiquant que vos données ont été envoyées apparaîtra à l'issue du test. Si ce message n'apparaît pas, contactez-nous.



Vous devez recommencer le test au minimum 3 fois. Le test doit être fait au matin, le soir, et avant de se coucher. Vous expirez donc au minimum 9 fois par jour dans le spiromètre.

Un technicien vous rendra visite afin de vous accompagner pour réaliser une mesure du monoxyde de carbone (eCO), ainsi qu'une mesure du monoxyde d'azote (NO).

POLLUSCOPE

## PARTENAIRES



POLLUSCOPE

## MERCI A VOUS

Sans votre participation active la réalisation de ce projet ne serait pas possible ! C'est pourquoi nous tenons à vous remercier au nom de tous les partenaires du projet Polluscope.

Site du projet : [polluscope.uvsq.fr](http://polluscope.uvsq.fr)  
Email de Contact : [polluscope@uvsq.fr](mailto:polluscope@uvsq.fr)  
Téléphone : 01.34.82.12.59 - 07.86.67.28.49





USER LEAFLET

## Welcome

The Canarin nano is a portable sensor measuring personal exposition to air pollution at a high resolution. Data is automatically uploaded to cloud services. The Canarin status and values can be monitored in real time.

Here's your unique Canarin ID :

nano\_1\_

## Sensors Cautions

Never store the sensor in a backpack or a purse, as those place can have dust or tiny debris that could damage the PM sensor. Prefer using the hook if you want to measure while being mobile, or store it in the provided plastic box

None of the intake or exhausts of the sensors should be obstructed. **The Canarin should never enter in direct contact with heavy dust, chemicals, or water.**

IMPORTANT INSTRUCTIONS

## Dysfunction

Because the Canarin nano is a prototype sensor, a tool for research and not a commercial product, it's prone to possible bugs or dysfunctions. If you observe an abnormal behaviour, feel free to do the following : turn on and off, push the reset button, push the battery reset button.

**Please signal any dysfunction as it will greatly help improve the Canarin nano development.**

IMPORTANT INSTRUCTIONS

## Validation

Assembled by :

Date of assembly & testing :

Sensors included (if missing, the cell is filled):

PMs :

VOCs / H2 / CO2 :

PHT / VOCs index :

GPS Antenna Type : Dipole  Active

USER LEAFLET

## Acknowledgements

This device is the result of the work and collaboration of multiple students from multiple Universities.



Professor Clemente Perin  
 Director, Sorbonne Université-Mexico  
 Enrique Aguilar  
 Juan De Vitor  
 Professor, Sorbonne Université  
 Leung Si Sang Chan  
 Ben Desamero

USER LEAFLET

## Charging

Charging is done using the provided micro USB cable and USB wall charger. When plugged, independently of the On/Off switch state, the Canarin nano will light up and start functioning.

The switch stops all connection to the battery and should be off only when the Canarin nano is stored. **The battery will only be recharged if the On/Off switch is on ON position.**

IMPORTANT INSTRUCTIONS

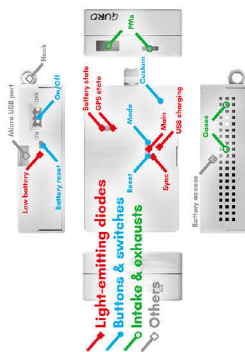
## If Discharged

If you attempt to turn the Canarin on or plug it while the battery is discharged, the main led will blink red and yellow. After the blinking sequence, the Canarin will go in a sleep mode and only the low battery LED will stay turned on with a steady red. Once plugged, it will start working again after approximately 30 minutes.

**Never leave the Canarin on while discharged. Turn it off or to recharge it immediately.**

IMPORTANT INSTRUCTIONS

## Lexicon



TECHNICAL INFORMATION

## Technical Specs

**Processor** : STM32F205RGT6 ARM CortexM3

**Sensors** : PMs, TVOC, Gas resistance, PHT

**Size** : 9.65 x 6.24 x 3.19 cm

**Sampling** : every minute, adjustable

**Storage** : 1 + 8 MB ; 48 days of sampling

**Connectivity** : 3G modem u-bloxSARA U-series

**GPS** : NEO-M8N

**Battery** : 4000 mAh ; 14 hours of sampling

TECHNICAL INFORMATION

## Personal Data

Each time the Canarin is given to a volunteer, a new authentication QR code is generated. Therefore you will be the only one user able to view and download your exposure and GPS location.

**The data you are producing will be analysed by scientists, but is anonymised and comply with GDPR EU law on data protection and privacy.**

DATA ACCESS

## Thanks

Thanks for participating in a research project. Your personal investment and efforts will help produce valuable scientific informations.



USER LEAFLET

## Web portal

To access the web portal and visualise your data : [www.canarin.net](http://www.canarin.net)

Optionnally, you can scan the QR code below :

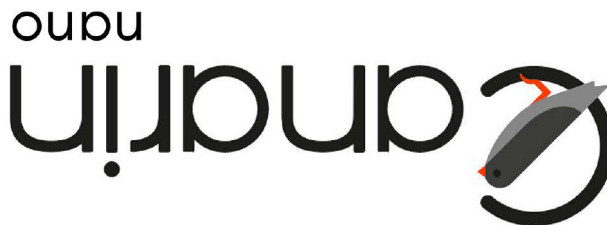


## Authentication

On Canarin.net, first provide the Canarin ID, then authenticate using your personal unique password.

Optionnally, the QR code will allow your copy & paste your personal unique password.

DATA ACCESS



## 13. BIBLIOGRAPHIE

1. Kampa, M.; Castanas, E. Human Health Effects of Air Pollution. *Environ Pollut* **2008**, *151*, 362–367, doi:10.1016/j.envpol.2007.06.012.
2. Air Pollution, Ecosystems and Biodiversity - Accessible en ligne : <https://unece.org/air-pollution-ecosystems-andbiodiversity#:~:text=Ecosystems%20are%20impacted%20by%20air,ability%20to%20function%20and%20grow.&text=Nutrient%20overloads%20in%20aquatic%20ecosystems,of%20oxygen%2C%20and%20of%20life>.
3. Thurston, G.D.; Kipen, H.; Annesi-Maesano, I.; Balmes, J.; Brook, R.D.; Cromar, K.; De Matteis, S.; Forastiere, F.; Forsberg, B.; Frampton, M.W.; et al. A Joint ERS/ATS Policy Statement: What Constitutes an Adverse Health Effect of Air Pollution? An Analytical Framework. *Eur Respir J* **2017**, *49*, doi:10.1183/13993003.00419-2016.
4. Henschel, S. Health Risks of Air Pollution in Europe – HRAPIE Project. 65.
5. Burnett, R.; Chen, H.; Szyszkowicz, M.; Fann, N.; Hubbell, B.; Pope, C.A.; Apte, J.S.; Brauer, M.; Cohen, A.; Weichenthal, S.; et al. Global Estimates of Mortality Associated with Long-Term Exposure to Outdoor Fine Particulate Matter. *Proc Natl Acad Sci USA* **2018**, *115*, 9592–9597, doi:10.1073/pnas.1803222115.
6. Beach, B.; Hanlon, W.W. Coal Smoke and Mortality in an Early Industrial Economy. *The Economic Journal* **2018**, *128*, 2652–2675, doi:10.1111/eoj.12522.
7. Boyd, D.R. The Human Right to Breathe Clean Air. *Annals of Global Health* **2019**, *85*, 146, doi:10.5334/aogh.2646.
8. New WHO Global Air Quality Guidelines Aim to Save Millions of Lives from Air Pollution - Accessible en ligne : <https://www.who.int/news/item/22-09-2021-new-who-global-air-quality-guidelines-aim-to-save-millions-of-lives-from-air-pollution>.
9. Pollution de l'air Ambiant : Nouvelles Estimations de Son Impact Sur La Santé Des Français - Accessible en ligne : <https://www.santepubliquefrance.fr/presse/2021/pollution-de-l-air-ambiant-nouvelles-estimations-de-son-impact-sur-la-sante-des-francais>.
10. Rapports Du Groupe d'experts Intergouvernemental Sur l'évolution Du Climat - Accessible en ligne : <https://www.ipcc.ch/reports/>.
11. The World Leaders Summit at COP 26 - Accessible en ligne : <https://unfccc.int/cop26/world-leaders-summit>.
12. Thurston, G.D.; Balmes, J.R.; Garcia, E.; Gilliland, F.D.; Rice, M.B.; Schikowski, T.; Van Winkle, L.S.; Annesi-Maesano, I.; Burchard, E.G.; Carlsten, C.; et al. Outdoor Air Pollution and New-Onset Airway Disease. An Official American Thoracic Society Workshop Report. *Annals ATS* **2020**, *17*, 387–398, doi:10.1513/AnnalsATS.202001-046ST.
13. Liu, C.; Li, J.; Zhang, Y.; Wang, L.; Deng, J.; Gao, Y.; Yu, L.; Zhang, J.; Sun, H. Widespread Distribution of PET and PC Microplastics in Dust in Urban China and Their Estimated Human Exposure. *Environment International* **2019**, *128*, 116–124, doi:10.1016/j.envint.2019.04.024.
14. Younan, D.; Petkus, A.J.; Widaman, K.F.; Wang, X.; Casanova, R.; Espeland, M.A.; Gatz, M.; Henderson, V.W.; Manson, J.E.; Rapp, S.R.; et al. Particulate Matter and Episodic Memory Decline Mediated by Early Neuroanatomic Biomarkers of Alzheimer's Disease. *Brain* **2020**, *143*, 289–302, doi:10.1093/brain/awz348.
15. Pandey, J.S.; Kumar, R.; Devotta, S. Health Risks of NO<sub>2</sub>, SPM and SO<sub>2</sub> in Delhi (India). *Atmospheric Environment* **2005**, *39*, 6868–6874, doi:10.1016/j.atmosenv.2005.08.004.
16. Achakulwisut, P.; Brauer, M.; Hystad, P.; Anenberg, S.C. Global, National, and Urban Burdens of Paediatric Asthma Incidence Attributable to Ambient NO<sub>2</sub> Pollution: Estimates from Global Datasets. *The Lancet Planetary Health* **2019**, *3*, e166–e178, doi:10.1016/S2542-5196(19)30046-4.
17. Liste Des Composés Organiques Volatils (COV) Mesurés Dans l'atmosphère - Accessible en ligne : <https://www.environnement.gouv.qc.ca/air/cov/liste.htm>.
18. Vivaldo, G.; Masi, E.; Taiti, C.; Caldarelli, G.; Mancuso, S. The Network of Plants Volatile Organic Compounds. *Sci Rep* **2017**, *7*, 11050, doi:10.1038/s41598-017-10975-x.
19. Les Composés Organiques Volatils (COV) - Accessible en ligne : <https://www.ademe.fr/entreprises-monde-agricole/reduire-impacts/reduire-emissions-polluants/dossier/composes-organiques-volatils-cov/definition-sources-demission-impacts>.
20. Mustafa, M.F.; Liu, Y.; Duan, Z.; Guo, H.; Xu, S.; Wang, H.; Lu, W. Volatile Compounds Emission and Health Risk Assessment during Composting of Organic Fraction of Municipal Solid Waste. *Journal of Hazardous Materials* **2017**, *327*, 35–43, doi:10.1016/j.jhazmat.2016.11.046.
21. Domingo, J.L.; Rovira, J.; Vilavert, L.; Nadal, M.; Figueras, M.J.; Schuhmacher, M. Health Risks for the Population Living in the Vicinity of an Integrated Waste Management Facility: Screening Environmental Pollutants. *Science of The Total*

*Environment* **2015**, 518–519, 363–370, doi:10.1016/j.scitotenv.2015.03.010.

22. Mo, Z.; Lu, S.; Shao, M. Volatile Organic Compound (VOC) Emissions and Health Risk Assessment in Paint and Coatings Industry in the Yangtze River Delta, China. *Environmental Pollution* **2021**, 269, 115740, doi:10.1016/j.envpol.2020.115740.

23. Cui, P.; Schito, G.; Cui, Q. VOC Emissions from Asphalt Pavement and Health Risks to Construction Workers. *Journal of Cleaner Production* **2020**, 244, 118757, doi:10.1016/j.jclepro.2019.118757.

24. Aschengrau, A.; Winter, M.R.; Gallagher, L.G.; Vieira, V.M.; Butler, L.J.; Fabian, M.P.; Carwile, J.L.; Wesselink, A.K.; Mahalingaiah, S.; Janulewicz, P.A.; et al. Reproductive and Developmental Health Effects of Prenatal Exposure to Tetrachloroethylene-Contaminated Drinking Water. *Environ. Sci.: Processes Impacts* **2020**, 22, 555–566, doi:10.1039/C9EM00590K.

25. Bahadar, H.; Mostafalou, S.; Abdollahi, M. Current Understandings and Perspectives on Non-Cancer Health Effects of Benzene: A Global Concern. *Toxicology and Applied Pharmacology* **2014**, 276, 83–94, doi:10.1016/j.taap.2014.02.012.

26. Oberdörster, G. Pulmonary Effects of Inhaled Ultrafine Particles. *International Archives of Occupational and Environmental Health* **2000**, 74, 1–8, doi:10.1007/s004200000185.

27. Janssen, N.A.H.; Hoek, G.; Simic-Lawson, M.; Fischer, P.; van Bree, L.; ten Brink, H.; Keuken, M.; Atkinson, R.W.; Anderson, H.R.; Brunekreef, B.; et al. Black Carbon as an Additional Indicator of the Adverse Health Effects of Airborne Particles Compared with PM<sub>10</sub> and PM<sub>2.5</sub>. *Environmental Health Perspectives* **2011**, 119, 1691–1699, doi:10.1289/ehp.1003369.

28. Anenberg, S.C.; Schwartz, J.; Shindell, D.; Amann, M.; Faluvegi, G.; Klimont, Z.; Janssens-Maenhout, G.; Pozzoli, L.; Van Dingenen, R.; Vignati, E.; et al. Global Air Quality and Health Co-Benefits of Mitigating Near-Term Climate Change through Methane and Black Carbon Emission Controls. *Environmental Health Perspectives* **2012**, 120, 831–839, doi:10.1289/ehp.1104301.

29. Wu, Y.; Li, R.; Cui, L.; Meng, Y.; Cheng, H.; Fu, H. The High-Resolution Estimation of Sulfur Dioxide (SO<sub>2</sub>) Concentration, Health Effect and Monetary Costs in Beijing. *Chemosphere* **2020**, 241, 125031, doi:10.1016/j.chemosphere.2019.125031.

30. Khaniabadi, Y.O.; Polosa, R.; Chuturkova, R.Z.; Daryanoosh, M.; Goudarzi, G.; Borgini, A.; Tittarelli, A.; Basiri, H.; Armin, H.; Nourmoradi, H.; et al. Human Health Risk Assessment Due to Ambient PM<sub>10</sub> and SO<sub>2</sub> by an Air Quality Modeling Technique. *Process Safety and Environmental Protection* **2017**, 111, 346–354, doi:10.1016/j.psep.2017.07.018.

31. Bernardini, F.; Attademo, L.; Trezzi, R.; Gobbicchi, C.; Balducci, P.M.; Del Bello, V.; Menculini, G.; Pauselli, L.; Piselli, M.; Sciarra, T.; et al. Air Pollutants and Daily Number of Admissions to Psychiatric Emergency Services: Evidence for Detrimental Mental Health Effects of Ozone. *Epidemiol Psychiatr Sci* **2020**, 29, e66, doi:10.1017/S2045796019000623.

32. Sweileh, W.M.; Al-Jabi, S.W.; Zyoud, S.H.; Sawalha, A.F. Outdoor Air Pollution and Respiratory Health: A Bibliometric Analysis of Publications in Peer-Reviewed Journals (1900 – 2017). *Multidiscip Respir Med* **2018**, 13, 15, doi:10.1186/s40248-018-0128-5.

33. Baromètre 2021, La Perception Des Risques et de La Sécurité Par Les Français l'analyse - Accessible en ligne : [https://www.irsn.fr/FR/IRSN/Publications/barometre/Documents/IRSN-Barometre\\_2021-Analyse.pdf](https://www.irsn.fr/FR/IRSN/Publications/barometre/Documents/IRSN-Barometre_2021-Analyse.pdf).

34. Cori, L.; Donzelli, G.; Gorini, F.; Bianchi, F.; Curzio, O. Risk Perception of Air Pollution: A Systematic Review Focused on Particulate Matter Exposure. *IJERPH* **2020**, 17, 6424, doi:10.3390/ijerph17176424.

35. King, K.E. Chicago Residents' Perceptions of Air Quality: Objective Pollution, the Built Environment, and Neighborhood Stigma Theory. *Popul Environ* **2015**, 37, 1–21, doi:10.1007/s11111-014-0228-x.

36. Hansstein, F.V.; Echegaray, F. Exploring Motivations behind Pollution-Mask Use in a Sample of Young Adults in Urban China. *Global Health* **2018**, 14, 122, doi:10.1186/s12992-018-0441-y.

37. Guo, Y.; Liu, F.; Lu, Y.; Mao, Z.; Lu, H.; Wu, Y.; Chu, Y.; Yu, L.; Liu, Y.; Ren, M.; et al. Factors Affecting Parent's Perception on Air Quality—From the Individual to the Community Level. *IJERPH* **2016**, 13, 493, doi:10.3390/ijerph13050493.

38. Emissions in the Automotive Sector - Accessible en ligne : [https://ec.europa.eu/growth/sectors/automotive/environment-protection/emissions\\_en](https://ec.europa.eu/growth/sectors/automotive/environment-protection/emissions_en).

39. *Les politiques publiques en faveur des véhicules à très faibles émissions*; France Stratégie, 2018; p. 140;.

40. Air Quality Reduction of National Emissions The EU and International Air Pollution Policy - Accessible en ligne : <https://ec.europa.eu/environment/air/>.

41. Air Quality - Existing Legislation - Accessible en ligne : [https://ec.europa.eu/environment/air/quality/existing\\_leg.htm](https://ec.europa.eu/environment/air/quality/existing_leg.htm).

42. Projet de Loi de Finances Pour 2021 : Développement Durable : Transition Énergétique, Climat et Recherche - Accessible en ligne :

- <http://www.senat.fr/rap/a20-142-4/a20-142-46.html#:~:text=La%20France%20a%20%C3%A9t%C3%A9%20condamn%C3%A9e,de%20mortalit%C3%A9%20pr%C3%A9matur%C3%A9e%20en%20France.>
43. ISO 16000-37:2019 - Accessible en ligne : <https://www.iso.org/obp/ui/#iso:std:iso:16000:-37:ed-1:v1:en>.
  44. Ambient Air - Standard Gravimetric Measurement Method for the Determination of the PM10 or PM2,5 Mass Concentration of Suspended Particulate Matter - Accessible en ligne : <https://www.en-standard.eu/csn-en-12341-ambient-air-standard-gravimetric-measurement-method-for-the-determination-of-the-pm10-or-pm2-5-mass-concentration-of-suspended-particulate-matter/>.
  45. Gilliam, J.H.; Hall, E.S. Reference and Equivalent Methods Used to Measure National Ambient Air Quality Standards (NAAQS) Criteria Air Pollutants - Volume I. 108.
  46. Atmo Hdf, Publications et Études - Accessible en ligne : <https://www.atmo-hdf.fr/publications.html>.
  47. Airparif Bilan et Impact Du Covid En 2020 En Île-de-France - Accessible en ligne : <https://www.airparif.asso.fr/dossiers-fiches-thematiques/2021/bilan-2020-de-la-qualite-de-lair-en-ile-de-france>.
  48. Januel, E.; Dessimond, B.; Colette, A.; Annesi-Maesano, I.; Stankoff, B. Fine Particulate Matter Related to Multiple Sclerosis Relapse in Young Patients. *Front. Neurol.* **2021**, *12*, 651084, doi:10.3389/fneur.2021.651084.
  49. LANCE: NASA Near Real-Time Data and Imagery | Earthdata - Accessible en ligne : <https://earthdata.nasa.gov/earth-observation-data/near-real-time>.
  50. TEMPO Index - Accessible en ligne : <http://tempo.si.edu/index.html>.
  51. Atkinson, J. First Satellite in Global Air Quality Constellation Launches - Accessible en ligne : <http://www.nasa.gov/feature/langley/first-satellite-in-global-air-quality-constellation-launches>.
  52. Feenstra, B.; Papapostolou, V.; Hasheminassab, S.; Zhang, H.; Boghossian, B.D.; Cocker, D.; Polidori, A. Performance Evaluation of Twelve Low-Cost PM2.5 Sensors at an Ambient Air Monitoring Site. *Atmospheric Environment* **2019**, *216*, 116946, doi:10.1016/j.atmosenv.2019.116946.
  53. Liu, H.-Y.; Schneider, P.; Haugen, R.; Vogt, M. Performance Assessment of a Low-Cost PM2.5 Sensor for a near Four-Month Period in Oslo, Norway. *Atmosphere* **2019**, *10*, 41, doi:10.3390/atmos10020041.
  54. Bauerová, P.; Šindelářová, A.; Rychlík, Š.; Novák, Z.; Keder, J. Low-Cost Air Quality Sensors: One-Year Field Comparative Measurement of Different Gas Sensors and Particle Counters with Reference Monitors at Tušimice Observatory. *Atmosphere* **2020**, *11*, 492, doi:10.3390/atmos11050492.
  55. Sayahi, T.; Butterfield, A.; Kelly, K.E. Long-Term Field Evaluation of the Plantower PMS Low-Cost Particulate Matter Sensors. *Environmental Pollution* **2019**, *245*, 932–940, doi:10.1016/j.envpol.2018.11.065.
  56. Feinberg, S.; Williams, R.; Hagler, G.S.W.; Rickard, J.; Brown, R.; Garver, D.; Harshfield, G.; Stauffer, P.; Mattson, E.; Judge, R.; et al. Long-Term Evaluation of Air Sensor Technology under Ambient Conditions in Denver, Colorado. **2019**, 21.
  57. Tryner, J.; L'Orange, C.; Mehaffy, J.; Miller-Lionberg, D.; Hofstetter, J.C.; Wilson, A.; Volckens, J. Laboratory Evaluation of Low-Cost PurpleAir PM Monitors and in-Field Correction Using Co-Located Portable Filter Samplers. *Atmospheric Environment* **2020**, *220*, 117067, doi:10.1016/j.atmosenv.2019.117067.
  58. Zheng, T.; Bergin, M.H.; Johnson, K.K.; Tripathi, S.N.; Shirodkar, S.; Landis, M.S.; Sutaria, R.; Carlson, D.E. Field Evaluation of Low-Cost Particulate Matter Sensors in High- and Low-Concentration Environments. *Atmos. Meas. Tech.* **2018**, *11*, 4823–4846, doi:10.5194/amt-11-4823-2018.
  59. Moreno-Rangel, A.; Sharpe, T.; Musau, F.; McGill, G. Field Evaluation of a Low-Cost Indoor Air Quality Monitor to Quantify Exposure to Pollutants in Residential Environments. *J. Sens. Sens. Syst.* **2018**, *7*, 373–388, doi:10.5194/jsss-7-373-2018.
  60. Zikova, N.; Hopke, P.K.; Ferro, A.R. Evaluation of New Low-Cost Particle Monitors for PM2.5 Concentrations Measurements. *Journal of Aerosol Science* **2017**, *105*, 24–34, doi:10.1016/j.jaerosci.2016.11.010.
  61. Badura, M.; Batog, P.; Drzeniecka-Osiadacz, A.; Modzel, P. Evaluation of Low-Cost Sensors for Ambient PM 2.5 Monitoring. *Journal of Sensors* **2018**, *2018*, 1–16, doi:10.1155/2018/5096540.
  62. Crilley, L.R.; Shaw, M.; Pound, R.; Kramer, L.J.; Price, R.; Young, S.; Lewis, A.C.; Pope, F.D. Evaluation of a Low-Cost Optical Particle Counter (Alphasense OPC-N2) for Ambient Air Monitoring. *Atmos. Meas. Tech.* **2018**, *11*, 709–720, doi:10.5194/amt-11-709-2018.
  63. Languille, B.; Gros, V.; Bonnaire, N.; Pommier, C.; Honoré, C.; Debert, C.; Gauvin, L.; Srairi, S.; Annesi-Maesano, I.; Chaix, B.; et al. A Methodology for the Characterization of Portable Sensors for Air Quality Measure with the Goal of Deployment in Citizen Science. *Science of The Total Environment* **2020**, *708*, 134698, doi:10.1016/j.scitotenv.2019.134698.

64. Samad, A.; Melchor Mimiaga, F.E.; Laquai, B.; Vogt, U. Investigating a Low-Cost Dryer Designed for Low-Cost PM Sensors Measuring Ambient Air Quality. *Sensors* **2021**, *21*, 804, doi:10.3390/s21030804.
65. Papapostolou, V. Development of an Environmental Chamber for Evaluating the Performance of Low-Cost Air Quality Sensors under Controlled Conditions. *Atmospheric Environment* **2017**, *9*.
66. Lesser, D.; Katra, I.; Dorman, M.; Harari, H.; Kloog, I. Validating and Comparing Highly Resolved Commercial “Off the Shelf” PM Monitoring Sensors with Satellite Based Hybrid Models, for Improved Environmental Exposure Assessment. *Sensors* **2020**, *21*, 63, doi:10.3390/s21010063.
67. Hasenfratz, D.; Saukh, O.; Sturzenegger, S.; Thiele, L. Participatory Air Pollution Monitoring Using Smartphones. 5.
68. Dutta, J.; Chowdhury, C.; Roy, S.; Middya, A.I.; Gazi, F. Towards Smart City: Sensing Air Quality in City Based on Opportunistic Crowd-Sensing. In Proceedings of the Proceedings of the 18th International Conference on Distributed Computing and Networking; ACM: Hyderabad India, January 5 2017; pp. 1–6.
69. Sivaraman, V.; Carrapetta, J.; Hu, K.; Luxan, B.G. HazeWatch: A Participatory Sensor System for Monitoring Air Pollution in Sydney. In Proceedings of the 38th Annual IEEE Conference on Local Computer Networks - Workshops; IEEE: Sydney, Australia, October 2013; pp. 56–64.
70. Guo, B.; Yu, Z.; Zhou, X.; Zhang, D. From Participatory Sensing to Mobile Crowd Sensing. **2014**, 6.
71. Zeng, Y.; Xiang, K. Adaptive Sampling for Urban Air Quality through Participatory Sensing. *Sensors* **2017**, *17*, 2531, doi:10.3390/s17112531.
72. Galinina, O.; Mikhaylov, K.; Huang, K.; Andreev, S.; Koucheryavy, Y. Wirelessly Powered Urban Crowd Sensing over Wearables: Trading Energy for Data. *IEEE Wireless Commun.* **2018**, *25*, 140–149, doi:10.1109/MWC.2018.1600468.
73. Capponi, A.; Fiandrino, C.; Kantarci, B.; Foschini, L.; Kliazovich, D.; Bouvry, P. A Survey on Mobile Crowdsensing Systems: Challenges, Solutions, and Opportunities. *IEEE Commun. Surv. Tutorials* **2019**, *21*, 2419–2465, doi:10.1109/COMST.2019.2914030.
74. Maag, B.; Zhou, Z.; Thiele, L. W-Air: Enabling Personal Air Pollution Monitoring on Wearables. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2018**, *2*, 1–25, doi:10.1145/3191756.
75. Jerrett, M.; Donaire-Gonzalez, D.; Popoola, O.; Jones, R.; Cohen, R.C.; Almanza, E.; de Nazelle, A.; Mead, I.; Carrasco-Turigas, G.; Cole-Hunter, T.; et al. Validating Novel Air Pollution Sensors to Improve Exposure Estimates for Epidemiological Analyses and Citizen Science. *Environmental Research* **2017**, *158*, 286–294, doi:10.1016/j.envres.2017.04.023.
76. Piedrahita, R.; Xiang, Y.; Masson, N.; Ortega, J.; Collier, A.; Jiang, Y.; Li, K.; Dick, R.P.; Lv, Q.; Hannigan, M.; et al. The next Generation of Low-Cost Personal Air Quality Sensors for Quantitative Exposure Monitoring. *Atmos. Meas. Tech.* **2014**, *7*, 3325–3336, doi:10.5194/amt-7-3325-2014.
77. Wong, M.; Yip, T.; Mok, E. Development of a Personal Integrated Environmental Monitoring System. *Sensors* **2014**, *14*, 22065–22081, doi:10.3390/s141122065.
78. Zhuang, Y.; Lin, F.; Yoo, E.-H.; Xu, W. AirSense: A Portable Context-Sensing Device for Personal Air Quality Monitoring. In Proceedings of the Proceedings of the 2015 Workshop on Pervasive Wireless Healthcare; ACM: Hangzhou China, June 22 2015; pp. 17–22.
79. Dutta, P.; Aoki, P.M.; Kumar, N.; Mainwaring, A.; Myers, C.; Willett, W.; Woodruff, A. Common Sense: Participatory Urban Sensing Using a Network of Handheld Air Quality Monitors. In Proceedings of the Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems - SenSys '09; ACM Press: Berkeley, California, 2009; p. 349.
80. Flow, by Plume Labs: The First Smart Air Quality Tracker - Accessible en ligne : <https://plumelabs.com/en/flow/>.
81. Atmotube – a Series of Portable Air Quality Monitors - Accessible en ligne : <https://notanothereone.com/en/atmotube#>.
82. Snyder, E.G.; Watkins, T.H.; Solomon, P.A.; Thoma, E.D.; Williams, R.W.; Hagler, G.S.W.; Shelow, D.; Hindin, D.A.; Kilaru, V.J.; Preuss, P.W. The Changing Paradigm of Air Pollution Monitoring. *Environ. Sci. Technol.* **2013**, *47*, 11369–11377, doi:10.1021/es4022602.
83. Park, J.; Jo, W.; Cho, M.; Lee, J.; Lee, H.; Seo, S.; Lee, C.; Yang, W. Spatial and Temporal Exposure Assessment to PM<sub>2.5</sub> in a Community Using Sensor-Based Air Monitoring Instruments and Dynamic Population Distributions. *Atmosphere* **2020**, *11*, 1284, doi:10.3390/atmos11121284.
84. Schneider, P.; Castell, N.; Vogt, M.; Dauge, F.R.; Lahoz, W.A.; Bartonova, A. Mapping Urban Air Quality in near Real-Time Using Observations from Low-Cost Sensors and Model Information. *Environment International* **2017**, *106*, 234–247, doi:10.1016/j.envint.2017.05.005.
85. Mueller, M.; Meyer, J.; Hueglin, C. Design of an Ozone and Nitrogen Dioxide Sensor Unit and Its Long-Term Operation within a Sensor Network in the City of Zurich. *Atmos. Meas. Tech.* **2017**, *10*, 3783–3799, doi:10.5194/amt-10-3783-2017.

86. Yuval; Magen Molho, H.; Zivan, O.; Broday, D.M.; Raz, R. Application of a Sensor Network of Low Cost Optical Particle Counters for Assessing the Impact of Quarry Emissions on Its Vicinity. *Atmospheric Environment* **2019**, *211*, 29–37, doi:10.1016/j.atmosenv.2019.04.054.
87. Tse, R.; Monti, L.; Prandi, C.; Aguiari, D.; Pau, G.; Salomoni, P. On Assessing the Accuracy of Air Pollution Models Exploiting a Strategic Sensors Deployment. In Proceedings of the Proceedings of the 4th EAI International Conference on Smart Objects and Technologies for Social Good - Goodtechs '18; ACM Press: Bologna, Italy, 2018; pp. 55–58.
88. Kumar, P.; Morawska, L.; Martani, C.; Biskos, G.; Neophytou, M.; Di Sabatino, S.; Bell, M.; Norford, L.; Britter, R. The Rise of Low-Cost Sensing for Managing Air Pollution in Cities. *Environ Int* **2015**, *75*, 199–205, doi:10.1016/j.envint.2014.11.019.
89. Système National Des Données de Santé - Accessible en ligne : <https://www.snds.gouv.fr/>.
90. Berke, E.M.; Choudhury, T.; Ali, S.; Rabbi, M. Objective Measurement of Sociability and Activity: Mobile Sensing in the Community. *The Annals of Family Medicine* **2011**, *9*, 344–350, doi:10.1370/afm.1266.
91. Allen, R.W.; Barn, P. Individual- and Household-Level Interventions to Reduce Air Pollution Exposures and Health Risks: A Review of the Recent Literature. *Curr Envir Health Rpt* **2020**, *7*, 424–440, doi:10.1007/s40572-020-00296-z.
92. Chaix, B.; Bean, K.; Leal, C.; Thomas, F.; Havard, S.; Evans, D.; Jégo, B.; Pannier, B. Individual/Neighborhood Social Factors and Blood Pressure in the RECORD Cohort Study: Which Risk Factors Explain the Associations? *Hypertension* **2010**, *55*, 769–775, doi:10.1161/HYPERTENSIONAHA.109.143206.
93. RGPD - Accessible en ligne : <https://www.cnil.fr/fr/reglement-europeen-protection-donnees>.
94. Reback, J.; Jbrockmendel; McKinney, W.; Van Den Bossche, J.; Augspurger, T.; Cloud, P.; Hawkins, S.; Gfyoung; Roeschke, M.; Sinhrks; et al. *Pandas-Dev/Pandas: Pandas 1.3.3*; Zenodo, 2021;
95. Hunter, J.D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95, doi:10.1109/MCSE.2007.55.
96. Waskom, M.; Botvinnik, O.; O’Kane, D.; Hobson, P.; Lukauskas, S.; Gemperline, D.C.; Augspurger, T.; Halchenko, Y.; Cole, J.B.; Warmenhoven, J.; et al. *Mwaskom/Seaborn: V0.8.1 (September 2017)*; Zenodo, 2017;
97. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.
98. STIC-ASIA: SEA HAZEMON - Accessible en ligne : <https://interlab.ait.ac.th/HAZEMON/>.
99. Aguiari, D.; Delnevo, G.; Monti, L.; Ghini, V.; Mirri, S.; Salomoni, P.; Pau, G.; Im, M.; Tse, R.; Ekpanyapong, M.; et al. Canarin II: Designing a Smart e-Bike Eco-System. In Proceedings of the 2018 15th IEEE Annual Consumer Communications & Networking Conference (CCNC); IEEE: Las Vegas, NV, January 2018; pp. 1–6.
100. Tse, R.; Pau, G. Enabling Street-Level Pollution and Exposure Measures: A Human-Centric Approach. In Proceedings of the Proceedings of the 6th ACM International Workshop on Pervasive Wireless Healthcare - MobiHealth '16; ACM Press: Paderborn, Germany, 2016; pp. 1–4.
101. Bédard, A.; Sofiev, M.; Arnavielhe, S.; Antó, J.M.; Garcia-Aymerich, J.; Thibaudon, M.; Bergmann, K.C.; Dubakiene, R.; Bedbrook, A.; Onorato, G.L.; et al. Interactions Between Air Pollution and Pollen Season for Rhinitis Using Mobile Technology: A MASK-POLLAR Study. *J Allergy Clin Immunol Pract* **2020**, *8*, 1063-1073.e4, doi:10.1016/j.jaip.2019.11.022.
102. Sensing the Air Quality: Research on Air Quality Sensors - Accessible en ligne : <https://aqicn.org/sensor/>.
103. Introduction - UDOO Neo Docs - Accessible en ligne : <https://www.udoo.org/docs-neo/Introduction/Introduction.html>.
104. Global IoT SIM Card & M2M Connectivity | Things Mobile - Accessible en ligne : <https://www.thingsmobile.com/business>.
105. Instances T2 Amazon EC2 – Amazon Web Services (AWS) - Accessible en ligne : <https://aws.amazon.com/fr/ec2/instance-types/t2/>.
106. What Is a Container? | App Containerization | Docker - Accessible en ligne : <https://www.docker.com/resources/what-container>.
107. Canarin.Net - Accessible en ligne : <https://canarin.net/>.
108. Participative Observatory for the Surveillance of Individual Exposure to Air Pollution Relating to Health - POLLUSCOPE - Accessible en ligne : <https://anr.fr/Project-ANR-15-CE22-0018>.
109. Accueil - Cohorte RECORD - Accessible en ligne : <https://www.hal.inserm.fr/RECORD>.

110. Micro-Sensors | Ambient - Accessible en ligne : <https://www.envea.global/solutions/ambient-monitoring/micro-sensors/>.
111. Cairsens® NO<sub>2</sub> | Micro-Sensors | Ambient | ENVEA - Accessible en ligne : <https://www.envea.global/s/ambient-en/micro-sensors-a/cairsens-no2/>.
112. MIR France - Agent officiel MIR pour la France et l’Afrique Francophone - Accessible en ligne : <http://www.mirfrance.com/>.
113. ARCHOS Oxygen 101 4G, Tablets - Overview - Accessible en ligne : [https://www.archos.com/us/products/tablets/oxygen/archos\\_oxygen1014g/index.html](https://www.archos.com/us/products/tablets/oxygen/archos_oxygen1014g/index.html).
114. Kluyver, T.; Ragan-Kelley, B.; Pérez, F.; Granger, B.; Bussonnier, M.; Frederic, J.; Kelley, K.; Hamrick, J.; Grout, J.; Corlay, S.; et al. Jupyter Notebooks – a Publishing Format for Reproducible Computational Workflows. In Proceedings of the Positioning and power in academic publishing: Players, agents and agendas; Loizides, F., Schmidt, B., Eds.; 2016; pp. 87–90.
115. Hafyani, H.E. Big Data Series Analytics in the Context of Environmental Crowd Sensing. In Proceedings of the 2020 21st IEEE International Conference on Mobile Data Management (MDM); IEEE: Versailles, France, June 2020; pp. 246–247.
116. Grafana: The Open Observability Platform - Accessible en ligne : <https://grafana.com/>.
117. Dessimond, B.; Annesi-Maesano, I.; Pepin, J.-L.; Srairi, S.; Pau, G. Academically Produced Air Pollution Sensors for Personal Exposure Assessment: The Canarin Project. *Sensors* **2021**, *21*, 1876, doi:10.3390/s21051876.
118. Optical Particle Counters (Centre for Atmospheric Science - The University of Manchester) - Accessible en ligne : <http://www.cas.manchester.ac.uk/restools/instruments/aerosol/opc/>.
119. He, M.; Kuerbanjiang, N.; Dhaniyala, S. Performance Characteristics of the Low-Cost Plantower PMS Optical Sensor. *Aerosol Science and Technology* **2020**, *54*, 232–241, doi:10.1080/02786826.2019.1696015.
120. Levy Zamora, M.; Xiong, F.; Gentner, D.; Kerkez, B.; Kohrman-Glaser, J.; Koehler, K. Field and Laboratory Evaluations of the Low-Cost Plantower Particulate Matter Sensor. *Environ. Sci. Technol.* **2019**, *53*, 838–849, doi:10.1021/acs.est.8b05174.
121. Cheng, Y.-H.; Lin, M.-H. Real-Time Performance of the MicroAeth® AE51 and the Effects of Aerosol Loading on Its Measurement Results at a Traffic Site. *Aerosol Air Qual. Res.* **2013**, *13*, 1853–1863, doi:10.4209/aaqr.2012.12.0371.
122. Lee, J. Performance Test of MicroAeth® AE51 at Concentrations Lower than 2 Mg/M3 in Indoor Laboratory. *Applied Sciences* **2019**, *9*, 2766, doi:10.3390/app9132766.
123. Dons, E.; Int Panis, L.; Van Poppel, M.; Theunis, J.; Wets, G. Personal Exposure to Black Carbon in Transport Microenvironments. *Atmospheric Environment* **2012**, *55*, 392–398, doi:10.1016/j.atmosenv.2012.03.020.
124. Virkkula, A.; Mäkelä, T.; Hillamo, R.; Yli-Tuomi, T.; Hirsikko, A.; Hämeri, K.; Koponen, I.K. A Simple Procedure for Correcting Loading Effects of Aethalometer Data. *Journal of the Air & Waste Management Association* **2007**, *57*, 1214–1222, doi:10.3155/1047-3289.57.10.1214.
125. Holmes, M.; Edwards, C.; Vick, B.; Costello, E. Longitudinal Trends in Air Quality and Home-Recorded Spirometry in Pulmonary Fibrosis Patients in the Dallas-Fort Worth Area. In Proceedings of the A52. SARCOIDOSIS AND EXPOSURE RELATED INTERSTITIAL LUNG DISEASE; American Thoracic Society, May 2019; pp. A1865–A1865.
126. Wijbenga, N.; Hoek, R.A.S.; Mathot, B.J.; Seghers, L.; van Weezel, J.J.; den Ouden, J.; Wijnsbeek, M.S.; Aerts, J.G.J.V.; Hellemons, M.E.; Moor, C.C. Evaluation of a Home Monitoring Application for Follow Up after Lung Transplantation—A Pilot Study. *JPM* **2020**, *10*, 240, doi:10.3390/jpm10040240.
127. Degryse, J.; Buffels, J.; Van Dijck, Y.; Decramer, M.; Nemery, B. Accuracy of Office Spirometry Performed by Trained Primary-Care Physicians Using the MIR Spirobank Hand-Held Spirometer. *Respiration* **2012**, *83*, 543–552, doi:10.1159/000334907.
128. Dictionnaire Médical de l’Académie de Médecine – Version 2021 - Accessible en ligne : <https://www.academie-medecine.fr/le-dictionnaire/index.php?q=Tiffeneau%20%28%C3%A9preuve%20de%29>.
129. Medical International Research (MIR) - Accessible en ligne : <https://www.spirometry.com/>.
130. Seabold, S.; Perktold, J. Statsmodels: Econometric and Statistical Modeling with Python. In Proceedings of the 9th python in science conference; 2010.
131. Spyder 5 - Accessible en ligne : <https://www.spyder-ide.org/>.
132. Qt - Accessible en ligne : <https://www.qt.io/>.
133. Accueil — Le Réseau National de Surveillance Aérobiologique — RNSA - Accessible en ligne : <https://www.pollens.fr/>.

134. Ansys Academic | Simulation Software for Educators, Researchers and Students - Accessible en ligne : <https://www.ansys.com/academic>.
135. IP Ratings | IEC - Accessible en ligne : <https://www.iec.ch/ip-ratings>.
136. Spécifications Noctua NF-P14s Redux-1500 PWM - Accessible en ligne : <https://noctua.at/en/nf-p14s-redux-1500-pwm/specification>.
137. Current Ranger | LowPowerLab - Accessible en ligne : <https://lowpowerlab.com/guide/currentranger/>.
138. Fusion 360 for Students | What Will You Design? - Accessible en ligne : <https://www.autodesk.com/campaigns/education/student-design>.
139. Animation : BANC D'ÉSSAI - Vue En Éclaté - Accessible en ligne : <https://youtu.be/0Rdb-XGBKHE>.
140. Animation : DIFFUSEUR - Visualisation de La Convergence - Accessible en ligne : <https://youtu.be/fNiCWtj0SLM>.
141. Animation : BANC D'ÉSSAI - Visualisation de La Convergence - Accessible en ligne : <https://youtu.be/hPFWy4yLSyk>.
142. Paris, France Weather Conditions | Weather Underground - Accessible en ligne : <https://www.wunderground.com/weather/fr/paris/IPARIS18204>.
143. SBDR Algorithm GitHub - Accessible en ligne : <https://github.com/BorisDess/SBDR>.
144. Aguiari, D.; Tse, R.; Giusto, D.; Pau, G. Monitoring Cultural Heritage Buildings via Low-Cost Edge Computing/Sensing Platforms: The Biblioteca Joanina de Coimbra Case Study. 5.
145. Apple M1 PassMark - Accessible en ligne : <https://www.cpubenchmark.net/cpu.php?cpu=Apple+M1+8+Core+3200+MHz&id=4104>.
146. Teensy 4.1 - Accessible en ligne : <https://www.pjrc.com/store/teensy41.html>.
147. STM32H7 - Arm Cortex-M7 and Cortex-M4 MCUs (480 MHz) - STMicroelectronics - Accessible en ligne : <https://www.st.com/en/microcontrollers-microprocessors/stm32h7-series.html>.
148. Benchmark Product List - EEMBC - Embedded Microprocessor Benchmark Consortium - Accessible en ligne : <https://www.eembc.org/products/>.
149. Portenta H7 - Arduino Boards & Modules - Arduino - Accessible en ligne : <https://store.arduino.cc/portenta-h7>.
150. MIPI Display Serial Interface (MIPI DSI) - Accessible en ligne : <https://mipi.org/specifications/dsi>.
151. LLC, L. LVGL - Light and Versatile Embedded Graphics Library - Accessible en ligne : <https://lvgl.io/>.
152. NRF24 Series - Nordic Semiconductor - Accessible en ligne : <https://www.nordicsemi.com/Products/Low-power-short-range-wireless/nRF24-series>.
153. Morawska, L.; Thai, P.K.; Liu, X.; Asumadu-Sakyi, A.; Ayoko, G.; Bartonova, A.; Bedini, A.; Chai, F.; Christensen, B.; Dunbabin, M.; et al. Applications of Low-Cost Sensing Technologies for Air Quality Monitoring and Exposure Assessment: How Far Have They Gone? *Environment International* **2018**, *116*, 286–299, doi:10.1016/j.envint.2018.04.018.
154. EEVblog UCurrent - Precision NA Current Measurement Assistant [v3] ID: 882 - \$59.95: Adafruit Industries, Unique & Fun DIY Electronics and Kits - Accessible en ligne : <https://www.adafruit.com/product/882>.
155. Song, Y.; Tan, H. Practical Pairing-Free Sensor Cooperation Scheme for Cloud-Assisted Wireless Body Area Networks. *Cybersecur* **2020**, *3*, 21, doi:10.1186/s42400-020-00061-7.
156. Bauerová, P.; Šindelářová, A.; Rychlík, Š.; Novák, Z.; Keder, J. Low-Cost Air Quality Sensors: One-Year Field Comparative Measurement of Different Gas Sensors and Particle Counters with Reference Monitors at Tušimice Observatory. *Atmosphere* **2020**, *11*, 492, doi:10.3390/atmos11050492.