THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National Polytechnique de Toulouse (Toulouse INP)

Discipline ou spécialité :

Génie industriel

Présentée et soutenue par :

M. SYLVAIN POUPRY

le mercredi 13 décembre 2023

Titre :

Contribution à la conception et à la mise en œuvre d'un système de surveillance de la qualité de l'air : application à la surveillance de la qualité de l'air dans les vallées des Gaves

---

École doctorale :

Systèmes (EDSYS)

Unité de recherche :

Laboratoire Génie de Production de l'ENIT (E.N.I.T-L.G.P.)

Directeurs de Thèse :

M. KAMAL MEDJAHER

M. CÉDRICK BELER

Rapporteurs :

M. BRUNO CASTANIER, UNIVERSITE D'ANGERS

M. GREGORY ZACHAREWICZ, ENS Mines- IMT Alès

Membres du jury :

MME ZINEB SIMEU-ABAZI, UNIVERSITE GRENOBLE ALPES, Présidente

M. CÉDRICK BELER, ECOLE NATIONALE D'INGENIEURS DE TARBES, Membre

M. ERIC DUVIELLA, IMT LILLE DOUAI, Membre

MI. KAMAL MEDJAHER, ECOLE NATIONALE D'INGENIEURS DE TARBES, Membre

Contribution à la conception et à la mise en œuvre d'un système de surveillance de la qualité de l'air :

Application à la surveillance de la qualité de l'air dans les vallées des Gaves

Sylvain POUPRY

Résumé du manuscrit

Cette thèse vise à mettre en place une démarche scientifique de surveillance et d'action de prévention de la qualité de l'air dans des zones dépourvues de stations de mesure conventionnelles. En effet, en raison d'un budget limité, des communes modestes ne disposent pas d'un moyen fiable et continu de collecter des mesures pour avoir un suivi de la qualité de l'air et de mener des actions sur le territoire en conséquence. Pour relever ce défi, cette thèse propose une approche innovante en utilisant la redondance modulaire à l'aide de micro-capteurs, également appelés Low Cost Sensors (LCS). L'approche proposée consiste en une architecture permettant l'auto-surveillance et l'auto-reconfiguration de ses composants. Ces caractéristiques permettent de générer des données fiables pour surveiller, prédire et agir sur la qualité de l'air au sein d'un territoire non couvert par des stations conventionnelles.

La première phase de cette thèse débute sur les dimensions spatiale et temporelle de l'architecture afin de couvrir l'ensemble du processus d'acquisition et de traitement des données. Par la suite, le travail continue avec le développement de stations de mesure composées de LCS capables de collecter des données sur la qualité de l'air. Cependant, les LCS présentent certaines limitations, notamment au niveau du matériel et de la fiabilité des données. Une architecture basée sur la redondance est proposée pour augmenter la disponibilité du matériel et la fiabilisation des données collectées. La fiabilisation des données consiste à avoir des résultats précis et cohérents sur une longue période, malgré les perturbations ou les dysfonctionnements éventuels des composants.

Ensuite, une méthode de co-design itératif est mise en œuvre pour l'élaboration d'une station de mesure. À chaque déploiement sur le terrain, le dispositif est amélioré pour permettre non seulement une collecte de données sur un horizon temporel étendu, mais aussi un suivi de l'état de santé intrinsèque de la station. Des fonctionnalités pour la détection et le diagnostic des composants défaillants sont également intégrées, ainsi qu'un indice de fiabilité associé aux observations.

En conclusion, bien que prometteuse, la méthode nécessite des améliorations, notamment en matière de prédiction des défaillances des capteurs, d'évolutivité et de cybersécurité. L'exploration d'algorithmes d'apprentissage automatique, tels que les auto encodeurs, pourrait contribuer à renforcer la prédiction des défaillances.

Mots clés : Internet des objets, micro-capteur, qualité de l'air, station de mesure, redondance active, capteur intelligent, détection, diagnostic, fiabilité.

Publications de l'auteur

Revues internationales :

— Poupry, S., Medjaher, K., & Béler, C. (2023). Data reliability and fault diagnostic for air quality monitoring station based on low cost sensors and active redundancy. In Measurement (Vol. 223, p. 113800). issn : 0263-2241.
https ://doi.org/10.1016/j.measurement.2023.113800 [55]

Conférences internationales avec actes et comité de lecture :

— Poupry, S., Béler, C., & Medjaher, K. (2022). Development of a reliable measurement station for air quality monitoring based on low-cost sensors and active redundancy. 2nd IFAC Workshop on Integrated Assessment Modelling for Environmental Systems IAMES 2022. Tarbes, France, 1-3 June 2022. IFAC-PapersOnLine, 55(5), 7-12.
https ://doi.org/10.1016/j.ifacol.2022.07.631 [52]

— Poupry, S., Béler, C., & Medjaher, K. (2022, June). Towards data reliability based on triple redundancy and online outlier detection. In PHM Society European Conference, Turin, Italy, 6-8 July 2022. Vol. 7, No. 1, pp. 410-420.
https ://doi.org/10.36001/phme.2022.v7i1.3366 [53]

— Poupry, S., Medjaher, K., & Béler, C. (2023, September). Automatic detection of hardware failures in an air quality measuring station with low cost sensors. In PHM Society Asia-Pacific Conference (Vol. 4, No. 1).
https ://doi.org/10.36001/phmap.2023.v4i1.3706 [54]

Remerciements

Ma décision de commencer et de poursuivre cette thèse a été alimentée par la curiosité et l'opportunité d'explorer de nouveaux domaines. À l'origine, la recherche ne figurait pas dans mes plans, mais l'attrait de l'inconnu m'a conduit sur ce chemin. Mes découvertes dans la construction d'un réseau de mesure complet, depuis le choix du capteur jusqu'à la programmation d'une interface web, ont été une agréable surprise, révélant des possibilités insoupçonnées. Les défis de cette thèse résidaient dans la montée en compétences dans plusieurs domaines simultanément. Que ce soit la qualité de l'air, l'administration réseau ou la programmation dans différents langages, chaque étape nécessitait un apprentissage intensif. La persévérance, la discipline, la curiosité et l'écoute attentive des experts dans leurs domaines respectifs ont été essentielles pour surmonter ces obstacles. La leçon fondamentale que j'en retire est la valeur de la détermination et de l'engagement. L'impact de mes travaux pourrait se manifester dans la possibilité de créer un réseau de mesure accessible avec un budget modeste. Cette réalisation a le potentiel d'intéresser les particuliers soucieux de la qualité de l'air. Au-delà de la thèse, je vois cela comme le commencement d'un projet plus vaste avec de nombreux défis et opportunités passionnants à venir. Cette thèse a été une expérience enrichissante qui m'a permis de travailler sur un projet d'envergure avec une collectivité locale. Elle a nourri ma fierté et m'a ouvert à des aspects tels que la sensibilisation auprès des citoyens et des élus. Au-delà des résultats scientifiques, c'est une aventure humaine et professionnelle qui restera gravée dans ma mémoire.

Les travaux présentés dans cette thèse ont été réalisés au sein du Laboratoire Génie de Production de Tarbes. J'ai été très bien accueilli au sein du groupe de recherche PRiSM et j'en remercie tous les membres qui ont toujours su se rendre disponibles pour m'éclairer sur tel ou tel sujet. J'adresse en particulier des remerciements à Kamal MEDJAHER, mon directeur de thèse, pour son soutien indéfectible, sa confiance en mon travail et son implication personnelle lors des phases de rush. Son expertise et son sens du détail, tant sur le fond que sur la forme, m'ont montré la rigueur scientifique à atteindre. Je retiendrai également les enseignements issus de son expérience au niveau des interactions avec des chercheurs et, en particulier, la manière de s'adresser aux autres. Je remercie Cédrick BELER pour m'avoir donné envie de faire de la recherche et pour avoir transmis sa passion du code (Smalltalk) et des nouvelles technologies. Nous avons beaucoup échangé et débattu. Leur soutien et leur collaboration ont été inestimables pour la réussite de ce projet.

Je tiens à exprimer ma gratitude envers ceux qui ont rendu ce travail possible : Walter NICO, dont les compétences et la disponibilité ont été essentielles dans la création de l'architecture réseau ; Pierre MASSONNAT, qui m'a soutenu et aidé à comprendre la « pensée scientifique ». Je remercie également Raymond HOUE NGOUNA et Xavier DESFORGES, qui m'ont écouté dans mes problématiques et n'ont pas hésité à échanger leurs informations. J'ai été encouragé à présenter mes recherches pour apprendre à être un bon orateur, tant en français qu'en anglais. Je me suis bien senti intégré, les échanges étaient d'égal à égal et je les en remercie pour tout cela.

Je remercie Philippe CLERMONT, qui m'a suivi depuis le début de ma reconversion professionnelle lors de mon entrée en formation continue. Je collabore aujourd'hui avec lui pour la formation en créativité. Une petite pensée pour Pascal CLAVERIE, mon ancien patron, qui sans son accord et son soutien, mon voyage intellectuel n'aurait jamais commencé.

Je veux te remercier, à toi, Élodie : tu as cru en moi. Ta foi inébranlable en mes capacités m'a donné la force et le courage d'arriver à bout de ces études. Dans le meilleur comme dans le pire, tu as été à mes côtés et tu as su me rassurer dans mes moments de doute. Pour tout cela, je te remercie de tout mon cœur, ma femme ! Et sans oublier Élaïa, mon petit bébé arrivé pendant la rédaction de la thèse, qui avec son petit sourire a déjà su me faire oublier les moments difficiles.

En guise de point final, un petit mot de sagesse de ma petite Élaïa : Chaque nuit, nous espérions une nuit complète de sa part, mais elle nous a vite montré que l'espoir ne remplace pas une bonne nuit de sommeil !

«L'espoir c'est pour les faibles»
— Élaïa, 24/08/2022

Table des matières

1 Introduction générale **15**
1.1 Contexte . . . 16
1.1.1 Inventaire des émissions . . . 18
1.1.2 Sources de pollution du secteur résidentiel . . . 19
1.1.3 Sources de pollution des activités agricoles . . . 20
1.1.4 Sources de pollution du secteur transport . . . 21
1.1.5 Sources de pollution du secteur industrie . . . 21
1.1.6 Les attentes de la CCPVG . . . 22
1.2 Problématiques . . . 24
1.3 Questions de recherche . . . 24
1.3.1 Acquisition de mesures fiables de la qualité de l'air . . . 24
1.3.2 Détection des défaillances et prédiction de l'état de santé des stations de mesure . . . 25
1.3.3 La couverture en plusieurs points du territoire à partir de mesures de stations . . . 25
1.3.4 La validation des acquisitions et la modélisation de la pollution de l'air . . . 25
1.4 Objectifs de recherche . . . 26
1.5 Contribution par rapport à l'état de l'art . . . 26
1.6 Conclusion du chapitre introduction . . . 27
1.7 Structure de la thèse . . . 28

2 État de l'art **30**
2.1 Introduction . . . 31
2.2 La surveillance de la Qualité de l'air . . . 31
2.2.1 Le LCSQA . . . 32
2.2.2 Les Normes de la qualité de l'air . . . 33
2.2.3 Modèlisation de la qualité de l'air : l'approche descendante et assimilation de données . . . 34
2.2.4 Les AASQA . . . 36
2.2.5 Le réseau de surveillance d'ATMO Occitanie . . . 36
2.2.6 Représentativité des stations de mesures . . . 40
2.2.7 Les Technologies de surveillance de la qualité de l'air . . . 42
2.3 État de l'art sur les micro-capteurs ou LCS . . . 45
2.3.1 Définition . . . 45
2.3.2 Les capteurs de particules . . . 46
2.3.3 Les capteurs de gaz . . . 47
2.3.4 Surveillance de la qualité de l'air avec des micro-capteurs . . . 50

2.3.5 Conclusion . . . 53
2.4 État de l'art sur l'acquisition, le traitement et la fiabilisation des données . 54
2.4.1 Acquisition des données afin d'alimenter un modèle pour faire des prédictions : l'exemple de la météorologie . . . 54
2.4.2 Acquisition des observations : augmenter le maillage . . . 55
2.4.3 La fiabilisation des données en un point d'observation . . . 56
2.4.4 La triple redondance modulaire et les algorithmes de vote . . . 57
2.4.5 Les algorithmes de détection des défaillances . . . 58
2.5 Conclusion du chapitre . . . 59

3 Proposition d'une architecture de monitoring territorial : une approche ascendante 60
3.1 Introduction . . . 61
3.2 Présentation générale de l'architecture . . . 61
3.3 Les dimensions spatiale et temporelle de l'architecture proposée . . . 62
3.3.1 Estimer un paramètre physique en un point précis . . . 65
3.3.2 Méthodes d'interpolations spatiales . . . 65
3.3.3 De la mesure locale à l'impact régional : l'approche ascendante . . . 73
3.3.4 Le réseau de mesure dans la dimension spatiale . . . 74
3.4 Observation et monitoring en un point de prélèvement . . . 76
3.4.1 Exploration des méthodes d'interpolation dans le cadre d'un réseau temporel de surveillance de la qualité de l'air . . . 77
3.4.2 Utilisation de la fenêtre roulante pour l'amélioration de la qualité des données et la réduction du bruit . . . 81
3.5 Conception d'une station de mesure en un point de prélèvement . . . 82
3.6 Conclusion du chapitre . . . 84

4 Données, Traitements et Décisions 85
4.1 Introduction . . . 85
4.2 Données de surveillance de la qualité de l'air . . . 86
4.3 Fiabilisation des données issues des stations de mesures . . . 90
4.3.1 Description du processus d'acquisition et de traitement des données depuis les SmS jusqu'au serveur local . . . 91
4.3.2 Formalisation mathématique de la fiabilisation des données . . . 93
4.4 Conclusion du chapitre Données, Traitements et Décision . . . 102

5 Mise en oeuvre et déploiement de l'architecture 103
5.1 Introduction . . . 104
5.1.1 La phase stratégique . . . 105
5.1.2 La phase de déploiement . . . 106
5.2 Le développement en co-design de la station . . . 107
5.2.1 La réalisation et le choix des technologies . . . 108
5.2.2 Le choix des LCS . . . 108
5.2.3 La conception de la carte électronique (PCB) . . . 109
5.2.4 Intégration du SmS : Assemblage, Boîtier et Connexions . . . 110
5.2.5 Le fonctionnement du capteur intelligent . . . 111
5.2.6 Interface graphique et aide à la décision . . . 112
5.3 Améliorations itératives des prototypes : les enseignements tirés du codesign 113
5.3.1 Premier prototype . . . 113

5.3.2 Fiabilisation des mesures : l'influence des hyperparamètres 118
5.3.3 Deuxième prototype avec fiabilisation des données 120
5.4 Conclusion du chapitre 126

6 Conclusion générale **128**
6.1 Résumé des chapitres 128
6.2 Rappel des objectifs principaux de la thèse 130
6.3 Contributions de la thèse 130
6.4 Implications et applications 131
6.5 Perspectives de recherche 132

A Annexes **137**
A.1 La pollution atmosphérique 137
A.1.1 Origine des polluants 137
A.1.2 Cycle de vie des émissions 139
A.1.3 La dégradation des polluants 140
A.1.4 La dispersion des polluants 140
A.1.5 Les phénomènes météorologiques particuliers 142
A.1.6 Les principaux polluants surveillés 144
A.1.7 Conclusion sur la pollution atmosphérique 145
A.2 Le PHM dans l'industrie 146
A.3 Vue d'ensemble de l'architecture avec les logiciels utilisés 147
A.4 Modes de fonctionnement du capteur intelligent 148
A.5 Détails techniques du SmS 150
A.5.1 Le PCB en détail 150
A.5.2 Les schémas électriques détaillés pour chaque composants 152
A.5.3 Les réalisations : du procept au prototype final 156

Table des figures

1.1 Situation géographique et relief de la CCPVG (geoportail.fr). 16
1.2 TMJA des routes de la CCPVG (ccpvg.fr). 21
1.3 Situation géographique de la CCPVG par rapport aux stations de mesures conventionnelles. 23
1.4 Structure et résumé des questions de recherche abordées dans ce manuscrit de thèse. 29

2.1 Schéma de fonctionnement du dispositif national français [39]. 32
2.2 Extrait du tableau des normes sur la qualité de l'air [60]. 33
2.3 Extrait du tableau des normes sur la qualité de l'air [35]. 34
2.4 Calcul de l'indice à partir d'un modèle méso-échelle [35]. 35
2.5 Le cadre réglementaire de la surveillance de la qualité de l'air [3]. 37
2.6 Comment sont modélisées les cartes de la Qualité de l'Air [39]. 39
2.7 Photographies d'une station fixe et d'une station mobile du réseau ATMO [36]. 40
2.8 Chaîne de mesure de la qualité de l'air pour les AASQA [40]. 43
2.9 Cartographie annuelle de la pollution en PM_{10} avec et sans micro-capteurs (www.atmosud.org). 44
2.10 Projet citoyen Mobicit'air (ATMO-auvergnerhonealpes.fr). 44
2.11 Principe de fonctionnement d'un capteur. 45
2.12 Illustration des critères fidélité, justesse et exactitude. 46
2.13 Fonctionnement du capteur de poussière SDS011 [18]. 46
2.14 Fonctionnement d'un capteur électrochimique [16]. 47
2.15 Principe de fonctionnement d'un capteur MO_x[76]. 48
2.16 Principe de fonctionnement capteur IR [8]. 48
2.17 Principe de fonctionnement capteur PID [50]. 49

3.1 Vue d'ensemble de l'architecture proposée. 62
3.2 La dimension spatiale de l'architecture. 63
3.3 La dimension temporelle de l'architecture. 64
3.4 Illustration des données dans l'espace bidimensionnel à un instant t. 67
3.5 Représentation des méthodes d'interpolation linéaire simple, cubique, Nearest et RBF 68
3.6 Représentation des méthodes d'interpolation linéaire simple et cubique ajustées, Nearest et RBF 70
3.7 Les opportunités de l'approche ascendante dans l'approche descendante. Illustration inspirée du modèle méso-échelle d'ATMO [35]. 71
3.8 Représentation de l'approche ascendante avec trois stations de mesure . . . 73
3.9 Le réseau de mesure dans la dimension spatiale du territoire. 74

3.10 Cartes de concentrations issues des points de prélèvement. 75
3.11 La dimension temporelle multi-paramètre des stations de l'architecture. . . 77
3.12 Interpolation de degré 1 sur des données brutes de PM10 78
3.13 Interpolations de degré 2,4,8 et 16 sur des données brutes de PM10 79
3.14 Interpolation de degré 100 sur des données brutes de PM10 80
3.15 Une fenêtre roulante pour un sous-échantillonnage des observations et des détections dans les données. 82
3.16 Station et observation des paramètres physiques. 83
3.17 La conception modulaire redondante de la station de mesure. 84

4.1 Le processus d'acquisition et de traitement des données de la surveillance de la qualité de l'air en détail. 87
4.2 De l'erreur relative fournie par le fabricant jusqu'aux écarts autorisés. . . . 90
4.3 Focus sur la fiabilisation des données issues des stations de mesures. 91
4.4 Le processus d'acquisition et de traitement des données de l'architecture de surveillance. 92
4.5 Construction de la matrice temporelle d'observation du SmS numéro x. . . 94
4.6 Extraction de la sous-matrice du SmS numéro x au moment k pendant τ. . 95
4.7 Vérification des sous-matrices pour la détection. 97
4.8 Concaténation des $\varphi^{\tau}_{i,x}$ provenant des SmS de la station. 98

5.1 Localisation de la CCPVG et des stations conventionnelles les plus proches. 104
5.2 Une station de mesure avec trois SmS assemblés et le boîtier d'alimentation. 106
5.3 Protocole de mesure du SmS. 111
5.4 Outil d'aide à la décision de la CCPVG pour la surveillance de la qualité de l'air. 113
5.5 Station déployée à titre de preuve de concept. 114
5.6 Données brutes de PM_{10} pendant la phase de déploiement. 115
5.7 Données brutes de $PM_{2.5}$ lors d'une journée typique à Argelès-Gazost. . . . 116
5.8 Données brutes de TVOC. 117
5.9 Données brutes de SO_2. 117
5.10 Données brutes provenant du MISC6814. 118
5.11 Résultat de la méthode en utilisant la détection des valeurs aberrantes avec $\tau = 10\ min$ et $R\varepsilon_2 = 30\%$. 119
5.12 Résultat de la méthode en utilisant la détection des valeurs aberrantes avec $\tau = 20\ min$ et $R\varepsilon_2 = 30\%$. 120
5.13 Résultat de la méthode en utilisant la détection des valeurs aberrantes avec $\tau = 10\ min$ et $R\varepsilon_2 = 35\%$. 121
5.14 Résultat de la méthode en utilisant la détection des valeurs aberrantes avec $\tau = 20\ min$ et $R\varepsilon_2 = 35\%$. 121
5.15 Visualisation du $S.I.D$ pour le paramètre Humidité lors de multiples déconnexions du SmS 3 sur une période d'une demi-journée. 123
5.16 Visualisation du $S.I.D$ pour le paramètre PM_{10} lors de multiples déconnexions du SmS 3 et le blocage à la valeur zéro du LCS du SmS 2. 123
5.17 Visualisation du $S.I.D$ pour le paramètre PM_{10} lors de multiples défaillances matérielles du LCS du SmS numéro 2 et des pics de pollution sur trois jours. 125

6.1 Fonctionnement d'un autoencodeur [70] 133
6.2 Mesures thermiques du PCB pour modélisation des flux thermiques. 134

6.3 Simulation des flux thermiques dans le boîtier initial modélisé en 2D. . . . 135
6.4 Prototype de boîtier imprimé en 3D. 135

A.1 Gaz constituant l'atmosphère [15]. 138
A.2 Cycle de vie des émissions [1]. 139
A.3 Variabilité spatiale et temporelle des constituants clés de l'atmosphère [13]. 140
A.4 Profil vertical de température et principales couches de l'atmosphère [66]. . 141
A.5 Structure de la troposphère [66]. 142
A.6 Étapes d'implémentation du PHM [22]. 146
A.7 Vue d'ensemble de l'architecture avec les logiciels utilisés. 148
A.8 Diagramme algorithmique du SmS. 149
A.9 Schémas électrique complet réalisé avec le logiciel Kicad. 151
A.10 La spécificité des capteurs de gaz analogiques par l'ajout de résistance de charge pour évaluer la résistance interne du capteur dont la valeur évolue en fonction des concentrations. 152
A.11 MICS 6814 en connexion analogique avec un pont diviseur de tension. . . . 152
A.12 Contrôle de l'alimentation des capteurs de gaz et du ventilateur en 5V. . . 153
A.13 Le capteur de gaz MQ4 connecté en analogique et alimenté en 5 V. 154
A.14 Le capteur de poussière SDS011 en connexion UART avec l'ESP32. 154
A.15 Capteur de gaz SPG30 en connexion I^2C. 155
A.16 Le BMP 280 en connexion I^2C avec l'ESP 32. 155
A.17 Le capteur de SO_2 en connexion analogique. 156
A.18 Le design du PCB final pour le deuxième prototype. 157
A.19 La preuve du concept avant mise en fabrication du 1er prototype. 158
A.20 1er prototype déployé. 158

Liste des tableaux

2.1 Labélisation de station en fonction des combinaisons possibles pour le couple station/polluant [40]. 41
2.2 Méthodes d'analyse pour la surveillance réglementaire des polluants [40]. . 42
2.3 Une sélection d'études portant sur les micro-capteurs. 51
2.4 Sélection d'études portant sur des application de déploiement de LCS pour la surveillance de la qualité de l'air. 53

3.1 Résumé des méthodes d'interpolation les plus utilisées 72

5.1 Sélection des LCS : première itération. 108
5.2 Sélection des LCS : seconde itération. 109

Liste des acronymes

AASQA	Association Agrée de Surveillance de la Qualité de l'Air
BASTER	Base de Données TEmps Réelle
BDREP	Base de Données du Registre des Émissions Polluantes
BPCO	Broncho-Pneumopathie Chronique Obstructive
CCPVG	Communautés de Communes Pyrénées Vallées des Gaves
CHIMERE	modèle déterministe de chimie-transport
CPIE	Centre Permanent d'Initiatives pour l'Environnement
CLA	Couche Limite Atmosphérique
COVNM	Composés Organiques Volatils Non Méthaniques
ENIT	École Nationale d'Ingénieurs de Tarbes
ESMERALDA	EtudeS Multi RégionALes De l'Atmosphère
GUI	Interface Utilisateur Graphique
HAP	Hydrocarbures Aromatiques Polycycliques
IoT	Internet des Objets
LCS	Low-Cost Sensor ou micro-capteur
ML	Machine Learning
PCAET	Plan Climat Air Énergie Territorial
PDU	Plans de Déplacements Urbains
PID	Détecteur à Photo-Ionisation
PPA	Plans de Protection de l'Atmosphère
ppm	partie par million
PRSQA	Programme Régional de Surveillance de la Qualité de l'Air
SCOT	Schémas de Cohérence Territoriale
TECV	Loi relative à la transition énergétique pour la croissance verte
TMJA	Trafic Moyen Journalier Annuels

Nomenclature

PM_{10}	particules de 10 micromètres
T	température
H	humidité
Pe	pression
P	nombre de zones d'intérêt, $P \in \mathbb{N}$
Z_j	zone d'intérêt numéro j, $j \in \{1, ..., P\}, j \in \mathbb{N}$
N	nombre de paramètres physiques surveillés, $N \in \mathbb{N}$
φ_i	paramètre physique numéro i, $i \in \{1, ..., N\}, i \in \mathbb{N}$
Q	nombre de station dans $Z_j, Q \in \mathbb{N}$
S_k	station numéro k, $k \in \{1, ..., Q\}, k \in \mathbb{N}$
R	nombre de SmS dans S_k, $R \in \mathbb{N}$
x	SmS numéro x avec $x \in \{1, ..., R\}, x \in \mathbb{N}$
Te_x	période d'échantillonnage du SmS n°x
τ	période d'échantillonnage de la station tel quel $\tau > 2Te_x$
τ	taille de la fenêtre d'observation
k	k^{eme} période d'échantillonnage avec $t = k * Te_x$
M_x	nombre d'observations du SmS x
Y_x	matrice temporelle d'observations du SmS x, $(M_x * N)$
$y_x^{\varphi_i}(k)$	paramètre φ_i observé par le SmS n°x at time k
$y_{x,k}$	observations du SmS x à l'instant k
q_x	nombre d'observations du SmS x pendant τ
$y_{x,k}^{\tau}$	sous-matrice Y_x de taille $(q_x * N)$ à l'instant k pendant τ
$\varphi_{i,x}^{\tau}$	colonne de $y_{x,k}^{\tau}$ contenant les observations de φ_i
$\varphi_{i,s}^{\tau}$	concaténation de $\varphi_{i,x}^{\tau}$ de tous les SmS
$S_{s,k}^{\tau}$	synthèse des observations des SmS entre t et $t + \tau$
$D_{s,k}^{\tau}$	matrice de détection entre t et $t + \tau$
$R\varepsilon_i$	erreur relative permise pour le LCS observant φ_i
$R\varepsilon_S$	vecteur d'erreurs relatives permises (PREV)
ThM_i	seuil maximum pour l'observation de φ_i
Thm_i	seuil minimum pour l'observation de φ_i
$Th_{s,k}^{\tau}$	matrice de seuils
$R_{i,x}^{\tau}$	vecteur Reliable du SmS x pour φ_i
$O_{i,x}^{\tau}$	vecteur Outlier du SmS x pour φ_i
$R_{i,S}^{\tau}$	concaténation de tous les $R_{i,x}^{\tau}$
$O_{i,S}^{\tau}$	concaténation de tous les $O_{i,x}^{\tau}$
$I_{s,k}^{\tau}$	matrice d'indicateurs
$Id_{i,x}^{\tau}$	indice de défaillance du SmS x pour φ_i
$Ih_{i,x}^{\tau}$	indice de santé du SmS x pour φ_i
$Ic_{i,S}^{\tau}$	indice de confiance pour φ_i au niveau de la station
Pc	ratio de confiance permis
$\mathcal{H}$	fonction échelon de Heaviside
$S_{s,k}^{\tau}$	synthèse de toutes les observations pendant τ
Y_{S_k}	matrice temporelle d'observations de la station S_k
$S.I.D.$	matrices Synthèse, Indices et Détection en sortie de l'Agrégateur

Chapitre 1

Introduction générale

Sommaire

1.1 Contexte 16
1.1.1 Inventaire des émissions 18
1.1.2 Sources de pollution du secteur résidentiel 19
1.1.3 Sources de pollution des activités agricoles 20
1.1.4 Sources de pollution du secteur transport 21
1.1.5 Sources de pollution du secteur industrie 21
1.1.6 Les attentes de la CCPVG 22
1.2 Problématiques 24
1.3 Questions de recherche 24
1.3.1 Acquisition de mesures fiables de la qualité de l'air 24
1.3.2 Détection des défaillances et prédiction de l'état de santé des stations de mesure 25
1.3.3 La couverture en plusieurs points du territoire à partir de mesures de stations 25
1.3.4 La validation des acquisitions et la modélisation de la pollution de l'air 25
1.4 Objectifs de recherche 26
1.5 Contribution par rapport à l'état de l'art 26
1.6 Conclusion du chapitre introduction 27
1.7 Structure de la thèse 28

1.1 Contexte

Cette thèse a pour origine un stage de recherche effectué pendant le projet de fin d'étude pour l'obtention du diplôme d'ingénieur de POUPRY Sylvain. Un partenariat avec l'École Nationale d'Ingénieurs de Tarbes (ENIT) et la Communautés de Communes Pyrénées Vallées des Gaves (CCPVG) a été créé pour étudier la qualité de l'air sur son territoire.

La CCPVG est au cœur de la vallée des Gaves. C'est un territoire situé au sud-ouest de la France, dans la région Occitanie et au sud du département Hautes-Pyrénées. Frontalière de l'Espagne, la CCPVG est un territoire intercommunal d'environ 1000 km^2 et couvert de forêts et de milieux semi-naturels ainsi que de hautes montagnes. La CCPVG a été créé le 1er janvier 2017. Elle est composée de :

- 46 communes ;
- 15 556 habitants (7% de la population du département) ;
- 91% du territoire couvert par des forêts et milieux semi-naturels (61% dans le 65) soit 19 500 ha ;
- 20 548 logements dont 61% sont des résidences secondaires ;
- près de 3 800 emplois tertiaires (38% des emplois) ;
- près de 360 exploitations agricoles principalement dédiées à l'élevage : ovins, caprins et équins (60%), bovins (30%) ;
- environ 80 industries principalement dans le secteur alimentaire (42% des établissements).

Elle est principalement composée de villes rurales et de deux villes centre (représentant plus de 50% de l'agglomération multi communale) : Pierrefitte-Nestalas et Argelès-Gazost. Ce territoire abrite une grande partie du cœur du Parc national des Pyrénées. Dépourvu d'habitants permanents, ce parc national fait l'objet d'une réglementation spécifique afin de préserver la biodiversité et le caractère exceptionnel des patrimoines naturel, culturel et paysager du Parc national. Il est traversé par la chaîne des Pyrénées et possède notamment plusieurs pics de hautes altitudes tels que le Vignemale (3298 m) où le Cabaliros (2334 m) illustrés par la Figure1.1.



FIGURE 1.1 – Situation géographique et relief de la CCPVG (geoportail.fr).

Ce territoire possède une économie basée sur le tourisme (il passe de 16 000 habitants à 28 500 en hautes saisons), l'agriculture, l'industrie et le transport. Il est régulièrement confronté à des pics de pollutions en raison de sa climat-topologie particulière qui est favorable à l'augmentation des concentrations des polluants. Avec moins de 20 000 habitants, la CCPVG ne se voit pas imposer l'obligation de réalisation d'un plan climat-air-énergie territorial (PCAET) par la Loi relative à la transition énergétique pour la croissance verte (TECV). Cependant, les élus communautaires ont formulé le souhait de voir leur projet de territoire conduit au regard des questions liées à la transition énergétique. C'est pourquoi ils ont décidé de prescrire l'élaboration d'un PCAET sur leur territoire par délibération en date du 27 septembre 2017. Pour connaître le profil de pollution de la CCPVG, un inventaire des émissions de pollution est édité chaque année par une Association Agrée de Surveillance de la Qualité de l'Air (AASQA). ATMO Occitanie est l'AASQA chargée de surveiller la région Occitanie. Elle fournit trois outils (Cf. Le réseau de surveillance d'ATMO Occitanie) :

1. L'inventaire des émissions de polluants est un résumé global des émissions de polluants sur une année au niveau départemental. Il est publié après l'année révolue ;
2. La modélisation de la pollution est établie à partir d'extrapolations de mesures depuis les stations de mesure. Il permet de cartographier la pollution ;
3. La mesure temps réel de la pollution depuis ses stations de mesure.

En résumé, le territoire est caractérisé par une topologie montagneuse et une activité touristique qui est à l'origine de phénomènes particuliers concernant la pollution atmosphérique. Ainsi, régulièrement les citoyens observent des concentrations accrues de polluants qui peuvent persister pendant plusieurs jours. Ces accumulations ne sont pas seulement le résultat de facteurs locaux, mais peuvent également être influencées par des conditions météorologiques et des courants d'air spécifiques associés aux terrains montagneux.

La CCPVG est particulièrement préoccupée par ce que l'on appelle « la pollution de fond ». Contrairement aux épisodes de pollution aiguë qui peuvent être intenses mais de courte durée, la pollution de fond est persistante et s'accumule sur de longues périodes. Elle peut être moins visible et moins perceptible immédiatement, mais ses effets sur la santé sont plus insidieux. En effet, une exposition prolongée à ce type de pollution peut entraîner des problèmes de santé chroniques, augmentant ainsi le risque de maladies respiratoires et cardio-vasculaires parmi les citoyens.

En outre, la CCPVG aspire à mieux comprendre les phénomènes de pollution qui se manifestent sur son teritoire. Elle cherche à déterminer la fréquence, l'intensité et la nature de ces épisodes de pollution afin de poser un diagnostic clair sur la qualité de l'air. En ayant une vision précise des tendances et des schémas de pollution, la CCPVG pourra mieux anticiper et répondre aux défis posés. L'objectif ultime est d'élaborer des stratégies efficaces pour protéger la santé de ses habitants, en tenant compte à la fois des effets immédiats de la pollution et de ses conséquences à long terme, notamment pour les populations les plus vulnérables comme les enfants et les personnes âgées.

Avec sa topographie unique marquée par un relief montagneux et les diverses activités humaines qui s'y déroulent, le territoire est régulièrement confronté, en particulier chaque hiver, à des épisodes de pics de pollutions aux particules fines. Ces pics ne sont pas le résultat d'une seule cause, mais d'une combinaison complexe de facteurs. Il existe plusieurs raisons principales à ces pics de pollution, notamment :

1. Les conditions météorologiques :
   — dans les situations stables où il y a peu ou pas de vent (accumulation de polluants) ;
   — lorsque l'air froid plaque les polluants à proximité du sol en période hivernale (pics particules et oxydes d'azote) ;
   — lorsqu'il fait chaud et ensoleillé en période estivale (conditions propices à la formation d'ozone et de particules fines secondaires) ;
2. L'apport massif d'une pollution sous l'effet du vent ;
3. L'augmentation saisonnière des émissions de polluants en lien avec certaines activités telles que le chauffage domestique ou les feux pastoraux.

Ces pics sont bien connus de la population et sont visibles. Cependant, sur toute la période de l'année la population est soumise à la pollution dite de fond qui a un impact notable sur la santé. Il est donc nécessaire de mesurer la qualité de l'air tout au long de l'année et notamment de suivre son évolution. La qualité de l'air résulte d'un équilibre complexe entre les apports directs de polluants émis dans l'air, les émissions polluantes, et les phénomènes auxquels ces polluants vont être soumis une fois dans l'atmosphère (transport, dispersion, dépôt ou réactions chimiques). Pour caractériser la qualité de l'air, il est important de faire état des émissions. Même si elle n'est pas directement liée, la qualité de l'air est définie par la mesure des concentrations ponctuelles de polluants et les émissions qui sont des moyennes sur l'année entière.

1.1.1 Inventaire des émissions

Élaboré et publié chaque année par ATMO Occitanie, l'inventaire régional des émissions, joue un rôle crucial dans la compréhension de la dynamique de la pollution atmosphérique au sein de la région. Non seulement cet inventaire offre une vue d'ensemble des activités génératrices d'émissions, mais il permet également d'évaluer la part contributive de chacune de ces activités, offrant ainsi des perspectives précieuses pour les interventions et les politiques futures. Selon l'inventaire régional en 2019, voici les principaux éléments clés :

— le secteur du transport est majoritaire dans les émissions d'oxydes d'azote représentant à lui seul 60% des émissions. La principale source de cette pollution provient des véhicules diesels, qui sont connus pour être d'importants émetteurs de ce type de polluant ;
— le secteur résidentiel est très émetteur de particules fines. Plus précisément, il contribue à 39% des émissions de particules de diamètre inférieur à 10 µm(PM_{10}) et à 52% des émissions de particules de diamètre inférieur à 2.5 µm($PM_{2.5}$). La majeure partie de ces émissions provient des chauffages au bois non performants, pouvant libérer une quantité importante de particules dans l'air en raison d'une combustion incomplète ;
— le secteur agricole possède le monopole des émissions de l'ammoniac (NH_3) de l'ordre de 99%. Cette situation est probablement due aux rejets organiques de l'élevage, qui libèrent de l'ammoniac dans l'atmosphère ;

— les secteurs industriel et du transport jouent un rôle non négligeable dans les émissions de gaz à effet de serre, avec des contributions respectives de 21% et 22%.

Les données présentées par l'inventaire d'ATMO Occitanie en 2019 mettent en évidence le poids considérable de certains secteurs dans les émissions polluantes du territoire. Cet inventaire ne se contente pas de fournir des chiffres ; il dresse un tableau précis des origines des polluants, offrant ainsi un diagnostic initial clair des principales sources d'émissions liées aux activités humaines. Les sous-sections suivantes approfondissent ces origines, en se penchant en détail sur chaque type de polluant. Il est essentiel d'avoir une compréhension approfondie de ces polluants car connaître leur nature permet d'adapter et d'optimiser les systèmes de surveillance, assurant ainsi une détection et une réponse plus efficaces face aux variations de la qualité de l'air sur le territoire.

1.1.2 Sources de pollution du secteur résidentiel

Le secteur résidentiel est le premier secteur émetteur de Composés Organiques Volatils Non Méthaniques ($COVNM$) (75%) et des gas à effet de serre (GES) (30%). La pratique du chauffage au bois émet la quasi-totalité des PM_{10} et $PM_{2.5}$ pour ce secteur. Dans la globalité des émissions, il est responsable de 56% des particules fines ($PM_{2.5}$) sur le territoire.

Le chauffage au bois

C'est une des énergies les moins chères au kWh. Il contribue également aux objectifs de réduction des émissions de gaz à effet de serre : le bilan est quasiment neutre dans le cas du bois énergie car il provient des forêts locales. Lors d'utilisation d'appareils de chauffage au bois, si la combustion est complète, elle va émettre du CO_2, de l'eau mais aussi des PM et du NO_X. Lors d'une combustion incomplète, en plus de ces molécules, d'autres polluants seront émis dans l'air (intérieur et extérieur) : monoxyde de carbone, COV, $HAP16$, dioxines, furanes, des suies, des goudrons, etc. Le chauffage au bois réalisé dans de mauvaises conditions va donc émettre divers polluants. Plusieurs facteurs sont à l'origine de ces émissions : les appareils peu ou pas performants (foyers fermés avant 2007 et foyers ouverts), sur ou sous dimensionnés, la combustion de bois humide ou de récupération, une combustion lente, un entretien peu régulier voire inexistant, etc. Afin de réduire ces émissions de polluants, il est possible d'agir sur trois principaux axes en simultané : les appareils de chauffage, le combustible et les habitudes de chauffage.

Le brûlage de déchets végétaux

Le brûlage de déchets verts (brûlage des déchets d'entretien des jardins des particuliers) a des conséquences sur la santé en plus de créer des troubles du voisinage. A savoir, 50 kg de déchets verts brûlés émettent 920 fois plus de PM_{10} qu'un trajet de 20 km jusqu'à une déchetterie (pour un véhicule diesel ancien). Cette combustion très peu performante est source multiple de polluants (les déchets verts étant souvent très humides). Ils s'additionnent à la pollution de fond. Malgré la réglementation qui interdit ce type de brûlage, il arrive souvent de voir des feux de déchets végétaux domestiques.

1.1.3 Sources de pollution des activités agricoles

Le secteur agricole est contributeur à 99% aux émissions d'ammoniac (NH_3), il contribue aussi à hauteur de 16% des oxydes d'azotes (NO_x) et de 20% des GES. Les pratiques ancestrales d'entretiens de la montagne et des pâturages sont propres au territoire et suscitent des débats entre les citoyens et les agriculteurs.

Les feux pastoraux

Également dénommés écobuages ou brûlage dirigé, les feux pastoraux sont réalisés par l'homme depuis plusieurs millénaires à des fins essentiellement agricoles et pastorales. Ils consistent à contrôler par incinération des espèces végétales non consommées par le bétail qui envahissent progressivement les pâturages, diminuant les ressources à disposition des troupeaux et banalisant le paysage. Le feu pastoral vise ainsi à :

— contrôler les refus (strate herbacée sèche ou morte) et la strate arbustive envahissante ;
— renouveler et diversifier les ressources pastorales dans le temps et dans l'espace ;
— favoriser l'accessibilité à la ressource pastorale, son appétence et sa précocité.

La mécanisation, très difficile voire impossible en milieu montagnard, a permis la persistance de cette technique. Le feu est utilisé pour produire de l'herbe sur de grandes surfaces avec le moins de travail possible. Les éleveurs et les bergers utilisent ces espaces pour le parcage de leur bétail. Les écobuages doivent être réalisés par temps clairs sans vent pour des raisons de sécurité évidentes. Or, ces conditions sont favorables aux inversions thermiques et donc à la stagnation des polluants. Beaucoup d'habitants des vallées s'en plaignent mais pour l'instant peu d'alternatives sont viables (des engins mécanisés existent mais à des coûts importants). Les feux pastoraux sont profondément ancrés dans la culture et les traditions pyrénéennes. La problématique de l'écobuage est complexe et nécessite du temps et de la concertation. Des réflexions sont engagées notamment par les services de l'État sur le département des Hautes-Pyrénées. Ils restent une source majeure d'émission de polluants dans les périodes où ceux-ci ne peuvent être dispersés correctement.

Les élevages et cultures

L'ammoniac (NH_3) émis sur le territoire est quasi-exclusivement émis par le secteur agricole. Les deux sources d'émissions d'ammoniac sur le territoire du Parc sont :

— les cheptels, par les composés azotés issus des déjections, à hauteur de 35% ;
— les cultures et les prairies, par fertilisation des sols, à hauteur de 65%.

Le nombre des ventes des produits régionaux est lié aux émissions d'ammoniac, ce qui fait qu'il est un excellent indicateur pour améliorer les estimations de ce polluant.

L'utilisation des tracteurs et fonctionnement des bâtiments

Le fioul ou le diesel sont la principale source d'énergie pour chauffer les serres et faire fonctionner les exploitations agricoles. Le secteur agricole diminue son impact grâce à l'utilisation de Gazole Non Routier (GNR) pour les engins agricoles.

1.1.4 Sources de pollution du secteur transport

La topologie du territoire ne permet pas d'avoir actuellement d'autres transport que routier. Les vallées sont encaissées et bien souvent il n'y a qu'une seule route. Le secteur transport contribue à 59% des émissions d'oxyde d'azote (NOx). Les émissions du transport routier sont liées à la combustion du carburant, aux émissions de particules dues aux freinages, usures des pneus et de l'usure de la route. Il n'existe pas pour le moment d'alternatives utilisables comme dans les grandes villes (tramway électrique ou métro). Les transports en communs tels que les bus sont présents. Les émissions sont évaluées à partir du Trafic Moyen Journalier Annuel (TMJA). En raison de sa topographie et de ses lieux de loisirs (stations de ski), les flux les plus importants sont concentrés sur les axes reliant les villes les plus peuplées, voir Figure1.2.

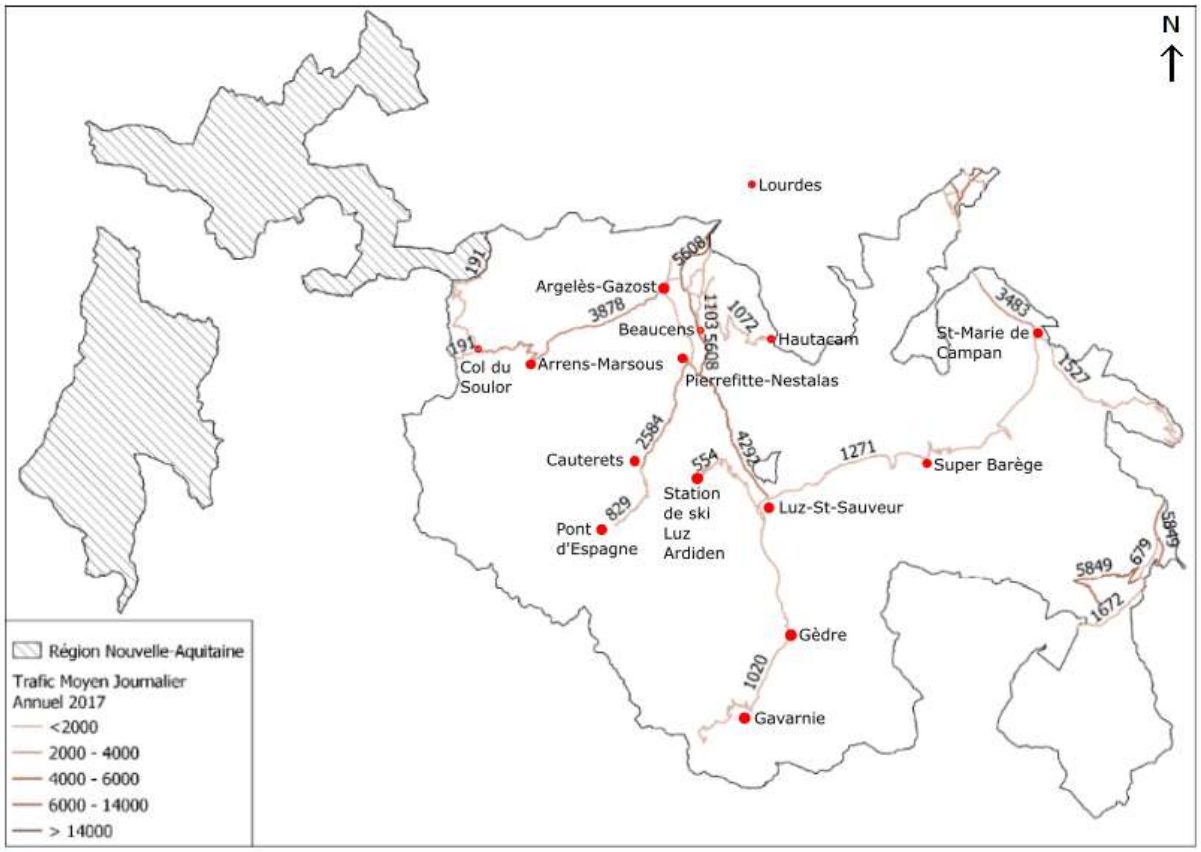


FIGURE 1.2 – TMJA des routes de la CCPVG (ccpvg.fr).

1.1.5 Sources de pollution du secteur industrie

Les émissions du secteur industriel proviennent de différentes sources, telles que les industries manufacturières, les industries chimiques et les carrières. La principale source de données utilisée dans l'inventaire régional est la Base de Données du Registre des Emissions Polluantes (BDREP) qui est un registre déclaratif pour l'ensemble des exploitants . Leurs émissions sont calculées en fonction de l'activité déclarée. Deux industries soumises à déclaration sont prises en compte sur ce territoire dans le calcul des émissions du secteur industriel. Ce dernier représente également 21% des émissions de $COVNM$ et 20% des émissions de GES sur le territoire en 2017. Les émissions importantes de $COVNM$

proviennent de l'utilisation de solvants dans l'industrie, notamment l'imprimerie, l'application de peinture ou encore la protection du bois. Le secteur industriel contribue de manière significative aux émissions de particules PM_{10} (42%) et $PM_{2.5}$ (36%) sur le territoire de la CCPVG. Les émissions de PM_{10} et $PM_{2.5}$ proviennent principalement des carrières présentes sur le territoire. Les activités d'extraction, de broyage et de transport de la roche, notamment, sont particulièrement émettrices de poussières. Ainsi sur le territoire de la CCPVG, les carrières émettent 78% des émissions totales de PM_{10} liées à l'industrie et environ 33% des émissions totales de PM_{10} du territoire, en 2017.

1.1.6 Les attentes de la CCPVG

La CCPVG dispose de trois outils pour évaluer la qualité de l'air sur son territoire : l'inventaire des émissions de polluants, la modélisation et la mesure temps réel. Toutefois, chacun de ces outils présente des limites.

L'inventaire des émissions de polluants, établi à l'échelle du département, ne permet pas d'avoir une vision actualisée de la situation de la CCPVG. De plus, les statistiques ne sont pas spécifiques à la zone géographique de la CCPVG. Bien que l'inventaire régional des polluants d'ATMO Occitanie soit utile pour définir le profil des sources de pollution des activités, il ne fournit pas d'informations précises sur la qualité de l'air. En effet, la qualité de l'air est basée sur la concentration des polluants en temps réel et non sur une moyenne lissée tout au long de l'année. De plus, cet inventaire est un résumé global issu des extrapolations des stations de mesures de Tarbes et Lourdes, situées hors du territoire de la CCPVG, voir Figure 1.3. Sa publication est tardive en raison d'un temps d'analyse non négligeable, de l'ordre de six à huit mois.

La modélisation, quant à elle, ne permet pas de prédire l'état de la qualité de l'air en tenant compte des sources de pollutions locales pour permettre aux acteurs locaux de réagir. Elle est établie à partir d'extrapolations de mesures sur un nombre réduit de capteurs, avec seulement deux stations de mesures dans le département et pas une seule dans la zone géographique de la CCPVG. De plus, le territoire est en relief, ce qui soulève des questions sur la prise en compte réelle des sources de pollution locales par le modèle de pollution, étant donné la complexité des phénomènes météorologiques.

Le troisième outil dont dispose la CCPVG pour évaluer la qualité de l'air est la mesure en temps réel. Cependant, la Figure 1.3 illustre parfaitement la situation géographique : le territoire couvre près de 1000 km^2 et il n'y a pas de station de mesure. Cela pose un problème pour la mesure en temps réel de la qualité de l'air.

Face à ces limites, la CCPVG a décidé de mettre en œuvre une démarche scientifique pour la surveillance de la qualité de l'air sur son territoire, par l'intermédiaire du projet BOLDAIR à l'origine de cette thèse. Elle a besoin de faire un premier diagnostic pour mieux se rendre compte des spécificités des activités en fonction des saisons, en vue de prévoir les pics de pollutions et de protéger les personnes les plus vulnérables. Le suivi de l'évolution de la qualité de l'air tout au long de l'année permettra de prendre des mesures pour réduire les émissions, notamment en sensibilisant, en régulant l'écobuage, en aménageant les transports et en éduquant les citoyens. Le projet BOLDAIR sera ainsi un outil d'aide à la décision permettant d'avoir un appui scientifique pour agir efficacement et améliorer la qualité de l'air sur le territoire de la CCPVG.

Pour résumer, la CCPVG a besoin d'avoir et de disposer d'un outil d'aide à la décision pour :

— avoir un état initial de la situation ;

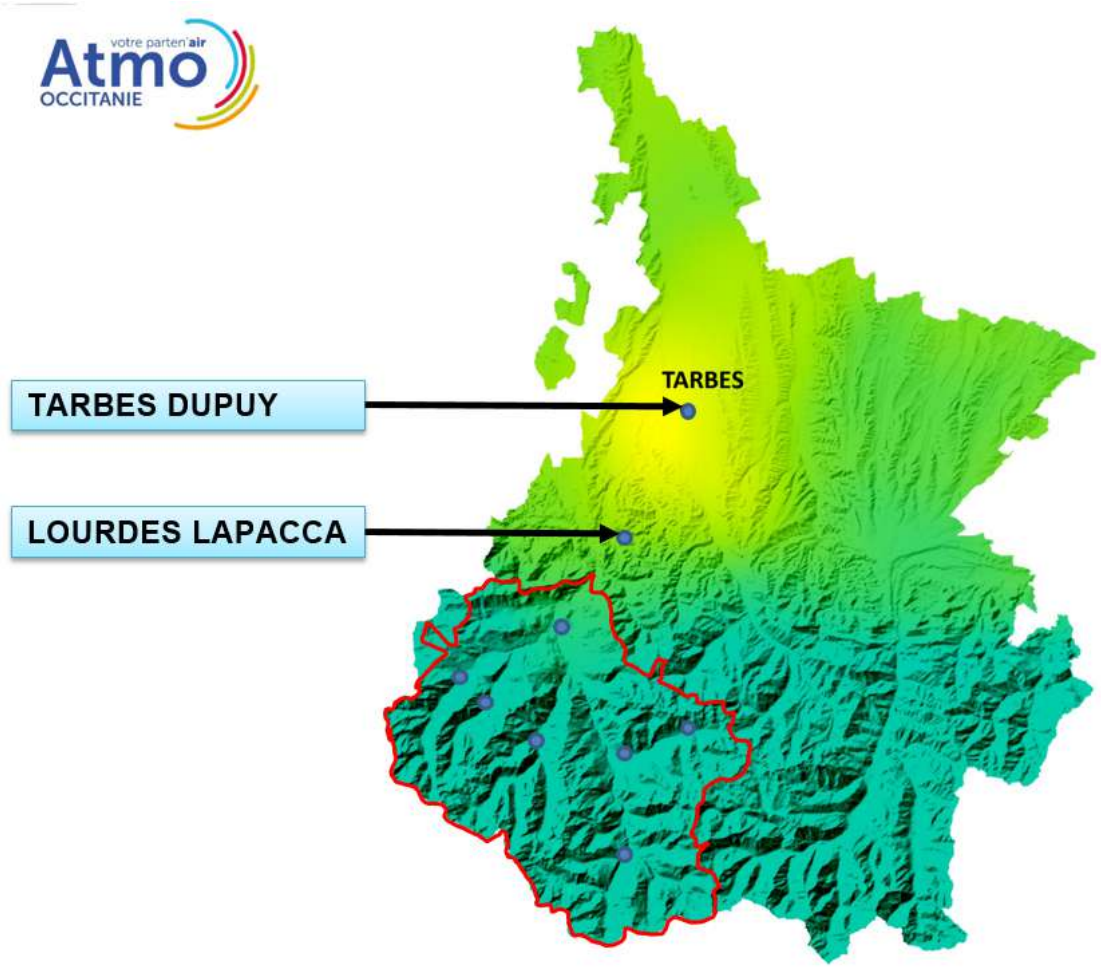


FIGURE 1.3 – Situation géographique de la CCPVG par rapport aux stations de mesures conventionnelles.

— monitorer les concentrations des polluants ;
— anticiper les pics de pollution.

En tenant compte des attentes de la CCPVG, il est nécessaire d'avoir une approche holistique pour aborder les enjeux de la pollution atmosphérique. C'est dans cette perspective que le projet BOLDAIR a été conçu, définissant des objectifs clairs pour l'ENIT, s'efforçant de répondre aux défis posés par les enjeux environnementaux et économiques du territoire.

1.2 Problématiques

La réalisation d'un projet de cette envergure nécessite de prendre en compte plusieurs défis spécifiques associés à la surveillance de la qualité de l'air sur le territoire de la CCPVG :

- la contrainte financière est à prendre en compte, principalement due à la taille modeste des communes, limitant ainsi le budget alloué ;
- la gestion et le contrôle des données au niveau local sont essentiels pour garantir la maîtrise des informations recueillies et l'indépendance de la CCPVG ;
- le territoire à couvrir est vaste, s'étendant sur plusieurs milliers de kilomètres carrés, ce qui nécessite de multiplier les points de prélèvement ;
- la topologie du territoire, avec ses caractéristiques uniques, peut influencer la dispersion des polluants et donc les méthodes de surveillance ;
- la mesure continue de la qualité de l'air, ce qui signifie une surveillance sans interruption tout au long de l'année ;
- enfin, l'absence de personnel qualifié pour la maintenance du réseau de mesure est une préoccupation majeure pour assurer la haute disponibilité du service.

De façon générale, des territoires similaires à la CCPVG observent des pollutions qui échappent souvent à la détection. Les trois outils principaux fournis par les AASQA (surveillance en temps réel, modélisation et inventaire des émissions) ne sont pas adaptés à ce type de territoire. Les stations conventionnelles, souvent coûteuses, se trouvent majoritairement dans les grandes agglomérations, laissant ainsi des zones moins denses non couvertes. De plus, lorsqu'il s'agit de modélisation, les territoires avec une topologie particulière, comme celui de la CCPVG, posent des défis spécifiques, rendant la mise en place de modèles réalistes difficile. En outre, puisque les données sont principalement détenues par les AASQA, les collectivités locales se retrouvent dépendantes de leur politique de gestion des données. Face à ces enjeux, il devient impératif de définir des objectifs de recherche clairs pour développer des solutions sur mesure et répondre aux besoins spécifiques du territoire.

1.3 Questions de recherche

La surveillance et la prédiction améliorées de la qualité de l'air sur un vaste territoire, tout en restant dans les limites des contraintes budgétaires et du personnel, constituent un défi majeur. L'adoption de technologies avancées telles que l'Internet des objets (IoT) et les micro-capteurs ou Low-Cost Sensor (LCS) dans une architecture de réseau de micro-capteurs vise à répondre à ce défi. L'objectif est de compléter et d'amplifier le travail des AASQA en multipliant les points de prélèvement pour affiner la cartographie des pollutions locales et suivre en temps réel les mouvements des polluants tout au long de l'année. Cette thèse vise ainsi à lever quatre verrous scientifiques majeurs et présentés ci-dessous.

1.3.1 Acquisition de mesures fiables de la qualité de l'air

L'exploitation d'un réseau de LCS ouvre un champ d'interrogations quant à la surveillance de la qualité de l'air en tenant compte des particularités locales et sur l'ensemble du processus d'acquisition. En effet, ces systèmes présentent des caractéristiques telles que

des défaillances aléatoires et brutales, ainsi qu'une faible précision qui nécessitent une attention soutenue. Comment garantir l'obtention de données fiables sur une longue période malgré ces contraintes ? La surveillance de la santé du système de mesure, la prédiction des défaillances et la haute disponibilité deviennent des préoccupations essentielles. La fiabilité est définie dans cette thèse comme l'obtention de résultats précis et cohérents sur une longue période de temps, malgré les perturbations ou les dysfonctionnements éventuels.

1.3.2 Détection des défaillances et prédiction de l'état de santé des stations de mesure

Les défaillances potentielles des LCS soulèvent des préoccupations quant à la fiabilité de la surveillance. Comment assurer une détection précoce des dysfonctionnements et anticiper l'état de santé des composants pour garantir une surveillance continue ? Les stations de mesure doivent fournir, outre les mesures, des indicateurs fiables sur leur état de santé, permettant des interventions rapides et assurant aux autorités une base solide pour leurs décisions.

1.3.3 La couverture en plusieurs points du territoire à partir de mesures de stations

La surveillance de la qualité de l'air sur un territoire étendu nécessite une répartition stratégique des stations de mesure pour garantir une couverture spatiale optimale. Toutefois, les contraintes budgétaires et logistiques peuvent limiter le nombre de ces stations. La question clé est donc la suivante : comment, à partir d'un nombre limité de mesures ponctuelles, peut-on extrapoler et estimer la qualité de l'air sur l'ensemble du territoire, et ce pour l'ensemble des polluants ? Les méthodes d'interpolation spatiale, telle que l'interpolation linéaire ou le krigeage, offrent des pistes prometteuses, mais nécessitent une adaptation précise aux spécificités du territoire et des données disponibles.

1.3.4 La validation des acquisitions et la modélisation de la pollution de l'air

L'exploitation efficace des données météorologiques et de pollution est cruciale pour établir un modèle prédictif crédible de la qualité de l'air. Face aux ressources limitées, comment peut-on optimiser l'intégration de ces données pour prévoir avec précision les variations futures de la qualité de l'air ? La modélisation, dans des contextes de territoires avec des ressources limitées, soulève des enjeux majeurs liés à la couverture spatiale et temporelle des prédictions.

Face à ces interrogations complexes et interdépendantes, il est essentiel de structurer une approche méthodologique rigoureuse. Pour adresser de manière concrète et systématique ces questions de recherche, cette thèse définit des objectifs précis, qui guideront chaque étape de l'étude et orienteront les méthodologies employées.

1.4 Objectifs de recherche

Cette thèse vise à mettre en place une démarche rigoureuse qui permettra d'apporter des éléments de connaissances scientifiques, tangibles et quantifiés aux citoyens et aux acteurs du territoire en charge de la surveillance et de la gestion de la qualité de l'air. Sur la base de ces éléments, ces derniers pourront mieux appréhender le problème de pollution, estimer, prédire et évaluer sa gravité, et définir des plans d'action préventifs (sensibilisation des habitants, incitation à l'utilisation de chauffage moins émetteur de polluants, incitation des éleveurs à moins recourir aux écobuages, etc.), mais aussi et surtout proactifs (régulation du trafic routier et de la pollution industrielle par anticipation de la dégradation future de la qualité de l'air). Les objectifs définis ici visent à faciliter la prise de décision pour les acteurs locaux, tout en sensibilisant et impliquant la communauté dans son ensemble. Les objectifs spécifiques sont :

1. Surveillance en temps réel : déployer une instrumentation adaptée sur le territoire pour obtenir une cartographie précise et en temps réel de la qualité de l'air. Cela implique le développement ou l'adaptation de technologies de surveillance, en tenant compte des spécificités locales ;
2. Prédiction et proactivité : élaborer des modèles prédictifs fiables pour anticiper les variations de la qualité de l'air. Ces modèles permettront d'initier des actions proactives, telles que la régulation du trafic ou la limitation de certaines activités polluantes lors des pics de pollution prévus ;
3. Engagement sociétal : impliquer activement les acteurs locaux (élus, habitants, industries, etc.) dans le processus de surveillance et de gestion. Cela comprend la sensibilisation aux enjeux, la formation à l'utilisation des outils de surveillance et la promotion de pratiques respectueuses de l'environnement.

En conclusion, cette thèse vise à doter les communes des outils et des informations nécessaires pour réagir de manière éclairée aux défis posés par la pollution de l'air, tout en intégrant la dimension locale (écobuages, agriculture, mauvaises pratiques et dégradations des végétaux).

1.5 Contribution par rapport à l'état de l'art

Les approches conventionnelles pour surveiller la qualité de l'air ont souvent reposé sur des stations de mesure fixes, coûteuses et nécessitant une expertise pour leur utilisation et leur maintenance. Bien que ces stations fournissent des données précises, elles peuvent ne pas couvrir de manière adéquate de vastes territoires, en particulier dans des régions éloignées et montagneuses. De plus, avec des ressources limitées, l'implantation de telles stations est difficile voire impossible.

L'utilisation d'un réseau de capteurs sur un large territoire, comme les capteurs à faible coût (LCS), offre une solution potentielle à ce problème. Cependant, cela vient avec ses propres défis :

— couverture d'un vaste territoire avec un budget modeste ;
— assurance de la cohérence et de la continuité des données ;
— absence de stations conventionnelles pour référence ;
— manque de personnel spécialisé et pas de politique de gestion des mesures ;
— surveillance et prédiction de la qualité de l'air avec des LCS.

L'approche proposée vise à répondre à ces défis en utilisant une combinaison de technologies modernes et d'analyses avancées. Les réseaux de LCS sont une opportunité pour surveiller la qualité de l'air, car ils sont moins coûteux et plus faciles à déployer et nécessitent peu de personnel qualifié. En effet, leur force réside dans la possibilité de les multiplier et ainsi détecter la pollution locale. Une architecture est proposée afin d'augmenter le territoire en ajoutant des capacités de calcul, de communication et d'analyse. Cela implique l'installation d'un réseau de mesure des concentrations de pollution pour surveiller l'environnement. Cependant, les études publiées sont insuffisantes en ce qui concerne la durée expérimentale du déploiement des LCS surtout lorsque le sujet de l'étude doit être suivi en continu sur de longues périodes.

l'approche proposée consiste à utiliser la redondance modulaire au niveau d'une station implantée sur un point de prélèvement. Le fait de regrouper et de comparer les mesures des LCS entre eux augmentent la fiabilité des données et permet de détecter leurs défaillances éventuelles. Cette approche permet également d'extraire des indicateurs pour prédire l'état de santé du point de prélèvement et par extension de l'architecture complète. En plus de surveiller le territoire, cette architecture peut s'auto-diagnostiquer et s'auto-configurer afin de permettre une haute disponibilité. Cette intégration permettra une plus grande confiance dans l'adaptabilité du système de prédiction basé sur des composants en réseau (capteurs, traitement et actionneurs). En résumé, cette approche permet de surmonter les inconvénients des déploiements standards avec les LCS et offre une solution performante pour surveiller l'environnement.

Comparée à d'autres travaux dans ce domaine, cette approche se distingue par son utilisation innovante de la redondance des LCS et sa combinaison de surveillance et de prédiction. Les implications sociétales de cette recherche sont vastes. En fournissant des outils plus accessibles et plus fiables pour la surveillance de la qualité de l'air, cette démarche peut aider les territoires ruraux à prendre des décisions éclairées pour protéger la santé de leurs habitants, même avec des ressources limitées. Il existe cependant des limitations. La fiabilité à long terme des LCS, même avec redondance, reste à déterminer. De plus, la prédiction dépend fortement de la qualité et de la quantité des données d'entrée. Des travaux futurs pourraient chercher à renforcer la robustesse des LCS et à intégrer davantage de sources de données pour la prédiction. En conclusion, cette recherche contribue à combler un vide important dans la surveillance de la qualité de l'air en proposant une solution innovante et rentable pour les territoires avec des ressources limitées.

1.6 Conclusion du chapitre introduction

La qualité de l'air dans le territoire de la CCPVG est un enjeu complexe et multidisciplinaire. Pour des territoires ayant une topologie et des activités spécifiques, mesurer la pollution nécessite une approche innovante pour la surveillance à long terme. Ce chapitre introduit ces défis et met en évidence les limites des méthodes conventionnelles de mesure de la qualité de l'air. En réponse, il esquisse les contours d'une approche qui utilise la redondance des LCS, l'intégration du PHM, et les technologies modernes comme l'IoT. L'objectif ultime est de fournir aux acteurs locaux les outils et les connaissances nécessaires pour prendre des décisions éclairées, protéger la santé des habitants et impliquer la communauté dans son ensemble. La thèse s'inscrit dans cette direction, visant à explorer et à résoudre les questions de recherche et à atteindre les objectifs définis. La contribution de cette recherche à l'état de l'art et ses implications sociétales sont soulignées, tout en reconnaissant les limitations et les domaines nécessitant des travaux futurs. En somme,

ce chapitre pose les bases pour une étude approfondie et une démarche scientifique rigoureuse, ouvrant la voie à une amélioration significative de la surveillance et de la gestion de la qualité de l'air dans les territoires ayant des ressources limitées.

1.7 Structure de la thèse

La thèse est organisée en cinq chapitres interdépendants, chacun abordant une facette spécifique du défi global de la surveillance de la qualité de l'air. Après cette introduction générale, le Chapitre 2 explore l'état de l'art, fournissant un contexte et une compréhension des méthodes existantes. Le Chapitre 3 présente l'architecture proposée, qui constitue la colonne vertébrale du système de surveillance, tandis que le Chapitre 4 explore dans le traitement des données, établissant les fondements mathématiques pour la fiabilité des mesures. Enfin, le Chapitre 5 applique la théorie et les méthodes développées aux cas spécifiques de la CCPVG, illustrant la mise en œuvre pratique. Pour terminer, dans la conclusion générale, l'ensemble des développements seront comparés aux objectifs initiaux de la thèse puis un résumé des travaux ainsi que les perspectives de recherche concluront ce manuscrit.

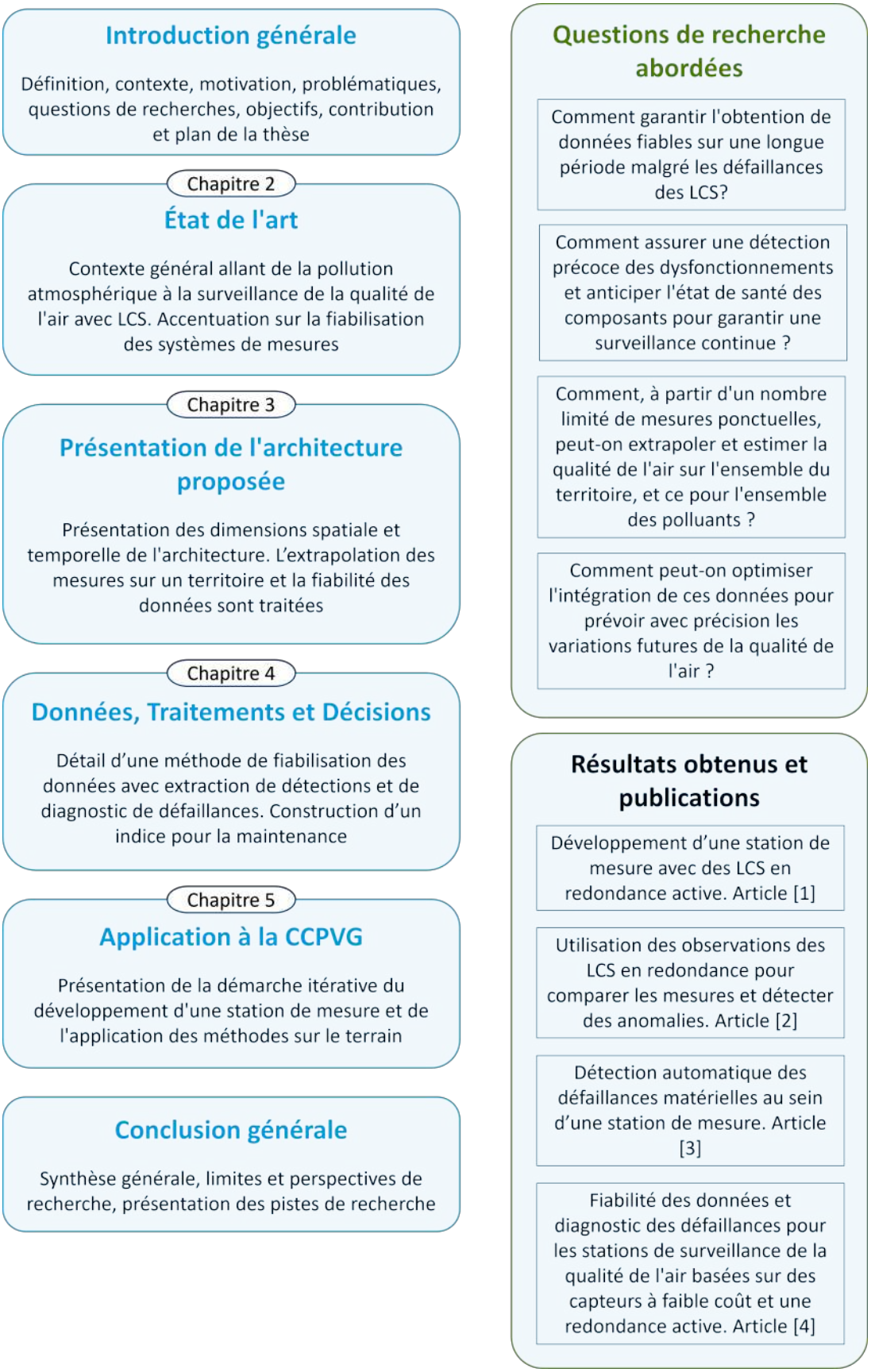


FIGURE 1.4 – Structure et résumé des questions de recherche abordées dans ce manuscrit de thèse.

Chapitre 2

État de l'art

Sommaire

2.1 Introduction 31
2.2 La surveillance de la Qualité de l'air 31
2.2.1 Le LCSQA 32
2.2.2 Les Normes de la qualité de l'air 33
2.2.3 Modèlisation de la qualité de l'air : l'approche descendante et assimilation de données 34
2.2.4 Les AASQA 36
2.2.5 Le réseau de surveillance d'ATMO Occitanie 36
2.2.6 Représentativité des stations de mesures 40
2.2.7 Les Technologies de surveillance de la qualité de l'air 42
2.3 État de l'art sur les micro-capteurs ou LCS 45
2.3.1 Définition 45
2.3.2 Les capteurs de particules 46
2.3.3 Les capteurs de gaz 47
2.3.4 Surveillance de la qualité de l'air avec des micro-capteurs . . . 50
2.3.5 Conclusion 53
2.4 État de l'art sur l'acquisition, le traitement et la fiabilisation des données 54
2.4.1 Acquisition des données afin d'alimenter un modèle pour faire des prédictions : l'exemple de la météorologie 54
2.4.2 Acquisition des observations : augmenter le maillage 55
2.4.3 La fiabilisation des données en un point d'observation 56
2.4.4 La triple redondance modulaire et les algorithmes de vote . . . 57
2.4.5 Les algorithmes de détection des défaillances 58
2.5 Conclusion du chapitre 59

2.1 Introduction

La mesure des concentrations de polluants atmosphériques est primordiale pour évaluer la qualité de l'air, un indicateur directement influencé par la pollution atmosphérique. Cette dernière correspond à l'introduction de polluants dans l'air, tandis que la qualité de l'air reflète l'état de l'atmosphère en fonction de ces polluants. Des concentrations élevées de polluants se traduisent par une mauvaise qualité de l'air, avec des implications potentiellement graves pour la santé humaine, l'équilibre écologique et les économies locales. Ce phénomène est particulièrement accentué dans les régions dotées d'une topologie spécifique, où la dynamique de dispersion des polluants peut poser des défis singuliers tels que la non prise en compte des pollutions locales. Cependant, ces territoires, bien qu'ayant un besoin de surveillance, sont souvent confrontés à une contrainte majeure : la mise en place d'un système de surveillance performant avec un budget modeste.

Ce chapitre, consacré à l'état de l'art, explore les méthodes et technologies de surveillance de la qualité de l'air, depuis les stations de mesure conventionnelles vers les réseaux de LCS. Ces LCS se distinguent notamment par leur flexibilité de déploiement, permettant une surveillance à grande échelle sans nécessiter de personnel hautement qualifié ni d'investissements lourds. Cette caractéristique les rend particulièrement adaptés aux territoires dotés de contraintes budgétaires, tout en offrant une dimension spatiale significative. La fiabilité des données est également au cœur des préoccupations, mettant en lumière des techniques d'acquisition robustes, des méthodes de validation et des systèmes de redondance, essentiels pour garantir la fiabilité des données collectées. Cette exploration vise à offrir une perspective claire sur les avancées, outils et méthodologies actuels adaptés à la surveillance de la qualité de l'air dans des territoires ayant des particularités topologiques et des ressources financières modestes. La complexité de la pollution atmosphérique, incluant les origines des polluants, leur cycle de vie ainsi que leur interaction avec l'atmosphère, est approfondie en annexe pour une compréhension plus détaillée.

2.2 La surveillance de la Qualité de l'air

La pollution de l'air est l'une des principales préoccupations de santé publique à l'échelle mondiale. Elle est responsable de 4,2 millions de décès par an dus à la pollution ambiante (air extérieur) [23]. Pour améliorer la qualité de l'air, l'Organisation Mondiale de la Santé (OMS) propose des recommandations [71] qui ont été adoptées par diverses organisations au niveau national ou continental (UE, CHINE, USA), telles que les Normes nationales de qualité de l'air ambiant aux États-Unis (NAAQS) ou ATMO France en France dans le cas présent. Parmi ces recommandations, les nouvelles directives de l'OMS, approuvées en septembre 2021 [74], fixent des seuils de pollution encore plus stricts que ceux définis en 2005 [72]. Afin de se conformer à ces recommandations, les gouvernements s'appuient sur des organisations telles que l'Agence européenne pour l'environnement (AEE), chargée de surveiller la qualité de l'air en Europe, AirNow aux États-Unis, le Département de l'écologie et de l'environnement de chaque province en Chine et le Conseil central de contrôle de la pollution en Inde, etc. [25]. Ces organisations sont responsables de la surveillance de la qualité de l'air et doivent fournir des informations fiables et régulières au public et aux autorités.

En France, la surveillance de la qualité de l'air est assurée par un réseau d'associations coordonnées par la fédération ATMO France. Le ministère de l'Écologie et du Développement Durable ainsi que le Laboratoire Central de Surveillance de la Qualité de l'Air

(LCSQA) supervisent cette surveillance, tandis que l'Agence de l'environnement et de la maîtrise de l'énergie (ADEME) est chargée du support technique.

2.2.1 Le LCSQA

Créé en 1991 par le ministère chargé de l'environnement, le LCSQA est un support scientifique et technique pour l'État et les AASQA. Depuis 2011, il est responsable de la coordination du dispositif de surveillance de la qualité de l'air. Il est en partenariat avec l'École des Mines de Douai, le Laboratoire National de métrologie et d'Essais ainsi que l'Institut National de l'Environnement industriel et des RISques. Une de ses missions est de garantir une surveillance optimale sur l'ensemble du territoire. La Figure 2.1 illustre les interactions entre les différents acteurs.

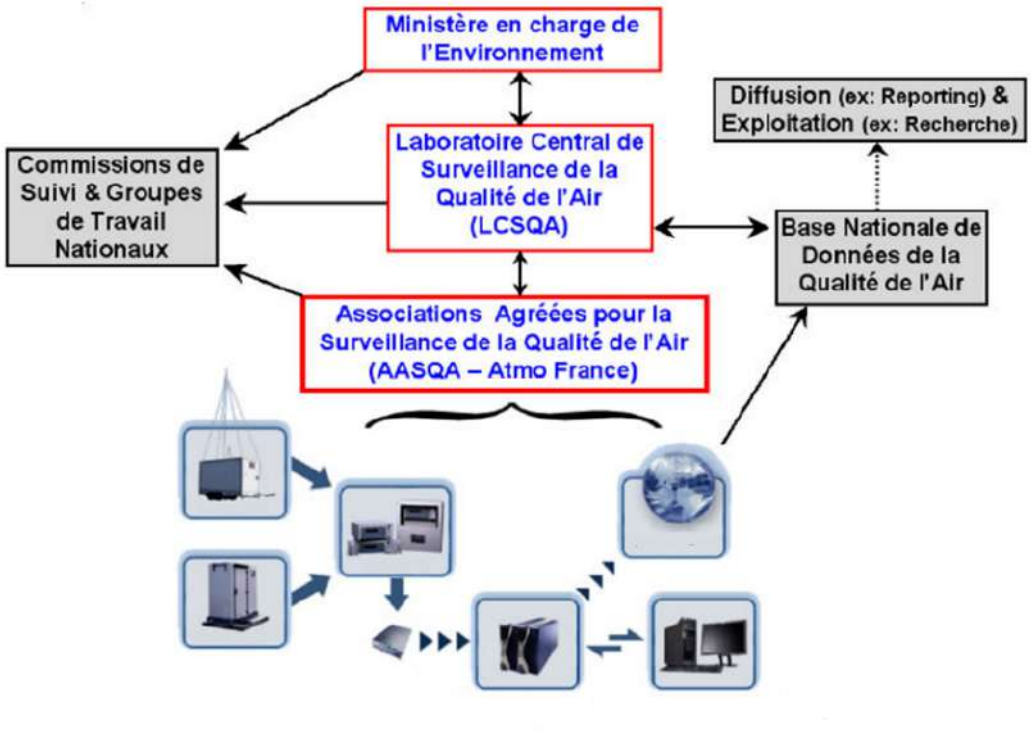


FIGURE 2.1 – Schéma de fonctionnement du dispositif national français [39].

Pour garantir la mission de surveillance, le LCSQA a mis en place une chaîne d'étalonnage pour assurer la cohérence des mesures de la qualité de l'air effectuées par les AASQA. Les principaux polluants mesurés sont le PM_{10}, $PM_{2.5}$, NO, NO_2, SO_2, CO et O_3, qui permettent de calculer l'indice ATMO pour les agglomérations de plus de 100 000 habitants et l'indice IQA pour les autres villes. Le LCSQA fournit également les cadres nécessaires à la mise en application des lois et propose des Plans Climat Air Energie National et Territorial. En vertu de la loi de transition énergétique pour la croissance verte (TECV), un Plan Climat Air Energie Territorial (PCAET) doit être élaboré pour les territoires de plus de 20 000 habitants [37]. Cette loi, publiée au Journal Officiel le 18 août 2015, a pour objectif de permettre à la France de contribuer plus efficacement à la lutte contre le dérèglement climatique et à la préservation de l'environnement.

Le réseau de stations de mesure fixe permet d'établir une situation et de suivre l'évolution des efforts de réduction des émissions, ainsi que d'agir au niveau de la population

en cas de pics de pollution en mettant en place des dispositions préfectorales. Le LCSQA coordonne et normalise les données pour permettre aux AASQA de surveiller la qualité de l'air sur tout le territoire national et de prédire son évolution, leur permettant ainsi d'agir pour protéger la population.

2.2.2 Les Normes de la qualité de l'air

L'Organisation Mondiale de la Santé (OMS) et la Communauté Européenne fixent les normes de la qualité de l'air. Les directives européennes sont la référence réglementaire légale. Les directives de l'OMS sont complémentaires et recommandées sans pour autant être réglementaires. Les enjeux sont sanitaires et elles sont définies afin de protéger la population notamment les personnes vulnérables. Les PM_{10}, $PM_{2.5}$, NO_2, SO_2 et O_3 sont étroitement surveillées pour leur effet de nuisance. En effet, rien que pour les particules fines, 45 120 décès prématurés sont recensés en 2013 selon Santé publique France. La Figure 2.2 présente un extrait du tableau des normes « Qualité de l'air ». En rouge, l'origine des valeurs seuils, car les normes peuvent être aussi nationale. Les objectifs de qualité sont les valeurs à atteindre à long termes. La valeur cible est un objectif à atteindre selon un délai défit. La valeur limite correspond à la valeur ou une concentration plus élevée constitue une menace pour la santé ou l'environnement.

PARTICULES (PM_{10})		
Objectif de qualité	30 µg/m³ **(FR)**	en moyenne annuelle
Valeurs limites pour la protection de la santé humaine	50 µg/m³ **(UE)**	en moyenne journalière à ne pas dépasser plus de 35 jours par an
	40 µg/m³ **(UE)**	en moyenne annuelle
Seuil d'information et de recommandation	50 µg/m³ **(FR)**	en moyenne sur 24 heures
Seuil d'alerte	80 µg/m³ **(FR)**	en moyenne sur 24 heures

PARTICULES ($PM_{2.5}$)		
Objectif de qualité	10 µg/m³ **(FR)**	en moyenne annuelle
Valeur cible pour la protection de la santé humaine	20 µg/m³ **(FR)**	en moyenne annuelle
Valeur limite 2015 pour la protection de la santé humaine	25 µg/m³ **(UE)**	en moyenne annuelle

FIGURE 2.2 – Extrait du tableau des normes sur la qualité de l'air [60].

Le cadre légal définit différents seuils de qualité de l'air, qui se divisent en seuils d'information et de recommandation pour la prévention de la pollution, ainsi que des seuils d'alerte qui entraînent des mesures d'urgence. Ces normes sont en place pour protéger la population et l'environnement, et sont utilisées par les autorités locales pour prendre des mesures, telles que l'encouragement des transports en commun et la réduction des activités industrielles.

Ces seuils servent également à calculer l'indice de la qualité de l'air, qui permet de connaître rapidement la qualité de l'air dans une région donnée. L'indice ATMO, utilisé en France, a été harmonisé avec celui de l'agence européenne pour l'environnement depuis le 1er janvier 2021, et comprend désormais les particules fines PM_2.5. L'indice est divisé en six classes pour faciliter la compréhension de la qualité de l'air par les citoyens.

		Bon	Moyen	Dégradé	Mauvais	Très mauvais	Extrêmement mauvais
Moyenne journalière	PM2.5	0-10	10-20	20-25	25-50	50-75	>75
Moyenne journalière	PM10	0-20	20-40	40-50	50-100	100-150	>150
Max horaire journalier	NO2	0-40	40-90	90-120	120-230	230-340	>340
Max horaire journalier	O3	0-50	50-100	100-130	130-240	240-380	>380
Max horaire journalier	SO2	0-100	100-200	200-350	350-500	500-750	>750



FIGURE 2.3 – Extrait du tableau des normes sur la qualité de l'air [35].

Le calcul de l'indice ATMO est effectué quotidiennement à partir des concentrations dans l'air de cinq polluants réglementés par l'Union européenne : le dioxyde de soufre SO_2, le dioxyde d'azote NO_2, l'ozone O_3 et les particules de diamètre inférieur à 10 micromètres PM_{10} et 2,5 micromètres $PM_{2.5}$. Il vise à donner une vue synthétique de la pollution de fond dans des conditions quotidiennes. Les AASQA calculent cet indice à partir de données de modélisation qui intègrent la météo, d'autres plates-formes de modélisation de la pollution, l'inventaire régional des émissions et les mesures des stations de fond de la qualité de l'air.

Il convient de noter que l'indice ATMO ne prend pas en compte les mesures influencées par la pollution, telles que le trafic et l'activité industrielle. Le modèle utilisé est un modèle méso-échelle, ce qui signifie que l'indice ATMO est déduit à partir d'un maillage caractérisé par des niveaux de pollution et est croisé avec la situation géographique de la commune ou de l'intercommunalité.

2.2.3 Modèlisation de la qualité de l'air : l'approche descendante et assimilation de données

La modélisation est utilisée pour étudier la qualité de l'air à l'échelle régionale, notamment en ce qui concerne la dispersion des polluants. Des modèles pour la prédiction de la qualité de l'air sont conçus pour simuler la dispersion et la transformation chimique des polluants atmosphériques sur une échelle allant de quelques mètres à plusieurs milliers de kilomètres. Il existe différents niveaux d'échelle en météorologie et dans la modélisation de la qualité de l'air :

1. Micro-échelle : de l'ordre de mètres à quelques kilomètres (par exemple, tourbillons de vent) ;
2. Méso-échelle : de quelques kilomètres à 1000 km (comme les orages ou les systèmes convectifs de méso-échelle) ;
3. Synoptique ou grande échelle : de l'ordre de 1000 à 5000 km (comme les dépressions ou les anticyclones).

Les modèles de méso-échelle prennent en compte les interactions complexes entre la surface terrestre et l'atmosphère, ainsi que les processus atmosphériques qui se produisent à cette échelle. Ce sont des modèles atmosphériques utilisés pour simuler ou prévoir le temps et les écoulements sur une échelle intermédiaire. L'échelle de méso-échelle couvre

des distances d'environ 1 à 1000 kilomètres. Ces modèles sont particulièrement utiles pour prédire les phénomènes météorologiques à cette échelle, tels que les orages, les tempêtes locales ou les brises marines. Ces modèles sont également primordiaux pour évaluer les problèmes de pollution tels que les épisodes de smog ou l'ozone troposphérique. L'indice ATMO permettant de définir la qualité de l'air est déduit du maillage caractérisé par les niveaux de pollution et est croisé avec la situation géographique comme le montre la Figure 2.4.

Adoptant une perspective descendante, ces modèles fournissent un éclairage sur la manière dont les variations à grande échelle influencent les conditions locales. Autrement dit, ces modèles partent d'une vue synoptique (par exemple, d'échelle régionale) pour arriver à une résolution plus locale. Ils commencent par des conditions initiales et de bord à une grande échelle et se précisent progressivement pour obtenir des prédictions à une échelle plus fine. Comme le montre la Figure 2.4, le modèle méso-échelle est utilisé comme point de départ pour créer un maillage plus détaillé. Lorsque l'on effectue un zoom sur une région couverte par un maillage de 1 km, des cellules de 100 mètres sont définies à l'intérieur de ces cellules plus grandes pour une meilleure caractérisation des variations de la qualité de l'air dans des zones spécifiques.

L'assimilation de données est une technique essentielle pour la pertinence des modèles méso-échelle. En intégrant les observations réelles, comme les concentrations mesurées de polluants, l'assimilation de données optimise les prédictions du modèle. De plus, elle facilite une mise à jour en temps réel des prévisions en fonction des changements environnementaux et sert d'outil de validation pour s'assurer de la justesse du modèle [46]. L'interpolation est utilisée dans les modèles méso-échelle de qualité de l'air pour estimer les valeurs de certaines variables (par exemple, les concentrations de polluants) à des points non échantillonnés à partir d'observations ou de données à des points échantillonnés. Plusieurs méthodes d'interpolation peuvent être employées, et le choix de la méthode dépend souvent de la nature des données et des spécificités du problème à traiter.

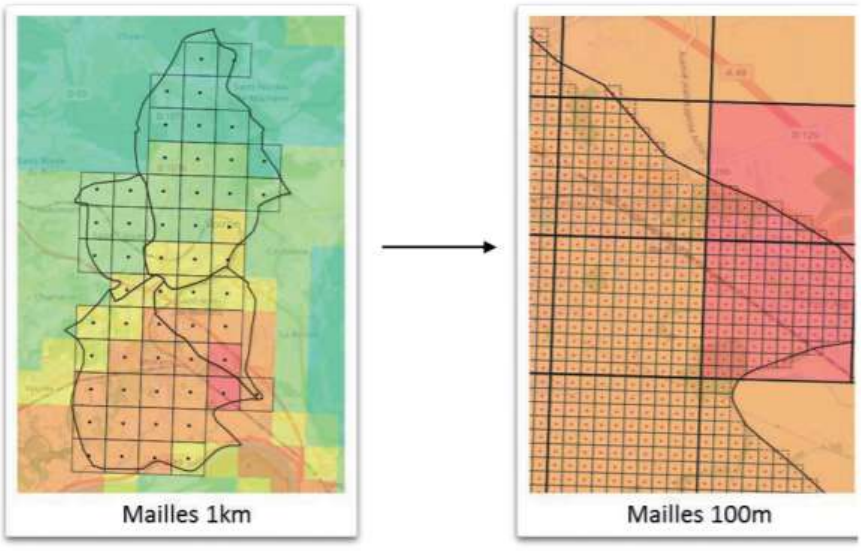


FIGURE 2.4 – Calcul de l'indice à partir d'un modèle méso-échelle [35].

La modélisation méso-échelle permet de comprendre et de prédire la dispersion et la transformation des polluants à différentes échelles. Son efficacité dépend également de la qualité et de la précision des données recueillies. Le système de surveillance de la qualité

de l'air en France est organisé à plusieurs niveaux pour garantir la qualité des données et les intégrer dans des politiques publiques et des actions régionales spécifiques. Au niveau international, l'Organisation Mondiale de la Santé (OMS) établit des lois et des réglementations. Au niveau européen, des directives sont établies pour les États membres. Au niveau national, la France est divisée en régions, ce qui permet de surveiller et de distribuer les responsabilités et les missions. Les AASQA remplissent ce rôle au niveau régional et départemental. Ils sont les acteurs locaux de la surveillance de l'air et fournissent les références scientifiques pour toutes les actions effectuées par les préfectures.

2.2.4 Les AASQA

La loi sur l'Air et l'Utilisation Rationnelle de l'Énergie (LAURE) du 30 décembre 1996 est à l'origine de l'agrément des AASQA créées en 1970 pour la surveillance de la qualité de l'air. Elle vise « le droit à chacun de respirer un air qui ne nuise pas à sa santé ». Ces associations ont pour mission :

— de surveiller et prévoir la qualité de l'air ;
— d'informer et sensibiliser la population ;
— d'accompagner les acteurs locaux ;
— de contribuer à l'innovation sur les territoires.

Ce sont des associations de loi 1901 agrées par l'état [12], indépendantes financièrement et dirigées de façon quadripartite (État, Collectivités, Entreprises, Associations et personnalités qualifiées santé-environnement).

Au nombre de 18 (une par région), elles sont au plus près des acteurs locaux et permettent de partager des diagnostics à l'échelle de la région. Les AASQA sont des acteurs de la transition écologique au plus près des citoyens et des décideurs, présents durablement dans les territoires et en mesure d'évaluer et de suivre, de manière indépendante, l'efficacité des plans d'actions sur la qualité de l'air [3]. Les enjeux sont croisés (air, climat et énergie) et traitent aussi la qualité de l'air extérieur et intérieur. Les données fournies par ces dernières se veulent être fiables et transparentes pour servir de références. Une normalisation des informations est faite pour construite un indicateur de la qualité de l'air à l'échelle nationale (indice ATMO). ATMO Occitanie est l'AASQA de la région Occitanie. Elle surveille, entre autres, le département des Hautes-Pyrénées où est située la CCPVG.

2.2.5 Le réseau de surveillance d'ATMO Occitanie

ATMO Occitanie surveille la qualité de l'air de 13 départements dont 2 grandes métropoles : Toulouse et Montpellier. La région Occitanie est la 1ère région viticole et la 2ème région agricole de France. Elle est composée d'un littoral méditerranéen, un massif montagneux (les Pyrénées) et des territoires ruraux. ATMO Occitanie accompagne les acteurs locaux pour concilier développement économique, développement démographique et qualité de l'air. C'est à travers le Programme Régional de Surveillance de la Qualité de l'Air (PRSQA) que les règlementations européennes et nationales sont suivies au niveau régional. Grâce aux suivis de plans et programmes territoriaux Plans Climat Air Energie Territorial (PCAET), Plans de Déplacements Urbains (PDU), Plans de Protection de l'Atmosphère (PPA), Schémas de Cohérence Territoriale (SCOT)), elle accompagne les collectivités pour lutter contre la pollution et l'air et assurer leur transition énergétique. La Figure 2.5 illustre les différents plan et programmes en liaison avec la loi LAURE et

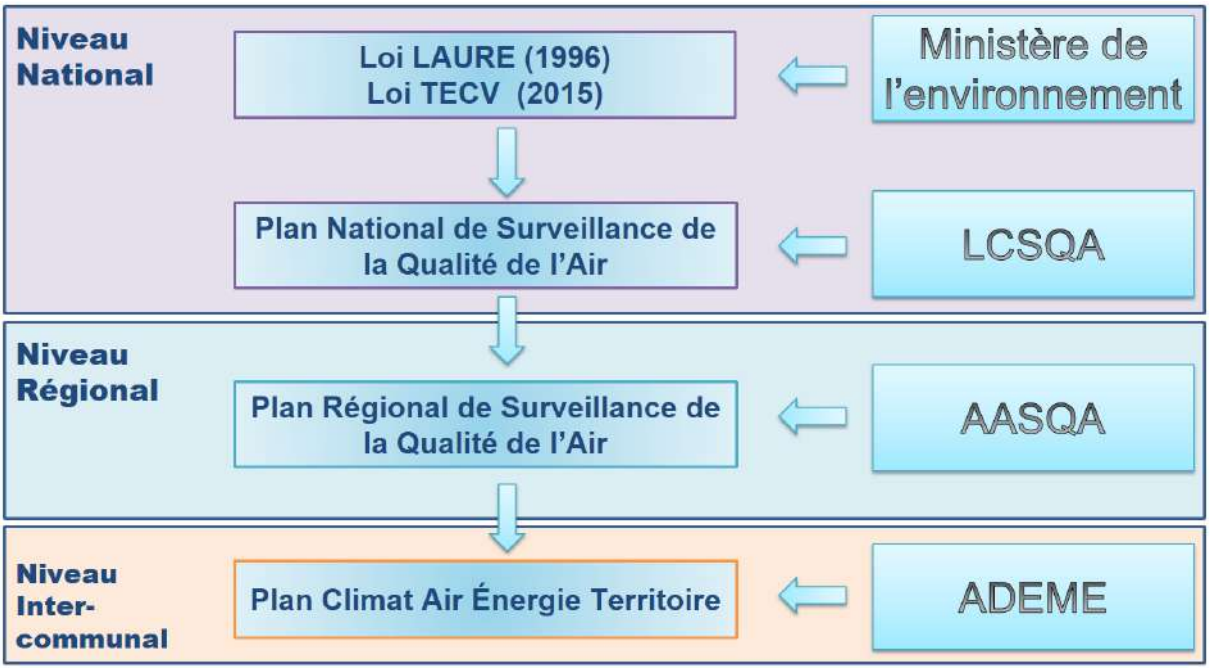


FIGURE 2.5 – Le cadre réglementaire de la surveillance de la qualité de l'air [3].

du niveau géographique.

La surveillance de la qualité de l'air est basée sur trois outils complémentaires et indépendants :

1. **L'inventaire des émissions** qui localise géographiquement les activités sources de pollution sur le territoire ;
2. **La modélisation** pour la prévision et fournir une cartographie de la pollution à différentes échelles dans le temps ;
3. **Le dispositif de mesures fixes ou mobiles** sur le territoire permet de mesurer la pollution temps réel.

Ainsi, la première mission de l'AASQA est de surveiller les polluants au niveau régional grâce aux mesures en temps réel et à l'inventaire régional des émissions. Le découpage de la région en département permet de suivre l'évolution de la pollution par département.

L'inventaire régional des émissions

Il s'inscrit dans le cadre de l'arrêté du 24 août 2011 relatif au Système National d'Inventaires d'Emissions et de Bilans dans l'Atmosphère (SNIEBA) et au Pôle de Coordination nationale des Inventaires Territoriaux (PCIT). Il est associé avec le Ministère en charge de l'Environnement, l'INERIS, le CITEPA et AASQA. ATMO Occitanie réalise et maintient à jour un Inventaire Régional Spatialisé des émissions de polluants atmosphériques et GES sur l'ensemble de la région Occitanie. L'inventaire des émissions référence une trentaine de substances avec les principaux polluants réglementés (NO_x, particules en suspension, NH_3, SO_2, CO, benzène, métaux lourds, HAP, COV, etc.) et les gaz à effet de serre (CO_2, $N2_O$, CH_4, etc.). Cet inventaire est notamment utilisé par les partenaires d'ATMO Occitanie comme outil d'analyse et de connaissance détaillée de la qualité de l'air sur leur territoire ou relative à leurs activités particulières. Les quantités annuelles d'émissions de polluants atmosphériques et GES sont ainsi calculées pour l'ensemble de

la région Occitanie, à différentes échelles spatiales (EPCI, communes, etc.), et pour les principaux secteurs et sous-secteurs d'activité. La méthodologie de calcul des émissions consiste en un croisement entre des données primaires (statistiques socio-économiques, agricoles, industrielles, données de trafic, etc.) et des facteurs d'émissions issus de bibliographies nationales et européennes. Elle est régie par l'équation n°(2.1) avec :
E : émission relative à la substance « s » et à l'activité « a » pendant le temps « t » ;
A : quantité d'activité relative à l'activité « a » pendant le temps « t » ;
F : facteur d'émission relatif à la substance « s » et à l'activité « a » ;

$$E_{s,a,t} = A_{a,t} \times F_{s,a} \tag{2.1}$$

Voici un exemple pour mieux comprendre : Les émissions de PM_{10} « Es » dues au chauffage bois « a » pendant l'année « t » sont égaux aux nombres de fois que le chauffage est allumé « Aa » pendant l'année « t » multiplié par la masse de polluant émise par tonne de bois « Fs » servant au chauffage « a ». Le résultat est en Tonne par An (T/An).

L'inventaire des émissions permet d'alimenter les données pour la modélisation et il est également un outil de diagnostic et d'aide à la décision pour les politiques publiques (études d'impact, scénarisation et plan climat).

La modélisation

La modélisation est un outil permettant de simuler le transport des polluants mais aussi leur dispersion. A partir de diverses données, elle permet de cartographier la concentration des polluants de l'échelle régionale à l'échelle de la rue. La modélisation peut aussi prévoir les mouvements de pollution et donc la cartographier à une date ultérieure, c'est ce qui est communément appelé une prévision. La modélisation de la pollution permet alors :

— d'évaluer la situation annuelle de la pollution de l'air sur un territoire au regard de la réglementation et d'identifier les zones à enjeux ;
— d'évaluer l'exposition des populations ;
— de prévoir la qualité de l'air du jour et les jours suivants, et ainsi anticiper la survenue d'épisodes de pollution de l'air.

La Figure 2.6 illustre les données entrant dans le système **EtudeS Multi RégionALes De l'Atmosphère** (ESMERALDA). La plate-forme inter-régionale ESMERALDA résulte de l'étroite collaboration de neuf AASQA [39]. **CHIMERE** est un modèle déterministe de chimie-transport. C'est un code informatique qui réunit un ensemble d'équations représentant le transport et la chimie d'espèces chimiques et qui permet de quantifier l'évolution d'un panache de polluants en fonction du temps sur différents domaines (de l'urbain au continental). Le modèle de prévision numérique planétaire **ARPEGE** (Action de Recherche Petite Echelle Grande Echelle) est un élément essentiel pour la prévision opérationnelle du temps à Météo-France.

Ainsi, l'ADEME a développé **BASTER**, une base de données qui recueille en temps quasi réel les mesures réalisées par les AASQA. Une liaison directe est établie entre BASTER et les modèles de qualité de l'air utilisés dans le cadre du système **PREV'AIR**. En résumé, l'ADEME fournit BASTER, une base de données temps réel et recueille les mesures réalisées par les AASQA. Ensuite PREV'AIR interroge la base pour faire fonctionner ses modèles et fournit les cartographies aux AASQA. La modélisation ne peut se

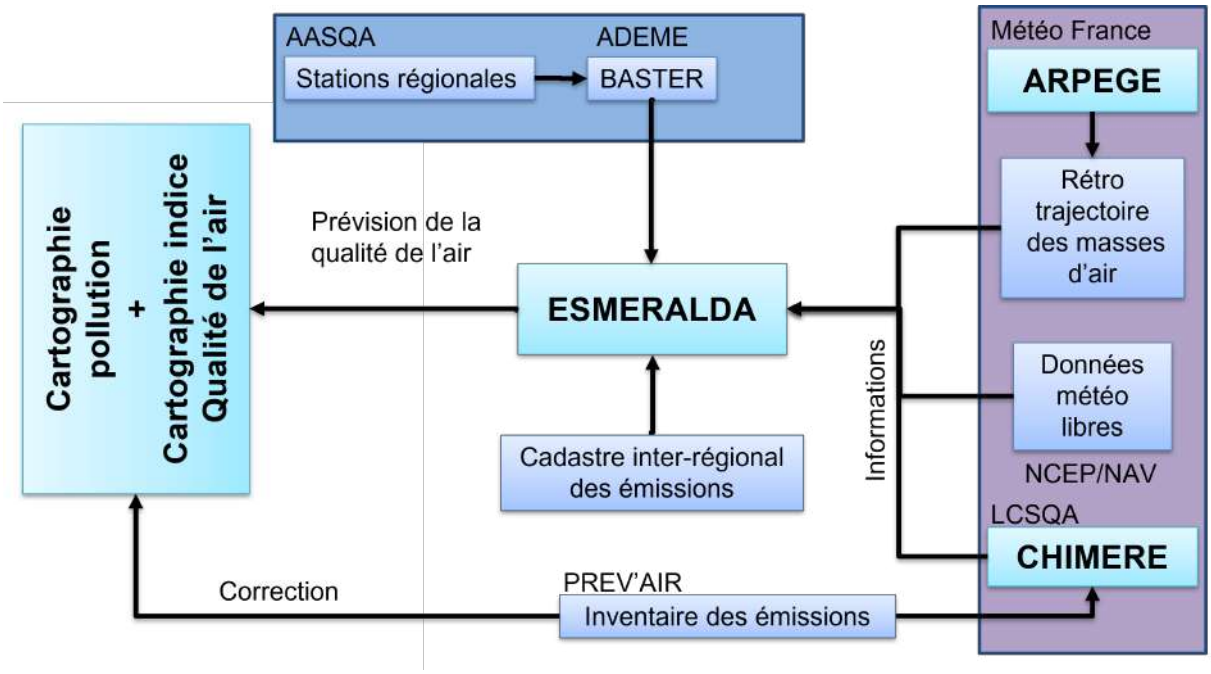


FIGURE 2.6 – Comment sont modélisées les cartes de la Qualité de l'Air [39].

faire sans les mesures des stations de la qualité de l'air, elles permettent notamment de corriger les cartes de pollutions et forment ce que l'on appelle une boucle de rétroaction.

Les stations de mesure fixes et mobiles

Au nombre de 44, les stations de mesure fixe évaluent la qualité de l'air sur l'ensemble de la région Occitanie [36]. Des stations de mesure mobiles sont aussi déployées pour couvrir ponctuellement certaines parties du territoire éloignées des stations ou sous influence d'une pollution à surveiller. En effet, les stations sont concentrées près des grandes villes (6 à Montpellier 16 à Toulouse, 3 à Nîmes et 3 à Perpignan). Les autres sont situées dans les préfectures de chaque département. Les stations de mesure fixes et mobiles sont installées par ATMO Occitanie. La mesure est continue ou bien temporaire. Elle se concentre sur les polluants surveillés par la réglementation en se focalisant notamment sur les zones urbaines et les agglomérations présentant une forte activité humaine. Le but est de surveiller et de vérifier si les valeurs de concentration sont respectées mais aussi d'alerter en cas de dépassement. Elles permettent de suivre l'évolution de la qualité de l'air dans le temps. La mesure peut être en direct ou différée (pluviométrie et analyse de particules de métal dans l'air). Les concentrations sont mesurées toutes les heures (en microgrammes par mètre cubes). Les polluants mesurés sont l'Ozone (O_3), les Oxydes d'Azotes (NO_x), les particules en suspension (PM_{10}) et les particules fines ($PM_{2.5}$).

Les détails techniques des stations de mesure fixe et mobile ont été décrits dans cette sous-section. La sous section suivante va traiter de la classification des stations de mesure en fonction du type de pollution à surveiller et de son lieu d'implantation.

2.2.6 Représentativité des stations de mesures

Une station de mesure de la qualité de l'air a pour objectif de contrôler la pollution atmosphérique et de surveiller les concentrations de polluants dans l'air ambiant. Qu'elles soient fixes ou mobiles (remorque ou camion-laboratoire), comme illustré dans la Figure 2.7, elles analysent en continu l'air. Les stations de mesure sont classées selon deux critères : l'environnement d'implantation et l'influence du type de pollution.



FIGURE 2.7 – Photographies d'une station fixe et d'une station mobile du réseau ATMO [36].

L'environnement d'implantation

L'environnement d'implantation définit le type de zone sur une échelle de plusieurs kilomètres [18]. Une station sera classifiée comme étant urbaine si l'environnement est une zone bâtie en continue (bâtiments de deux étages minimum). Une station sera périurbaine lorsque la zone n'est pas bâtie en continue (combinée avec des zones non construites). La station rurale est classée en trois sous zones :

- proche zone urbaine : elle est située dans les 10 km de la périphérie de la zone urbaine ;
- régionale : elle est éloignée entre 10 et 50 km d'une source de pollution ;
- nationale : elle est le plus éloignée possible de toute source de pollution (> 50 km).

Les stations rurales régionale participent à la surveillance de l'exposition de la population et des écosystèmes à la pollution atmosphérique de fond, notamment photochimique, à l'échelle régionale. Les stations nationales, implantées dans les zones rurales, participent à la surveillance de la pollution atmosphérique de fond issue des transports de masses d'air sur de longues distances et notamment à l'échelle nationale et continentale. Elles sont soumises à des directives (2004/107/CE (article 4, point 9) et 2008/50/CE (article 6, point 5)) très strictes et doivent répondre à leurs exigences. Selon les dispositions de la Directive

2008/50/CE (annexe III – § 2), « les points de prélèvement, visant à assurer la protection de la végétation et des écosystèmes naturels, sont implantés à plus de 20 km des agglomérations ou à plus de 5 km d'une autre zone bâtie, d'une installation industrielle, d'une autoroute ou d'une route principale sur laquelle le trafic est supérieur à 50 000 véhicules par jour ». L'objet de ces critères est de garantir une représentativité spatiale minimale (au moins 1000 km^2) des points de prélèvement.

Le type d'influence

Le type de pollution présent dans la zone de la station va déterminer l'influence sur la mesure [40]. La station sera dite sous **influence industrielle** lorsque la présence de polluants issue de l'industrie (centrale thermique, raffinerie, aéroport, construction, etc.) a une influence conséquente sur les concentrations. Elle sera sous **influence trafic** lorsque le point de prélèvement est à proximité d'un axe routier important. Enfin, elle sera classifiée de **fond** lorsqu'elle n'est pas soumise aux deux influences précédentes. Elle est représentative d'une zone d'au moins plusieurs km^2. Elle est principalement utilisée pour surveiller l'exposition moyenne de la pollution de fond de la population et/ou de la végétation. Il est possible qu'une station de mesure soit soumise à plusieurs influences. La Table 2.1 illustre les possibilités de classifications.

		Environnement d'implantation				
		Urbain	Péri-Urbain	Rural		
				Proche d'une zone urbaine	Régional	National
Type d'influence	Fond	U_F	PU_F	RP-F	RR_F	RN_F
	Trafic	U_F	PU_T	RP_T	**OS**	**OS**
	Industrielle	U_I	PU_I	RP_I	**OS**	**OS**

U : Urbain
RP : Rural Proche (de zone urbaine)
F : Fond
PU : Péri_Urbain
RR : Rural Régional
T : Trafic
OS : Observation spécifique
I : Industrielle
RN : rural National

TABLE 2.1 – Labélisation de station en fonction des combinaisons possibles pour le couple station/polluant [40].

Ces stations de mesures ont une classification et une labélisation en fonction de critères vus précédemment. Cependant, Il est important de noter qu'en fonction de leurs objectifs, des différences existent. En effet, surveiller une pollution de fond implique des mesures de concentrations spécifiques en comparaison à des stations de mesure sous influence du trafic. Ces stations sont soumises à des directives européennes et à des normes éditées par l'AFNOR.

2.2.7 Les Technologies de surveillance de la qualité de l'air

La Directive 2008/50/CE impose des méthodes de mesures normalisées pour chaque polluant surveillé. L'intérêt de ces méthodes est d'avoir une grande précision des concentrations mesurées pour les comparer aux objectifs de qualité de l'air (cf. 4.2 Normes de la QA). Elles sont appliquées à toutes les stations de mesure de la qualité de l'air gérées par les AASQA. Le but est d'avoir les mêmes réglages des analyseurs de gaz pour toutes les stations.

Les stations de mesure conventionnelles

Ces stations suivent la Directive 2008/50/CE. Elles doivent surveiller les gaz réglementés suivant des méthodes normalisées car c'est sur la base de ces mesures que les services de l'état, les collectivités, les médias et le public sont informés de la qualité de l'air (pic de pollution surtout). C'est aussi à partir de ces mesures que peuvent être déclenchées des procédures réglementaires destinées à améliorer la qualité de l'air : réduction des émissions industrielles, limitation de vitesse ou circulation alternée, etc. Les données de qualité de l'air sont exploitées au niveau national et international pour les missions suivantes :

— études épidémiologiques ;
— suivi de la pollution atmosphérique ;
— analyse au niveau régional ;
— transmission et exploitation par la commission européenne.

Directives et Normes Européennes

Les diverses applications des données mettent en évidence la nécessité de renforcer la fiabilité du système de surveillance, tant sur la durée que dans l'espace, pour assurer l'exactitude des mesures et leur pérennité. Cela permet à la population, aux autorités publiques, aux instances européennes, etc., de s'appuyer sur ces données de qualité de l'air en toute confiance. La Table 2.2 présente les méthodes d'analyse pour la surveillance réglementaire des polluants [40].

	Polluants					
	NO_x	O_3	SO_2	CO	PM_{10}	$PM_{2.5}$
Méthode de référence	EN 1421	EN 14625	EN14212	EN 14626	EN 12341	EN 14907
Principe de mesure	Chimiluminescence	Absorption UV	Fluorescence UV	Rayonnement IR non dispersif	Gravimétrie sur filtre	

TABLE 2.2 – Méthodes d'analyse pour la surveillance réglementaire des polluants [40].

La chaîne d'acquisition des polluants

Une station de mesure possède sur son toit des cannes de prélèvement de l'air automatique. Elles permettent de récupérer l'air ambiant au moyen d'une pompe et de l'acheminer aux différents analyseurs à l'intérieur de la station. Chaque analyseur mesure le polluant en suivant la méthode de référence lui correspondant. La fréquence de mesure est inférieure à l'heure, en effet, plusieurs mesures sont effectuées et elles sont ensuite analysées et transmisse toutes les heures à la station d'acquisition. La station d'acquisition récolte les données de toutes les stations de la région, les stocke et les transmet via la 4G toutes les heures au bureau des techniciens. Ce bureau valide une première fois les données et écarte

les incohérences géographiques et typologique de la station. Une validation complémentaire est effectuée par le bureau des ingénieurs pour le suivi de l'évolution de la qualité de l'air. Ensuite le pôle ATMO, où sont regroupé d'autre ingénieurs, vérifie la conformité des mesures avec les réglementations européennes et alerte suivant les dépassement les pouvoirs concernés. Ils éditent également les bilans de la qualité de l'air, les modélisations de la pollution, la cartographie à l'échelle régionale et urbaine ainsi que l'inventaire des émissions. La Figure 2.8 illustre la chaîne de mesure.

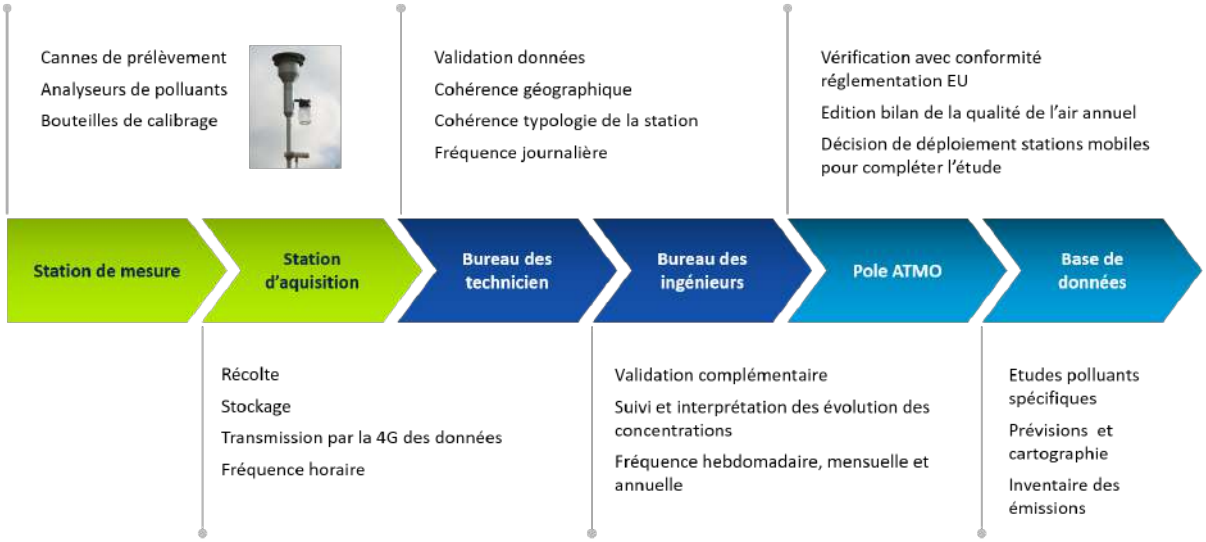


FIGURE 2.8 – Chaîne de mesure de la qualité de l'air pour les AASQA [40].

Les réglementations européennes et les normes alourdissent la chaîne de mesure, et l'ajout de nouvelles stations de mesure est difficile et coûteux. Lorsque la surveillance de la qualité de l'air ne nécessite pas une précision aussi élevée que celle des stations traditionnelles (niveaux de concentration plutôt que mesures réglementaires), des alternatives de mesures complémentaires peuvent être envisagées, à savoir, les réseaux de micro-capteurs (LCS).

Les réseaux de LCS

Les stations conventionnelles, a cause de leur coût et de leur manque de mobilité et malgré le déploiement de stations mobiles, ne peuvent répondre à un besoin croissant des points de prélèvements pour surveiller la qualité de l'air localement. Des initiatives citoyennes commencent à émerger grâce aux progrès technologiques et la disponibilité des capteurs à faible coût. Ces réseaux de capteurs permettent de cartographier la pollution avec des niveaux de pollution en temps réel et avec une densité spatiale augmentée au regard des stations traditionnelles [29]. Dans ce qui suit sont donnés quelques exemples de projets menés par différentes AASQA :
ATMOSUD déploie 5 stations de mesure composées de micro-capteur pour affiner les cartes de pollution fournie par elle –même. La Figure 2.9 illustre la précision spatiale apportée par les mesures provenant des LCS. La cartographie annuelle de la pollution en PM_{10} située à droite de la Figure bénéficie d'une précision accrue par rapport à celle de gauche grâce à l'installation de LCS, représentés par des points noirs.

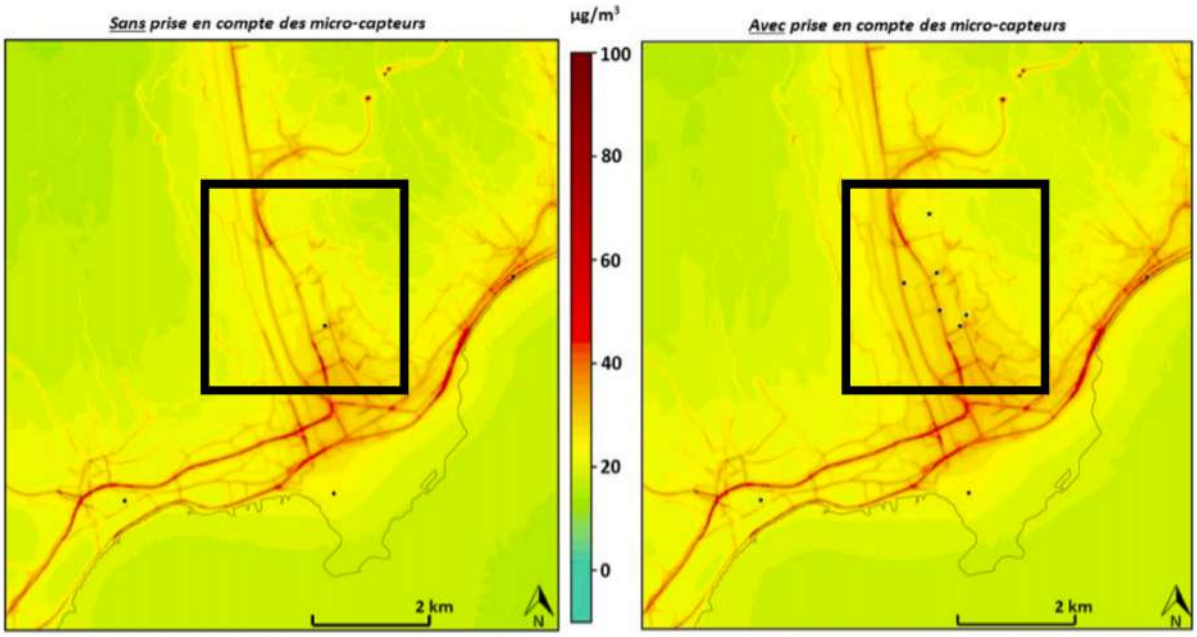


FIGURE 2.9 – Cartographie annuelle de la pollution en PM_{10} avec et sans micro-capteurs (www.atmosud.org).

ATMO Auvergne-Rhône-Alpes a mené une campagne de sensibilisation auprès des citoyens à l'aide de micro-capteurs : le projet Mobicit'air illustré dans la Figure 2.10. Elle permet à 30 volontaires de mesurer la qualité de l'air avec un dispositif portable sur une période de 18 mois. Ce projet citoyen vise à sensibiliser sur la qualité de l'air et rassembler autour de la réduction de pollution.



FIGURE 2.10 – Projet citoyen Mobicit'air (ATMO-auvergnerhonealpes.fr).

ATMO Hauts-de-France équipait 50 volontaires de la région avec des micro-capteurs, dans le cadre du projet SoNumAir pendant 15 jours. Le bilan de cette étude est que les citoyens contournent le problème : ils diminuent leur exposition à la pollution mais ne cherchent pas à réduire la production.

ATMO Grand Est connecte des micro-capteurs à son réseau de surveillance de la pollution afin de mieux détecter les pics de pollutions et de réguler le trafic en conséquence.

Les réseaux de capteurs à bas coûts sont une réelle opportunité pour la surveillance de la qualité de l'air et sont de plus en plus utilisé par les AASQA pour une complémentarité du réseau traditionnel de mesure. La section suivante est dédiée aux micro-capteurs, leur définition et les opportunités qu'ils peuvent offrir pour la surveillance de la qualité de l'air. Certaines de leurs limites seront également abordées.

2.3 État de l'art sur les micro-capteurs ou LCS

Les progrès actuels technologiques et la disponibilité des micro-capteurs permettent aux initiatives citoyennes de surveiller la qualité de l'air à faible coût [29]. Ils permettent d'augmenter la densité de mesure spatio-temporelle pour en faire des cartes de gradient de pollutions en temps réel à l'échelle locale. Dans cette section, les différentes technologies de micro-capteur seront présentées. Ensuite, l'état de l'art de la surveillance de la qualité de l'air avec de tels dispositifs sera abordé pour enfin discuter de leur limites.

2.3.1 Définition

Selon le dictionnaire Larousse, un **capteur** est *un organe qui élabore, à partir d'une grandeur physique, une autre grandeur physique, souvent de nature électrique, utilisable à des fins de mesure ou de commande*. De façon générale, un capteur converti la grandeur physique à mesurer appelée également mesurande en un signal électrique de mesure pour des traitements numériques ultérieurs (Figure 2.11).



FIGURE 2.11 – Principe de fonctionnement d'un capteur.

De manière générale, un capteur doit répondre à plusieurs caractéristiques correspondant à sa qualité métrologique :

- étendue de mesure : gamme de mesure du capteur (ex : de -10 V à +20 V son étendue est de 30 V) ;
- résolution : plus petite variation de la grandeur mesurée ;
- sensibilité : variation du signal en sortie en fonction de l'entrée ;
- exactitude : coïncidence avec la valeur dite « vraie » on la définit à travers l'erreur de mesure ;
- justesse : erreur globale résultant d'un grand nombre de mesure ;
- fidélité : capacité à donner des mesures sans erreurs accidentelles ;
- répétabilité : fidélité de mesure lorsque la même méthode est utilisée, dans les mêmes conditions, par le même opérateur en utilisant le même équipement pendant un court laps de temps ;

— reproductibilité : fidélité de mesure lorsque la même méthode est utilisée, dans des conditions différentes, par différents opérateurs avec des équipements différents.

Ces caractéristiques sont utiles pour la comparaison qualitative des capteurs entre eux. D'autres caractéristiques s'ajoutent également telles que la fiabilité et la qualité qui sont déterminants pour la disponibilité du capteur. La Figure 2.12 illustre les différences entre fidélité, justesse et exactitude. La définition du micro-capteur ou LCS, provient essentiellement du fait que le prix du capteur est à minima inférieur à 10% du prix d'un capteur de référence utilisé par les stations traditionnelles [57]. Les micro-capteurs de mesure de la pollution sont ceux qui seront utilisés pour le déploiement de l'architecture. Ils se déclinent en deux types de capteurs : les capteurs de gaz et les capteurs de particules.



FIGURE 2.12 – Illustration des critères fidélité, justesse et exactitude.

2.3.2 Les capteurs de particules

C'est un capteur optique (Optical Particle Counter) basé sur la diffraction d'une lumière Infrarouge. Un flux d'air forcé passe à travers le capteur grâce à un mini ventilateur. Les particules en passant devant la lumière vont diffracter une partie du faisceau vers le détecteur photosensible. L'intensité de la diffraction est liée à la taille de la particule mais aussi à sa densité. Par un calcul, on en déduit la concentration en micro-grammes par mètre cubes (µg/m3).

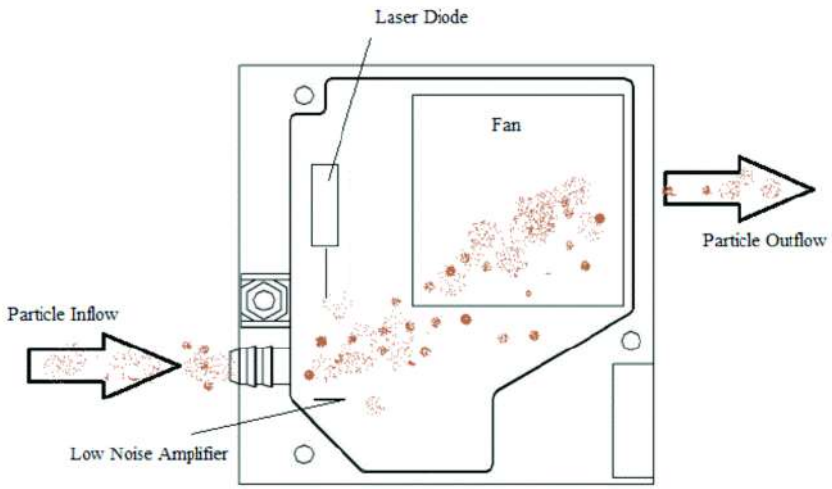


FIGURE 2.13 – Fonctionnement du capteur de poussière SDS011 [18].

Un exemple type de capteur de poussière est le SDS011 (Figure 2.13) utilisant un laser pour mesurer la concentration. Ils peuvent détecter les PM_{10} et $PM_{2.5}$. Cependant, ils sont sensibles aux conditions météorologiques (brouillard, vent fort). Ils peuvent être saturés lorsqu'ils sont placés trop près de sources de poussières (carrières, routes non goudronnées). Ils sont aussi sensibles à la condensation.

2.3.3 Les capteurs de gaz

Il existe quatre technologies de capteurs de gaz, électrochimique, semi-conducteurs, à infrarouge et détecteur à photoionisation (PID). Tous renvoient en valeur une concentration en ppm ou en µg/m3. Ils sont utilisés pour le suivi qualitatif de l'air, c'est-à-dire que la précision et les interférences entre gaz polluant importe peu puisque l'objectif est de mesurer la globalité. Les méthodes de mesures sont plus simples car ces capteurs utilisent des propriétés physiques et/ou des réactions chimiques.

Les capteurs électrochimiques

Le principe de fonctionnement se base sur une réaction électrochimique. Le gaz transmet des charges électriques à l'électrolyte solide, ce qui va induire une différence de potentiel ou un courant en fonction de la concentration. La Figure 2.14 détaille le fonctionnement d'un tel capteur.

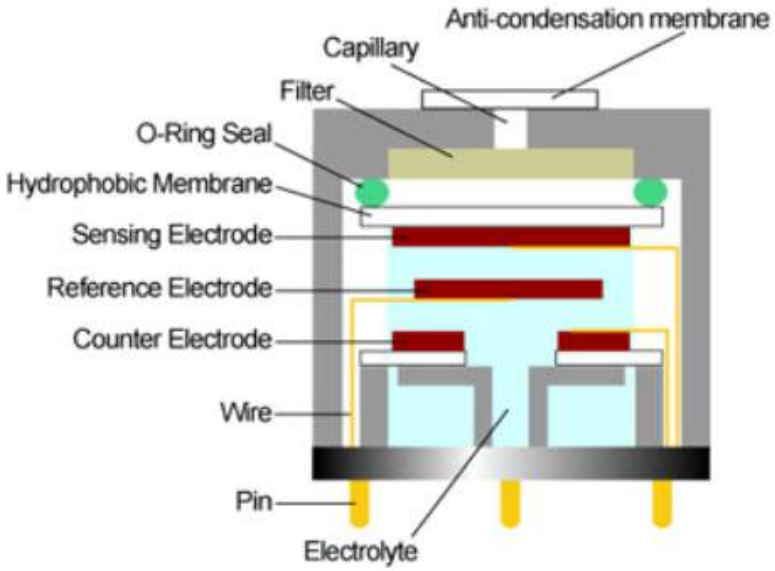


FIGURE 2.14 – Fonctionnement d'un capteur électrochimique [16].

Le principal problème de ces capteurs est que des gaz peuvent interférer dans la mesure. Ils ont tendance à dériver à long terme car l'électrolyte est consommé par la réaction d'oxydo-réduction. Le choix des matériaux du capteur peut amener à des détections de faibles concentrations mais augmente le coût.

Les capteurs semi-conducteurs MO_x

Appelé aussi capteur métal oxyde, son fonctionnement est basé sur le changement de conductivité de cette couche sensible lors de l'exposition au gaz. La résistance de la

couche d'oxyde sera mesurée et permet d'en déduire la concentration du gaz. Il est à noter qu'une résistance de chauffage est intégrée afin de cibler le gaz mesuré. Il doit donc être préchauffé pour effectuer la mesure. La Figure 2.15 présente le fonctionnement d'un tel capteur.

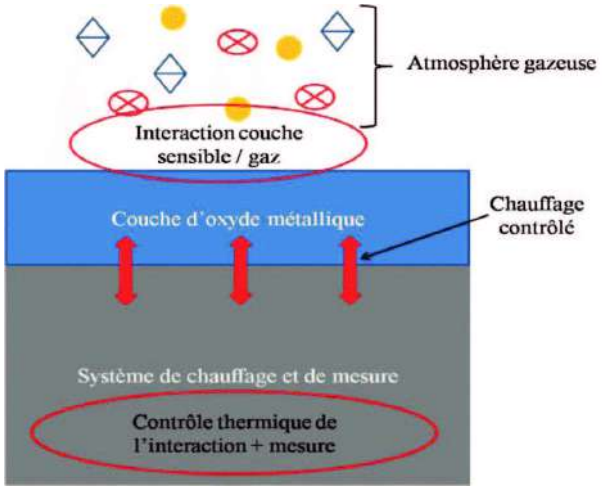


FIGURE 2.15 – Principe de fonctionnement d'un capteur MO_x[76].

C'est un capteur très sensible mais il peut interférer avec d'autres gaz s'ils ne sont pas filtrés. Une résistance chauffante est nécessaire à son fonctionnement ce qui augmente sa consommation d'énergie. Il a tendance à accumuler le gaz mesuré ce qui entraîne une dérive dans le temps. Il est également sensible à la température et à l'humidité.

Les capteurs à infrarouge (Absorption IR, NDIR)

La détection de certains gaz s'appuie sur leur capacité à absorber certaines longueurs d'onde notamment dans le domaine de l'infrarouge. Cette atténuation d'énergie du faisceau laser est mesurée et est proportionnelle à la concentration (loi de Beer-Lambert). la Figure 2.16 illustre le fonctionnement à l'intérieur du boîtier.

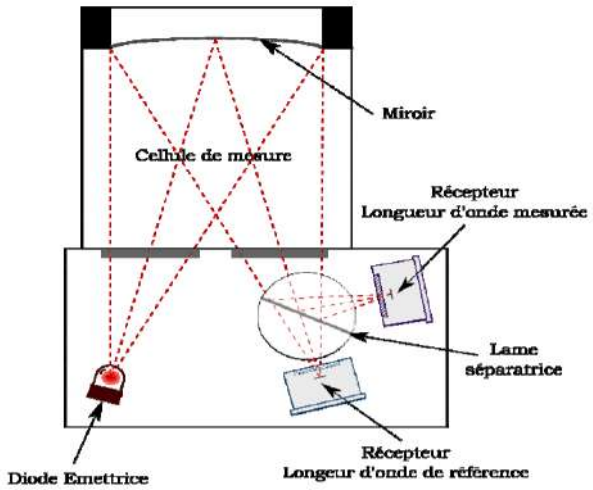


FIGURE 2.16 – Principe de fonctionnement capteur IR [8].

Les capteurs à infrarouge sont peu sensibles à la température et à l'humidité, et ils sont principalement utilisés pour mesurer la concentration de CO_2 en raison de leur précision exceptionnelle. Cependant, il devient difficile de distinguer quels gaz sont mesurés (sélectivité) lorsqu'il s'agit de mesurer des concentrations d'hydrocarbures.

Les capteurs PID ou détecteur à photo-ionisation

Un flux d'air forcé amène l'air dans une chambre d'ionisation avec une lampe UV et de deux électrodes (Figure 2.17). Le but est de créer des ions à partir de l'intensité du rayonnement de la lampe UV. Un courant est créé et il est proportionnel à la concentration des molécules ionisées.

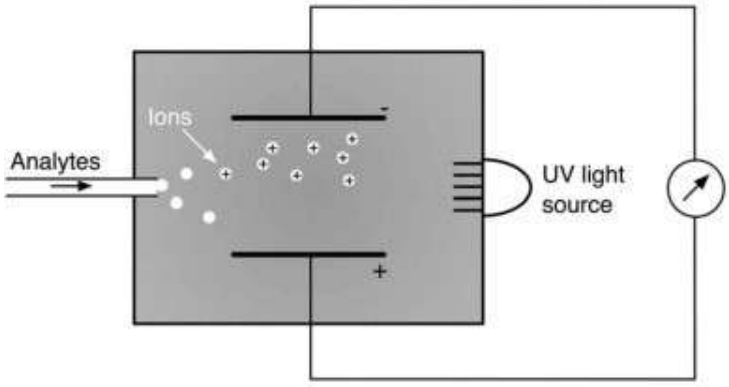


FIGURE 2.17 – Principe de fonctionnement capteur PID [50].

La mesure par PID dépend de l'ionisation de certains gaz. Il est utilisé notamment pour les COV pour sa sensibilité. A cause du vieillissement de la lampe, il subit une dérive.

Les micro-capteurs de gaz et de particules permettent de mesurer la concentration de polluants atmosphériques. Ils sont basés sur plusieurs technologies, dont les plus utilisées sont les électrochimiques, optiques et thermiques. Les micro-capteurs présentent un certain nombre d'avantages, notamment leur faible coût, leur taille compacte et leur facilité d'utilisation. Cependant, ils ne sont pas aussi précis que les instruments de mesure traditionnels. Malgré leurs limites, les micro-capteurs présentent un certain potentiel pour la surveillance de la qualité de l'air, notamment en termes d'amélioration de la couverture.

2.3.4 Surveillance de la qualité de l'air avec des micro-capteurs

Actuellement, la majorité des mesures de la pollution proviennent de stations fixes situées dans des endroits ciblés tels que les aéroports, les zones militaires ou les universités. Le réseau de mesure est occasionnellement renforcé par des stations de mesure mobiles. Cependant, leur densité est très faible par rapport à l'étendue du territoire à couvrir. Les capteurs à bas coût offrent la possibilité d'accroître la densité des mesures spatiales. La question soulevée est de savoir si ces capteurs représentent une véritable opportunité pour la surveillance de la qualité de l'air, et plusieurs travaux de recherche sont présentés en vue d'y répondre.

Les études portant sur les micro-capteurs

Plusieurs articles scientifiques relatent le développement de stations de mesure de la qualité de l'air, notamment le déploiement de capteurs à bas coût. La question principale est de savoir si ces derniers peuvent apporter une plus-value par rapport aux stations de mesure traditionnelles. La littérature scientifique a été parcourue et six études sur ces capteurs ont été sélectionnées, puis synthétisées dans la table 2.3.4. Les critères de sélection sont établis en se basant sur des études qui explorent le déploiement sur le terrain des capteurs, dans le but d'observer leur comportement réel et les éventuelles solutions mises en place. Ces études concentrent les avantages de l'utilisation des micro-capteurs sur les points suivants :

- déploiement aisé ;
- possibilité de mesure temps réel, précision temporelle ;
- idéal pour niveaux de concentration et calcul indice QA ;
- adapté pour la surveillance de pics de pollution ;
- densité de mesure dans le temps et l'espace excellente ;
- maintenance faible ;
- faible coût (rapport x10 à x100).

Les principaux inconvénients trouvés sont :

- incertitudes et répétabilité ;
- qualité variable d'un capteur à l'autre ;
- reproductibilité des mesures (mesures différentes dans le même environnement) ;
- interférences avec d'autres gaz (la concentration mesurée ne sera pas uniquement celle du gaz attendu) ;
- dérive dans le temps de certains capteurs (due à leur usure) ;
- précision ne permettant pas d'égaler celle des stations de mesure traditionnelles.

La qualité métrologique des LCS est discutable mais des solutions sont disponibles pour les contourner et offrir une réelle alternative. La Table 2.3.4 présente une sélection des études les plus pertinentes pour améliorer les performances des réseaux de LCS.

Réf	Titre de l'article	Contributions
[32]	Evaluating the performance of low cost chemical sensors for air pollution research	Les LCS sur étagère varient en qualité et les interférences croisées à faibles concentrations posent des problèmes, tout comme leur mauvaise reproductibilité. Les méthodes statistiques et les capteurs en grappe offrent des solutions potentielles pour résoudre ces problèmes.
[9]	Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates ?	Obtenir des capteurs robustes et reproductibles est un défi. Il est nécessaire de les évaluer sur le terrain dans différentes conditions environnementales. Les capteurs peu coûteux ne conviennent pas aux applications nécessitant une grande précision. Cependant, ils peuvent fournir des informations générales sur la qualité de l'air.
[44]	Applications of low-cost sensing technologies for air quality monitoring and exposure assessment : How far have they gone ?	Examen approfondi des technologies de détection à faible coût pour la surveillance de la qualité de l'air. Les LCS conviennent à des applications spécifiques, mais des travaux supplémentaires sont nécessaires pour exploiter leur plein potentiel. Les capteurs de $PM_{0.1}$ à bas coût ne sont pas encore disponibles, et il n'y a aucune corrélation entre les particules $PM_{0.1}$ et les $PM_{2.5}$.
[64]	An improved low-power measurement of ambient NO2 and O3 combining electrochemical sensor clusters and machine learning.	En combinant l'utilisation de capteurs et en analysant ces données grâce au Machine Learning (ML), on peut approcher la précision des instruments étalons, comme l'indique la valeur de l'Erreur Quadratique Moyenne en racine (RMSE). Plus la durée de collecte de données est étendue, plus les prédictions du ML tendent à coïncider avec les mesures de ces instruments de référence.
[42]	The use of electrochemical sensors for monitoring urban air quality in low-cost, highdensity networks.	Cet article démontre la faisabilité des réseaux de LCS de type electrochimique avec des regroupements et évolutifs avec des post-traitements pour diverses périodes allant jusqu'à 2,5 mois et à bas coût. Les données ont une densité de mesure (dans le temps et dans l'espace) impossible à atteindre avec les méthodes de mesure standard actuelles.
[4]	Development of a normalized multisensors system for low cost online atmospheric pollution detection.	Des techniques d'étalonnage et de normalisation ont été développées pour corriger les écarts de référence et de sensibilité, garantissant ainsi une réponse homogène des capteurs. Cette approche permet d'informer le public en temps réel et avec une grande précision spatiale sur la pollution en combinant les prévisions et les indices de pollution.

TABLE 2.3 – Une sélection d'études portant sur les micro-capteurs.

Les solutions de la littérature

Ces études révèlent des solutions de pré- et post-traitement permettant de compenser les incertitudes et les problèmes de répétabilité associés à ces capteurs. Les solutions vont de l'emplacement des capteurs aux solutions avec des algorithmes.

1. Placement des capteurs :
   — regrouper les capteurs et traiter les données avec le ML. La période de collecte de données doit être longue pour que ces algorithmes prédisent avec précision les observations ;
   — utiliser de manière adéquate les capteurs électrochimiques pour fournir une sensibilité élevée avec un bruit faible ;
   — mise au point de techniques d'étalonnage et de normalisation pour corriger les écarts de référence et de sensibilité afin d'assurer une homogénéité dans la réponse des capteurs.
2. Fiabilité des capteurs :
   — une évaluation des performances sur le terrain dans diverses conditions environnementales est nécessaire ;
   — l'utilisation du ML espacerait le temps requis pour les calibrations physiques et fiabilise les données une fois le capteur calibré dans son environnement ;
   — le regroupement de capteurs combiné à un post-traitement améliore considérablement les résultats mais aussi leur disponibilité.
3. Reproductibilité des mesures :
   — les capteurs en grappe peuvent améliorer considérablement la reproductibilité.
4. Interférences avec d'autres gaz :
   — les méthodes statistiques offrent des perspectives considérables pour supprimer les interférences, si elles sont sauvegardées par le laboratoire.

Ces capteurs à bas coût sont des opportunités pour surveiller la qualité de l'air. Cependant ils ne peuvent remplacer les stations de mesure traditionnelles. Ils sont actuellement utilisés pour des applications de surveillance de sites industriels ou pour affiner le maillage dans certains endroits non couverts par les stations de mesure.

Solutions avancées et déploiements

Les applications de surveillance de la qualité de l'air se multiplient. Quatre initiatives en collaboration avec les AASQA ont été recensées (voir section 2.2.7 Les réseaux de LCS), cependant, leur portée se limite principalement à la collecte d'informations. À partir d'une sélection d'articles, quatre développements ont été choisis, présentant des idées allant au-delà de la simple surveillance de l'air, comme le montre la table 2.3.4. Ces contributions exploitent et valident l'utilisation de LCS, chacune proposant un prisme d'amélioration de tels systèmes :

1. Alerter la population des zones polluées ;
2. Prendre les mesures en temps réel et les afficher sur une carte ;
3. Utiliser le modèle météo pour prévoir les mouvements de pollution ;
4. Utiliser l'intelligence artificielle à partir des données pour quantifier l'incertitude des capteurs et fiabiliser les mesures.

Réf	Titre article	Contributions
[19]	Distributed System as Internet of Things for a New Low-Cost, Air Pollution Wireless Monitoring on Real Time	— Capteurs sans fil IDE Arduino. — Les données sont envoyées sur les bases du (WHO), application à la surveillance.
[75]	IOT-Based Air Pollution Monitoring and Forecasting System	— Déploiement d'un réseau de capteurs qui relèvent des données en temps réel. — Effectue des prédictions de mouvement de la pollution et l'analyse de l'environnement. — Utilisé lors des accidents de pollution industrielle.
[44]	Air Pollution Monitoring System Using Zigbee and GPS Module	— Réseau de capteurs utilisant la technologie Zigbee et un module GPS. — Affichage des mesures sur une carte de Google Map en temps réel pour une surveillance géographique. — Des alertes peuvent être utilisée suivant le niveau de pollution.
[64]	Wireless sensor network based pollution monitoring system in metropolitan cities	— Réseau de capteurs sans fil. — Les données sont analysées pour le Data Mining afin d'estimer la pollution par probabilités. — Il prédit la pollution et son mouvement pour des alertes d'évacuation ou de confinement.

TABLE 2.4 – Sélection d'études portant sur des application de déploiement de LCS pour la surveillance de la qualité de l'air.

2.3.5 Conclusion

Les capteurs à bas coûts ou micro-capteurs sont de plus en plus utilisés. De l'initiative citoyenne à la surveillance de sites industriels ou sensibles, ils suscitent un véritable engouement. L'avantage de cette technologie est qu'elle est portative et peut être miniaturisée jusqu'à aller dans une poche. Ils séduisent par leur simplicité d'utilisation. Ils peuvent également être utilisés de manière complémentaire aux stations de mesure de référence sur une courte période de temps. Utilisés ainsi, ils doivent être déployés dans des cadres précis tels que la surveillance de pics de pollution, des mouvements exceptionnels de pollution atmosphérique, lancement d'alerte suivants des seuils établis, améliorer une cartographie de pollution modélisée avec un maillage plus fin et enfin faire du diagnostic

en cas de sinistre par exemple. La qualité de ces capteurs s'améliore d'année en année bien qu'ils restent sensibles aux interférences (température, humidité, autres gaz, etc.). Ils ont besoin d'être couplés avec un appareil de référence sur la zone d'étude notamment pour l'étalonnage. Il est important de noter que la plupart des déploiements de micro-capteurs n'ont pas de réelle stratégie de positionnement. La stratégie couramment utilisée repose sur la duplication d'un même système, ce choix étant motivé par des considérations évidentes de conception et de coûts. De plus, il n'y a pas de stratégie en place pour la gestion des défaillances. Cependant, dans une perspective à long terme, cette dimension devient essentielle pour garantir la continuité du service du système de surveillance de la qualité de l'air.

2.4 État de l'art sur l'acquisition, le traitement et la fiabilisation des données

D'une manière plus générale, le but des stations de mesure conventionnelles consiste à alimenter un modèle afin de pouvoir faire des prédictions. Pour garantir la qualité de ces prédictions, il est important de fiabiliser les points de prélèvement. Cette section est dédiée à cette tâche. Actuellement, les stations de mesure conventionnelles sont des références scientifiques et sont exploitées par des personnes qualifiées. Le principal problème de l'utilisation de ces stations consiste en leur coût d'exploitation et de maintenance. En conséquence, il y a peu de points de prélèvement et des vastes territoires ne sont pas couverts par les stations et des modèles sont utilisés pour extrapoler les mesures des stations de mesure pour faire des estimations de pollution en fonction des mouvement dynamiques de l'atmosphère.

2.4.1 Acquisition des données afin d'alimenter un modèle pour faire des prédictions : l'exemple de la météorologie

Les prévisions météorologiques reposent sur l'analyse et la modélisation de données obtenues à partir de diverses sources (stations de surface, ballons-sondes, radars et satellites). Une première acquisition permet aux prévisionnistes d'évaluer l'état atmosphérique actuel et déterminer les tendances des grands mouvements de l'atmosphère. L'analyse des données permet de déterminer les tendances météorologiques à court terme.

Les modèles numériques de prévision simulent le comportement de l'atmosphère et sont également utilisés pour faire des prédictions. L'assimilation des données dans ces modèles est une étape importante. Elle consiste à utiliser des données réelles pour ajuster les équations par la comparaison des données d'observation aux prévisions du modèle. Les différences sont calculées et un lissage est effectué pour obtenir une solution plus précise. Une méthode courante pour cela est l'analyse variationnelle à trois dimensions ou 3D-VAR. Une autre méthode est l'analyse variationnelle à quatre dimensions ou 4D-VAR, qui consiste à ajuster les observations à une simulation numérique et à refaire l'analyse à partir de l'heure initiale [67].

Les prévisions d'ensemble, utilisées par les différents services météorologiques nationaux d'importance tels que le Centre Européen de Prévision Météorologique à Moyen Terme (CEPMMT), le Centre National Américain de Prévision Environnementale (NCEP), le Centre Météorologique Canadien (CMC) et Météo-France, combinent plusieurs modèles avec des analyses légèrement différentes pour obtenir la solution la plus probable. Les pré-

visionnistes comparent leurs estimations et les sorties de ces modèles pour juger de leur validité et les corriger au besoin. Il en est de même pour le modèle AROME pour la prévision des mouvements de pollution. Ces différents modèles ont ainsi besoin en entrée des données d'acquisition les plus fiables et robustes possible.

2.4.2 Acquisition des observations : augmenter le maillage

Les AASQA fournissent trois outils tels que l'inventaire régional des émissions, la modélisation et les stations de mesure. Les stations de mesures sont la base de toutes les décisions et des modélisations, elles doivent suivre des réglementations et des normes. Cependant, leur précision et leur valeur scientifique se font au détriment de leur mobilité et de leur coût d'installation et de fonctionnement élevé. Les territoires tels que celui de la CCPVG n'ont pas de stations de mesure, en conséquence, les élus peuvent difficilement se fier aux trois outils proposés par les AASQA car ce sont des résultats issus des extrapolations des stations de mesure les plus proches. Ces solutions ne sont pas satisfaisantes, notamment lorsque la topologie du territoire est complexe. Par conséquent, les phénomènes de pollution locaux ne sont pas couverts par le réseau de surveillance de la qualité de l'air et sont difficilement pris en compte dans les modèles de dispersion ou par les techniques d'assimilation de données [56]. En d'autres termes, il n'y a pas suffisamment de points de prélèvement pour estimer correctement l'état de la pollution sur une grande zone présentant une topologie variable. Ainsi, pour augmenter le nombre de points d'observations, deux choix s'appliquent :

— le déploiement de stations mobiles des AASQA ;
— la création de son propre réseau de mesure.

Le premier choix est une solution ponctuelle et ne permet pas le suivi des actions entreprises sur le territoire à long terme. Pour disposer d'un réseau de mesure sur le long terme avec un budget limité et sans avoir à installer des capteurs coûteux et exigeants, une solution alternative et complémentaire existe et consiste en la mise en réseau de LCS. Favorisés par l'émergence de l'Internet des Objets (IoT), les LCS sont faciles à mettre en place et moins coûteux. En effet, il y a une véritable popularité pour cette solution, notamment dans les initiatives citoyennes [29, 9] où chaque citoyen peut posséder une station de mesure et participer à la surveillance des concentrations de polluants pour la qualité de l'air intérieur dans les hôpitaux [61] ou dans leurs maisons intelligentes [68] pour protéger la santé humaine. D'autres initiatives citoyennes comprennent la collecte de paramètres environnementaux sur de vastes zones extérieures [2], ou à la fois à l'intérieur et à l'extérieur sur une vaste zone en utilisant le protocole ESP-MESH [27]. Cependant, les LCS sont moins précis et globalement moins fiables que les stations conventionnelles. Néanmoins, en raison de leur polyvalence et de leur accessibilité, il devient possible d'avoir une grille de mesure plus fine et, par conséquent, une meilleure précision spatiale. Ces deux solutions peuvent coexister : les réseaux de LCS apportent davantage de données (avec moins de précision) et les stations conventionnelles fournissent des données de référence (avec une meilleure fiabilité) pouvant aider à l'étalonnage de l'ensemble de l'architecture. Les réseaux de LCS représentent une véritable opportunité et plusieurs améliorations ont été publiées dans la littérature scientifique [58], [44]. Pour couvrir de vastes zones géographiques, des architectures déployées basées sur la mise en réseau de LCS sont utilisées. Elles permettent par la multiplication des points de prélèvement de couvrir un territoire complet. Par exemple, une architecture distribuée, à plusieurs niveaux basée sur un réseau sans fil utilisant la technologie ZigBee avec une application web et des LCS a été mise

en place sur plusieurs sites en Équateur pendant 4 mois [19]. Cette preuve de concept de surveillance en temps réel démontre la possibilité de développer une plateforme de surveillance de la qualité de l'air pour comparer et vérifier la conformité avec les normes de l'OMS. Une autre étude va plus loin en proposant une prédiction sur 24 heures en utilisant un réseau de neurone récurrent Long short-term memory (LSTM) pour effectuer une analyse et une prédiction sur les données mesurées [34]. Ces études fournissent des solutions pour la surveillance en temps réel de vastes zones (d'un campus à trois villes) avec des investissements modestes. Elles se concentrent notamment sur le post-traitement des données des LCS et permettent de compenser leurs inconvénients. Pour aller plus loin, les applications des LCS pour la surveillance de la qualité de l'air sont utilisées à des fins spécifiques (surveillance, ajustement fin du maillage en complément des stations conventionnelles, détection de la pollution, etc.) et sont suffisamment adaptées pour cartographier les gradients de concentration [44]. Des voies supplémentaires sont envisagées pour augmenter le potentiel de cette technologie, notamment l'utilisation de l'apprentissage automatique et/ou des techniques avancées de traitement des données pour compléter les données des stations de mesure conventionnelles [4].

La littérature scientifique montre qu'il est tout à fait possible de surveiller un territoire avec des LCS mis en réseau. Cependant, une critique qui peut être formulée à l'égard de ces travaux concerne leur temporalité. Ils sont déployés sur des périodes courtes (d'un mois à quelques mois). Un déploiement sur le toit d'un laboratoire pendant 11 mois [6] est l'étude la plus longue actuellement enregistrée et se rapproche le plus de l'objectif de continuité des mesures défini dans l'introduction de cette thèse. Les déploiements des LCS utilisent un seul dispositif par point à surveiller, ce qui est insuffisant pour évaluer la fiabilité des mesures. En effet, les LCS présentent plusieurs problèmes de fiabilités traités dans la sous-section suivante. Ainsi, la fiabilité des mesures en un point d'échantillonnage constitue le point de départ de la surveillance de la qualité de l'air.

2.4.3 La fiabilisation des données en un point d'observation

Les LCS posent un défi en termes de robustesse et de reproductibilité des mesures. Ils fournissent des mesures grossières et, pour améliorer leurs performances, il est nécessaire de les étalonner sur le terrain [9, 20] ou de les corriger par post-traitement [73]. Plus précisément, lors de la surveillance de la pollution dans un contexte périurbain, les interférences croisées avec d'autres polluants influencent considérablement le signal généré par le mesurande, ce qui impacte la performance de la mesure de concentration des LCS gaz. Ils mesurent plusieurs gaz simultanément et leurs résultats correspondent à la somme des concentrations de tous les gaz interagissant avec la partie sensible du capteur [32]. Ces incertitudes sont d'autant plus importantes lorsque les concentrations sont faibles, c'est-à-dire lorsque les capteurs sont situés à l'extérieur. Ces études mettent en évidence les défis de l'utilisation de capteurs à faible coût dans des environnements à faible concentration. De tels contextes extérieurs peuvent parfois atteindre les limites de détection des capteurs, notamment pour le SO_2 [21]. Pour corriger cet inconvénient, les auteurs de ces articles proposent de regrouper les capteurs en clusters et d'utiliser des méthodes statistiques pour augmenter leur fiabilité. D'autres travaux sur l'amélioration de la durée de vie et de la fiabilité des mesures LCS consistent à les regrouper [42] ou à utiliser le ML en post-traitement pour compenser la dérive du capteur [64]. En bref, la combinaison de l'étalonnage, du clustering et du post-traitement à travers une longue période d'apprentissage aide à augmenter la qualité des données et la précision des me-

sures. Une première piste confirme l'utilisation des techniques précédentes (clustering et ML) sur 2,5 mois. Plusieurs stratégies prometteuses pour améliorer la fiabilité sans avoir recours à l'étalonnage ont été explorées. Une méthode utilise un cadre basé sur le ML pour détecter et réparer les anomalies dans les données brutes avant étalonnage, soulignant la nécessité d'assurer la fiabilité dans un réseau de capteurs à faible coût [59]. Une autre approche se concentre sur l'influence de la pression barométrique sur les lectures des capteurs de particules à faible coût. L'étude introduit un modèle de régression pour souligner la nécessité d'une correction de la pression barométrique pour optimiser ces mesures [11]. En termes de fiabilité matérielle, d'autres voies sont explorées pour améliorer la fiabilité du système à long terme grâce à la détection des défauts des capteurs. Une approche consiste à examiner et comparer les motifs électriques distincts dans les capteurs de pollution à faible coût pour identifier les problèmes [41]. Dans un autre domaine, une étude emploie deux techniques parallèles pour évaluer l'état des capteurs d'humidité en évaluant à la fois la capacité du sol et celle du capteur lui-même [47]. Une autre étude utilise une double redondance sur les LCS [10]. Les auteurs ont validé leur prototype en comparant leurs résultats à un autre dispositif commercial (5 % de différence) utilisant des LCS pendant cinq jours. Néanmoins, la double redondance n'est pas suffisante pour garantir la continuité du service. En effet, en cas de défaillance, il n'y a pas de correction d'erreur et la détection du composant défectueux ne peut pas être faite automatiquement. Ces travaux réduisent les inconvénients des LCS en termes de performance et de fiabilité de l'équipement à court terme.

En raison de l'objectif de continuité des mesures, il est important d'augmenter la disponibilité du point de prélèvement, à savoir d'une station de mesure composée de LCS. Afin de répondre à la clustérisation des LCS et à la fiabilisation de leurs observations par la mise en parallèle des LCS afin de pallier aux problème de répétabilité, l'approche proposée dans cette thèse consiste à développer une station basée sur le concept de la triple redondance modulaire.

2.4.4 La triple redondance modulaire et les algorithmes de vote

Pour répondre aux objectifs de surveillance de la qualité de l'air (continuité du service et fiabilité des mesures), chacune des études précédentes apporte un élément de réponse mais est individuellement insuffisante pour combler les lacunes des LCS. Par conséquent, l'architecture proposée dans le prochain chapitre tente de surmonter les lacunes suivantes : la fiabilité des mesures, la continuité du service et l'horizon de mesure. Cette proposition consiste à regrouper les LCS en clusters, inspirée de l'utilisation de la redondance modulaire triple active (TMR) pour rendre le système tolérant aux pannes, et en mettant en place une plateforme de surveillance inspirée des déploiements précédents. La TMR augmente les coûts matériels, mais ceux-ci sont maîtrisés grâce à l'utilisation des LCS et à leur intégration dans la conception. L'architecture basée sur les LCS permet de contrôler les coûts et multiplie ainsi le nombre de points d'échantillonnage. De plus, elle présente les caractéristiques suivantes : redondante, tolérante aux pannes, extensible et avec un horizon de surveillance évolutif. Sa résilience permet de collecter des données complètes et fiables pour des tâches de traitement ultérieur et de prise de décision.

L'architecture de la station de mesure s'inspire de la TMR qui est largement utilisée dans l'industrie pour assurer une disponibilité élevée et une fiabilité des applications critiques. Le principe repose sur trois modules identiques et indépendants fonctionnant en parallèle et ayant les mêmes entrées. La sortie de ces modules est soumise à une unité

de vote pour créer une sortie et générer une synthèse. Les agrégations effectuées par l'unité de vote sont généralement le vote à la majorité, le vote médian et le vote pondéré moyen [38].

Les algorithmes de vote à la majorité rendent le système tolérant aux pannes en sélectionnant la sortie correspondant à la majorité des sorties des modules. Sinon, la sortie est un code pour arrêter le système en toute sécurité. Les derniers développements de l'algorithme de vote à la majorité utilisent les données historiques pour optimiser le choix et rendre ainsi le système plus fiable. Cependant, ce type d'algorithme de vote à la majorité présente deux inconvénients majeurs. Le premier inconvénient est que lorsque les sorties des modules ne sont pas proches (ce qui est le cas pour les LCS), un seuil doit être défini afin de regrouper les valeurs proches de sorte que l'algorithme les considère identiques et, lorsque la majorité n'est pas atteinte, le système est arrêté. Le deuxième inconvénient vient du fait que les défaillances sont masquées. En effet, la sortie correspond à la valeur la plus fréquente, mais les valeurs rejetées ne sont pas exploitées.

Les algorithmes de vote pondéré moyen effectuent la moyenne des sorties des modules avec des poids attribués à chaque sortie. Ces poids sont calculés à partir de leurs écarts respectifs. Plus l'écart est grand, plus le poids est petit. De nouveaux algorithmes sont proposés [30] et sont plus fiables selon leurs auteurs que le vote à la majorité en présence d'erreurs. Leur principal avantage est de fournir une sortie quel que soit le nombre de modules présents (contrairement aux algorithmes de vote). Cependant, lorsque les erreurs sont importantes, le nombre de sorties incorrectes de l'algorithme augmente. Cette observation provient de l'utilisation de la moyenne. En effet, la moyenne est influencée par les extrêmes, et plus il y a de valeurs aberrantes dans les entrées, plus la moyenne sera influencée [33]. De plus, les performances de ces algorithmes dépendent du choix des poids.

Les algorithmes de vote médian sont plus efficaces que ceux basés sur la moyenne pondérée. En effet, l'utilisation de la médiane permet de réduire l'influence des valeurs aberrantes. Cependant, la fiabilité d'un tel algorithme est diminuée lorsque la majorité des valeurs sont des valeurs aberrantes [5].

Chaque type d'algorithme de vote fonctionne bien lorsque les erreurs d'entrée représentent une minorité de la sortie totale. Ils sont optimisés pour fournir une sortie avec le moins d'erreurs possible. La combinaison d'algorithmes et l'utilisation de la classification pour trouver la meilleure sortie augmentent la fiabilité, mais également la complexité de calcul et le temps de traitement [26]. Cependant, pour le traitement en temps réel, le temps de calcul doit être pris en compte. Quoi qu'il en soit et quelle que soit la combinaison d'algorithmes, ils restent influencés par les erreurs des entrées. Enfin, le point commun de ces algorithmes est le masquage des erreurs. C'est pourquoi, afin d'augmenter les performances des algorithmes, une étape de détection des erreurs basée sur les données est proposée dans la fiabilisation des observations issues des LCS et mise en œuvre en amont de l'agrégation dans l'unité de vote. En effet, selon les auteurs de cette étude [28], cette configuration permet de rendre un système TMR plus fiable.

A partir d'une configuration en TMR, les détections de défaillances peuvent être réalisées grâce à des algorithmes spécialisés.

2.4.5 Les algorithmes de détection des défaillances

Dans les méthodes de détection des défauts, le domaine le plus proche de l'application abordée dans cette thèse est celui des réseaux de capteurs sans fil (WSN). Il existe différentes techniques de détection de défauts et une comparaison qualitative des derniers

algorithmes de détection de défauts pour le déploiement de WSN est répertoriée dans cette référence [45]. Parmi toutes les techniques possibles, le choix est motivé par un diagnostic distribué d'auto-détection des nœuds capteurs, qui sont les SmS dans cette thèse. Pour un déploiement à grande échelle, chaque station de mesure doit identifier ses défauts afin de limiter la complexité informatique au niveau du réseau global. Ainsi, dans [49], un algorithme est proposé utilisant l'écart-type médian normalisé pour détecter et éliminer les valeurs aberrantes. Composé de deux phases, la première phase de l'algorithme vise à collecter les mesures pendant une durée d'estimation et à les associer à l'identification des SmS. La deuxième phase élimine les valeurs aberrantes à l'aide d'une méthode statistique utilisant l'Écart Absolu Médian Normalisé (MADN). Cependant, bien que la première phase soit inspirante pour l'étape de détection, le calcul du MADN suppose que la distribution des mesures soit normale, ce qui n'est pas le cas des mesures liées aux phénomènes naturels.

Ces méthodes de détection ont en commun la comparaison des observations entre les capteurs pour effectuer la détection des défaillances. En raison de l'incohérence de leur reproductibilité, l'intérêt d'un algorithme de détection des valeurs aberrantes, développé dans le chapitre 4, réside dans sa capacité à fiabiliser la mesure de la station, à repérer les capteurs défectueux et à générer des indices de défaillance et de confiance.

2.5 Conclusion du chapitre

Au cours de ce chapitre, la relation entre la pollution atmosphérique et la qualité de l'air a été mise en évidence. L'importance de mesurer les concentrations de polluants atmosphériques s'est révélée essentielle pour évaluer la qualité de l'air, avec une attention particulière portée aux défis rencontrés dans les territoires non couvert par les stations conventionnelles. Les différentes méthodes et technologies disponibles pour surveiller la qualité de l'air, y compris les avantages distinctifs offerts par les LCS, ont été explorées. Il en résulte que l'approche descendante associée à l'assimilation de données représente une méthode fiable. Toutefois, des limites ont été observées, particulièrement dans les zones éloignées des stations de mesure. Les LCS constituent une opportunité pour multiplier les points de prélèvement sur un territoire. Bien que la facilité de mise en œuvre et le bas coût des LCS les rendent attrayants, des préoccupations quant à la fiabilité des données ont été soulevées, soulignant la nécessité d'approches robustes pour l'acquisition et la validation des données.

A partir de cet état de l'art, le prochain chapitre se concentrera sur l'élaboration d'une architecture pour la surveillance de la qualité de l'air. Cette architecture intègrera les concepts et les technologies discutés, offrant une structure cohérente pour la mise en œuvre efficace de la surveillance de la qualité de l'air, en tenant compte à la fois des contraintes budgétaires et des spécificités des territoires concernés.

Chapitre 3

Proposition d'une architecture de monitoring territorial : une approche ascendante

Sommaire

3.1 Introduction 61
3.2 Présentation générale de l'architecture 61
3.3 Les dimensions spatiale et temporelle de l'architecture proposée 62
3.3.1 Estimer un paramètre physique en un point précis 65
3.3.2 Méthodes d'interpolations spatiales 65
3.3.3 De la mesure locale à l'impact régional : l'approche ascendante 73
3.3.4 Le réseau de mesure dans la dimension spatiale 74
3.4 Observation et monitoring en un point de prélèvement 76
3.4.1 Exploration des méthodes d'interpolation dans le cadre d'un réseau temporel de surveillance de la qualité de l'air 77
3.4.2 Utilisation de la fenêtre roulante pour l'amélioration de la qualité des données et la réduction du bruit 81
3.5 Conception d'une station de mesure en un point de prélèvement 82
3.6 Conclusion du chapitre 84

3.1 Introduction

Ce chapitre explore la question de la surveillance de la qualité de l'air en mettant en avant la dimension spatio-temporelle de l'architecture proposée. Le chapitre précédent a souligné les limites de l'approche descendante d'ATMO, notamment en ce qui concerne les territoires éloignés des stations conventionnelles. Il en ressort que des phénomènes de pollution locaux peuvent échapper à la surveillance des stations actuelles. L'une des solutions à ce problème, et qui constitue le cœur de ce chapitre, réside dans une approche « ascendante » par la multiplication des points de mesure sur un territoire. Cette approche est la dimension spatiale de l'architecture proposée dans cette thèse. Cette approche ascendante se distingue également par son recours possible à des méthodes d'interpolation, qui peuvent remplacer ou compléter les modèles utilisés par ATMO. Une combinaison de ces deux méthodes est envisagée, notamment pour les points situés aux limites des zones surveillées. Après une description générale, chaque dimension de l'architecture est mise en exergue, révélant les défis associés à un tel système de surveillance. L'un des principaux enjeux est la garantie d'une bonne connaissance de la qualité de l'air à chaque endroit du territoire. Des méthodes d'interpolation seront donc explorées pour traiter cette problématique. La distinction entre les approches descendante, basées sur des stations conventionnelles, et ascendante, centrale dans cette thèse, est également abordée. Enfin, en ce qui concerne la dimension temporelle, ce chapitre fournira des détails sur la conception d'une station de mesure, mettant en évidence les paramètres clés mesurés, l'historique des données et les prédictions pour des mesures futures.

3.2 Présentation générale de l'architecture

La Figure 3.1 montre une vue d'ensemble de l'architecture proposée. Les paramètres physiques à collecter sont déterminés par les autorités, en se basant sur des critères de protection de la population et de l'environnement. Plus précisément, la sélection des paramètres physiques relatifs aux concentrations de polluants est effectuée en collaboration avec des experts en qualité de l'air. À titre d'exemple, pour le suivi du trafic routier, les polluants PM_{10}, $PM_{2.5}$ et NO_x sont souvent sélectionnés. Quant aux paramètres physiques centrés sur la météorologie, ceux généralement retenus sont la température, la pression et l'humidité. La vitesse et la direction du vent peuvent également être prises en compte pour affiner l'analyse des mouvements de pollution. Les données contextuelles servent à une meilleure interprétation des mesures et à une compréhension approfondie des événements de pollution. Ces données peuvent inclure des caractéristiques topologiques, telle la géométrie d'une vallée, qui influence les courants d'air locaux et leur orientation selon l'ensoleillement des pentes. Des éléments saisonniers liés aux activités touristiques sont également pris en compte pour expliquer les variations de pollution. Au sein du système de surveillance, des experts en mesures environnementales et en maintenance des équipements participent à son bon fonctionnement. Le système propose une aide à la décision aux responsables, présentée sous la forme d'une interface graphique. Cette aide à la décision englobe des analyses en temps réel, des prédictions et des recommandations basées sur les données collectées, permettant aux responsables d'agir de manière informée pour améliorer la qualité de l'air. Les mesures correctives sur le terrain sont ensuite prises en se basant sur ces informations.

Le système de surveillance a pour objectif principal de surveiller la qualité de l'air sur un territoire étendu, en tenant compte de la protection de la population, de l'envi-

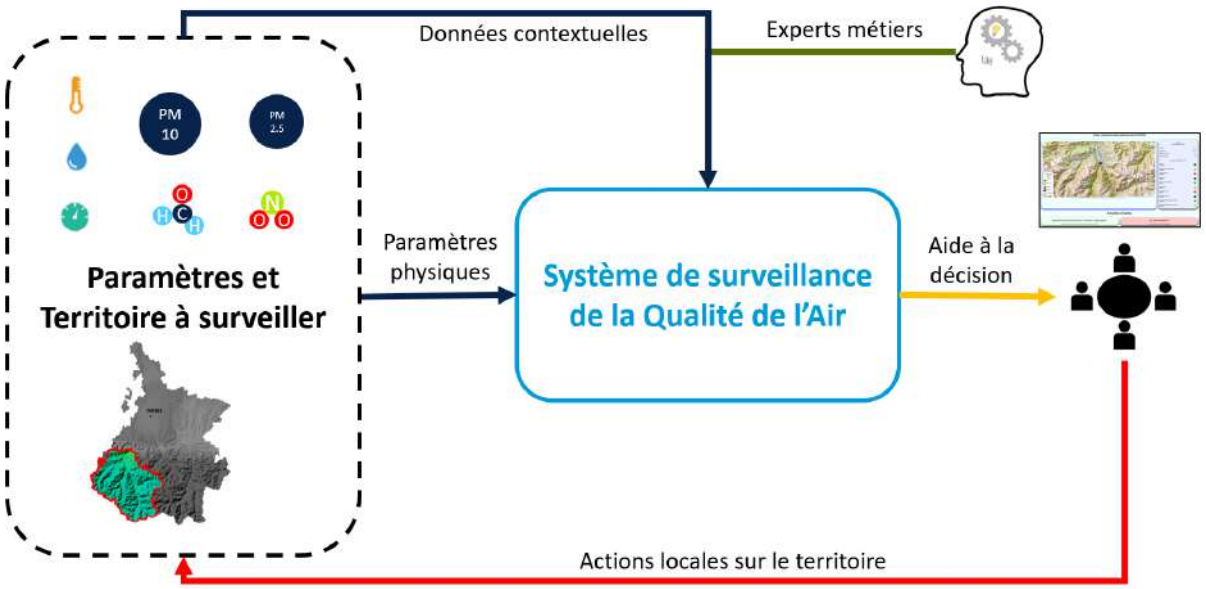


FIGURE 3.1 – Vue d'ensemble de l'architecture proposée.

ronnement et des zones à risque de pollution. Ce système vise à fournir un soutien aux élus locaux dans leur prise de décision. L'objectif est de permettre la mise en œuvre d'actions ciblées sur le territoire. Les données sont cruciales pour ces décisions. Cependant, les données actuellement disponibles sont souvent peu exploitées en raison de la distance qui sépare les stations conventionnelles des zones concernées. Pour garantir des données sur la qualité de l'air fiables et sur le long terme, quatre objectifs clés ont été identifiés :

1. Le système de surveillance doit être capable de couvrir de vastes zones par ajout de nouvelles stations ;
2. Le réseau de mesure doit être conçu pour fonctionner sur le long terme tout en ayant une haute disponibilité ;
3. Le réseau de mesure doit être extensible en distribuant les calculs directement au niveau des nœuds ;
4. L'architecture proposée doit non seulement fournir des outils décisionnels, mais aussi rationaliser l'ensemble du processus de mesure pour produire une base de données prête à être utilisée pour la prise de décision.

3.3 Les dimensions spatiale et temporelle de l'architecture proposée

L'approche ascendante en matière de modélisation de la qualité de l'air se distingue par sa focalisation initiale sur des sources spécifiques de pollution, où leurs émissions et leurs propriétés sont identifiées. Basée sur des données locales, cette méthode vise ensuite à extrapoler ces informations pour évaluer l'impact sur l'ensemble du territoire. Dans un premier temps, des données contextuelles servent à identifier et déterminer l'origine des émissions polluantes, qu'il s'agisse d'une usine, d'une zone touristique, etc. Ces données sont par la suite intégrées dans un cadre spatial élargi, permettant ainsi de mesurer la

dispersion et l'interaction des polluants à travers le territoire. Bien que l'approche ascendante offre une richesse d'informations et une précision notable, sa mise en œuvre nécessite beaucoup de points de prélèvement en rapport avec une connaissance fine du terrain et le concours de nombreux experts locaux. Cela la différencie de l'approche descendante, évoquée au chapitre précédent. Cependant, sa principale force réside dans la capacité à élaborer des stratégies adaptées pour les autorités, leur permettant d'agir plus efficacement en matière de réduction des émissions. Il est également essentiel de souligner que cette méthode est complémentaire de l'approche descendante : elle permet d'affiner les prédictions de cette dernière et révèle d'éventuelles divergences entre les deux techniques. En conclusion, l'approche ascendante est caractérisée par sa précision spatiale en ce qui concerne les sources de pollution individuelles et vise à établir une compréhension globale basée sur ces informations localisées.

Les approches ascendante et descendante sont importantes pour la compréhension de la qualité de l'air, il est essentiel d'examiner comment ces méthodologies se traduisent concrètement en termes de surveillance sur un territoire étendu. Pour surveiller la qualité de l'air sur un territoire vaste, il est nécessaire de prendre en compte les dimensions spatiale et temporelle du processus de surveillance. La dimension spatiale, illustrée par la Figure 3.2, est abordée en multipliant les points de prélèvement. Cette stratégie permet d'assurer une couverture adaptée à la zone concernée. Un réseau maillé est ainsi créé sur le territoire. Au maximum, chaque maille est dotée d'une station de mesure localisée en un point de prélèvement spécifique. La station est représentative de l'ensemble de la maille, servant ainsi de point de référence pour la surveillance de la qualité de l'air dans cette portion du territoire.

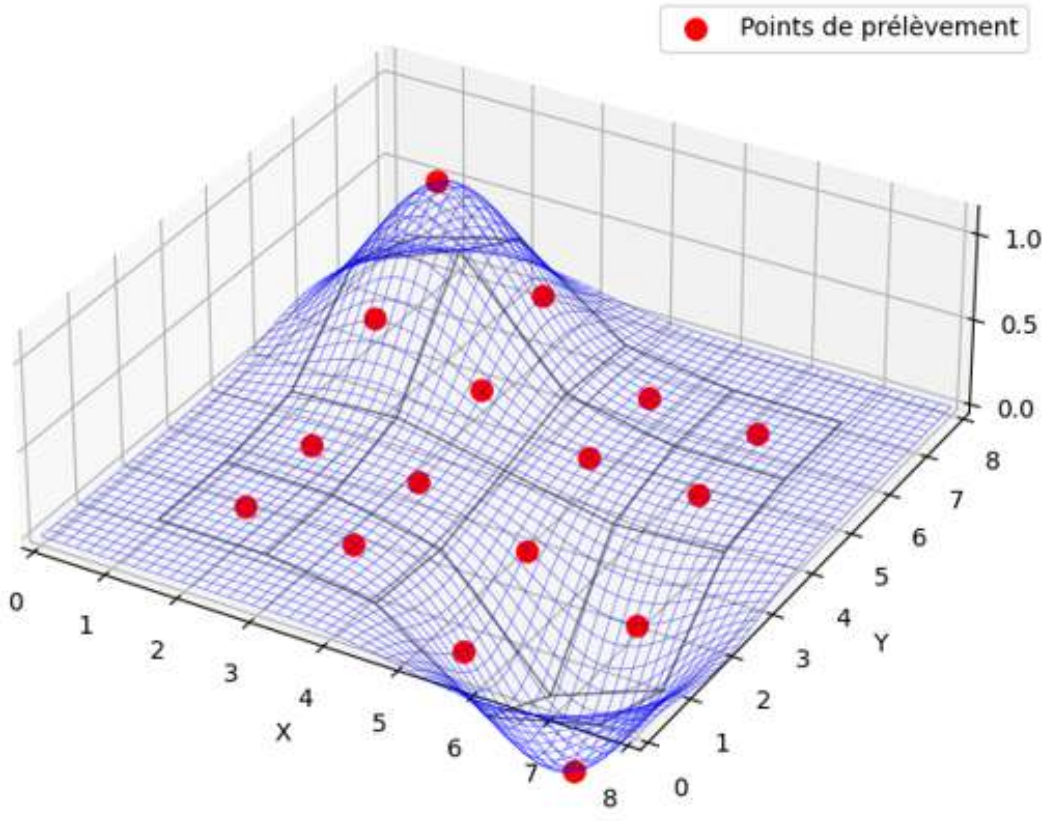


FIGURE 3.2 – La dimension spatiale de l'architecture.

La dimension temporelle, mise en évidence par la Figure 3.3, intègre plusieurs composantes distinctes qui contribuent à sa complexité. Cette figure représente une mesure d'un paramètre physique au fil du temps. Un point rouge y symbolise le moment présent, capturant ainsi les mesures en temps réel d'un paramètre physique tel qu'une concentration de polluant ou un paramètre météorologique. Les données antérieures sont représentées en bleu, indiquant l'archivage effectué pour les analyses rétrospectives et les études comparatives. En ce qui concerne le futur, trois courbes sont tracées selon des scénarios optimiste, réaliste et pessimiste. Ces courbes illustrent les prédictions futures ainsi que les incertitudes associées à ces prévisions, permettant ainsi des interventions plus ciblées et proactives au niveau d'un paramètre physique.

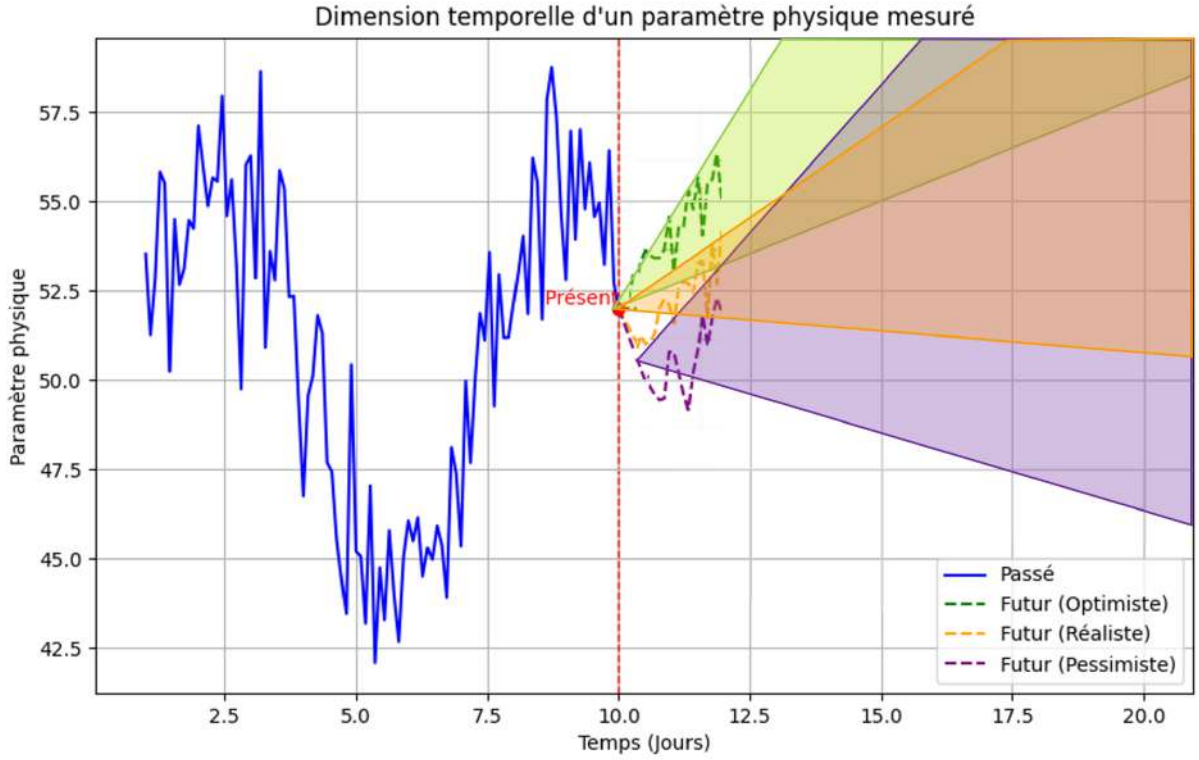


FIGURE 3.3 – La dimension temporelle de l'architecture.

Ainsi, la surveillance de la qualité de l'air nécessite une coordination des dimensions spatiale et temporelle. Alors que la première se focalise sur la couverture géographique et la répartition des stations de mesure, la seconde s'attache à la dynamique des paramètres physiques à différents moments. Il est donc indispensable de les considérer conjointement. Pour l'architecture proposée, il est nécessaire dans un premier temps de définir les paramètres physiques à mesurer. L'étape suivante consiste en une identification des points de prélèvement à positionner sur le territoire. Cette démarche se manifeste par l'établissement de zones d'intérêt disséminées à travers le territoire. Chaque zone d'intérêt peut accueillir une ou plusieurs stations de mesure. Le positionnement stratégique de ces stations permet la génération de cartographies de concentrations polluantes, particulièrement en ce qui concerne la qualité de l'air. Les zones d'intérêt peuvent soit agréger les données de plusieurs stations, soit être représentées par une station unique. Des cartes de pollution spécifiques à chaque paramètre physique sont créées par interpolation entre les points de mesure, répondant ainsi à la nécessité de connaître les niveaux de pollution en tout point du territoire. Le volet temporel s'intègre également à cette architecture ; les

paramètres physiques varient avec le temps, entraînant des fluctuations dans les cartes de pollution. Des techniques d'interpolation sont aussi employées au niveau de chaque station pour permettre des modélisations comportementales et des prédictions sur l'évolution des concentrations. L'approche spatio-temporelle de cette architecture et la distribution des stations à divers emplacements du territoire visent à améliorer la précision des cartes de pollution fournies par ATMO, en mettant un accent particulier sur les phénomènes de pollution locaux.

3.3.1 Estimer un paramètre physique en un point précis

En fonction des objectifs de mesure et de protection de l'environnement et de la population, les stations de mesure implantées dans des zones d'intérêts peuvent être très éloignées entres elles. L'évaluation précise de la qualité de l'air à un point spécifique peut alors présenter des défis significatifs. La problématique abordée dans cette sous-section concerne l'estimation d'un paramètre physique en un point précis, compte tenu des mesures disponibles aux alentours.

L'interpolation spatiale est couramment utilisée pour estimer des valeurs sur l'ensemble d'un espace géographique à partir d'un nombre limité de points de mesure [31]. Il existe plusieurs méthodes d'interpolation spatiale, chacune ayant ses propres avantages, inconvénients et conditions préalables en termes de nature des données et de densité de distribution. Les méthodes d'interpolations les plus utilisées sont : l'interpolation linéaire, l'interpolation cubique, l'interpolation par le plus proche voisin (Nearest), l'interpolation par fonction de base radiale (RBF), l'interpolation par pondération inverse de la distance (IDW), le krigeage et l'interpolation par splines.

La méthode d'interpolation influe de manière significative sur le résultat de la carte de pollution, de même, le choix de la méthode d'interpolation est directement liée aux points de mesures, à leur quantité et leur emplacement spatial. Une exploration des principales méthodes utilisées (à savoir l'interpolation linéaire, cubique, Nearest, RBF, IDW, krigeage et l'interpolation par splines) ainsi que leur visualisation sur la base de trois stations de mesure est proposée afin d'avoir un comparatif en fonction de la densité des données et des ressources en calcul. Cette comparaison a pour but de servir de point de départ pour des recherches plus approfondies sur les méthodes d'interpolation spatiale et leur application à la surveillance de la qualité de l'air.

3.3.2 Méthodes d'interpolations spatiales

L'interpolation spatiale est une technique mathématique utilisée pour estimer de paramètres physiques dans un espace géographique en se basant sur les mesures prises en des points de prélèvement. Ce processus est très utilisé dans la surveillance environnementale, la météorologie et l'urbanisme. Le surajustement et le sous-ajustement sont deux problèmes fréquemment rencontrés lors d'une interpolation. Le surajustement se produit lorsque le modèle d'interpolation est trop complexe, ce qui lui permet de capturer le bruit dans les données, au lieu de la tendance générale. En d'autres termes, le modèle se comporte très bien sur les données existantes mais a une faible capacité à généraliser à de nouvelles données ou à des conditions différentes. Le sous-ajustement se produit lorsque le modèle est trop simple pour capturer la complexité inhérente aux données. Un modèle sous-ajusté échouera à rendre compte des variations dans les données, donnant ainsi des estimations inexactes.

Il est donc essentiel de choisir une méthode d'interpolation qui soit adaptée à la nature des données et au contexte de l'étude pour éviter ces deux problèmes. La nature des données, plus exactement leur forme de variation, conditionne le choix de la méthode d'interpolation. Une variation linéaire des données sera mieux utilisée pour une interpolation linéaire, pour des variations plus complexes l'interpolation cubique est appropriée. Les interpolations Nearest et RBF sont adaptées sur des données discrètes ou non linéaires. Ainsi, le choix de la méthode d'interpolation la plus adaptée dépend de plusieurs facteurs liés aux données et aux exigences liées au contexte.

— **Densité et distribution des points de données :** dans le cas où les points sont peu denses, les méthodes d'interpolation linéaire ou cubique peuvent ne pas être en mesure de fournir des estimations précises. En effet, elles sont plus précises et efficaces lorsque la densité des points est importante. Les méthodes Nearest et RBF peuvent être plus appropriées.

— **Précision :** plus les estimations ont besoin de précision, plus la méthode utilisée devra être complexe et difficile à utiliser.

— **Ressources de calcul disponibles :** les méthodes d'interpolation avancées demandent davantage de ressources informatiques et requièrent plus de temps d'exécution qui peuvent affecter, entre autres, la réactivité. Le nombre de points à traiter, leur densité, ainsi que la fréquence de calcul sont autant de facteurs cruciaux à considérer.

Dans ce qui suit, et pour démontrer les différentes méthodes d'interpolation, une simplification est effectuée en considérant uniquement deux dimensions spatiales (x, y) et trois stations de mesure de pollution, bien qu'il soit reconnu que la dimension de la hauteur joue également un rôle significatif. Ces stations, nommées respectivement A, B et C, sont situées aux coordonnées suivantes : (1,1), (6,2) et (3,5). Les mesures de la qualité de l'air sont représentées sur une échelle allant de 0 à 10 pour faciliter la compréhension. À l'instant t considéré, les valeurs mesurées aux stations A, B et C sont respectivement de 10, 2 et 5. La Figure 3.4 illustre ces données dans l'espace bidimensionnel.

Sur cette base de trois stations, les méthodes explorées sont l'interpolation linéaire, cubique, Nearest et RBF.

L'interpolation linéaire est la méthode la plus simple et consiste à calculer des valeurs intermédiaires en utilisant une interpolation linéaire entre les points de mesure les plus proches. Cela implique de tracer une ligne droite entre deux points voisins et d'estimer les valeurs intermédiaires en fonction de la distance par rapport à ces points.

L'interpolation cubique est une méthode plus gourmande en calcul qui, contrairement à l'interpolation linéaire, ajuste un polynôme de degré trois (une courbe) entre les points. Cela permet à la méthode de capter des tendances plus complexes dans les données qui ne peuvent pas être représentées avec précision par une ligne droite. Cependant, un risque de sur-ajustement est présent, notamment en capturant le bruit et en rendant le modèle moins généralisable. Ce phénomène est accentué en présence de bruit dans les données ou de variations brusques.

L'interpolation spatiale Nearest est une méthode d'interpolation simple. Elle est basée sur l'idée que la valeur d'un point inconnu doit être la même que la valeur du point connu le plus proche. Elle ne fait aucune hypothèse sur la relation entre les valeurs mesurées

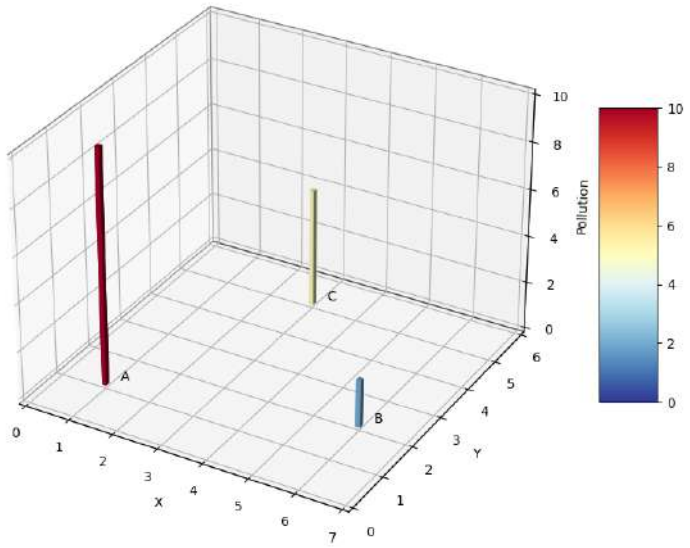


FIGURE 3.4 – Illustration des données dans l'espace bidimensionnel à un instant t.

et leurs emplacements, ce qui la rend applicable à une grande variété de données. Elle est peu consommatrice de ressources en calcul. Mais elle peut produire des résultats très imprécis si la densité de points de mesure est faible ou si les valeurs varient rapidement avec la distance. De plus, elle ne peut pas représenter les variations continues dans les données, car elle crée des régions autour de chaque point de mesure où la valeur estimée est constante.

L'interpolation par fonction de base radiale (RBF) utilise une fonction, qui dépend de la distance entre les points, pour estimer les valeurs. Elle est particulièrement utile pour les données qui présentent des changements réguliers et des tendances non linéaires. Elle prend en compte tous les points de données, pas seulement les plus proches, ce qui lui permet de capturer des tendances plus complexes. Elle est également moins susceptible de sur-ajuster les données que certaines autres méthodes, comme l'interpolation cubique. Cependant, l'interpolation RBF peut nécessiter plus de ressources informatiques que d'autres méthodes, en particulier pour les grands ensembles de données. De plus, le choix de la fonction utilisée peut influencer significativement les résultats.

Ces méthodes présentées ont été appliquées sur les données spatiales et sont illustrées dans la Figure 3.5. Les points rouges représentent les mesures de pollution des stations A, B, et C, et les barres colorées correspondent aux mesures de pollution à ces stations. Les surfaces colorées représentent les estimations de la pollution interpolées à partir de ces points.

Compte tenu de la nature des données, les interpolations linéaire et cubique n'ont pas réussi à fournir des estimations valides en dehors de la zone triangulaire formée par les trois points, entraînant ainsi l'apparition de régions dépourvues de valeurs. Les méthodes d'interpolation linéaire et cubique génèrent une surface d'interpolation en utilisant respectivement des fonctions linéaires et cubiques pour estimer les valeurs entre les points de données connus. Cependant, elles ne peuvent pas extrapoler des valeurs pour des points qui se trouvent en dehors de la zone définie par les points de données d'origine. En d'autres termes, cela signifie que toute zone de cette grille qui se situe en dehors du minimum et

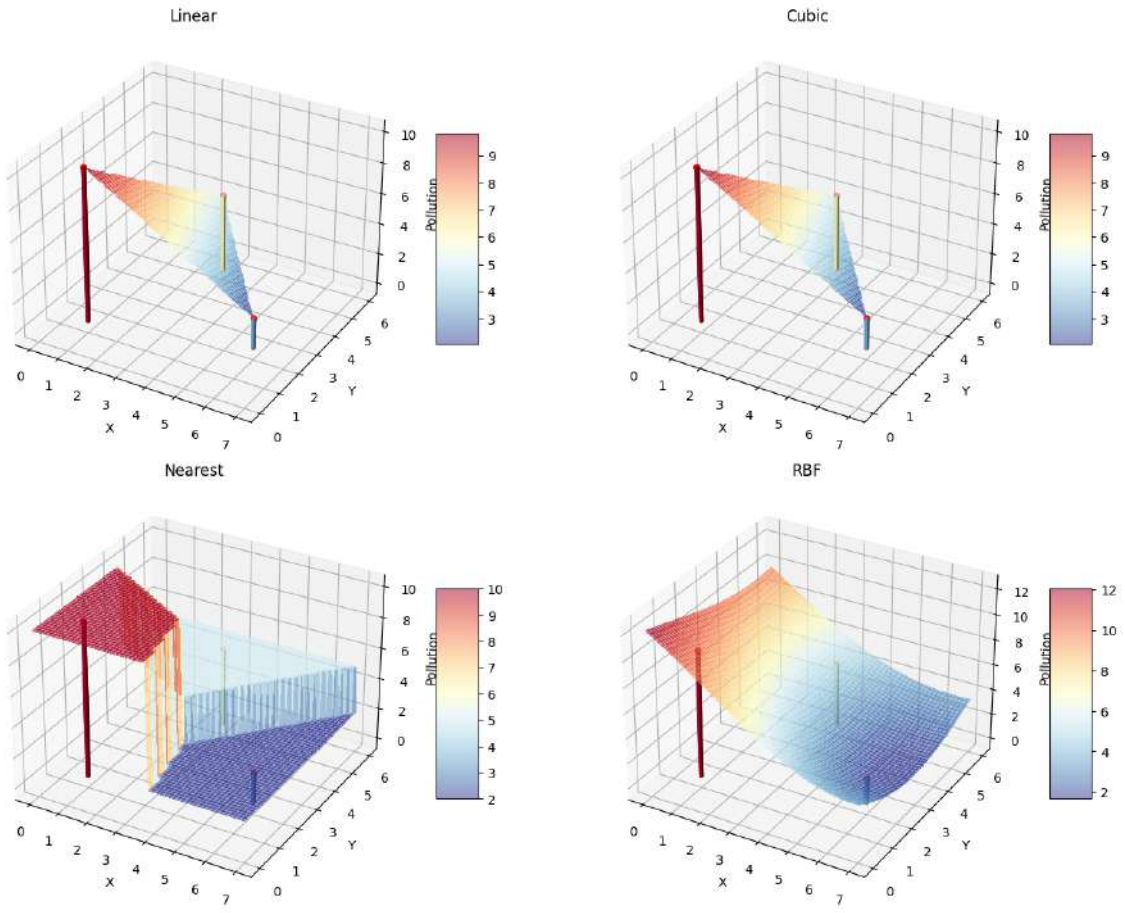


FIGURE 3.5 – Représentation des méthodes d'interpolation linéaire simple, cubique, Nearest et RBF

du maximum des points de données d'origine (stations A, B, C) n'aura pas de valeur interpolée avec les méthodes linéaire et cubique.

L'interpolation par le plus proche voisin (Nearest) a produit une surface avec des changements abrupts, ce qui reflète le fait qu'elle attribue simplement la valeur du point de mesure le plus proche à chaque point de la grille.

La fonction à base radiale (RBF) a produit une surface plus lisse, indiquant qu'elle est capable d'estimer des variations plus graduelles de la pollution.

Pour résoudre le problème des interpolations linéaire et cubique, il est possible d'ajuster les limites de la grille pour qu'elle ne s'étende que sur la zone couverte par les points de données afin de fournir une valeur pour chaque point de la grille. Une autre méthode consiste à ajouter des points de données aux limites. Ainsi, en ajoutant une valeur appropriée, les méthodes linéaire et cubique peuvent générer une surface d'interpolation qui couvre toute la grille. Bien que la résolution du problème des interpolations permette d'afficher des résultats, les conditions aux limites sont difficilement acceptables dans un cas réel. Il est donc intéressant de se pencher sur d'autres méthodes plus complexes pour éviter de fixer des conditions aux limites qui pourraient éloigner les résultats de la réalité. Voici une proposition de trois méthodes d'interpolation connues :

1. **L'interpolation par inverse de la distance pondérée (IDW) :** elle attribue des poids aux points de mesure en fonction de leur distance par rapport au point

d'interpolation. Les points les plus proches ont un poids plus élevé, tandis que les points plus éloignés ont un poids plus faible. Les valeurs interpolées sont calculées en utilisant une combinaison pondérée des valeurs des points de mesure ;

2. **Le krigeage :** c'est une méthode d'interpolation plus avancée qui utilise des techniques de géostatistique. Elle prend en compte à la fois la distance entre les points de mesure et la variation spatiale des données. Le krigeage estime les valeurs interpolées en minimisant l'erreur de prédiction en utilisant un modèle mathématique basé sur la covariance spatiale des données. C'est à dire en prenant en compte l'interdépendance des valeurs aux différents emplacements ;
3. **Les splines :** ce sont des méthodes d'interpolation couramment utilisées pour modéliser des surfaces continues à partir de données discrètes. Elles peuvent être ajustées à des surfaces régulières en utilisant des fonctions mathématiques lisses. Les splines peuvent être univariées (pour une seule dimension) ou bivariées (pour deux dimensions) en fonction des besoins.

La Figure 3.6 présente à nouveau une comparaison des méthodes avec pour l'interpolation linéaire et cubique des points aux limites avec la valeur égale à zéro. L'interpolation par splines n'a pu être effectuée car il faut au minimum 16 points pour commencer à créer une surface et il n'y a que 3 points dans le cas présent. Cette méthode sera préférée dans le cadre d'une forte densité de points. La méthode IDW estime les valeurs en un point donné en calculant la moyenne pondérée des valeurs des points voisins, où les poids sont inversement proportionnels à la distance au point d'estimation. Cette méthode ne provoque pas de surajustement contrairement à la méthode RBF et cubique où les estimations de pollution dépassent la valeur maximale des données. En effet, l'interpolation cubique utilise des polynômes de degré trois pour interpoler entre les points de données. Les polynômes cubiques peuvent avoir des oscillations locales importantes, surtout lorsqu'ils sont ajustés à des points de données qui changent rapidement. Cela peut conduire à des valeurs qui dépassent les limites des données d'origine. Quant à la méthode RBF, elle utilise une combinaison de fonctions radiales pouvant produire des oscillations importantes et ainsi créer des valeurs qui dépassent les limites des données d'origine. En d'autres termes, chaque fonction radiale sert à approximer la distribution des données autour d'un certain point central. Dans cet exemple, la fonction multiquadratique utilisée dans l'interpolation RBF est :

$$f(r) = \sqrt{(r/c)^2 + 1}$$

où r est la distance à un point donné et c est une constante positive appelée paramètre de forme. Et pour le variogramme sphérique utilisé dans l'interpolation par Kriging :

$$\gamma(h) = \begin{cases} c_0 + c_1 \left(\frac{3h}{2a} - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right) & \text{si } h \leq a \\ c_0 + c_1 & \text{si } h > a \end{cases}$$

où :

— h est la distance entre deux points ;
— a est la portée (la distance à laquelle la corrélation entre deux points devient nulle) ;
— c_0 est l'effet de pépite (la discontinuité à l'origine qui représente la variabilité à petite échelle) ;
— c_1 est la sill (la variabilité totale, qui est la somme de l'effet de pépite et de la variabilité structurelle).

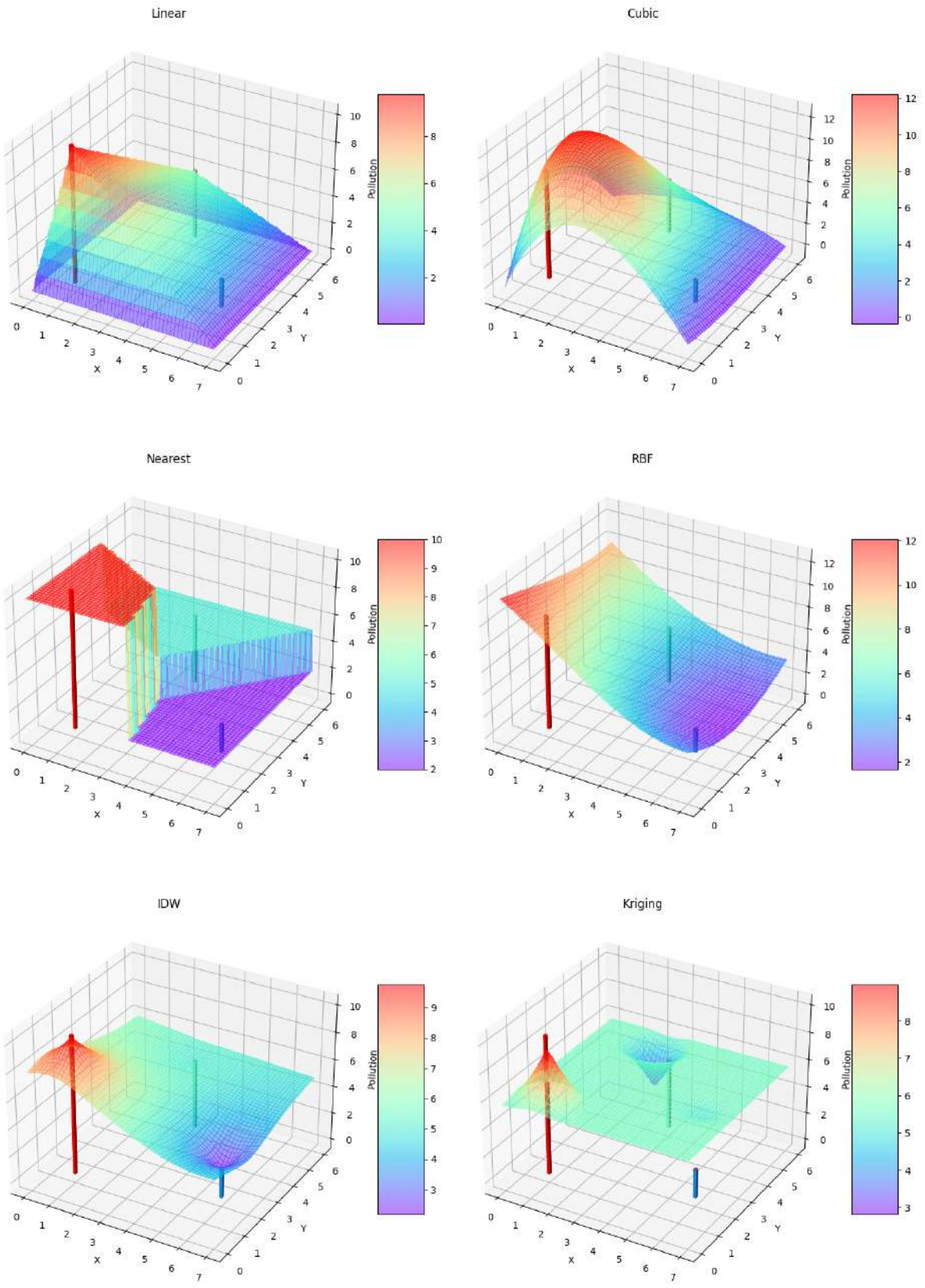


FIGURE 3.6 – Représentation des méthodes d'interpolation linéaire simple et cubique ajustées, Nearest et RBF

En résumé, les méthodes d'interpolation présentent divers avantages et inconvénients, conditionnés par la nature, la densité et la distribution des données. Ces méthodes sont synthétisées dans la table 3.3.2, où leurs atouts et limites sont énumérés. À titre illustratif, dans un scénario impliquant trois stations de mesure dispersées sur une zone donnée, les méthodes Nearest et RBF apparaissent comme étant les plus efficaces. Leur capacité à estimer des valeurs pour tous les points de la grille et à produire une surface d'interpolation cohérente avec les données initiales souligne l'importance du choix de la méthode d'interpolation sur le résultat final.

En conclusion, le choix de la méthode d'interpolation appropriée est influencé par plusieurs facteurs tels que la nature des données, la topographie du terrain, les conditions météorologiques, la résolution et la structure de la grille du modèle. Dans le contexte de la modélisation de la qualité de l'air par la méthode descendante, l'assimilation de données, qui va au-delà de simples techniques d'interpolation, permet d'optimiser les prévisions en combinant les observations aux sorties du modèle.

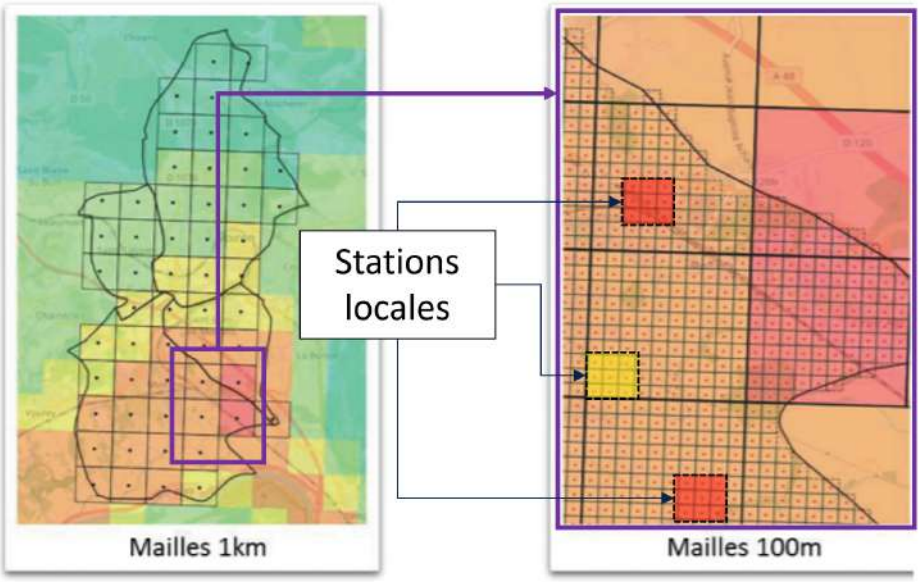


FIGURE 3.7 – Les opportunités de l'approche ascendante dans l'approche descendante. Illustration inspirée du modèle méso-échelle d'ATMO [35].

La Figure 3.7 aborde maintenant un aspect crucial : l'intégration complémentaire des approches ascendante et descendante. Cette figure, inspirée du modèle méso-échelle d'ATMO, illustre comment les deux approches peuvent se compléter pour affiner la précision des cartes de pollution et de qualité de l'air. Les modèles méso-échelle, couramment utilisés dans l'approche descendante, permettent d'affiner la granularité des cartes de pollution en découpant des mailles de 1 km en un maillage plus serré de 100 m, comme illustré par l'encadré violet de la Figure 3.7. L'ajout de stations de mesure en points locaux devient alors un complément précieux à cette démarche. Il offre la possibilité d'affiner la carte de pollution à des endroits stratégiques grâce à des mesures locales. Les deux approches se révèlent donc complémentaires : les conditions aux limites peuvent être celles fournies par ATMO, tandis que les interpolations sont recalculées en fonction des nouvelles mesures obtenues. Ainsi, l'approche ascendante est l'élément central de la conception de l'architecture proposée dans cette thèse.

Méthode	**Avantages**	**Inconvénients**	**Nature des données**	**Densité et distribution**	**Ressources en calcul**	**Ref**
Linéaire	Facilité d'implémentation et cohérence avec une haute densité de données.	Les données doivent suivre une tendance linéaire pour conserver la précision.	Données à tendance linéaire.	Adaptée pour une haute densité de données.	Faible	[43]
Cubique	Modélise des relations plus complexes que l'interpolation linéaire.	Peut surajuster les données, en particulier avec une densité de données faible.	Données à tendance non linéaire.	Adaptée pour une haute densité de données.	Moyen à élevé	[43]
Nearest	Facilité de mise en oeuvre. Pas d'hypothèses sur la tendance des données.	Résultats très imprécis si la densité de données est faible.	Tous type de données.	Adaptée pour une haute densité de données.	Faible	[62]
RBF	Peut modéliser des relations complexes. Ne nécessite pas une densité de données élevée.	Le choix de la fonction de base radiale peut avoir un impact significatif sur les résultats.	Tous type de données.	Adaptée pour une densité de données faible à moyenne.	Moyen	[63]
IDW	Prend en compte la distance des points voisins. Ne nécessite pas une densité de données élevée.	Le choix de la puissance peut avoir un impact significatif sur les résultats.	Tous type de données.	Adaptée pour une densité de données faible à moyenne.	Moyen	[14]
Krigeage	Peut modéliser des relations complexes. Prend en compte la corrélation spatiale des données.	Difficile à mettre en œuvre.	Tous type de données.	Adaptée pour une densité de données faible à moyenne.	Élevé	[48]
Spline	Peut modéliser des relations complexes et peut être plus précise que l'interpolation linéaire ou cubique.	Peut surajuster les données, en particulier avec une densité de données faible.	Données à tendance non linéaire.	Adaptée pour une haute densité de données.	Élevé	[24]

TABLE 3.1 – Résumé des méthodes d'interpolation les plus utilisées

3.3.3 De la mesure locale à l'impact régional : l'approche ascendante

L'architecture mise en avant dans cette thèse repose essentiellement sur l'approche ascendante, intégrant des stations de mesure localisées et des données contextuelles associées à la topographie du site. Cette approche permet dans le cas d'un réseau de mesure ayant des stations disposées localement d'extraire des informations supplémentaires par rapport à l'approche descendante. Ce concept est illustré sur la Figure 3.8. Dans cette Figure, les mesures des stations correspondent à des niveaux de polluant allant de 0 à 10. Pour illustrer l'aspect spatial, en bas de la Figure sont représentées les mesures des trois stations disposées sur un périmètre. L'axe vertical correspond au temps et les mesures sont prises à trois instants distincts nommés (t1, t2 et t3). Les graphiques représentant les mesures sont superposés de manière que l'évolution de mesure de pollution puisse être visualisée dans l'espace. En haut de la Figure sont représentées les mesures de chaque station dans le temps pour illustrer l'aspect temporel de l'architecture.

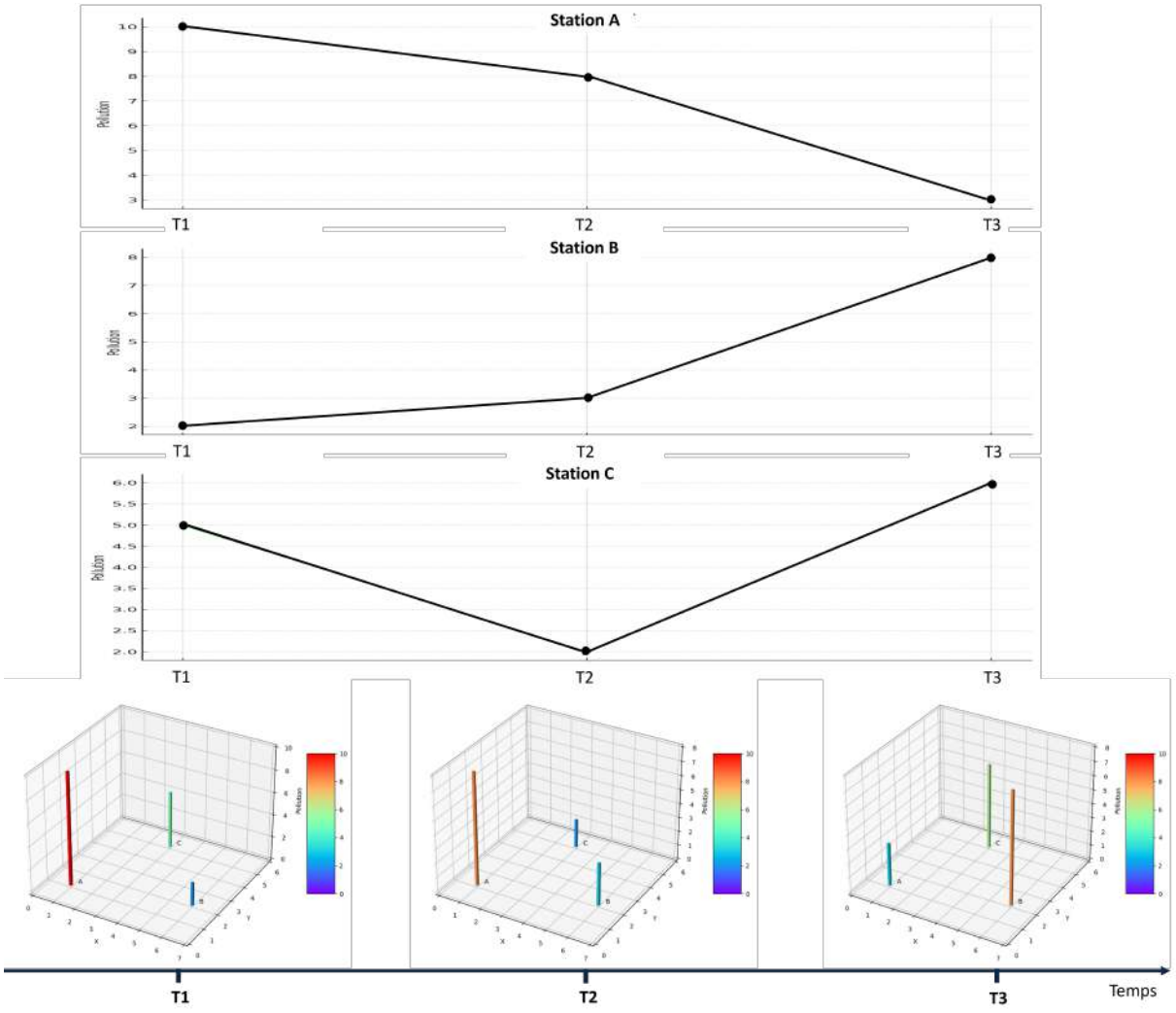


FIGURE 3.8 – Représentation de l'approche ascendante avec trois stations de mesure

Les dimensions spatiale et temporelle sont intrinsèquement liés lors de la surveillance d'un territoire portant l'ensemble des dimensions à 4. Ainsi, sur cet exemple, il est constaté que les mesures de la station C montre que la qualité de l'air diminue avec le temps puisque les niveaux de polluants augmentent. A l'inverse, la qualité de l'air s'améliore sur les trois instants et on ne peut pas conclure pour la station C. En combinant l'aspect topologique et les éléments météorologiques, l'augmentation de la pollution au niveau de la station B pourrait être expliquée soit par un phénomène concentratif ou bien un déplacement de la pollution de la station A vers la station C. Cette hypothèse peut être confirmée par d'autres mesures telles que la direction du vent et sa force.

Cet exemple souligne à la fois les avantages et les défis de l'approche ascendante. Sans une connaissance approfondie des facteurs locaux, il est difficile d'interpréter certains phénomènes ou d'identifier de nouveaux épisodes polluants. Néanmoins, cette méthode se prête bien à une augmentation du nombre de points de mesure, facilitant la détection des mouvements de pollution et des nouvelles émissions. Les enjeux et les possibilités de cette approche sont ainsi mis en exergue. Il convient cependant de souligner que la mise en place de stations de mesure sur l'ensemble du territoire se heurte à des contraintes physiques, budgétaires et humaines, peuvent affecter la précision des interpolations et donc l'efficacité de la surveillance de la qualité de l'air.

3.3.4 Le réseau de mesure dans la dimension spatiale

L'efficacité de la surveillance de la qualité de l'air dépend principalement de deux éléments clés : une couverture territoriale étendue et une mise à l'échelle du réseau de mesure. Cette extensibilité assure que le système maintient ses performances, même face à une augmentation du nombre de stations de mesure et de la demande en données. Ainsi, en prenant en compte la topologie du territoire et ses spécificités, l'architecture mise en place se doit d'être à la fois résiliente et fiable. Son adaptabilité et son évolutivité la rendent apte à satisfaire tant les besoins présents que futurs par l'ajout de nouvelles stations mais aussi par la prise en compte d'autres paramètres physiques.

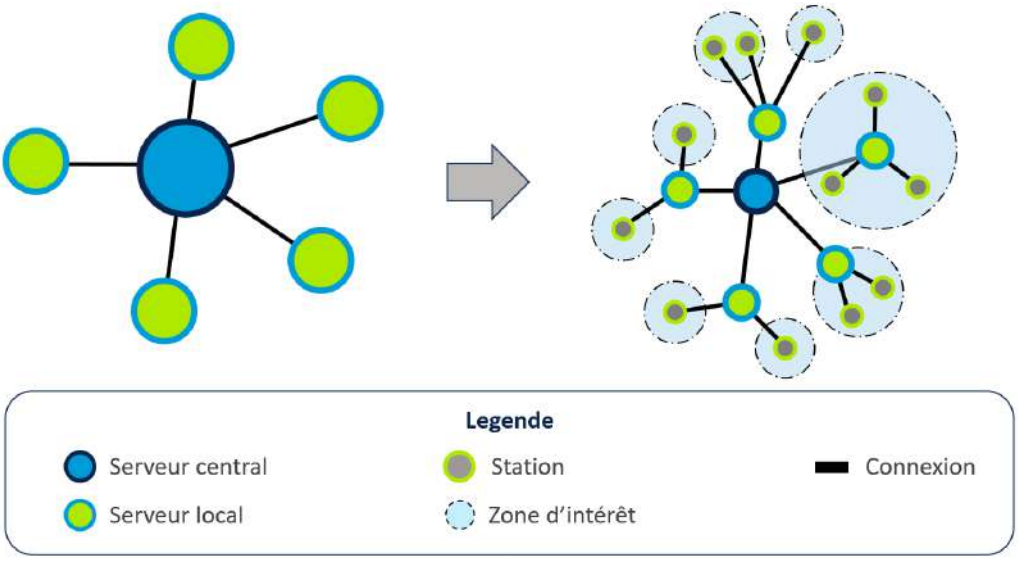


FIGURE 3.9 – Le réseau de mesure dans la dimension spatiale du territoire.

La couverture du territoire s'articule autour d'une hiérarchie de serveurs, depuis le niveau local jusqu'au niveau central, pour une surveillance coordonnée et complète de la qualité de l'air sur le territoire. Le serveur central, ayant la responsabilité de la surveillance globale, communique avec les serveurs locaux afin de créer une carte de concentration d'après toutes les zones d'intérêt et de toutes les stations de mesures réparties sur le territoire. Tout comme les serveurs locaux, il communique via le réseau internet lui permettant d'accéder à toutes les données sur l'ensemble du territoire. La Figure 3.9 illustre la dimension spatiale du réseau de mesure par les connexions entre les serveurs locaux et le serveur central. Les communications sont à double sens entre les serveurs permettant d'échanger des informations. Cela permet d'adapter les données en fonction des responsabilités du client. Par exemple, le serveur central par l'intermédiaire des outils d'aide à la décision va fournir une synthèse sur l'ensemble du territoire et, au niveau du serveur local, une synthèse plus précise de la zone d'intérêt afin de mieux appréhender les pollutions locales.

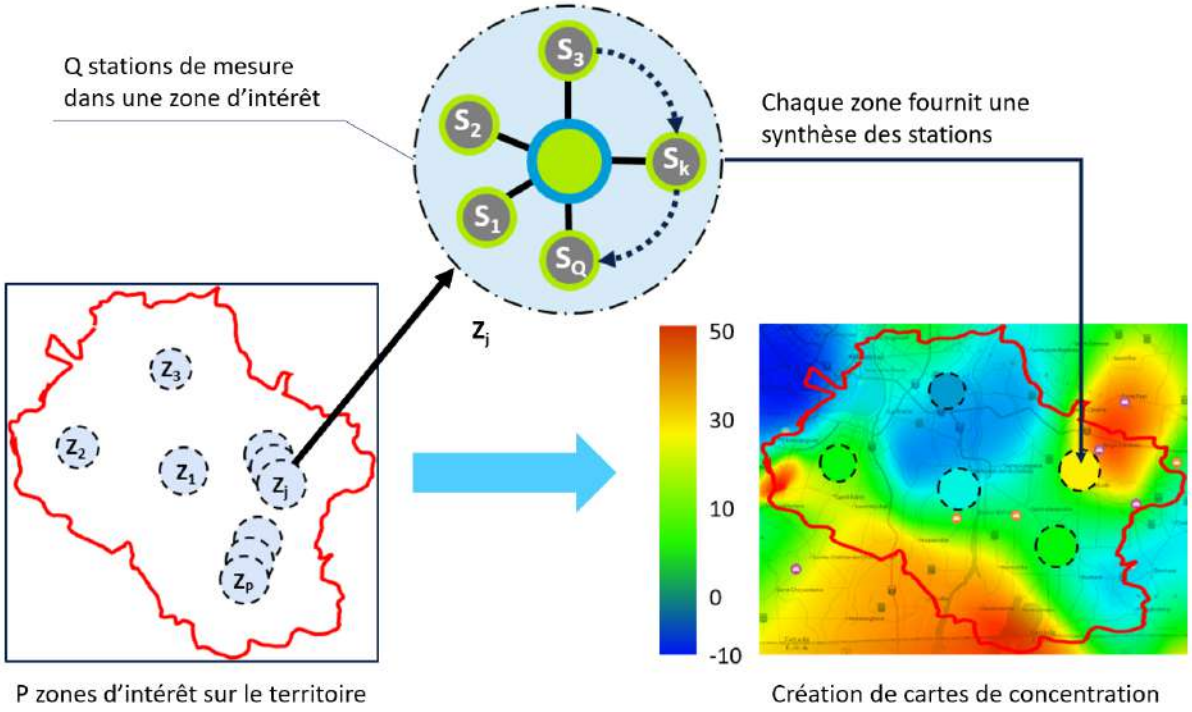


FIGURE 3.10 – Cartes de concentrations issues des points de prélèvement.

Dans la Figure 3.10, la dimension spatiale est prise en compte par l'identification et la création de zones d'intérêt sur le territoire. Au nombre de P, elles ont pour but de définir une zone géographique d'un kilomètre carré avec un objectif à protéger en fonction des polluants émis. Dans le cas d'un objectif de protection de la population, les mesures des concentrations de polluants spécifiques lèvent des alertes afin de signaler aux autorités des anomalies au niveau de la qualité de l'air et ainsi potentiellement agir sur ces émission pour les réduire. Ce périmètre peut être à l'échelle d'une ville, d'une forêt ou d'une zone industrielle par exemple. En effet, l'implantation des stations est normée et dépendante de l'objectif de la mesure. Ainsi, au cœur d'une zone d'intérêt Z_j se trouve le serveur local, qui agit comme un hub central recueillant les informations issues des Q stations de mesures. Il facilite la communication grâce à une interface de programmation

d'application (API), permettant une intégration fluide et un échange de données entre les stations et les outils d'aide à la décision utilisée par les décideurs. Le serveur local communique via internet. Il a la responsabilité des données issues des stations de mesure et également de la maintenance dispensée à ces dernières. La responsabilité du serveur se situe au niveau des détections et des diagnostics dans des supports d'aide à la décision pour les décisionnaires. Pour une surveillance à l'échelle du territoire, les données issues des serveurs locaux sont centralisées et agrégées sur un serveur central. Ce serveur a une responsabilité hiérarchique sur les serveurs locaux et permet en outre de créer des cartes de concentration pour représenter de manière visuelle les concentrations les plus élevées d'une grandeur physique φ_i liée à un polluant.

L'aspect spatial de l'architecture vise à augmenter le nombre de points de mesure pour maximiser la couverture du territoire. Les zones d'intérêts permettent de quadriller le territoire afin de cibler les éléments à protéger. Dans ces zones d'intérêt des points de prélèvements sont identifiés. Ils sont les lieux d'implantation des stations de mesure. Ces points de prélèvements doivent satisfaire deux critères essentiels. D'une part, ils doivent offrir la possibilité d'installer physiquement les stations de mesure. D'autre part, la stabilité des conditions environnementales doit être garantie assurant ainsi que les LCS fonctionnent dans des conditions comparables. Ce second critère définit le « périmètre de mesure » : un espace où les variables à mesurer demeurent stables à travers tous les points du périmètre, que ce soit un parking, un toit, une cour, ou un jardin public, par exemple.

Les stations de mesure sont mises en place à l'intérieur d'une zone d'intérêt pour mesurer les concentrations de polluants et les paramètres météorologiques en temps réel afin de fournir des données contextuelles essentielles à l'approche ascendante. Désignées par S_k, les stations mesurent les N paramètres physiques définis par les autorités, désignés par φ_i. Ces paramètres peuvent correspondre à des polluants ou à des variables météorologiques. Toutefois, la fiabilité de ces stations de mesure constitue une préoccupation majeure. Elle relève de la dimension temporelle de l'architecture et s'attache à garantir une continuité de service durable et des observations fiables des paramètres physiques considérés. L'absence d'une telle fiabilité compromettrait non seulement la mise en œuvre de l'approche ascendante, mais également la précision des interpolations effectuées.

Après avoir caractérisé la dimension spatiale qui permet d'obtenir une couverture territoriale étendue, il est essentiel de se tourner vers la dimension temporelle. En effet, les stations monitorent l'évolution de la qualité de l'air au fil du temps. Ces mesures en temps réel sont cruciales pour une surveillance continue et précise. La section suivante traite de l'aspect temporel de l'architecture pour la surveillance de la qualité de l'air.

3.4 Observation et monitoring en un point de prélèvement

Les stations de mesure sont le socle de la dimension temporelle de l'architecture avec des mesures en temps réel des concentrations de polluants et des paramètres météorologique sur un point de prélèvement. Elles jouent un rôle important dans la surveillance de la qualité de l'air. L'aspect temporel de la surveillance de la qualité de l'air par l'intermédiaire des stations de mesure englobe la période de surveillance, la résolution temporelle, la représentation des données dans le temps et la continuité de la surveillance.

La Figure 3.11 donne une vue schématique de cette architecture centrée sur la dimension temporelle. Chaque station de mesure permet de récolter les mesures en temps réel

de manière continue d'un paramètre physique tel qu'une concentration de polluant ou un paramètre météorologique. Les données historiques sont archivées pour des analyses rétrospectives et les études comparatives entre les stations pour les cartes de concentration. Également, des modèles de prédiction peuvent être appliqués pour anticiper les évolutions. Dans ce contexte, la résolution temporelle, l'étendue de la période de surveillance, la modélisation des données temporelles, ainsi que la continuité de la surveillance constituent ensemble les bases de la dimension temporelle de l'architecture proposée. Les sous-sections suivantes explorent les divers aspects en détail, afin de saisir toute la complexité et les enjeux associés à la dimension temporelle dans la surveillance de la qualité de l'air.

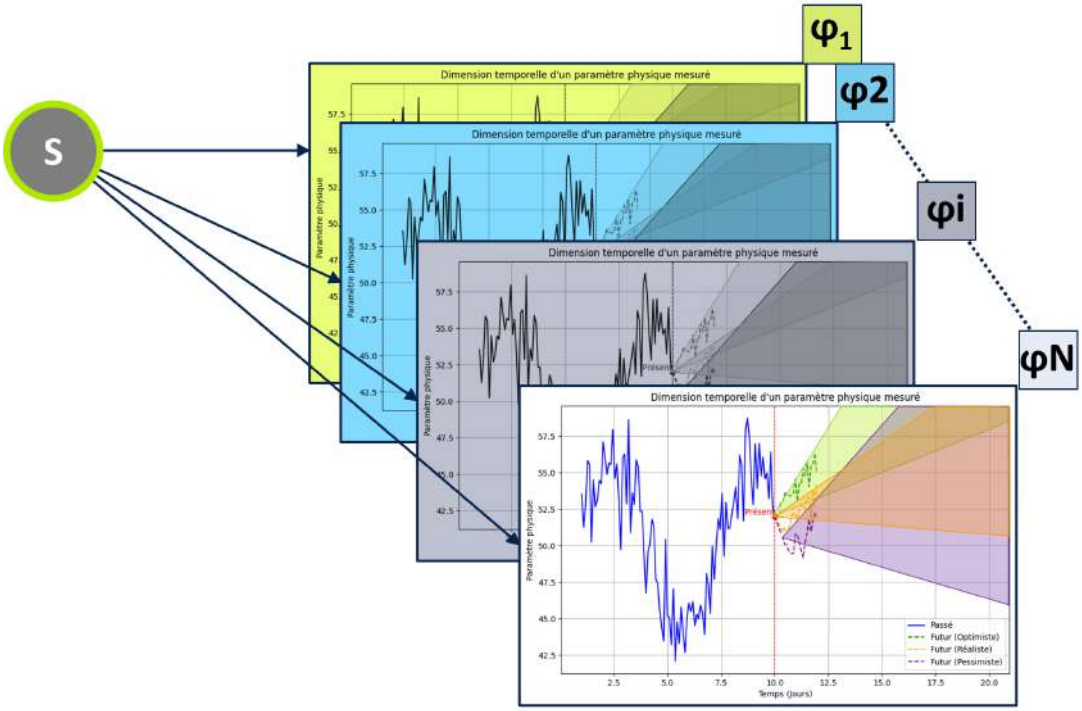


FIGURE 3.11 – La dimension temporelle multi-paramètre des stations de l'architecture.

3.4.1 Exploration des méthodes d'interpolation dans le cadre d'un réseau temporel de surveillance de la qualité de l'air

Les méthodes d'interpolation sont utilisées non seulement pour estimer les valeurs entre des points de mesure existants, mais également pour élaborer un modèle capable de servir de base à des prédictions futures de paramètres physiques. Il est donc crucial de sélectionner une méthode d'interpolation bien adaptée aux données à traiter. Cette sous-section a pour objectif d'explorer diverses méthodes d'interpolation appliquées à une série temporelle de concentrations de PM10. L'application de méthodes d'interpolation en dimension temporelle permet de modéliser les points de mesure en vue de réaliser des prédictions. Cette exploration est d'autant plus pertinente qu'elle s'inscrit dans un contexte opérationnel : le réseau de surveillance de la qualité de l'air et plus précisément des observations issues des stations de mesure. Différentes méthodes d'interpolation sont évaluées pour déterminer leur pertinence et leurs limitations, en particulier en ce qui concerne leur application à des données de concentration de pollution de PM10 à caractère temporel.

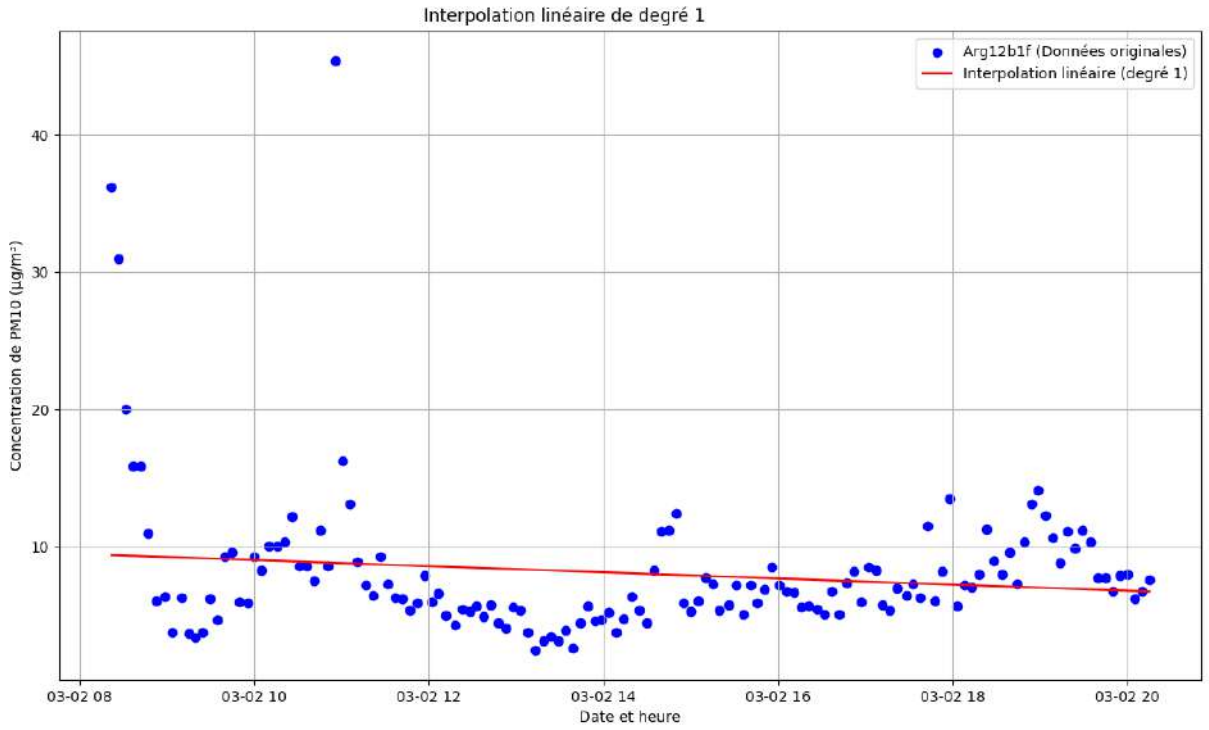


FIGURE 3.12 – Interpolation de degré 1 sur des données brutes de PM10

La Figure 3.12 montre une représentation d'une série temporelle des valeurs de PM_{10} mesurées par le capteur. Les points en bleu représentant les concentrations de PM_{10}. En rouge, une interpolation de degré 1 a été appliquée sur les données brutes. Elle permet d'avoir approximation initiale de la tendance générale des données mais ne saisit pas les fluctuations fines. Cette méthode, par sa simplicité, sert de point de référence pour les méthodes plus complexes qui suivront.

La Figure 3.13 montre une visualisation de plusieurs interpolations polynomiales de degrés supérieurs associées aux données brutes. Des interpolations polynomiales de degrés 2, 4, 8 et 16 ont été réalisées. Dans ces graphiques, l'augmentation du degré de l'interpolation révèle une adéquation croissante aux données. Néanmoins, lorsque le degré de l'interpolation polynomiale devient trop élevé, le modèle commence à s'adapter de manière trop précise aux données disponibles. Ce niveau de précision peut sembler bénéfique à première vue, mais il pose en réalité un problème : le modèle commence à s'ajuster non seulement aux caractéristiques fondamentales des données, mais aussi au bruit statistique qui peut être présent. Ce bruit peut être dû à des erreurs de mesure, des fluctuations naturelles à court terme, ou d'autres facteurs non représentatifs de la tendance générale que l'on souhaite étudier. En conséquence, le modèle perd sa capacité à généraliser à partir de nouvelles données et devient moins utile pour des prédictions ou des analyses futures.

Sur la Figure 3.14, une interpolation polynomiale de degré 100 a été employée. Si cette méthode épouse plus fidèlement chaque point de donnée, elle génère également des oscillations marquées entre ces points. Ce comportement est typique d'un surajustement. Dans ce cas, la flexibilité excessive du modèle lui permet d'absorber non seulement la tendance réelle des données, mais également le bruit statistique qui y est associé. L'impact négatif de ce surajustement se manifeste lorsque le modèle est utilisé pour des prédictions ou pour généraliser à de nouvelles données. Étant calibré de manière trop précise sur les données initiales, y compris leur bruit, le modèle perd en fiabilité et peut produire des

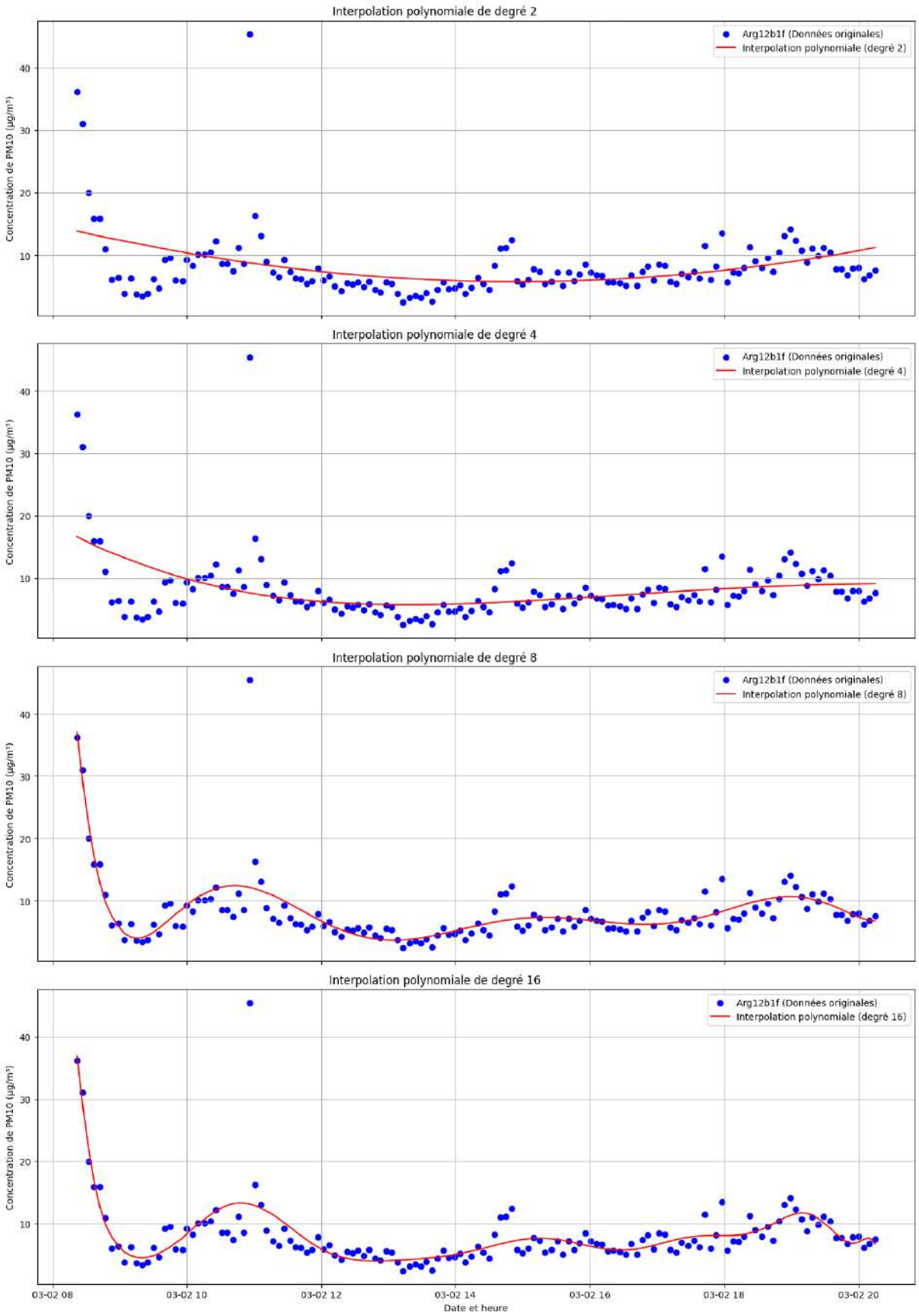


FIGURE 3.13 – Interpolations de degré 2,4,8 et 16 sur des données brutes de PM10

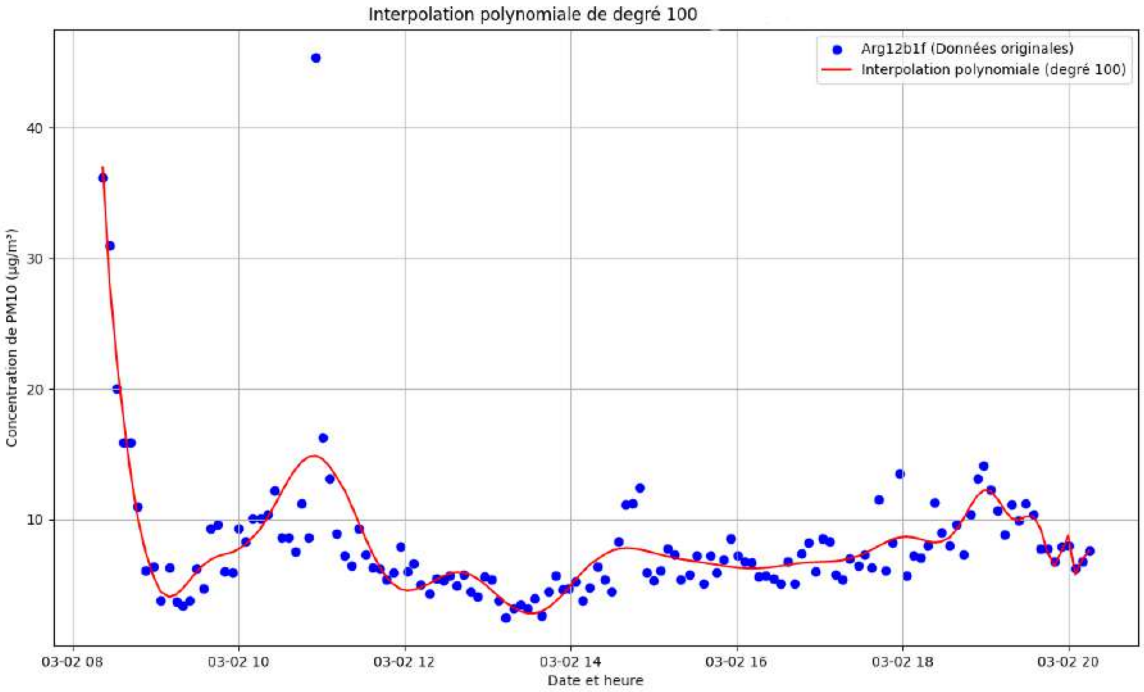


FIGURE 3.14 – Interpolation de degré 100 sur des données brutes de PM10

résultats erronés ou trompeurs. Cette perte de fiabilité est particulièrement problématique dans des applications sensibles comme la surveillance environnementale, où des décisions importantes sont prises sur la base de ces modèles.

Des tentatives d'interpolation avec des degrés encore plus élevés (150, 200) ont mis en lumière des problèmes numériques, tels que l'instabilité des calculs et l'incapacité à trouver une solution. Ces défis supplémentaires soulignent non seulement les limites techniques des méthodes numériques mais mettent aussi en évidence les risques inhérents à l'utilisation de modèles d'une complexité excessive pour des ensembles de données qui ne le justifient pas. Cette exploration des différentes méthodes d'interpolation souligne l'importance d'une sélection réfléchie de la méthode, surtout dans un environnement aussi complexe et dynamique que la surveillance environnementale. Le compromis entre la précision du modèle et sa complexité doit être soigneusement évalué en fonction des objectifs spécifiques du réseau de surveillance.

Le choix de la méthode d'interpolation peut avoir un impact significatif sur la qualité des informations que l'on peut tirer des données, particulièrement dans un réseau de surveillance temporel de la qualité de l'air. Une méthode d'interpolation linéaire simple, par exemple, pourrait ne pas capturer des variations saisonnières ou des événements anormaux ponctuels. Il est à noter que la qualité des interpolations ne dépend pas uniquement du choix de la méthode, mais aussi de la qualité des données elles-mêmes. Des données bruitées ou incomplètes peuvent compromettre la fiabilité de tout modèle interpolé, indépendamment de sa complexité. Par conséquent, une attention particulière devrait être accordée à la réduction du bruit et à l'amélioration de la qualité des données. Dans cette optique, un outil souvent utilisé pour le prétraitement des données est la fenêtre roulante, qui permet de réduire le bruit statistique ou les fluctuations aléatoires présentes dans les données. Dans cette thèse, il a été choisi d'utiliser une fenêtre roulante sans chevauchement pour garantir que chaque point soit conditionné uniquement par les données dans sa fenêtre d'observation, sans redondance. La taille de cette fenêtre est un autre paramètre

clé qui sera examiné dans la sous-section suivante.

3.4.2 Utilisation de la fenêtre roulante pour l'amélioration de la qualité des données et la réduction du bruit

Une technique couramment employée pour améliorer la qualité des données de séries temporelles consiste à utiliser une fenêtre roulante ou fenêtre glissante. Cette méthode vise à lisser les données, c'est-à-dire à éliminer ou réduire les variations à court terme, en calculant, pour chaque point, une mesure statistique (souvent la moyenne) sur un ensemble adjacent de points. Ainsi, chaque point de la série temporelle est remplacé par cette mesure statistique calculée à partir de ses voisins, ce qui permet de réduire les effets des anomalies ou des fluctuations à court terme. L'ensemble en question est défini par la taille de la fenêtre choisie. L'objectif principal de cette approche est la réduction du bruit, ce qui permet une meilleure mise en évidence des tendances sous-jacentes.

Avantages :

1. **Réduction du bruit** : cette technique permet de minimiser les fluctuations aléatoires à haute fréquence dans les données, rendant ainsi la tendance sous-jacente plus apparente ;
2. **Clarification des tendances** : en réduisant le bruit, il devient plus aisé d'identifier les tendances ou motifs pertinents pour l'analyse ;
3. **Optimisation des modèles** : des données lissées peuvent aboutir à des modèles d'interpolation ou de prédiction plus fiables ;
4. **Simplification des données** : le lissage diminue la complexité des données, facilitant leur visualisation et leur analyse ;
5. **Adaptabilité** : la taille de la fenêtre peut être ajustée pour différents niveaux de lissage, permettant ainsi une adaptabilité à divers besoins analytiques.

Inconvénients et précautions :

1. **Perte d'information** : le lissage peut éliminer des détails ou anomalies qui pourraient être cruciaux pour l'analyse ;
2. **Retard temporel** : cette méthode peut introduire un décalage dans la détection de changements soudains ou de points d'inflexion dans les données ;
3. **Choix de la taille de la fenêtre** : le choix de la taille de la fenêtre est un paramètre critique qui doit être soigneusement évalué, car il influence directement les résultats obtenus.

Dans le cadre de la surveillance de la qualité de l'air, l'utilisation d'une fenêtre roulante peut s'avérer particulièrement utile pour filtrer des variations temporaires qui ne sont pas significatives à long terme, comme des pics de pollution éphémères. La taille de la fenêtre va être choisie en fonction des objectifs de surveillance de pollution. Elle permet également de sous-échantillonner les observations des capteurs pour synchroniser leurs observations et également permettre des détections. Comme le montre la Figure 3.15, une fenêtre roulante de taille τ est utilisée pour sous-échantillonner les observations des capteurs. À chaque instant Ti (i allant de 1 à n) une synthèse des observations est effectuée en utilisant les données dans la fenêtre courante. Cette approche permet non seulement de synchroniser les observations des différents capteurs, mais aussi de détecter les anomalies

liées à la série temporelle à chaque instant. À chaque itération, la fenêtre d'observation est avancée d'un temps τ, ce qui permet un rééchantillonnage continu et une détection d'anomalies en temps réel sur l'ensemble de la base de données.

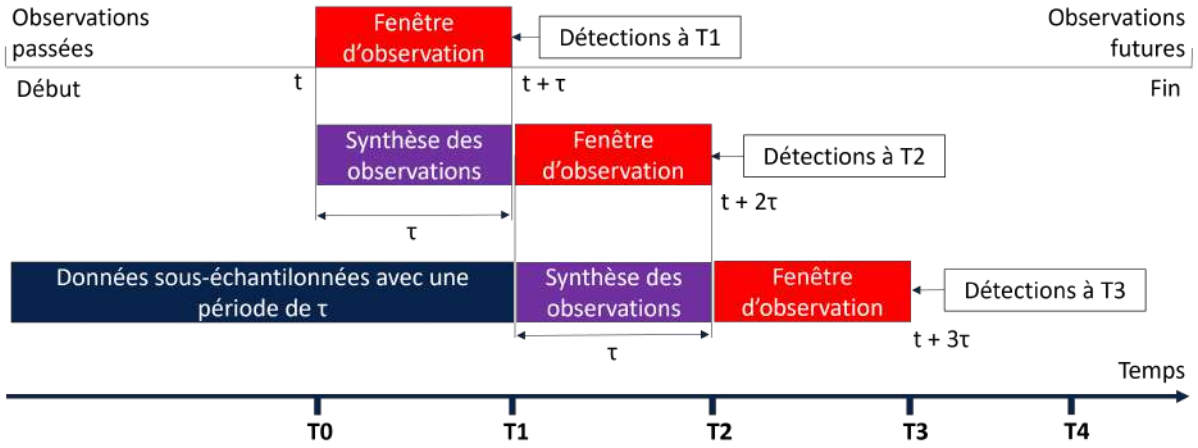


FIGURE 3.15 – Une fenêtre roulante pour un sous-échantillonnage des observations et des détections dans les données.

En conclusion, toutes les méthodes et techniques de prétraitement tirent leur efficacité de la qualité des données initiales. En guise de prétraitement et pour améliorer cette qualité, la technique de la fenêtre roulante est appliquée. Elle permet d'effectuer des observations en continu et de réduire le bruit. Derrière ces techniques et méthodologies se trouve le véritable cœur du système : la conception de la station de mesure. Afin de garantir une surveillance continue et une haute disponibilité, la station se positionne comme l'élément fondateur de l'architecture du réseau de surveillance, tant dans la dimension spatiale que temporelle. Elle doit être conçue pour être fiable, en tenant compte de la durée de vie relative et des possibles dysfonctionnements des LCS. La fiabilité de la première source de données est d'une importance cruciale car elle conditionne la précision et la performance des différents traitements en aval. Les observations se déroulent généralement sur un intervalle d'une heure, tandis que l'échantillonnage des LCS est réalisé à l'échelle d'une minute, en raison de leurs capacités techniques.

3.5 Conception d'une station de mesure en un point de prélèvement

Premier point d'observation des grandeurs physiques et situé sur un point de prélèvement, la station de mesure doit satisfaire des critères élevés de fiabilité et de disponibilité. Dans cette optique, sa conception est inspirée de la TMR, couramment utilisée dans les systèmes critiques, comme le montre le zoom de la station à gauche de la Figure 3.16. La station observe plusieurs paramètres physiques en suivant une dimension temporelle, comme décrit dans les sous-sections précédentes. L'Agrégateur constitue l'élément clé pour archiver les observations, recueillir les données en temps réel et proposer des prédictions. Les SmS en redondance modulaire active permettent de fiabiliser les données et d'augmenter la disponibilité de la station. Dans ce contexte spécifique, la redondance permet de regrouper les capteurs mesurant les mêmes paramètres sous les mêmes conditions, augmentant ainsi la robustesse des observations. Cette méthode se révèle particulièrement

adaptée aux environnements extérieurs. Dans ces conditions, les LCS sont susceptibles de se détériorer rapidement ou de subir des dysfonctionnements imprévus. Cela rend la maintenance complexe et justifie d'autant plus l'utilisation de la redondance modulaire pour assurer une surveillance constante et fiable. Bien que chaque module supplémentaire accroisse le coût de construction, il renforce aussi la fiabilité des détections et la tolérance aux défaillances du système. Les défis liés aux coûts supplémentaires sont résolus grâce à une conception et un assemblage effectués en laboratoire, ce qui permet une meilleure maîtrise des dépenses.

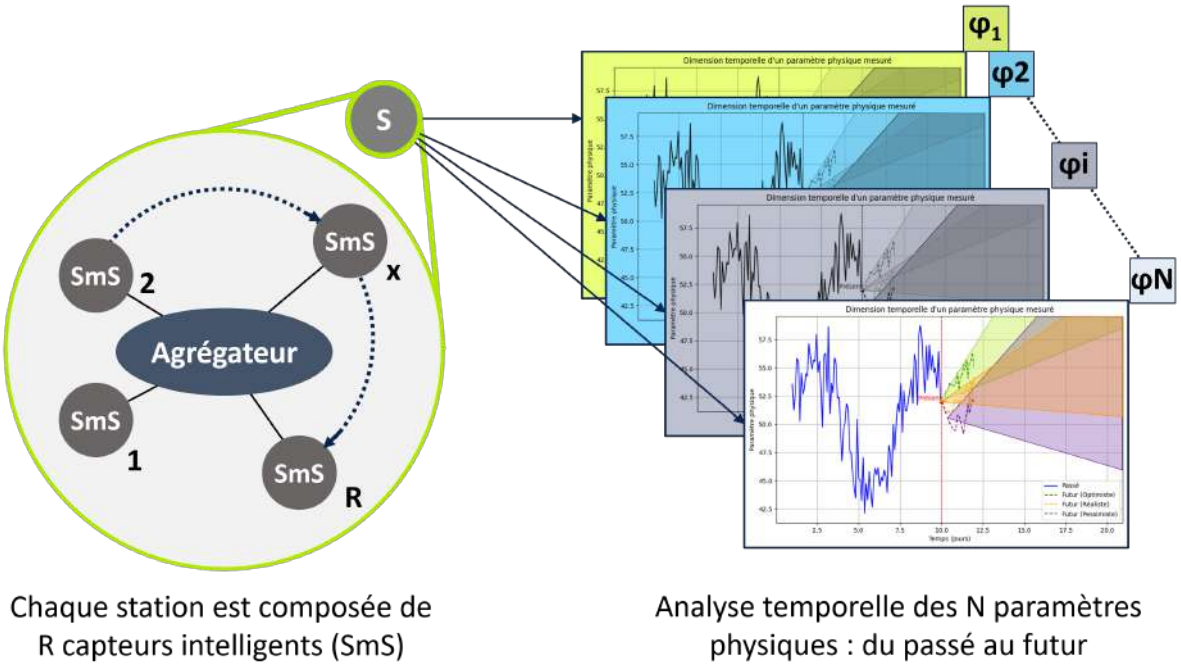


FIGURE 3.16 – Station et observation des paramètres physiques.

La Figure 3.17 détaille les composants de la station et son aspect modulaire. Chaque station est composée d'un Agrégateur et de R capteurs intelligents (SmS), les SmS sont connectés à l'Agrégateur via un réseau local (LAN) sur une distance de l'ordre de quelques mètres. L'Agrégateur effectue tous les traitements nécessaires au niveau de la station pour la rendre autonome. Il représente également le cœur du traitement décentralisé, une technique qui vise à exécuter les traitements aussi proches que possible des composants de la chaîne de mesure. Chaque SmS est composé d'un microcontrôleur et de plusieurs LCS reliés et soudés les uns aux autres par des pistes de l'ordre du millimètre. Le microcontrôleur effectue un prétraitement avec des formules de conversion pour interpréter les mesures de chaque LCS et ainsi construire l'observation des paramètres physiques. Cette configuration matérielle confère aux SmS des fonctionnalités avancées pour le traitement et la communication, leur permettant ainsi de s'interfacer efficacement avec l'Agrégateur et d'être en première ligne de la chaîne de mesure. Pour terminer, l'Agrégateur communique au serveur local les indicateurs Synthèses, Indices et Détection appelés $S.I.D$ (cf. Chapitre 4, page 102). Ces indicateurs correspondent à des synthèses, des indices et des détections calculés par l'Agrégateur à la fin de la fenêtre d'observation, ce qui permet de déterminer la santé de la station et la qualité des données. Une fenêtre d'observation est une fenêtre roulante de taille τ qui collecte les observations de chaque SmS avec sa

propre période d'échantillonnage te_x. À la fin de la fenêtre roulante, un indicateur $S.I.D$ est fourni sur la base des observations recueillies pendant τ.

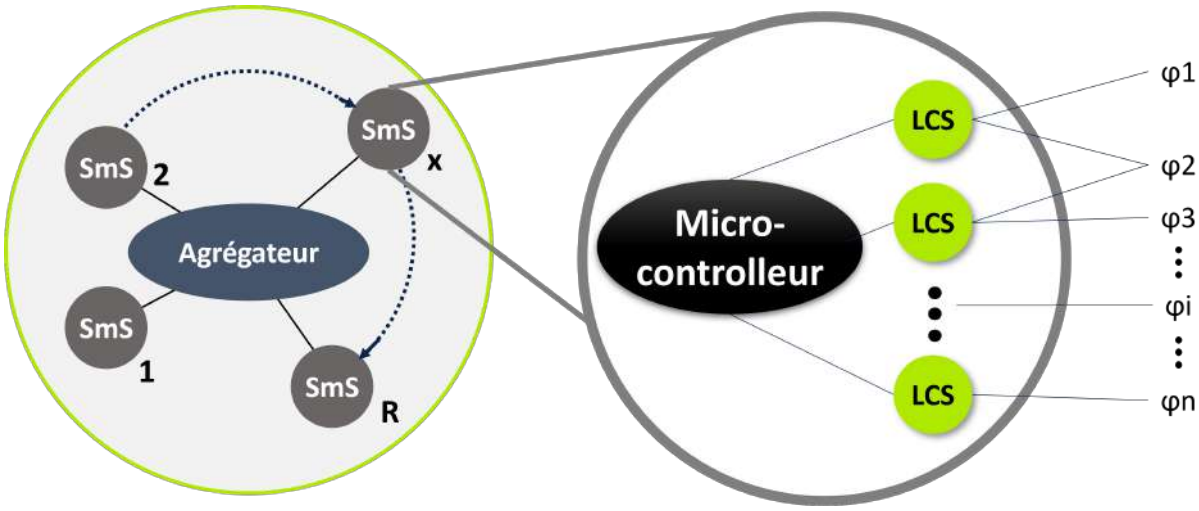


FIGURE 3.17 – La conception modulaire redondante de la station de mesure.

3.6 Conclusion du chapitre

Au cours de ce chapitre, l'architecture proposée pour une surveillance optimisée de la qualité de l'air a été détaillée, mettant en lumière sa dimension spatio-temporelle. La dimension spatiale présente ses propres défis : l'estimation des mesures en tout point du territoire, un approche « ascendante » avec des mesures issues des modèles de l'approche « descendante » et la couverture territoriale rendue possible grâce aux ajouts de stations de mesure. L'approche ascendante est dépendante de la méthode d'interpolation associée, ce qui en conséquence relie la dimension temporelle de l'architecture aux évolutions des paramètres physiques au cours du temps.

La dimension temporelle est caractérisée par des observations en temps réel des stations de mesure sur les points de prélèvement. Des méthodes d'interpolation et des techniques de réduction du bruit existent pour l'ensemble des traitements de la chaîne d'acquisition, servant à extraire des modèles sur les données passées et à établir une base pour des prédictions. La fiabilité et la qualité des données provenant des stations de mesure ont été identifiées comme fondamentales pour les traitements relatifs aux dimensions spatiale et temporelle de l'architecture.

L'architecture englobe la couverture du territoire, la répartition géographique des zones d'intérêt, et l'ensemble de la chaîne d'acquisition, en commençant par les LCS pour la conception des stations de mesure, jusqu'aux serveurs locaux et centraux présentant les interfaces utilisateur en fonction des décisionnaires. Ces éléments fournissent un aperçu des composantes essentielles pour une surveillance efficace.

Ainsi, ce chapitre a établi les fondations d'une architecture de surveillance fiable. Le chapitre suivant traitera plus en détail des données et des traitements, afin que les décisionnaires puissent disposer d'un support de décision, que ce soit pour l'action sur le territoire en ce qui concerne les polluants ou bien pour la disponibilité du réseau de surveillance de la qualité de l'air.

Chapitre 4

Données, Traitements et Décisions

Sommaire

4.1 Introduction 85
4.2 Données de surveillance de la qualité de l'air 86
4.3 Fiabilisation des données issues des stations de mesures . . . 90
4.3.1 Description du processus d'acquisition et de traitement des données depuis les SmS jusqu'au serveur local 91
4.3.2 Formalisation mathématique de la fiabilisation des données . . 93
4.4 Conclusion du chapitre Données, Traitements et Décision . . 102

4.1 Introduction

Le chapitre précédent a établi les fondations d'une architecture adaptée à la surveillance de la qualité de l'air, soulignant l'importance de la couverture territoriale et de la dimension temporelle. Alors que cette architecture facilite une collecte efficace des données, leur traitement et leur évaluation demeurent essentiels pour une surveillance précise et pertinente.

Ce chapitre explore le cœur du système qui consiste à obtenir des données fiabilisées de surveillance de la qualité de l'air. Il explore la chaîne d'acquisition, et les différents traitements pour la fiabilisation des données afin de fournir un support d'aide à la décision pour les autorités. L'objectif est de fournir des synthèses issues des stations de mesures accompagnées d'indices de confiance permettant à la maintenance d'agir efficacement sur le réseau de mesure pour assurer un bon niveau de fiabilité sur le long terme et pour les élus d'avoir des données fiables pour les différentes prises de décision.

L'examen portera d'abord sur les paramètres physiques choisis pour la collecte de données, en tenant compte des priorités établies par les décisionnaires. Le processus global de fiabilisation des données sera ensuite détaillé, permettant de créer des données contextuelles associées aux mesures de pollutions pour une meilleure prise en compte des aspects localisés des émissions.

4.2 Données de surveillance de la qualité de l'air

L'approche proposée dans cette thèse est un processus de traitement des données comprenant la chaîne d'acquisition, de traitement et d'évaluation des données composée de six blocs comme le montre la figure 4.1. Chaque bloc constitue une étape du processus d'acquisition et de traitement des données de surveillance depuis la collecte des grandeurs physiques, leur traitement et leur évaluation en collaboration avec des experts, jusqu'à leur utilisation comme support de décision.

1. Stations de mesures : les SmS de chaque station de mesure créent des matrices temporelles d'observation issues des mesures des paramètres physiques des LCS ;
2. Fiabilisation des données : au sein de l'Agrégateur, les observations des LCS sont ensuite fiabilisées pour en extraire des indicateurs ;
3. Évaluation : les indicateurs sont analysés et comparés avec des seuils définis par les experts. Ces indicateurs permettent la détection et le diagnostic pour la maintenance des stations, ainsi que le calcul d'un indice de confiance pour la synthèse ;
4. Décision : les diagnostics sont analysés et conduisent à une identification du composant défaillant auprès de l'équipe de maintenance. Cette équipe planifiera une intervention sur les stations de mesure ;
5. Modélisation : les synthèses validées sont utilisées afin de construire un modèle basé sur les historiques de données et les détections de pollution passées. Des détections peuvent être définies à partir des comportements passés ;
6. Prédiction - Diagnostic Qualité de l'air : en fonction des analyses des experts de la qualité de l'air, les détections mènent à un diagnostic et des actions à mener sont proposées aux élus.

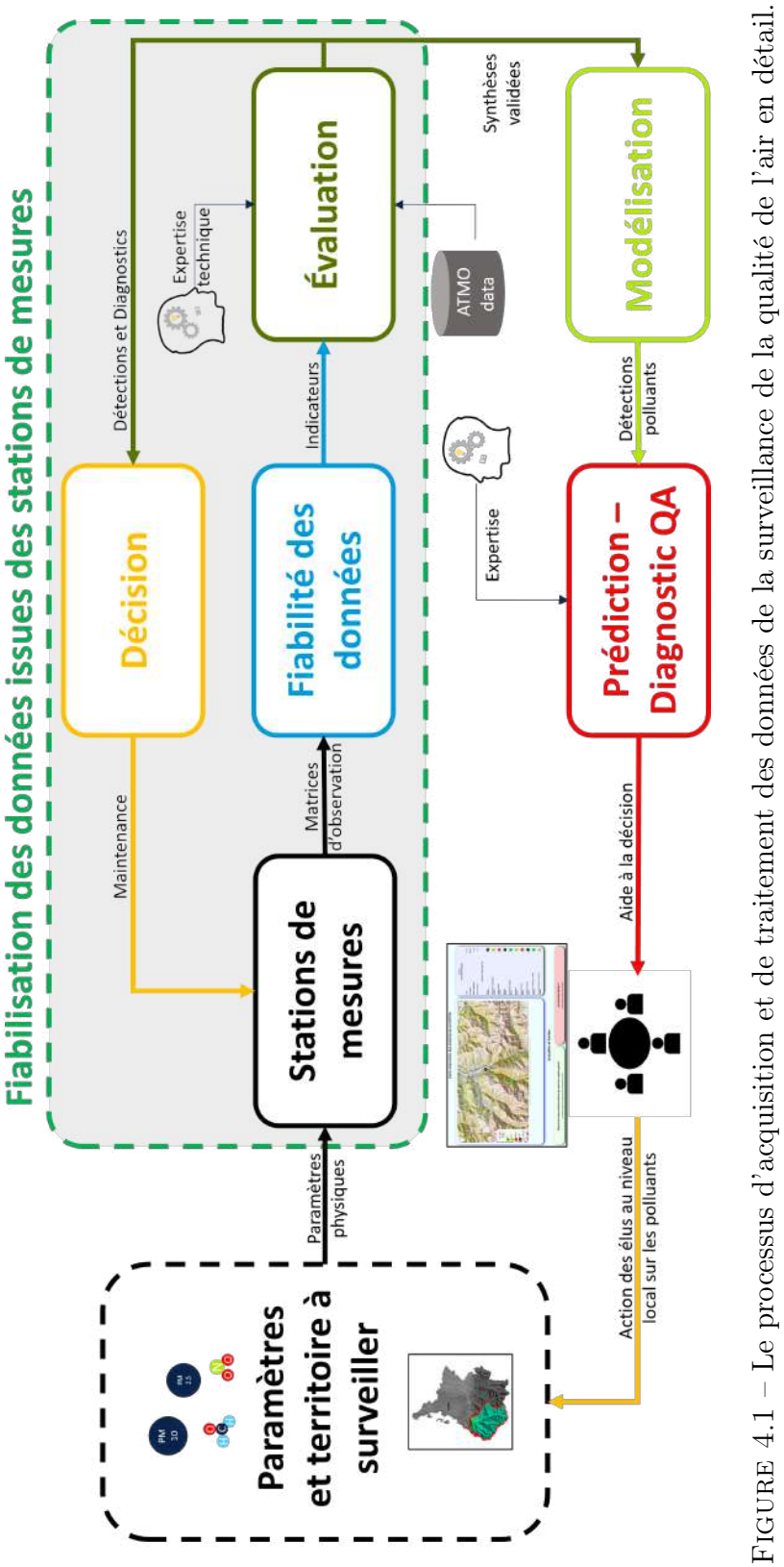


FIGURE 4.1 – Le processus d'acquisition et de traitement des données de la surveillance de la qualité de l'air en détail.

Ce processus permet un contrôle sur l'ensemble des données et permet d'avoir des supports d'aide à la décision en fonction des parties prenantes du système de surveillance de la qualité de l'air. Dans cette optique de contrôle et de fiabilisation des données, les définitions et les hypothèses de travail sont établies ci-après. L'objectif est de fournir une base théorique pour mieux comprendre à la fois l'architecture et le processus d'acquisition et de traitement des données.

L'architecture proposée se concentre à la fois sur la couverture spatiale grâce à de multiples points d'échantillonnage et sur les mesures temporelles en temps réel des polluants et des conditions météorologiques. Cette stratégie identifie des zones d'intérêt, comme des villes ou des caractéristiques géographiques spécifiques, pour une surveillance ciblée. Chaque zone dispose d'un serveur local qui collecte les données, communique via une API pour une intégration avec l'interface utilisateur des décideurs, et se connecte via internet. Au sein de chaque zone, plusieurs stations fournissent des données en temps réel sur les polluants et les paramètres météorologiques. Ces stations mettent en évidence l'accent temporel de l'architecture, avec des observations continues utilisant une fenêtre d'une heure en continu. En général, les LCS échantillonnent toutes les minutes en raison de leurs capacités. Sur la base de cette description, les paramètres suivants sont utilisés. Pour une liste plus détaillée des symboles et des définitions, une section Nomenclature est incluse au début de la thèse.

— P : nombre de zones d'intérêt ;
— Z_j : zone d'intérêt spécifique ;
— Q : stations de surveillance dans une zone ;
— S_k : station de mesure spécifique ;
— N : paramètres physiques définis par les autorités ;
— φ_i : paramètre physique spécifique.

Chaque station est composée d'un Agrégateur et de R SmS connectés via un réseau local (LAN). L'Agrégateur, incarnant le concept de l'informatique en périphérie (edge computing), traite les données à la source. Chaque SmS est composé d'un microcontrôleur et de plusieurs LCS. Les SmS réalisent des observations et transmettent les données brutes selon leurs périodes d'échantillonnage spécifiques, notées te_x. L'Agrégateur envoie des indicateurs, appelés $S.I.D$ (Synthèses, Indices et Détection), au serveur local, reflétant l'état de santé de la station et la qualité des données. Ces indicateurs sont basés sur une fenêtre d'observations roulante. En analysant les sorties des LCS, le système peut identifier les problèmes matériels. En raison de la détérioration rapide des LCS en extérieur, une redondance modulaire est mise en œuvre avec un nombre R de SmS pour garantir la santé de la station. Les paramètres utilisés à chaque niveau de station sont les suivants :

— R : nombre de capteurs intelligents dans une station ;
— $S.I.D$: indicateurs (Synthèses, Indices et Détection) provenant de l'Agrégateur ;
— τ : durée de la fenêtre roulante pour la collecte de données ;
— te_x : période d'échantillonnage spécifique de chaque SmS.

Après avoir défini les paramètres et les composantes clés de l'architecture de surveillance, les hypothèses de travail sont définies ci-après pour une meilleure compréhension de la méthode de fiabilisation des données proposée au sein de l'Agrégateur dans cette thèse.

1. La calibration métrologique des capteurs n'est pas l'objet de cette thèse. En raison des contraintes du laboratoire, les calibrations LCS du fabricant sont utilisées. Des ajustements sont toutefois effectués dans une « salle blanche » sur deux jours à l'aide de données connues. Il est donc supposé que les mesures obtenues à partir du SmS sont suffisamment précises.
2. La qualité de l'air en extérieur est étudiée avec un intérêt particulier porté à l'exposition sur le long terme, omettant ainsi les pics de pollution de courte durée. Étant donné la lenteur du système physique relatif à l'air, la fenêtre d'observation τ est fixée à 1 heure.
3. Après avoir effectué des calculs dans la fenêtre τ, la détection de fautes lors d'un déploiement en temps réel prend une heure. Ce délai est considéré comme acceptable pour les besoins de planification de l'équipe d'intervention.
4. L'emplacement du SmS n'influence pas les mesures. Tous les LCS au même point de prélèvement sont supposés enregistrer des mesures identiques.
5. Le SmS échantillonne toutes les minutes, tenant compte des exigences de chauffage des LCS recommandées par le fabricant. Pour une fenêtre d'observation τ, au moins 3 points sont nécessaires pour les comparaisons ($\tau > 2 \times Te_x$). Les SmS sont conçus pour être facilement remplaçables et reconfigurables.
6. L'Agrégateur est un composant autonome qui est continuellement connecté au SmS et au serveur local. Intégré dans le matériel et sur le même réseau que le SmS, cette configuration est un compromis entre évolutivité et résilience au niveau de la station de mesure.
7. Pour construire le vecteur d'erreurs relatives permises (PREV), $R\varepsilon_S$, il est supposé que tous les SmS sont identiques (même LCS). Les erreurs relatives sont basées sur les spécifications du fabricant. La Figure 4.2 illustre les écarts autorisés dans les mesures des LCS. Les valeurs dépassant deux fois l'erreur relative, représentées par $R\varepsilon_i$ pour un paramètre φ_i, sont considérées comme des valeurs aberrantes.
8. Les LCS seront déployés à différents moments, empêchant les défaillances simultanées. Cette stratégie vise à garantir qu'il y a toujours un LCS relativement plus récent fournissant les mesures les plus précises. En pratique, cette hypothèse est étayée en installant les capteurs à des intervalles échelonnés et en les achetant par petits lots.
9. En cas de déconnexions, des valeurs nulles sont attribuées pour préserver l'intégrité des données. Si aucune observation n'est enregistrée pour un paramètre spécifique, une valeur nulle est attribuée, cette dernière étant censée ne survenir que dans des conditions contrôlées.
10. Pour évaluer la fiabilité d'un LCS, un seuil, Pc, est introduit comme étant le ratio de confiance permis. Si le rapport entre les erreurs du capteur et les observations dépasse Pc, le LCS est considéré comme fiable. Ainsi, Pc est crucial pour déterminer l'indice de confiance.
11. Des hyper-paramètres significatifs tels que τ, Pc, et $R\varepsilon_S$ influencent les sorties de l'Agrégateur et surtout ce qui est considéré comme une pollution à long terme. Ces

paramètres initiaux sont ajustés à mesure que davantage de données sont collectées et validées.

Les paramètres utilisés pour la construction du $S.I.D$ sont :

— τ : durée de la fenêtre d'observation ;
— $R\varepsilon_S$: vecteur d'erreurs relatives permises ;
— $R\varepsilon_i$: Valeur d'erreur relative pour le paramètre physique φ_i ;
— Pc : ratio de confiance permis.

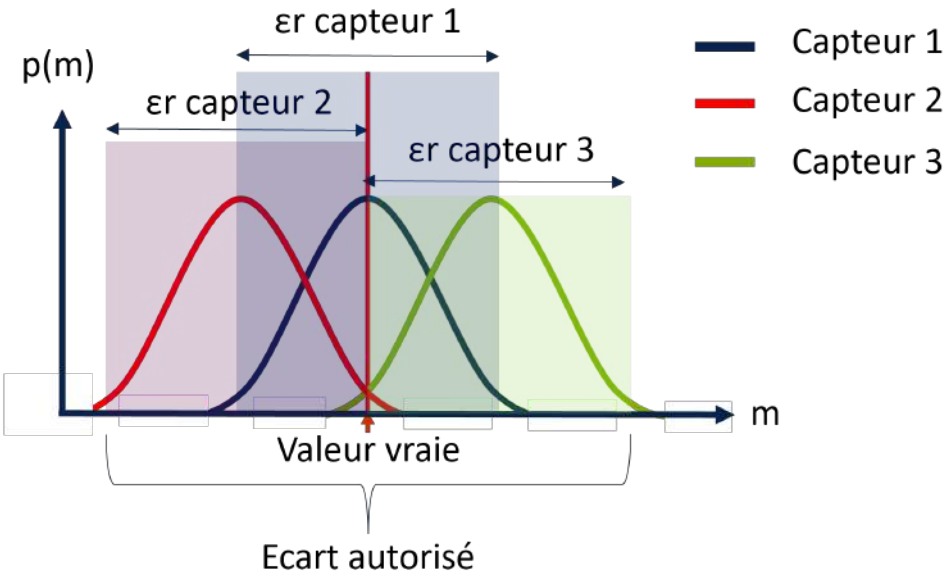


FIGURE 4.2 – De l'erreur relative fournie par le fabricant jusqu'aux écarts autorisés.

4.3 Fiabilisation des données issues des stations de mesures

Cette section se concentre sur la partie du processus de fiabilisation des données, en mettant spécifiquement l'accent sur les blocs *Stations de mesure*, *Fiabilité des données*, *Évaluation* et *Décision* encadrés en vert dans la Figure 4.3. Cet aspect concerne la surveillance du réseau de mesure, en analysant les données pour la détection et le diagnostic du système de surveillance. Dans le bloc *Évaluation*, la partie qui implique une comparaison des données avec les données issues d'ATMO, n'est pas abordée dans cette méthode.

Le bloc *Stations de mesure* concerne la phase d'acquisition des données, où les observations des SmS sont transmises à l'Agrégateur de station via des matrices d'observations temporelles. L'Agrégateur au sein du bloc *Fiabilité des données* analyse ces matrices pour générer des indicateurs $S.I.D$. Ces indicateurs sont ensuite transmis au serveur local, qui les évalue pour extraire les détections et les diagnostics en vue de leur intégration via des outils d'aide à la décision. La responsabilité de la prise de décision incombe à l'équipe de maintenance, qui planifie les interventions sur les LCS défectueux. De plus, une synthèse annotée, accompagnée d'un indice de confiance, est extraite dans le bloc *Évaluation* pour faciliter les comparaisons avec d'autres stations connectées au serveur local.

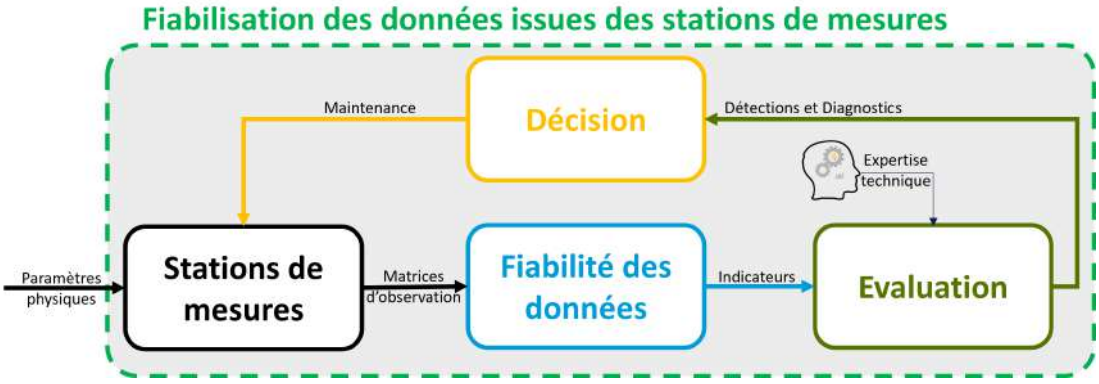


FIGURE 4.3 – Focus sur la fiabilisation des données issues des stations de mesures.

4.3.1 Description du processus d'acquisition et de traitement des données depuis les SmS jusqu'au serveur local

L'Agrégateur joue un rôle central au niveau de la station, agissant comme le lien de communication entre les serveurs locaux et le SmS. Une formalisation mathématique de la méthode mise en œuvre est présentée dans la sous-section suivante. Cette méthode comporte neuf étapes, toutes mises en œuvre dans le nœud Agrégateur en sein d'une station de mesure S_k. Cette méthode est utilisée au sein de chaque station et le PREV peut varier en fonction des LCS utilisés dans les R SmS de la station. La Figure 4.4 présente un résumé des différentes étapes de traitement nécessaires pour générer les indicateurs.

Tous les SmS envoient leurs observations à l'Agrégateur, qui les stocke dans une matrice d'observation temporelle (1). Une fenêtre roulante de taille τ est initialisée pour collecter les observations de tous les SmS afin d'améliorer la fiabilité des données (2). Des sous-matrices sont extraites au sein de la fenêtre d'observation, après un certain temps pour garantir qu'au moins une observation par SmS soit présente (3). Une fois extraites, celles-ci sont vérifiées en termes de taille et de valeurs, garantissant ainsi leur validité tout en détectant les pannes matérielles (4). Les données valides sont concaténées, à l'exclusion des données provenant des LCS bloqués (5). Une matrice de seuil est construite sur la base des erreurs relatives et des valeurs médianes issues de la concaténation (6). Cette matrice est essentielle pour la détection d'anomalies dans les séries temporelles. À l'aide de cette matrice, les observations aberrantes sont identifiées et les résultats sont divisés en deux vecteurs nommés *Reliable* et *Outlier* (7). Enfin, les tailles de ces vecteurs sont utilisées pour construire des indices de défaillance pour la surveillance des LCS. Un algorithme de vote basé sur la médiane synthétise les valeurs du vecteur *Reliable* en une observation qui est ajoutée à la matrice d'observation temporelle à l'échelle de la station (8). Les matrices résultantes sont appelées la matrice de Synthèse, la matrice d'Indices et la matrice de Détection, désignées par l'acronyme $S.I.D$. Elles sont transmises au serveur local, qui les évalue d'abord, en extrait les détections et les diagnostics, attribue un indice de confiance à la synthèse et communique le résultat aux autorités concernées via une interface graphique. Cela permet une surveillance en temps réel et une réponse rapide en cas d'incidents majeurs de pollution.

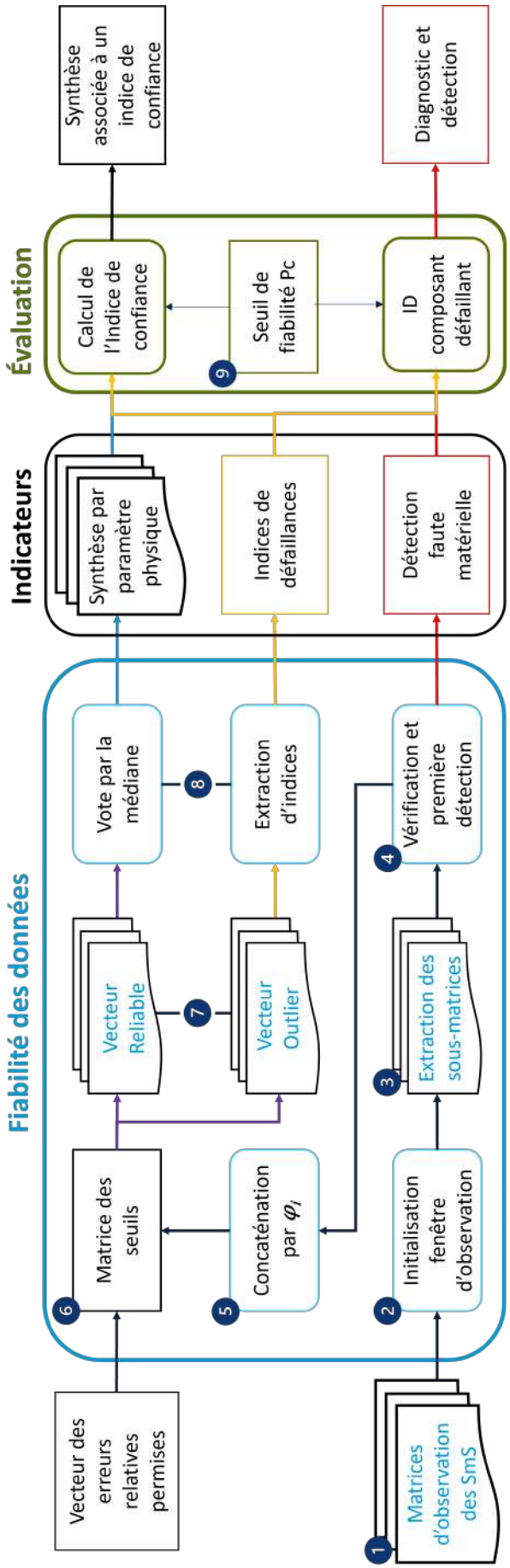


FIGURE 4.4 – Le processus d'acquisition et de traitement des données de l'architecture de surveillance.

4.3.2 Formalisation mathématique de la fiabilisation des données

Dans cette sous-section, la formalisation mathématique détaillée de la méthode est présentée, fournissant une compréhension approfondie de sa logique et de sa rigueur. Ce processus implique la formation de trois matrices clés : la matrice de synthèse, la matrice d'indicateurs et la matrice de détection, collectivement appelées $S.I.D$.

Étape 1 : construction des matrices d'observations temporelles

Tous les SmS de la station envoient leurs observations à l'Agrégateur pour les stocker dans une matrice temporelle notée Y_x. Cette matrice est également connue sous le nom de données brutes ou données historiques pour un SmS x. Les mesures physiques ne sont pas synchronisées car chaque SmS a sa propre période d'observation indépendante.

Pour chaque station de mesure, chaque SmS est associé à une matrice temporelle d'observations notée Y_x. Cette matrice contient M_x lignes qui représentent les observations historiques du SmS x, mesurant chacune N paramètres physiques, ce qui donne une dimension de $M_x \times N$. La Figure 4.5 montre un exemple de matrice d'observations pour le SmS x mesurant trois paramètres physiques. En raison de leur conception ne comportant pas d'horloge synchronisée, chaque SmS autonome et indépendant ne réalise pas de mesures physiques en synchronisation précise, ce qui entraîne un décalage des mesures prises par chaque SmS par rapport au même instant de temps. Par conséquent, le paramètre physique mesuré par le SmS numéro x au temps k est noté $y_x^{\varphi_i}(k)$, où $k \in 1, ..., M_x$ et $y_x^{\varphi_i}(k) = y_x^{\varphi_i}(t)$ avec $t = k \times te_x$. Une observation des N paramètres physiques au temps k est notée $y_{x,k}$ comme indiqué dans l'équation. (4.1). La matrice d'observations Y_x est décrite dans la matrice (4.2).

$$y_{x,k} = \{y_x^{\varphi_1}(k), \ldots, y_x^{\varphi_i}(k), \ldots, y_x^{\varphi_N}(k)\} \tag{4.1}$$

$$Y_x = \begin{bmatrix} y_x^{\varphi_1}(1) & \cdots & y_x^{\varphi_i}(1) & \cdots & y_x^{\varphi_N}(1) \\ y_x^{\varphi_1}(2) & \cdots & y_x^{\varphi_i}(2) & \cdots & y_x^{\varphi_N}(2) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ y_x^{\varphi_1}(k) & \cdots & y_x^{\varphi_i}(k) & \cdots & y_x^{\varphi_N}(k) \\ y_x^{\varphi_1}(k+1) & \cdots & y_x^{\varphi_i}(k+1) & \cdots & y_x^{\varphi_N}(k+1) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ y_x^{\varphi_1}(M_x) & \cdots & y_x^{\varphi_i}(M_x) & \cdots & y_x^{\varphi_N}(M_x) \end{bmatrix} \tag{4.2}$$

Étape 2 : initialisation de la fenêtre d'observation

Une fenêtre roulante de taille τ est initialisée à l'instant t, dans le but de collecter les observations de tous les SmS pendant cette période. La taille de τ est déterminée par la condition (4.3), qui regroupe les mesures du même paramètre physique des SmS pour augmenter la fiabilité des données. La fenêtre roulante ne chevauche pas la précédente. Cela permet de réduire le temps de calcul car seules les données correspondant à la période de temps τ sont traitées. Cette technique permet de confirmer les détections par comparaison et de vérifier les mises à jour matérielles par la validation des données collectées. Cette fenêtre roulante est conçue pour fonctionner à la fois pour les mesures en temps réel et hors ligne.

$$(k + q_x) \times Te_x - (k + 1) \times Te_x \leq \tau \tag{4.3}$$

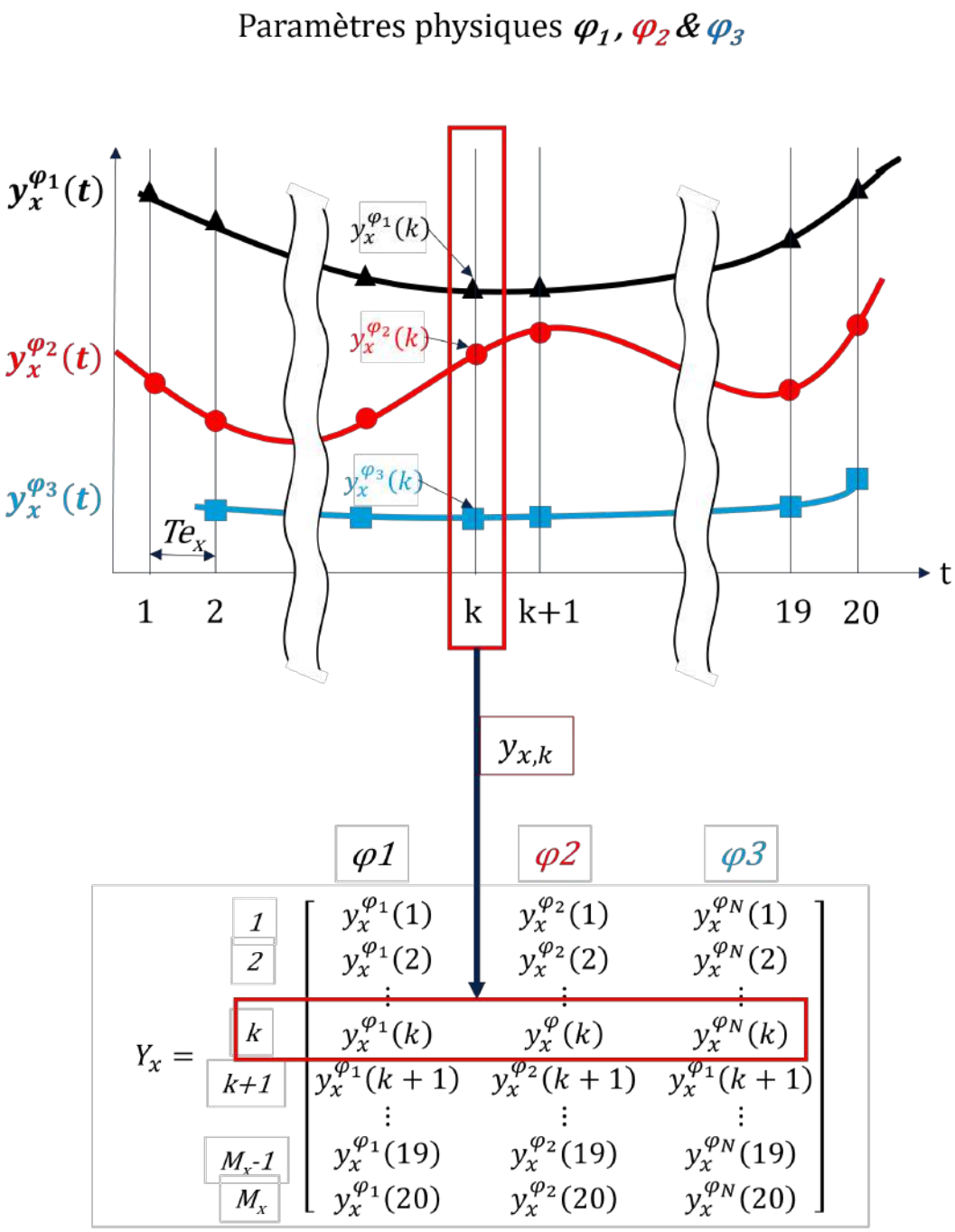


FIGURE 4.5 – Construction de la matrice temporelle d'observation du SmS numéro x.

Étape 3 : extraction des sous-matrices

Des sous-matrices sont extraites de la matrice d'observation temporelle de chaque SmS après un échantillonnage spécifique k, correspondant au début de la fenêtre d'observation t, pour garantir qu'il y a au moins une observation par SmS. Ces sous-matrices sont extraites et liées aux SmS respectifs.

Pour chaque SmS, les observations stockées dans la fenêtre d'observation sont utilisées pour s'assurer qu'il y a q_x observations. La fenêtre d'observation inclut q_x observations du SmS x en raison de sa propre période d'échantillonnage Te_x, qui commence à $t = (k+1)te_x$. Une sous-matrice est extraite de la matrice d'observations temporelles pour chaque SmS après un temps spécifique k, comme illustré sur la Figure 4.6. À l'instant $t = k \times Te_x$, la sous-matrice d'observations $y^{\tau}_{x,k}$ de taille $(q_x \times N)$ est extraite de la matrice temporelle Y_x. Les sous-matrices sont extraites selon la condition (4.3) pour avoir au moins une observation par SmS. En cas de $q_x = 0$, une sous-matrice vide est créée.

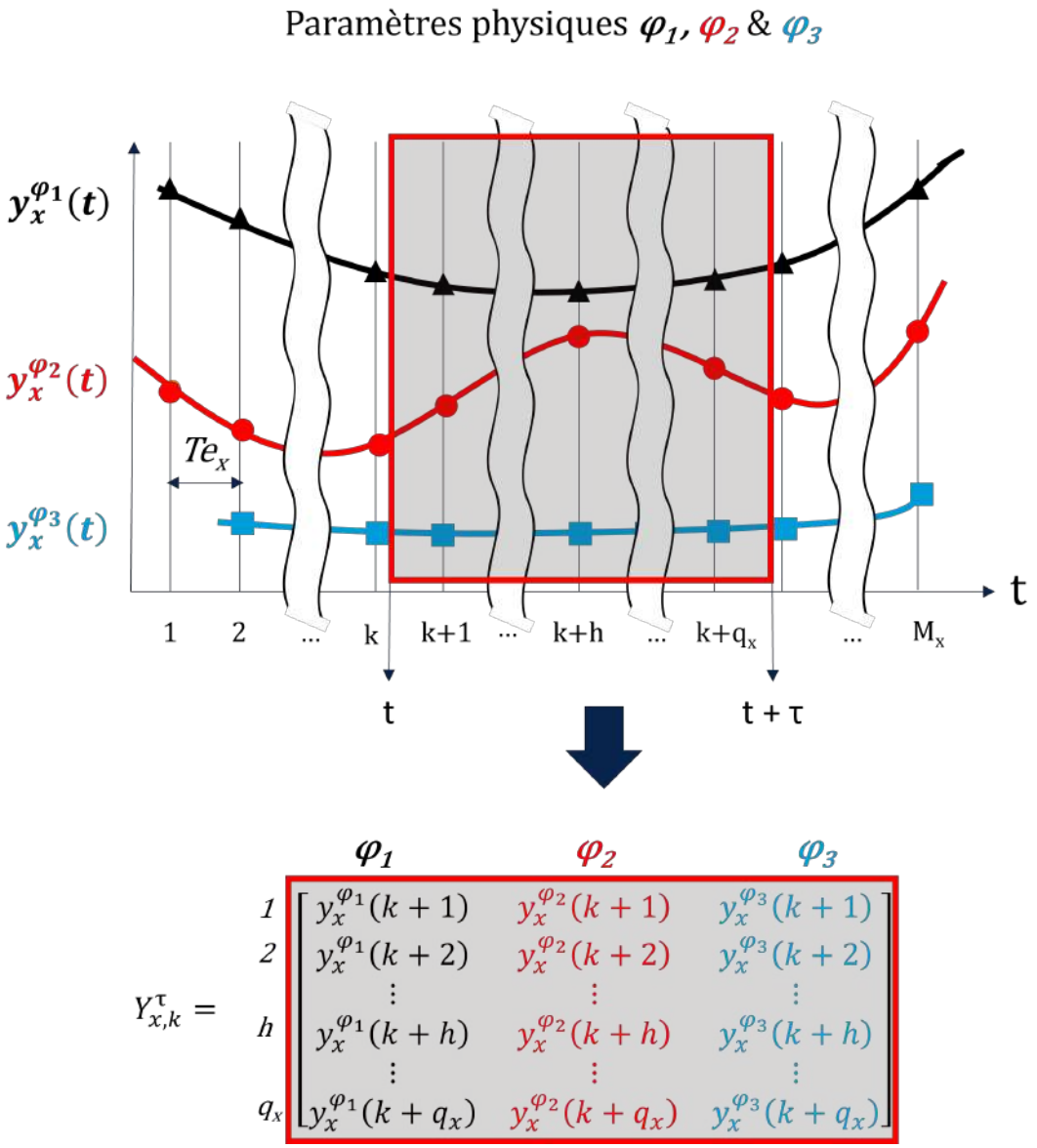


FIGURE 4.6 – Extraction de la sous-matrice du SmS numéro x au moment k pendant τ.

$$Y_x = \begin{bmatrix} y_x^{\varphi_1}(1) & \cdots & y_x^{\varphi_i}(1) & \cdots & y_x^{\varphi_N}(1) \\ y_x^{\varphi_1}(2) & \cdots & y_x^{\varphi_i}(2) & \cdots & y_x^{\varphi_N}(2) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ y_x^{\varphi_1}(k+1) & \cdots & y_x^{\varphi_i}(k+1) & \cdots & y_x^{\varphi_N}(k+1) \\ y_x^{\varphi_1}(k+2) & \cdots & y_x^{\varphi_i}(k+2) & \cdots & y_x^{\varphi_N}(k+2) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ y_x^{\varphi_1}(k+h) & \cdots & y_x^{\varphi_i}(k+h) & \cdots & y_x^{\varphi_N}(k+h) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ y_x^{\varphi_1}(k+q_x) & \cdots & y_x^{\varphi_i}(k+q_x) & \cdots & y_x^{\varphi_N}(k+q_x) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ y_x^{\varphi_1}(M_x) & \cdots & y_x^{\varphi_i}(M_x) & \cdots & y_x^{\varphi_N}(M_x) \end{bmatrix} \tag{4.4}$$

$$Y_{x,k}^{\tau} = \begin{bmatrix} y_x^{\varphi_1}(k+1) & \cdots & y_x^{\varphi_i}(k+1) & \cdots & y_x^{\varphi_N}(k+1) \\ y_x^{\varphi_1}(k+2) & \cdots & y_x^{\varphi_i}(k+2) & \cdots & y_x^{\varphi_N}(k+2) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ y_x^{\varphi_1}(k+h) & \cdots & y_x^{\varphi_i}(k+h) & \cdots & y_x^{\varphi_N}(k+h) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ y_x^{\varphi_1}(k+q_x) & \cdots & y_x^{\varphi_i}(k+q_x) & \cdots & y_x^{\varphi_N}(k+q_x) \end{bmatrix} \tag{4.5}$$

Étape 4 : vérification des sous-matrices et première détection

Après extraction, les sous-matrices sont vérifiées puis les données sont considérées comme valides si elles contiennent au moins deux points. Si elles contiennent plus de trois points, leurs valeurs sont comparées pour détecter d'éventuelles défaillances matérielles. Une matrice de détection est ensuite construite à partir des sous-matrices validées.

Une fois extraites, les sous-matrices sont vérifiées pour leur taille et leurs valeurs à l'aide de l'algorithme donné dans la Figure 4.7. Cette vérification permet de déterminer si un SmS donné a pu effectuer des observations. Si aucune observation n'est trouvée (sous-matrice vide), cela suggère une défaillance matérielle affectant le SmS x. La vérification est répétée pour toutes les sous-matrices afin de compter le nombre de SmS en service. Pour les sous-matrices contenant des observations, chaque colonne est extraite, représentant les observations d'un paramètre spécifique avec les capteurs du SmS identifié. Les données de ces colonnes sont ensuite analysées si elles contiennent plus de trois points. Deux points sont considérés comme valides car avec seulement deux points, il est impossible de déterminer avec certitude si les capteurs sont bloqués à une valeur. En revanche, si elles contiennent plus de trois points, leurs valeurs sont comparées pour détecter d'éventuelles similarités. Si les valeurs sont identiques, cela indique que le capteur est bloqué à une valeur désignée par α, et ces données sont exclues de la concaténation. Si les valeurs sont différentes, elles sont considérées comme valides et utilisées pour la concaténation.

$$D_{s,k}^{\tau} = \begin{bmatrix} nb_1 & \cdots & nb_i & \cdots & nb_N \\ m_1 & \cdots & m_i & \cdots & m_N \\ b_1 & \cdots & b_i & \cdots & b_N \end{bmatrix} \tag{4.6}$$

Sur la base de ces vérifications, une matrice de détection notée $D_{s,k}^{\tau}$ est construite. Cette matrice (4.6) contient des informations sur le nombre de SmS actifs, noté nb_i,

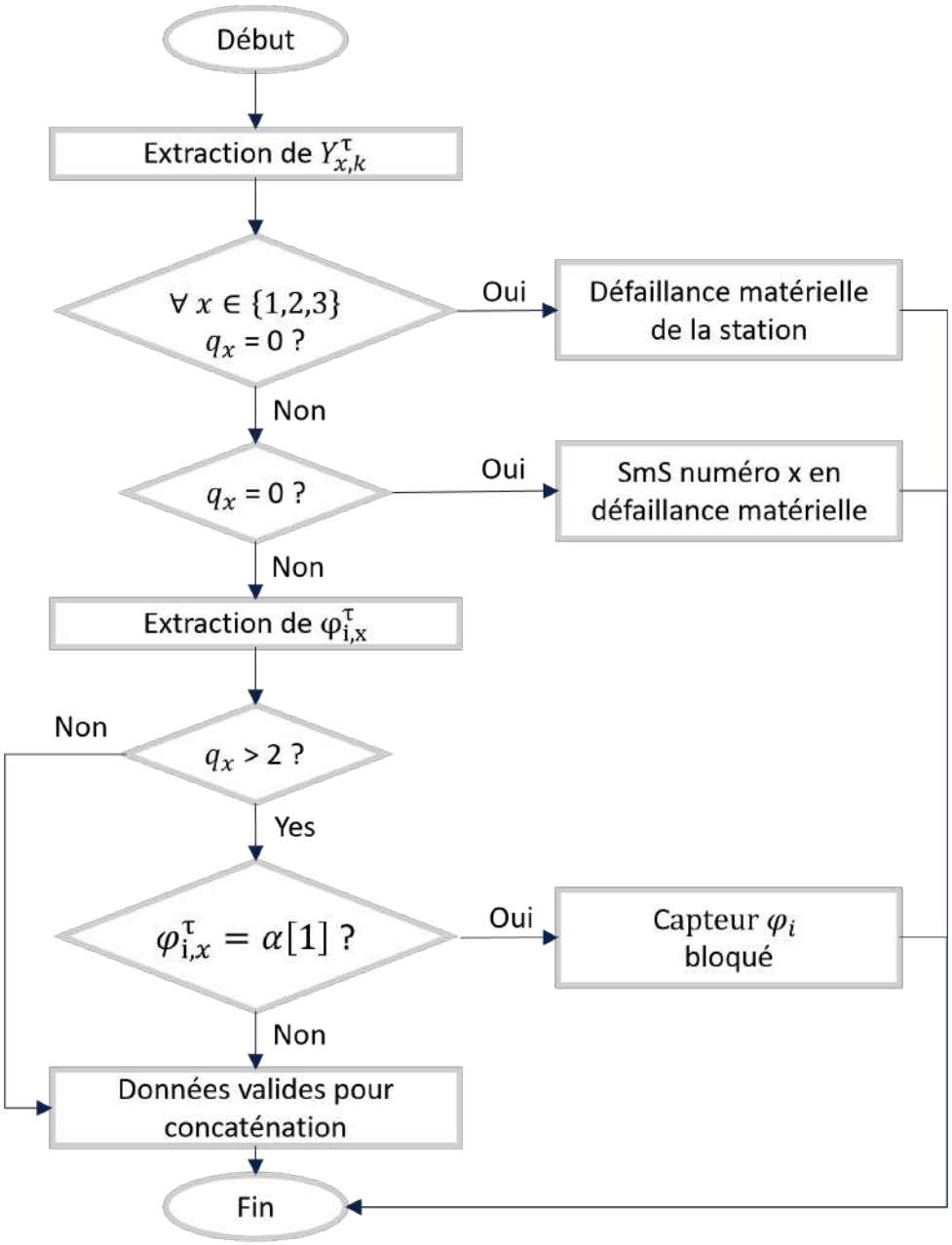


FIGURE 4.7 – Vérification des sous-matrices pour la détection.

pour un paramètre donné, ainsi que des listes d'identifiants pour les SmS sans observations, notés m_i, et ceux avec des valeurs bloquées, notés b_i, pendant une période τ. Une fois transmise au bloc d'évaluation, la matrice de détection aide à l'identification et à la transmission des processus de détection et de diagnostic via l'interface graphique. Elle facilite également le diagnostic précoce pour prendre des décisions. En comparant la matrice de détection avec la précédente, si disponible, seules les modifications sont signalées à l'interface graphique, évitant ainsi des alertes redondantes. La valeur de α est examinée, et elle peut représenter un diagnostic en fonction de sa valeur.

Étape 5 : concaténation des données par φ_i

Toutes les données valides sont concaténées et les données des LCS bloqués sont exclues en fonction du paramètre physique φ_i. Ensuite, ces données concaténées sont utilisées pour construire la matrice de seuils. Il s'agit d'une matrice représentant les différents seuils mobiles pour chaque paramètre physique.

Après l'étape de vérification, les données validées sont concaténées tandis que les données identifiées provenant d'un capteur bloqué sont exclues. Comme illustré sur la Figure 4.8, les données dans une colonne sont désignées par $\varphi^{\tau}_{i,x}$, représentant l'observation d'un paramètre physique φ_i provenant du SmS x. Toutes les données de tous les SmS sont concaténées dans une seule colonne, $\varphi^{\tau}_{i,S}$. Cette colonne représente les données d'observation d'un paramètre physique φ_i provenant de la station S. Cette étape de concaténation est appliquée à tous les paramètres. Une fois les concaténations terminées, les valeurs contenues dans les colonnes $\varphi^{\tau}_{i,S}$ sont utilisées pour construire la matrice des seuils.

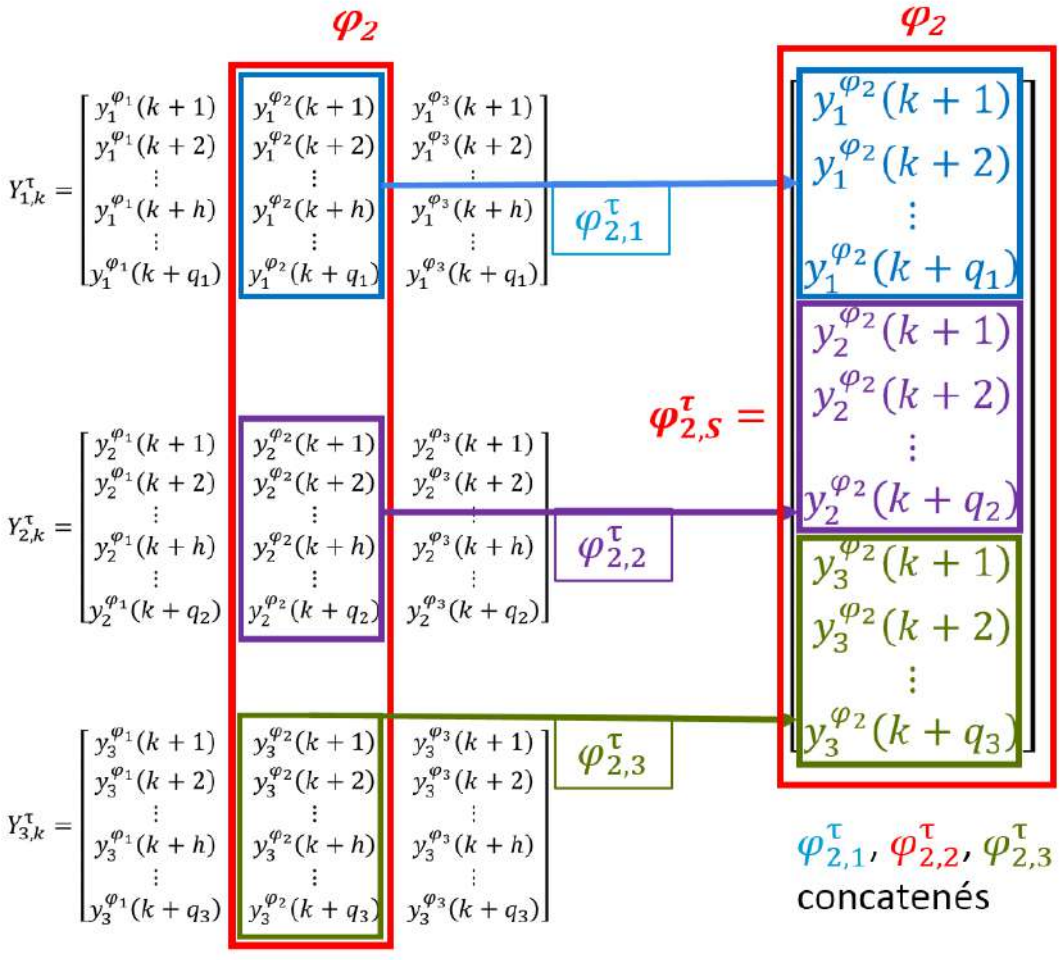


FIGURE 4.8 – Concaténation des $\varphi^{\tau}_{i,x}$ provenant des SmS de la station.

Étape 6 : construction de la matrice des seuils

La construction de la matrice de seuils utilise le PREV pour chaque paramètre observé par une station S_k. Les valeurs médianes des concaténations de paramètres sont ensuite utilisées avec le PREV pour calculer les seuils supérieurs et inférieurs pour chaque paramètre. La matrice de seuils résultante est notée $Th^{\tau}_{S,k}$. Cette étape est essentielle pour la détection des anomalies dans les séries temporelles, et la matrice de seuils est un composant clé de la méthode proposée.

Le PREV, désigné par $R\varepsilon_S$, est introduit en tant que vecteur d'erreurs relatives autorisées associées aux LCS. Ce vecteur (4.7) contient toutes les erreurs relatives admissibles $R\varepsilon_i$ associées aux LCS observant φ_i à l'échelle de la station S_k. Les seuils mobiles sont ensuite définis en fonction de ce vecteur. Initialement, les médianes des concaténations, notées $Med(\varphi^{tau}_{i,S})$, sont calculées, et ces résultats sont ensuite combinés avec le PREV pour déterminer les seuils supérieurs et inférieurs. Ainsi, le seuil maximal est noté ThM_i et est calculé à l'aide de l'équation (4.8), tandis que le seuil minimal pour φ_i est noté Thm_i et est calculé à l'aide de l'équation (4.9). La matrice des seuils est représentée par une matrice notée $Th^{\tau}_{S,k}$, et prend la forme de la matrice (4.10).

$$R\varepsilon_S = [R\varepsilon_1, \ldots, R\varepsilon_i, \ldots, R\varepsilon_N] \tag{4.7}$$

$$ThM_i = Med(\varphi^{tau}_{i,S}) \times \left(1 + \frac{R\varepsilon_i}{100}\right) \tag{4.8}$$

$$Thm_i = Med(\varphi^{tau}_{i,S}) \times \left(1 - \frac{R\varepsilon_i}{100}\right) \tag{4.9}$$

$$Th^{\tau}_{s,k} = \begin{bmatrix} Med_1 & \cdots & Med_i & \cdots & Med_N \\ ThM_1 & \ddots & ThM_i & \ddots & ThM_N \\ Thm_1 & \cdots & Thm_i & \cdots & Thm_N \end{bmatrix} \tag{4.10}$$

Étape 7 : Construction des vecteurs *Outlier* et *Reliable* par φ_i

Dans cette étape, la matrice de seuils est utilisée pour identifier les observations aberrantes. Chaque sous-matrice d'observations est comparée avec la matrice de seuils, pour construire les vecteurs *Outlier* et *Reliable*. Le vecteur *Reliable* contient les observations dans les limites des seuils, tandis que le vecteur *Outlier* contient celles qui les dépassent. Les dimensions de ces vecteurs sont telles que leur somme est égale au nombre d'observations d'origine. Les vecteurs *Outlier* et *Reliable* sont concaténés pour chaque SmS afin de créer des vecteurs au niveau de la station.

La matrice des seuils est utilisée pour sélectionner les observations qui sont trop éloignées de la médiane. Par conséquent, la matrice $Th^{\tau}_{s,k}$ est appliquée en tant que masque pour chaque sous-matrice d'observation $y^{\tau}_{x,k}$. Les observations contenues dans $\varphi^{\tau}_{i,x}$ sont classées dans un vecteur *Reliable* associé à la SmS x, noté $R^{\tau}_{i,x}$, et dans un vecteur *Outlier* noté $O^{\tau}_{i,x}$. La dimension de $R^{\tau}_{i,x}$ est $r_{i,x}$ et contient toutes les valeurs de $\varphi^{\tau}_{i,x}$ telles que $y^{\varphi_i}_x \in [Thm_i; ThM_i]$. De même, la dimension de $O^{\tau}_{i,x}$ est $o_{i,x}$ et contient toutes les valeurs de $\varphi^{\tau}_{i,x}$ telles que $y^{\varphi_i}_x \notin [Thm_i; ThM_i]$. La relation de dimension entre les vecteurs $R^{\tau}_{i,x}$, $O^{\tau}_{i,x}$ et $\varphi^{\tau}_{i,x}$ est telle que $q_x = r_{i,x} + o_{i,x}$. Les vecteurs *Reliable* et *Outlier* associés à la station de mesure sont notés $R^{\tau}_{i,S}$ et $O^{\tau}_{i,S}$, respectivement. Ces vecteurs ont une taille de

(R_i) et (O_i) respectivement, et sont les concaténations des vecteurs *Outlier* et *Reliable* de chaque SmS. Ainsi, les relations dimensionnelles sont exprimées par les équations (4.11).

$$R_i = \sum_{x=1}^{R} r_{i,x} \quad \text{et} \quad O_i = \sum_{x=1}^{R} o_{i,x} \tag{4.11}$$

Étape 8 : extraction des indices et vote par la médiane

La taille des vecteurs *Outlier* et *Reliable* pour chaque paramètre physique est utilisée pour construire des indices de défaillance permettant de surveiller les LCS. L'algorithme de vote basé sur la médiane est ensuite utilisé sur le vecteur *Reliable* pour générer une observation de synthèse pour la station. Cette observation résultante est une sous-matrice qui est concaténée avec l'observation précédente pour remplir la matrice d'observation temporelle à l'échelle de la station.

Les dimensions des matrices *Reliable* et *Outlier* par paramètre sont enregistrées dans la matrice (4.12) pour construire les indices, notés $I^{\tau}_{s,k}$. Chaque élément de la matrice est une paire de valeurs $(o_{i,x}, q_x)$ correspondant à la taille du vecteur *Outlier* pour le paramètre φ_i et au nombre d'observations de la SmS associée x. Cette matrice donne également la possibilité de construire des indices de défaillance pour chaque paramètre physique, ce qui permet de surveiller l'état de défaillance du LCS connecté au SmS respectif. L'indice de défaillance du LCS mesurant un paramètre physique φ_i du SmS numéro x est noté $Id^{\tau}_{i,x}$. Inversement, un indice de santé, noté $Ih^{\tau}_{i,x}$, peut être calculé, et la relation entre les deux est détaillée dans l'équation (4.13). Il convient de noter que lorsqu'il n'y a pas d'observations ($q_x = 0$), l'indice de défaillance est considéré comme étant à son maximum, c'est-à-dire égal à 1. En revanche, pour un état de santé nominal, cet indice serait de zéro.

$$I^{\tau}_{s,k} = \begin{bmatrix} (o_{1,1}, q_1) & \cdots & (o_{i,1}, q_1) & \cdots & (o_{N,1}, q_1) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ (o_{1,x}, q_x) & \cdots & (o_{i,x}, q_x) & \cdots & (o_{N,x}, q_x) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ (o_{1,R}, q_R) & \cdots & (o_{i,R}, q_R) & \cdots & (o_{N,R}, q_R) \end{bmatrix} \tag{4.12}$$

$$Id^{\tau}_{i,x} = \frac{o_{i,x}}{q_x} \quad \text{et} \quad Ih^{\tau}_{i,x} = \frac{r_{i,x}}{q_x} = 1 - Id^{\tau}_{i,x} \tag{4.13}$$

L'algorithme de vote implémenté dans l'Agrégateur utilise la médiane de toutes les observations fiables des paramètres physiques pour produire une synthèse pour la station pendant la fenêtre d'observation de taille τ. La méthode de vote basée sur la médiane consiste à trier les valeurs de $R^{\tau}_{i,S}$ par ordre croissant et à sélectionner la valeur à la position $q_x/2$ en cas de nombre impair de valeurs. Pour un nombre pair de valeurs, la moyenne des valeurs à la position $q_x/2$ et $(q_x + 1)/2$ est calculée. L'ensemble de valeurs résultant obtenu par le vote par la médiane représente une observation de la station à l'instant k, comprenant tous les paramètres physiques notés $S^{\tau}_{s,k}$, collectés au niveau de la station pour un intervalle d'échantillonnage spécifique τ. Cette étape sert de processus de rééchantillonnage des données, où cette observation est concaténée avec la précédente dans une matrice temporelle générale décrite comme Y_{S_k} et composée de $S^{\tau}_{s,k}$.

Étape 9 : construction des Indicateurs et transmission au serveur local

À la fin de ces huit étapes, la station enregistre N paramètres physiques distincts dans une fenêtre d'observation roulante de durée τ, produisant des indicateurs, trois matrices importantes, désignées par $S.I.D$:

1. Une matrice de Synthèse qui représente une observation temporelle, obtenue en appliquant un algorithme de vote médian sur les observations fiables des SmS tout en éliminant les valeurs bloquées, fournissant ainsi un résumé complet du paramètre mesuré pour chaque période τ ;
2. Une matrice d'Indices, qui est une extraction de toutes les erreurs et des états d'observation pour construire des indices de santé ou de défaillance pour chaque paramètre physique ;
3. Une matrice de Détection, générée en fonction des défaillances des SmS pendant la période d'observation et du nombre de SmS actifs dans les fenêtres d'observation.

Ces indicateurs sont ensuite transmis au serveur local où les blocs *Évaluation* et *Décision* sont mis en œuvre. Le bloc *Évaluation* est également décrit dans la Figure 4.4. Le taux de confiance permis Pc est le seuil situé dans le bloc *Évaluation*. Il est utilisé pour calculer l'indice global de confiance pour la station. Il peut également être utilisé avec la matrice de détection pour rendre les reconfigurations et les défaillances pertinentes et vérifiées. En fait, Pc couplé avec la matrice d'indicateurs peut être utilisé pour confirmer les résultats de la matrice de détection et effectuer des diagnostics à travers le composant d'identification des défaillances.

L'indice global de confiance pour φ_i, représenté par $Ic^{\tau}_{i,S}$, est calculé en utilisant les indices de santé de chaque SmS, couplés avec Pc, qui reflète le nombre d'observations fiables par rapport à leur nombre total. La fonction de Heaviside, notée $\mathcal{H}$, est utilisée dans le calcul de cet indice. Plus précisément, lorsque l'indice de santé d'un SmS x, noté $Ih^{\tau}_{i,x}$, dépasse Pc, le LCS correspondant sera considéré comme pertinent et l'indice de confiance global sera incrémenté de 1. L'indice de confiance global est une valeur qui varie entre zéro et R. La valeur R est le nombre de SmS au sein de la station. Ainsi, une valeur de l'indice de confiance proche de R permet de déterminer le nombre de SmS encore en fonctionnement et considérées comme fiables. Le calcul de l'indice est défini par l'équation (4.14).

$$\begin{gathered} Ic^{\tau}_{i,S} = \sum_{x=1}^{R} \mathcal{H}(Ih^{\tau}_{i,x} - Pc) \\ \text{avec} \\ \mathcal{H}(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases} \end{gathered} \tag{4.14}$$

Le $S.I.D$, $Ic^{\tau}_{i,S}$ et l'identification des composants défaillants facilitent la détection et le diagnostic précoce pour améliorer la disponibilité. Seules les modifications pertinentes sont communiquées à l'interface utilisateur pour éviter les alertes inutiles. Le bloc de décision reçoit ces rapports, permettant à l'équipe d'intervention de planifier ses actions en fonction de sa stratégie de maintenance.

Cette sous-section présente une nouvelle méthode pour formaliser une station de mesure dans un cadre mathématique. La station est conçue pour observer N paramètres physiques distincts dans une fenêtre d'observation continue d'une durée τ à travers ses

SmS, produisant le $S.I.D$. Ces indicateurs sont ensuite transmis au serveur local pour évaluation en utilisant le seuil Pc. Le bloc d'évaluation calcule l'indice global de la station et extrait des informations de détection et de diagnostic. Ensuite, tous ces résultats sont transmis au bloc de décision pour les parties prenantes concernées via une interface graphique. Ces parties prenantes comprennent les responsables élus chargés de surveiller la qualité de l'air et de gérer les interventions dans la ville, ainsi que le personnel de maintenance qui peut prendre les mesures correctives nécessaires en cas de défaillance ou de reconfiguration de l'équipement.

4.4 Conclusion du chapitre Données, Traitements et Décision

Ce chapitre a introduit un processus de traitement des données en huit blocs pour la surveillance de la qualité de l'air sur un territoire étendu. Le processus comprend les étapes suivantes : Stations de mesure, Fiabilisation des données, Évaluation, Décision, Modélisation, et Prédiction-Diagnostic de la qualité de l'air. Chaque bloc constitue une étape du processus depuis la collecte des grandeurs physiques jusqu'à leur utilisation comme support de décision.

Les données sont divisées en deux catégories principales : celles liées à la fiabilisation des données issues des stations de mesures et celles dédiées à la protection de la population et de l'environnement.

Après avoir exposé ce cadre général, la formalisation mathématique de la fiabilisation des données a été présentée dans ce chapitre. Une méthode en neuf étapes, mise en œuvre au sein de l'Agrégateur, a été détaillée, jouant un rôle central dans la fiabilisation des données issues des stations de mesures et de leur disponibilité. En commençant par la construction des matrices d'observations temporelles et en terminant par l'extraction des indices et le vote médian, cette méthode exhaustive permet une analyse précise et fiable des données. En somme cette méthode permet la gestion des données non synchronisées, la détection des anomalies et la génération d'indicateurs fiables.

Les matrices résultantes, la matrice de synthèse, la matrice d'indices et la matrice de détection ($S.I.D$), sont transmises au serveur local, facilitant une évaluation rapide de la fiabilité des observations des stations de mesures sous la responsabilité de ce dernier.

Dans ce chapitre, les bases d'un système robuste et fiable ont été développées pour la surveillance de la qualité de l'air, en mettant l'accent sur la fiabilisation des données et la haute disponibilité du réseau de mesure. Bien que ce chapitre soit axé sur la partie fiabilisation des données issues des stations de mesures, il constitue une étape essentielle dans le processus global de traitement des données, ouvrant la voie à des travaux futurs sur les autres aspects du processus à savoir la Modélisation et la Prédiction-Diagnostic de la qualité de l'air.

Chapitre 5

Mise en oeuvre et déploiement de l'architecture

Sommaire

5.1 Introduction 104
5.1.1 La phase stratégique 105
5.1.2 La phase de déploiement 106
5.2 Le développement en co-design de la station 107
5.2.1 La réalisation et le choix des technologies 108
5.2.2 Le choix des LCS 108
5.2.3 La conception de la carte électronique (PCB) 109
5.2.4 Intégration du SmS : Assemblage, Boîtier et Connexions 110
5.2.5 Le fonctionnement du capteur intelligent 111
5.2.6 Interface graphique et aide à la décision 112
5.3 Améliorations itératives des prototypes : les enseignements tirés du codesign 113
5.3.1 Premier prototype 113
5.3.2 Fiabilisation des mesures : l'influence des hyperparamètres . . . 118
5.3.3 Deuxième prototype avec fiabilisation des données 120
5.4 Conclusion du chapitre 126

5.1 Introduction

Le territoire cible choisi pour déployer l'architecture proposée est une organisation intercommunale territoriale appelée CCPVG, située dans le sud-ouest de la France, dans le département des Hautes Pyrénées en région Occitanie. Le territoire s'étend sur une superficie d'environ 1000 km², dont 91% est constitué de forêts et de terrains montagneux. Il comprend 46 communes réparties des plaines aux régions de haute montagne. Chacune ayant ses propres caractéristiques environnementales et défis liés à la qualité de l'air.

La topologie montagneuse de la région peut entraîner des concentrations de pollution localisées, qui passent souvent inaperçues par les stations conventionnelles exploitées par ATMO Occitanie, l'association chargée de la surveillance de la pollution dans la région. Notamment, la station conventionnelle la plus proche se trouve à 10 km de la CCPVG, comme illustré dans la Figure 5.1. Cette situation géographique met en évidence la nécessité d'un système de surveillance plus localisé et complet au sein du territoire CCPVG pour garantir une détection précise et rapide des problèmes de qualité de l'air.

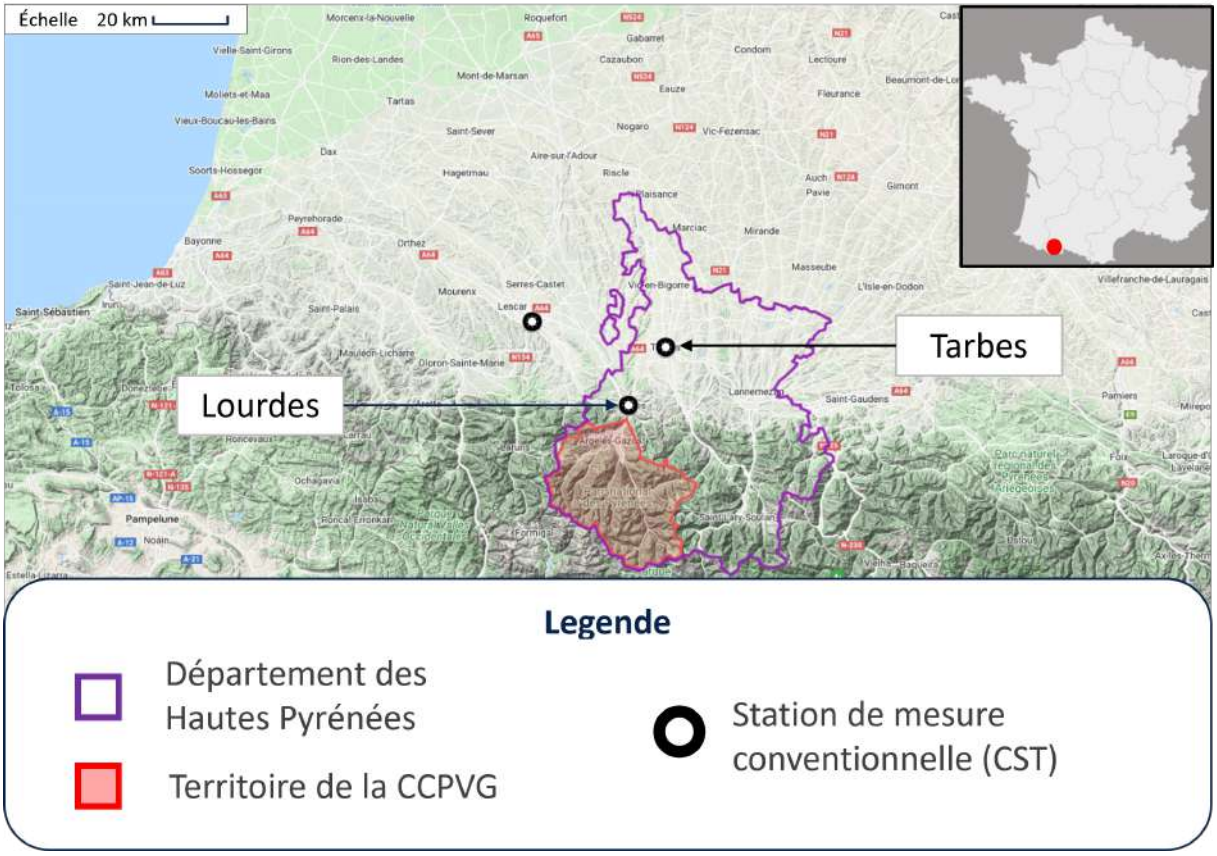


FIGURE 5.1 – Localisation de la CCPVG et des stations conventionnelles les plus proches.

C'est ainsi que le projet BOLDAIR est né pour mettre en œuvre une démarche scientifique pour la surveillance de la qualité de l'air sur son territoire. Ce projet est à l'origine de cette thèse pour combler le besoin de disposer d'une meilleure vision de la pollution réelle de la région, car l'estimation actuelle de l'ATMO était trop imprécise. En particulier, l'objectif était de détecter et de constater l'effet du tourisme local ainsi que des pratiques ancestrales de brûlage dans la région. Par conséquent, la proposition consistait à développer et déployer un réseau de mesure de la qualité de l'air en complément des stations de mesure conventionnelles. Cela permet aux élus et aux citoyens de surveiller

différents polluants, de prendre des décisions et d'agir en conséquence. En effet, l'architecture proposée visait à améliorer la connaissance de la situation du territoire et à permettre la surveillance des concentrations de polluants par les stations de mesure pour soutenir les décisions politiques.

La conception de l'architecture est fédérée, avec des serveurs locaux et des stations déployés pour couvrir de vastes zones géographiques, et un serveur principal responsable de l'ensemble du territoire. Ainsi, la principale différence entre les serveurs locaux réside dans la zone géographique, appelée zone d'intérêt, qu'ils couvrent et en rapport avec autorités locales responsables des infrastructures et de la population. Dans le cas de la CCPVG, le serveur local collecte les données des stations d'une vallée et propose des décisions au niveau de la ville, tandis que le serveur principal collecte les données de tous les serveurs locaux et propose des décisions pour l'ensemble du territoire. Leur interface varie en fonction des utilisateurs et de leur niveau de responsabilité. La phase d'observation du projet comprend deux étapes : la phase stratégique et la phase de déploiement. La phase stratégique consiste à identifier les zones d'intérêt selon les normes établies par le Laboratoire Central de Surveillance de la Qualité de l'Air (LCSQA). La phase de déploiement consiste à installer les stations et leurs SmS dans les périmètres de mesure sélectionnés. Le côté matériel et logiciel a été conçu de manière générique, permettant une intégration polyvalente de différents types de capteurs et de mesures.

5.1.1 La phase stratégique

Cette première phase consiste à choisir et identifier les zones d'intérêt et les périmètres de mesure afin de procéder à la mise en place des stations de mesure. Pour rappel, un périmètre de mesure est une zone où les conditions sont telles que les paramètres physiques à mesurer sont identiques en tout point du périmètre. Cette définition précise assure une uniformité dans la collecte des données.

Pour atteindre l'objectif de préservation de la population, la localisation des zones d'intérêt a été minutieusement étudiée. La sélection s'est concentrée sur les endroits où la densité de population est la plus élevée et de son emplacement stratégique de fond de vallée, ce qui a conduit à la sélection de cinq villes pour représenter les cinq vallées du territoire. Les périmètres de mesure, quant à eux, sont déterminés en fonction de la stabilité des paramètres à surveiller. Ils peuvent s'agir d'un toit, d'un parking ou d'un endroit spécifique où les conditions de mesure sont identiques quel que soit le point choisi dans le périmètre. L'environnement local joue un rôle clé dans cette sélection, avec des facteurs tels que le vent, les sites émetteurs de poussière, les arbres à proximité, etc., pris en compte pour assurer une surveillance précise.

Les paramètres physiques à surveiller pour la qualité de l'air ont été définis en collaboration avec des experts de la qualité de l'air et sont les suivants :

— T : Température (°C) ;
— H : Humidité (%Hr) ;
— Pe : Pression (hPa) ;
— SO_2 : Dioxyde de soufre ($\mu g/m^3$) ;
— NH_3 : Ammoniac ($\mu g/m^3$) ;
— COV : Composés organiques volatils ($\mu g/m^3$) ;
— CO_2 : Dioxyde de carbone ($\mu g/m^3$) ;
— PM_{10} : Particules en suspension ($\mu g/m^3$) ;
— $PM_{2.5}$: Particules fines ($\mu g/m^3$).

Pour cette étude de cas, la station est composée de trois SmS ($R = 3$). Chaque SmS contient cinq LCS, mesurant neuf paramètres ($N = 9$), ainsi qu'un Agrégateur. La Figure 5.2 montre les SmS assemblés et déployés dans la ville d'Argelès-Gazost, illustrant la réalisation concrète de cette phase stratégique. Sur la gauche, les trois SmS sont disposés sur une console extérieure et la quatrième boîtier correspond à l'alimentation électrique. La partie droite de la figure montre un agrandissement du SmS dans son boîtier. Le SmS est placé verticalement pour limiter les problèmes liés à l'humidité. Un ventilateur a été ajouté et sert à renouveler l'air dans le SmS à intervalles réguliers. Cette approche méticuleuse garantit une collecte de données fiable et robuste, essentielle pour une surveillance efficace de la qualité de l'air dans la région.

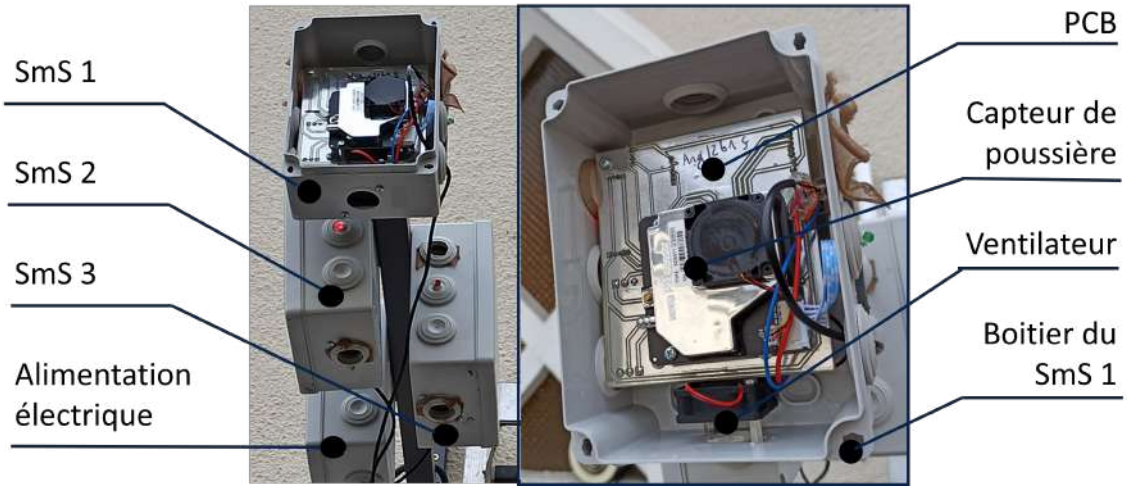


FIGURE 5.2 – Une station de mesure avec trois SmS assemblés et le boîtier d'alimentation.

À la suite de la phase stratégique, où les zones d'intérêt et les périmètres de mesure ont été soigneusement choisis et identifiés, la phase de déploiement est initiée. Elle consiste en l'installation des stations de mesure conformément à la configuration du terrain définie précédemment tout en veillant à la sélection des sites optimaux pour les capteurs et à l'alignement avec les objectifs de surveillance de la qualité de l'air dans la région.

5.1.2 La phase de déploiement

La phase de déploiement est l'étape suivante consistant à implémenter les stations de mesure dans les zones d'intérêt sur le terrain.

Installation du serveur local : le serveur local, représentant la zone d'intérêt de la ville d'Argelès-Gazost, est installé dans le bâtiment de la CCPVG. Conformément aux protocoles de sécurité, il est géré par l'administrateur réseau de la collectivité. Ce serveur est responsable de la zone d'intérêt, collectant et traitant les données.

Localisation des stations de mesure : les emplacements choisis pour les stations de mesure, et plus exactement leurs SmS, sont soigneusement sélectionnés pour leur accès à l'électricité et à Internet via le routeur du bâtiment. La couverture jusqu'au périmètre de mesure est évaluée en tenant compte de la configuration du terrain. Pour ce premier déploiement, la cour du bâtiment a été choisie en raison de sa protection aux conditions météorologiques : l'exposition directe au soleil a été évitée et les composants sont protégés par les vents violents pouvant impacter leurs fixations sans pour autant être coupé des mouvements de l'air.

Implantation de l'Agrégateur : les Agrégateurs sont basés sur des microprocesseurs de types Raspberry Pi. Ils jouent un rôle clé dans l'agrégation des observations des SmS. Leur rôle se limite à l'agrégation des observations des SmS et à l'extraction du $S.I.D$ avant de le transmettre au serveur local, qui procède à une évaluation et met également en œuvre une interface graphique (application web). Plus sensibles aux conditions extérieures par rapport aux SmS, ils sont généralement disposés à l'intérieur du bâtiment ou bien simulés sur le serveur local en fonction des besoins. Le choix entre une technologie physique (dispositif matériel) ou virtuelle (simulation logicielle) est crucial pour l'efficacité et la fiabilité.

La structure de la station de mesure utilise les SmS en tant que modules indépendants fonctionnant en parallèle selon une redondance modulaire triple. Le SmS, répliqué à travers le territoire, est la pièce fondamentale de l'architecture. Capable de supporter plusieurs modifications, il constitue la base de la conception grâce au choix des LCS. Sa réplication assure une redondance et une résilience.

La phase de déploiement nécessite une coordination étroite avec les autorités locales, les fournisseurs d'énergie et de réseau, et d'autres parties prenantes. Les défis tels que les conditions météorologiques, les restrictions locales et les exigences réglementaires sont soigneusement pris en compte pour les déploiements successifs.

En conclusion, le fonctionnement de cette architecture se déroule en deux phases : la phase stratégique pour la détermination des paramètres à surveiller et la phase de déploiement pour placer les SmS sur le terrain. La réalisation de ces deux phases est une orchestration complexe de technologie, de logistique et de collaboration. Elle traduit la vision stratégique en une réalité opérationnelle, jetant les bases d'une surveillance efficace et fiable de la qualité de l'air sur le territoire.

Après avoir établi les fondations de l'architecture et présenté les phases stratégique et de déploiement pour la mise en œuvre, l'attention se porte désormais sur le développement des stations de mesure et plus précisément le co-design itératif des SmS.

5.2 Le développement en co-design de la station

Dans cette section, le co-design est abordé comme un processus d'amélioration continue et simultané des aspects matériels (hardware) et logiciels (software) des SmS. Cette méthode a été choisie pour le développement de la station de mesure, qui a été conçue dans son ensemble à partir de zéro, nécessitant une sélection minutieuse de chaque composant pour répondre aux défis techniques.

L'approche itérative a permis d'adapter le design en fonction des besoins spécifiques du projet. La programmation d'un capteur intelligent s'est déroulée en plusieurs étapes, suivant une méthode d'essai et d'erreur, afin de maîtriser l'appropriation du matériel et les différentes interfaces. L'intégration des LCS au niveau du microcontrôleur était dépendante de la programmation de celui-ci, ainsi que des interactions avec les autres composants en termes de ressources. Cette complexité a nécessité une approche de co-design flexible et adaptative.

Au fur et à mesure de la réalisation et du test des premiers prototypes en extérieur, des défis supplémentaires ont émergé. La nécessité de renforcer les connexions filaires a conduit à la création de cartes électroniques spécifiques, suivie de tests supplémentaires pour identifier et corriger les défaillances dues aux conditions extérieures. Ce processus a contribué à renforcer et à robustifier le SmS tant au niveau matériel que logiciel.

Dans cette section, seuls les prototypes ayant démontré des résultats prometteurs seront présentés, illustrant le succès de cette méthode de co-design dans la conception d'une station de mesure efficace et fiable.

5.2.1 La réalisation et le choix des technologies

Pour la réalisation du prototype SmS, le choix s'est porté sur l'ESP32 DEVKITV1 en raison de sa polyvalence, de sa connectivité WiFi, de ses capacités de calcul, et de son intégration aisée avec l'IDE ARDUINO (Environnement de Développement Intégré). Ce choix a été influencé par la disponibilité de bibliothèques de programmation open source pour les LCS. Cependant, celles-ci ont souvent été conçues pour des cartes légèrement différentes, nécessitant quelques adaptations spécifiques.

L'une des principales difficultés rencontrées avec l'ESP32 réside dans son alimentation spécifique de 3.3V, contrairement à la majorité des cartes qui fonctionnent à 5V. Ce détail technique a requis une attention particulière lors de la conception. Néanmoins, les avantages de l'ESP32, tels que sa polyvalence et ses multiples possibilités en termes de capacité de calcul, de mémoire, et de communication, l'emportent largement sur cette contrainte. Le fait qu'il n'ait pas été nécessaire d'acheter d'autres cartes complémentaires en fait un choix économique et efficace.

5.2.2 Le choix des LCS

Après l'identification des N paramètres physiques à mesurer, les LCS correspondants ont été sélectionnés et intégrés au circuit imprimé (PCB) pour les connecter physiquement à l'ESP32 faisait de ce fait la base des SmS. Il est à noter que certains LCS possèdent la capacité multicanaux, permettant de mesurer plusieurs paramètres en même temps et offrant ainsi de nouvelles perspectives de mesure. Par exemple, l'utilisation d'un nez électronique (E-nose) permet d'analyser les odeurs à proximité des rejets publics, tandis que des détecteurs de fuites peuvent protéger les travailleurs contre les gaz inflammables ou neurotoxiques dans une station de traitement de l'eau. Les LCS sélectionnés ont été privilégiés pour leur utilisation éprouvée dans des contextes similaires.

Fabricant	Modèle	Paramètres mesurés	Type
Nova Fitness	SDS011 V1.3	PM_{10} / $PM_{2.5}$	UART
Grove	MICS 6814	NO_2 / NH_3 / CO	Analog
Hanwei	MQ4	CH_4	Analog
Sensirion	SGP30	$COVT$ / eCO_2	I2C
Spec sensors	ULPSM SO2	SO_2	Analog

TABLE 5.1 – Sélection des LCS : première itération.

La table 5.1 présente un récapitulatif des capteurs choisis pour le développement de la station. La première sélection des LCS a été guidée par plusieurs critères, dont leur disponibilité sur le marché, leur coût (en considération du nombre de stations à déployer), et la qualité de la documentation fournie par le fabricant. Lors de cette première itération, des incompatibilités de connexion ont été découvertes lors de l'intégration des LCS au SmS. De plus, des problèmes importants ont été rencontrés, compromettant le fonctionnement global du SmS. Des LCS technologiquement supérieurs ont dû être écartés en raison d'une

documentation insuffisante. Suite à ces défis rencontrés lors du déploiement du prototype initial, une réévaluation a été entreprise, conduisant à une seconde sélection de LCS.

Cette nouvelle sélection, présentée dans la table 5.2, est un compromis entre plusieurs facteurs : compatibilité avec l'ESP32, qualité des composants, et intégration. Le passage à une connexion I^2C, en comparaison avec une connexion analogique, a été dicté par plusieurs raisons. Premièrement, l'I^2C simplifie l'intégration car il intègre les formules de conversion au niveau du capteur. Deuxièmement, l'activation du WiFi sur l'ESP32 réaffecte certains pins initialement disponibles pour des entrées analogiques, limitant ainsi leur utilisation pour les LCS analogiques. Enfin, l'utilisation du WiFi augmente la consommation électrique, ce qui a aussi été un facteur déterminant dans le choix des capteurs, particulièrement pour ceux à faible consommation d'énergie.

Fabricant	Modèle	Paramètres mesurés	Type
Nova Fitness	SDS011 V1.3	PM_{10} / $PM_{2.5}$	UART
Grove	MICS 6814	NO_2 / NH_3 / CO	I2C
Sensirion	SGP30	COVT / eCO_2	I2C
Gravity	SEN0470	SO_2	I2C
BOSCH	BME280	T / P / H	I2C

Table 5.2 – Sélection des LCS : seconde itération.

La conception du SmS a également été pensée pour permettre une polyvalence et une intégration future aisée de nouveaux LCS. Les connexions UART et I2C facilitent cette intégration matérielle grâce à une connexion à trois fils. Du côté logiciel, le défi majeur consiste à intégrer de nouvelles bibliothèques sans créer de conflits avec les composants existants.

5.2.3 La conception de la carte électronique (PCB)

La conception du PCB a fait l'objet d'une attention particulière une fois les composants choisis. Les premiers prototypes ont été assemblés sur des plaquettes d'essai (breadboards) pour faciliter les connexions rapides entre les composants. Toutefois, cette méthode de connexion filaire a révélé des problèmes de communication entre les LCS et le microcontrôleur.

Pour la première itération robustifiée, un PCB a été conçu pour souder directement les modules LCS ainsi que l'ESP32, comme illustré en gros plan à droite dans la Figure 5.2. La méthode de gravure employée, dite « méthode anglaise », a consisté à fraiser une plaque recouverte de cuivre avec des outils mécaniques, sans l'utilisation de produits chimiques. La gravure se trouve donc sur un seul côté pour minimiser les coûts de fabrication.

La deuxième itération a été conçue avec des pistes plus larges, facilitant les opérations de soudage et de dessoudage. En anticipant la nécessité de maintenance du SmS par remplacement des composants défaillants, cette conception permet à un opérateur humain d'intervenir sans risquer de rendre le SmS irréparable. Les pistes ont une largeur de 0,8 mm avec une isolation équivalente, afin de faciliter les opérations de maintenance telles que le remplacement d'un module LCS.

Pour la troisième et dernière itération, une technique de trempage de la plaque de cuivre dans un bain d'étain a été employée pour augmenter sa résistance à l'oxydation. Les premiers déploiements sur le terrain ont révélé des signes d'oxydation du cuivre, créant

des connexions non souhaitées entre les pistes proches. En effet, lors du déploiement du premier prototype, le faible écart entre les pistes (isolation) était la cause principale de création de ponts électriques. Cela a conduit à leur écartement et à une augmentation de l'isolation dans les prototypes suivants.

Le logiciel KiCad EDA, une suite d'automatisation de conception électronique multi-plateforme et open source, a été utilisé pour créer les circuits électriques. Les empreintes de chaque composant ont été fidèlement reproduites à partir des dimensions réelles, et les symboles correspondants ont été créés. La fabrication du PCB a été réalisée dans le laboratoire d'électronique de l'université. Ce choix a permis une meilleure réactivité et un contrôle des PCB sur place dans le cadre d'un processus de conception itératif.

5.2.4 Intégration du SmS : Assemblage, Boîtier et Connexions

Le SmS est composé des modules LCS et de l'ESP32, soudés sur le PCB. Ce processus d'assemblage vise à garantir une connexion robuste et fiable entre les différents composants, en évitant les problèmes rencontrés avec des connexions filaires sur des breadboards lors des premiers prototypes.

Logé dans un boîtier électrique de type IP65, le SmS est conçu pour résister aux conditions environnementales. Ce boîtier est placé à une hauteur minimale de 2 m pour éviter les dégradations humaines et est éloigné de la façade du bâtiment via une console pour des conditions optimales de mesure. Toutefois, ce type de boîtier a présenté des problèmes, notamment en termes de renouvellement de l'air. Il est donc nécessaire d'avoir une circulation d'air forcée à l'intérieur du boîtier. Une étude approfondie du premier prototype a mené à la conception d'un boîtier avec deux ouvertures en haut, permettant un meilleur renouvellement de l'air. Des simulations ont également aidé à comprendre les phénomènes de spirale lors de la circulation d'air, conduisant à des ajustements dans la conception. L'étude du boîtier ainsi que la conception d'un autre type d'encapsulation remplissant les critères précités sont présentés dans les perspectives (cf. Chapitre 6.5 Perspectives de recherche, page 139).

Pour diminuer les problèmes d'oxydation du PCB causés par l'humidité présente dans l'environnement extérieur, une méthode de tropicalisation a été employée. Cela implique l'application d'une résine fine pour étanchéifier les composants et leurs connexions après soudage tout en évitant les parties sensibles des LCS. Ensuite, le positionnement vertical du PCB et l'utilisation d'entretoises assurent également une circulation d'air appropriée, réduisant les risques d'accumulation de condensation. L'eau condensée est évacuée par le bas du boîtier par gravité. De même, une attention particulière a été portée au niveau de modules de développement pour les positionner verticalement lors de la conception afin que les gouttes d'eau ne stagnent pas sur les circuits et ainsi réduire le phénomène d'oxydation.

La sécurité et la conformité aux normes actuelles du bâtiment (norme RT2012) ont été des considérations clés dans la conception de l'unité d'alimentation électrique. Un transformateur spécialisé 220 V AC - 5 V CC, relié à un disjoncteur de 2 A, assure une alimentation stable et sûre aux SmS et à l'Agrégateur.

L'intégration réussie des LCS nécessite une attention particulière à la fois au niveau matériel et logiciel. L'utilisation de l'ESP32 DEVKITV1 facilite la connexion avec différents types de LCS, qu'ils soient analogiques, bus Inter Integrated Circuit (I^2C) ou Universal Asynchronous Receiver Transmitter (UART). La connexion UART est nécessaire pour les capteurs de particules et ces derniers fournissent directement la concentration en

ppm. Ces capteurs sont numériques et comportent une unité de prétraitement intégrant les formules de compensation en mémoire. Pour les LCS disposant d'une connexion I^2C, ceux-ci offrent une flexibilité en permettant un remplacement facile et la détection de leur présence avec seulement la connexion de trois fils. Ce type de connexion permet à l'ESP32 de détecter la présence des LCS et ainsi permet de créer une variable personnalisée avec une valeur spécifique en cas de dysfonctionnement. L'ESP32, grâce à ses capacités de calcul, utilise les formules de conversion fournies par le fabricant pour prendre en compte efficacement les paramètres de température et d'humidité provenant des autres LCS lors du pré-traitement. Ensuite, il agrège et concatène les mesures. L'ESP32 pilote également deux relais pour contrôler la consommation d'énergie des capteurs de gaz et du ventilateur qui renouvelle l'air dans le boîtier. Il interagit également avec l'Agrégateur, rendant les LCS capable d'économiser de l'énergie, de détecter les anomalies et d'alerter en cas de défaillances. Cette gestion intelligente constitue un atout majeur dans la conception globale du SmS.

5.2.5 Le fonctionnement du capteur intelligent

La configuration matérielle du SmS permet de programmer deux modes de fonctionnement : le mode *Configuration* et le mode *Mesure*. Le logiciel du SmS permet de passer d'un mode à l'autre. Le mode *Configuration* est activé pour les connexions initiales à l'Agrégateur, les diagnostics et les tests des LCS en cas de défaillance. Ce mode est conçu pour une intégration facile et des diagnostics rapides sans accès physique. Le deuxième mode, qui est activé à partir du mode *Configuration*, permet les mesures à partir des LCS et économise l'énergie entre les mesures en mettant le SmS en mode veille profonde. La figure 5.3 illustre le protocole de mesure conçu pour respecter les directives du fabricant. L'échange d'air à l'intérieur du boîtier est important pour éviter une pollution cumulative sur la partie sensible du LCS. De plus, l'acquisition des mesures de température et d'humidité à différents moments du processus permet de compenser les valeurs des concentrations de gaz.

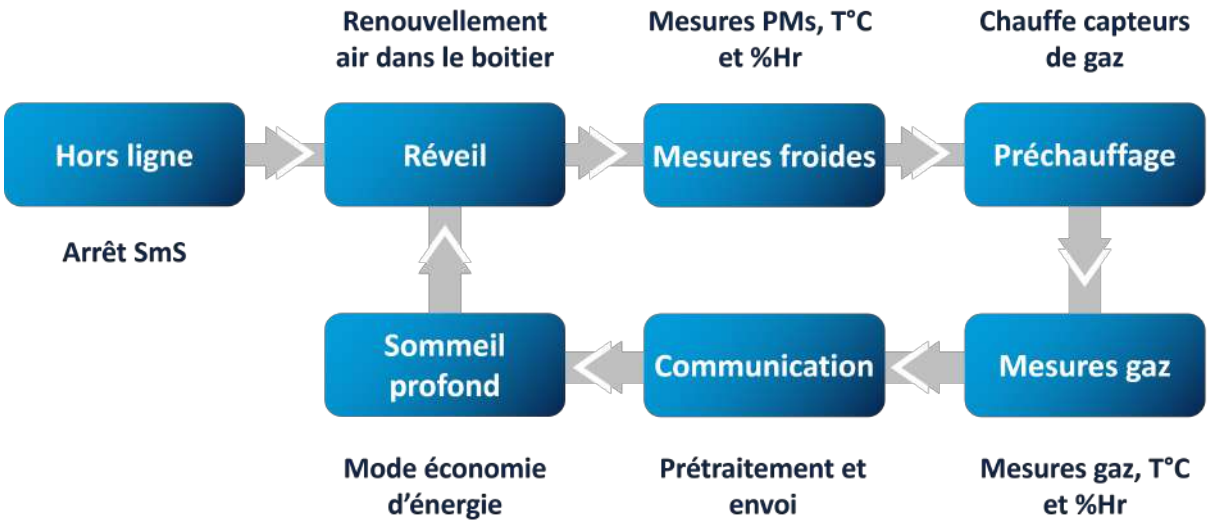


FIGURE 5.3 – Protocole de mesure du SmS.

La structure de la station de mesure utilise les SmS en tant que modules indépendants

fonctionnant en parallèle selon une redondance modulaire triple. Les SmS fonctionnent en permanence pour garantir une disponibilité élevée. Ainsi, l'architecture est résiliente et réduit l'impact des défaillances sur la station de mesure. Par conséquent, ils peuvent fournir en continu des mesures physiques de la qualité de l'air, qui est l'objectif principal. De plus, les LCS qui composent les SmS sont choisis pour être faciles à remplacer. La conception du PCB permet plusieurs modifications afin d'évoluer en fonction de la mesure d'autres paramètres pour prendre des décisions pour les autorités. Enfin, des mesures ont été prises pour éviter la dégradation rapide des circuits électroniques dans un environnement extérieur.

5.2.6 Interface graphique et aide à la décision

Afin de communiquer des informations utiles à la prise de décision, une application web a été développée. Cette application est hébergée sur les différents serveurs et propose des vues agrégées des informations provenant des différentes stations. L'outil d'aide à la décision du projet BOLDAIR, illustrée à la Figure 5.4, est hébergé sur le serveur central de la CCPVG et agit comme une plateforme centralisée pour la surveillance de la qualité de l'air. Conçue pour être à la fois intuitive et informative, cette interface permet aux citoyens et aux élus d'explorer en temps réel les données de la qualité de l'air collectées par des stations de mesure à travers le territoire. Les utilisateurs peuvent interagir avec une carte interactive, utilisant des bibliothèques open-source de cartographie telles que Leaflet [1] et OpenStreetMap (OSM) [2], pour sélectionner des stations spécifiques et obtenir des informations détaillées. En plus de servir de point central pour les actualités et les alertes sur la qualité de l'air, l'interface graphique offre des renseignements sur diverses initiatives liées à la pollution, telles que des ateliers avec le Centre Permanent d'Initiatives pour l'Environnement (CPIE) et le Plan Climat Air Énergie Territorial (PCAET). Le back-end, qui gère la collecte et la centralisation des données, soutient toutes ces fonctionnalités.

Le développement de la station en co-design a mis en lumière l'importance d'une approche intégrée et collaborative pour la conception et l'installation des SmS. En combinant les éléments hardware et software de manière itérative, cette méthode a permis de créer une solution adaptée, flexible et robuste. L'utilisation de l'ESP32 et des bibliothèques open source a été essentielle pour assurer la polyvalence et l'efficacité du prototypage du SmS. De même, les multiples possibilités de connexions de l'ESP32 a permit la sélection des LCS appropriés pour mesurer les paramètres physiques, offrant des possibilités d'extension et d'adaptation futures. La conception itérative du circuit imprimé a permis de faire évoluer la conception et de l'adapter aux nouveaux défis rencontrés lors des déploiements successifs. L'assemblage du SmS a nécessité une attention particulière à la protection du SmS et aux intégrations des LCS en évitant les conflits logiciels. Le co-design s'est avéré être une approche essentielle pour le succès de ce projet, favorisant l'innovation et la résilience. Il a permis de traduire une vision complexe en une réalité opérationnelle, apportant les bases d'une architecture de surveillance de la qualité de l'air ascendante du territoire. La section suivante explore plus en détail les prototypes d'une station de mesure composée de LCS.

1. Bibliothèque de cartographie interactive : `https://leafletjs.com/`
2. Service de cartographie : `https://www.openstreetmap.org/`

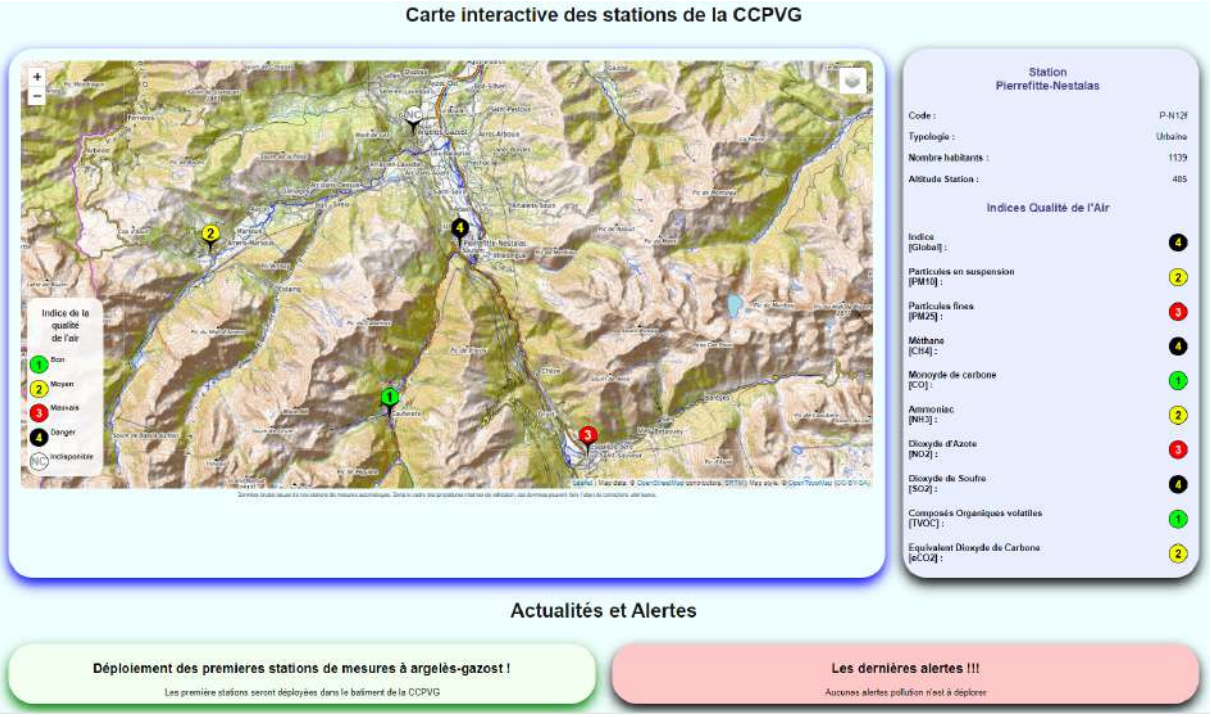


FIGURE 5.4 – Outil d'aide à la décision de la CCPVG pour la surveillance de la qualité de l'air.

5.3 Améliorations itératives des prototypes : les enseignements tirés du codesign

Dans cette section, l'exploration des différents prototypes met en lumière l'acquisition de connaissances à partir des développements successifs du système de surveillance de la qualité de l'air. Le projet a débuté à partir de zéro, et le parcours a été jalonné par des défis et des opportunités d'apprentissage. Les déploiements sur le terrain ont joué un rôle crucial dans l'amélioration des SmS, fournissant des enseignements qui ont influencé à la fois les aspects matériels et logiciels. Les stations de mesure, avec leurs composants et leur conception, ont évolué au fil de ces déploiements. Cette section met en évidence ces évolutions et souligne l'impact central des enseignements issus du codesign. Grâce à un processus d'itérations et d'ajustements constants, les expériences se sont consolidées et les prototypes ont été adaptés pour assurer une collecte de données fiable.

5.3.1 Premier prototype

La première itération d'une station de mesure consiste dans un premier temps à tester sur le terrain les SmS. L'achat des composants et la réalisation ont été effectués en petit lots pour vérifier l'hypothèse que les composants ne se dégradent pas en même temps. Les SmS envoient leurs données à l'Agrégateur qui les stocke, les structure et effectue les premiers traitements. Le serveur local reçoit les données depuis l'Agrégateur afin de les présenter sur un outil d'aide à la décision. Dans le cas du déploiement d'une seule station de mesure composée de trois SmS, et en fonction de la capacité de calcul et de la faible capacité de stockage des données, il a été décidé de mettre en œuvre l'Agrégateur dans le serveur central du bâtiment administratif. Ainsi, plutôt que de se connecter directement

à un Agrégateur distinct, les SmS communiquent avec le serveur local. Ce dernier joue alors le rôle d'Agrégateur et transmet les données en conséquence, comme illustré à la Figure 5.5

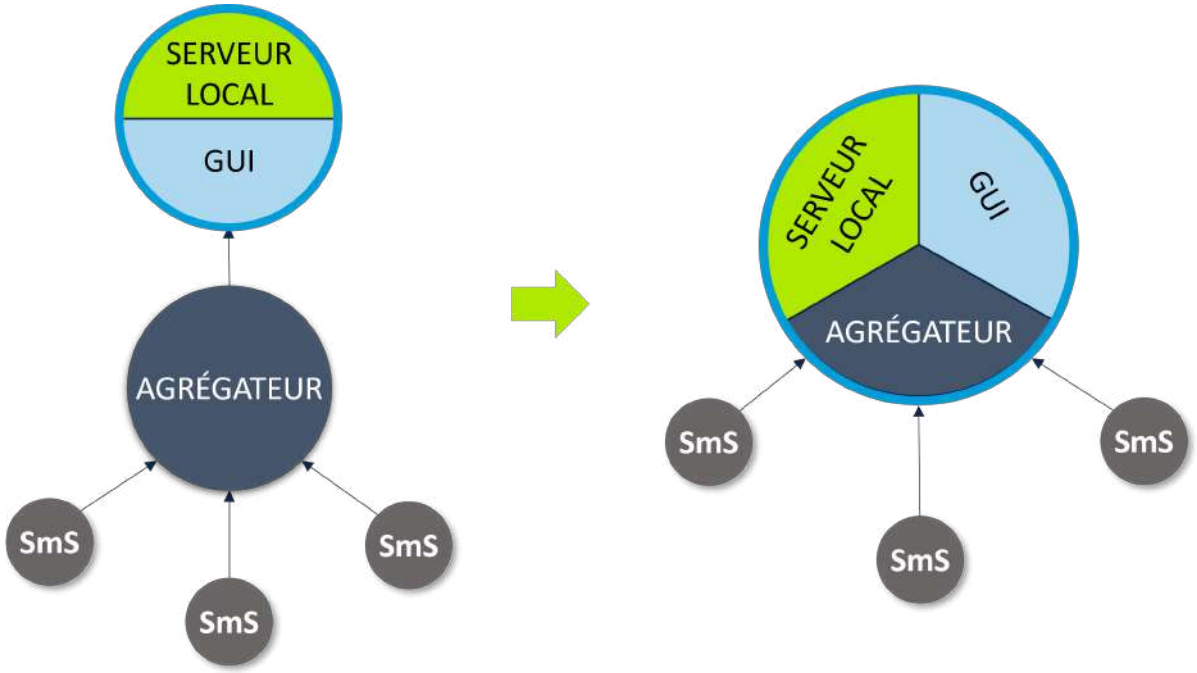


FIGURE 5.5 – Station déployée à titre de preuve de concept.

Cette modification ne change pas le principe de fonctionnement de la station de mesure, car les composants sensibles sont les SmS et le fonctionnement de l'Agrégateur ne dépend pas de son emplacement physique. Le serveur local est ainsi composé de deux parties : l'Agrégateur pour communiquer et traiter les données des SmS, et l'outil d'aide à la décision pour présenter des représentations graphiques des données. Ces deux parties communiquent entre elles exactement de la même manière que si elles étaient physiquement situées à deux endroits différents. Les communications entre les composants utilisent le protocole HTTP et les données échangées sont au format JSON. Au niveau du serveur local, deux serveurs fonctionnent en parallèle, un serveur web Apache2 [3] est utilisé pour la partie *FRONT END* (gestion de la communication via API entre les SmS et l'interface graphique). Pour le *BACK-END*, un serveur Node-RED [4] est couplé à une base de données INFLUXDB [5] (stockage des mesures et traitement des données). Les SmS sont installés sur une console extérieure et alimentés par le bâtiment administratif. Pour communiquer avec le serveur local, ils se connectent au réseau local WiFi (LAN) et envoient leurs données via le routeur du réseau Internet.

Collecte des données brutes

La connexion au serveur central permet d'accéder aux données brutes restructurées par l'Agrégateur. La collecte des données a été effectuée sur une période de 4 mois. Seules

3. Serveur web : `https://httpd.apache.org/`
4. Flux de données visuel : `https://nodered.org/`
5. Base de données temporelle : `https://www.influxdata.com/influxdb/`

les périodes les plus représentatives, où les défaillances des SmS et des LCS étaient les plus présentes et critiques pour la mise en œuvre de la station, sont sélectionnées dans cette sous-section. Toutes les figures présentées ci-dessous sont extraites de l'interface graphique et représentent les données issues des SmS regroupées par types de paramètres. Les données et l'algorithme proposé sont disponibles via le lien fourni dans [51].

Au moment de la première mise en service, lors du déploiement du SmS, plusieurs problèmes pratiques sont apparus (panne de courant, dysfonctionnements, etc.). Les notifications en rouge sur la Fig. 5.6 indiquent l'arrêt de fonctionnement d'un SmS et les notifications en bleu correspondent à la mise en marche du SmS par des interventions humaines. Malgré les arrêts, la station de mesure de la zone a pu fournir des mesures de concentration continues pour les données du paramètre physique PM_{10}. La redondance des SmS permet d'élargir l'horizon des observations en effectuant plusieurs interventions pour remplacer ceux qui présentent des dysfonctionnements. Il convient également de noter que la station peut perdre des jours de données en cas de panne de courant (plusieurs jours d'observation ont été perdus au début janvier lors du déploiement), d'indisponibilité du réseau internet ou lorsqu'elle atteint sa capacité mémoire maximale. Lorsque la station atteint sa capacité mémoire maximale, cela signifie que l'espace de stockage prévu pour enregistrer de nouvelles données est saturé. En conséquence, la station n'a nulle part où stocker les nouvelles données collectées, ce qui entraîne la perte de ces données non enregistrées.

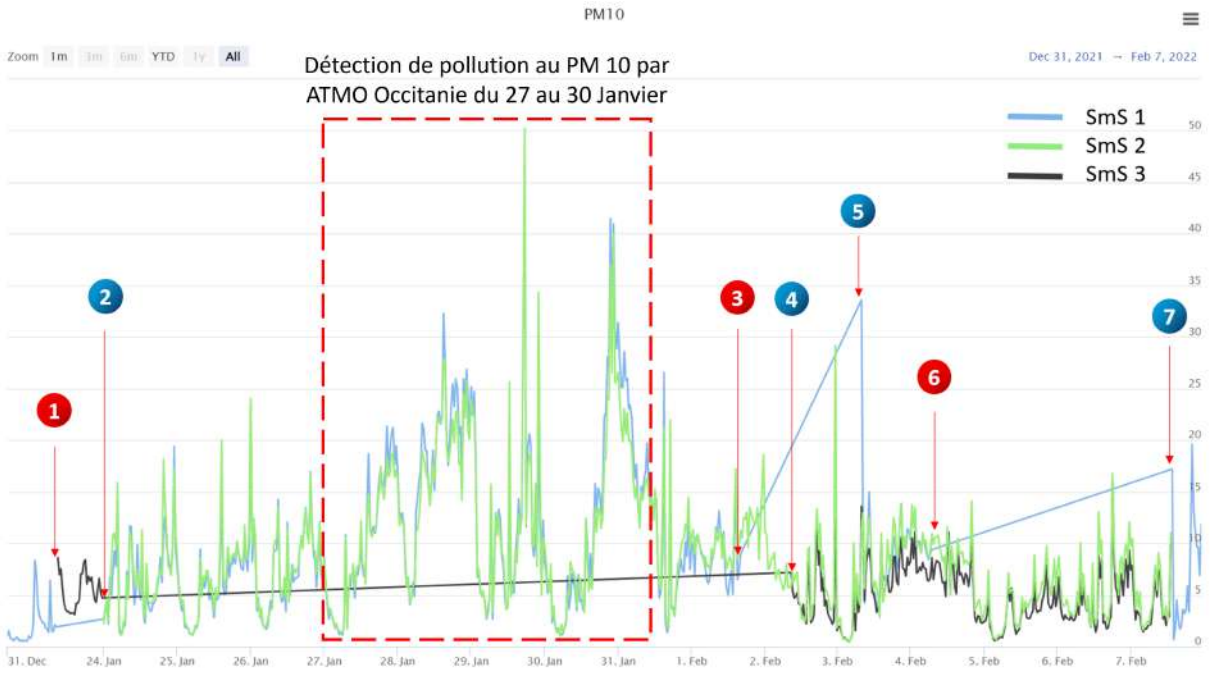


FIGURE 5.6 – Données brutes de PM_{10} pendant la phase de déploiement.

Les données brutes affichées dans la Fig. 5.6 sont insuffisantes en elles-mêmes, elles doivent être contextualisées et réajustées. En effet, les concentrations mesurées sont en ppm tandis que les normes de l'OMS sont en microgrammes par mètre cube ($\mu g/m^3$). Ces conversions nécessitent des adaptations en fonction de la température et de l'humidité. Malgré ces différences entre les mesures brutes et celles qui devraient être comparées aux normes, une augmentation de la concentration de PM_{10} peut être observée pendant la

période du 27 au 30 janvier soulignée par le rectangle rouge dans la Fig. 5.6. En contextualisant ces observations avec les alertes de pollution données par ATMO Occitanie, l'organisme chargé de surveiller la pollution de l'air, la station développée et déployée confirme l'alerte de pollution au PM_{10} par ses observations d'une augmentation des concentrations. Pour cette première itération, les SmS de la station effectuent des observations en temps réel avec une période moyenne de 5 minutes, ce qui permet de suivre l'évolution des concentrations de pollution dans une journée.

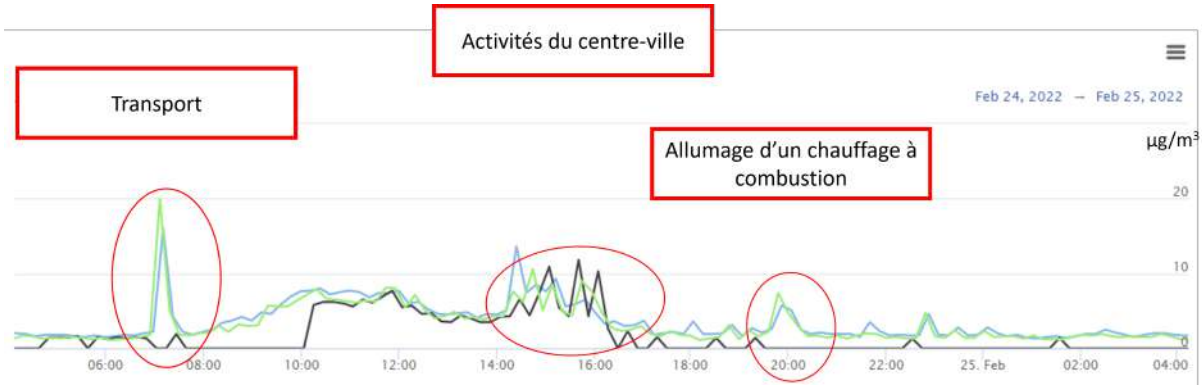


FIGURE 5.7 – Données brutes de $PM_{2.5}$ lors d'une journée typique à Argelès-Gazost.

La Fig. 5.7 montre les concentrations de particules fines ($PM_{2.5}$) pendant une journée typique. Cette pollution est principalement liée aux transports et aux activités du centre-ville. Pour une journée typique sans alerte de pollution, le premier pic observé correspond aux déplacements des personnes se rendant a leur travail (pic de 7h30). Ensuite, les émissions de particules fines augmentent au cours de la journée, avec des pics observés entre 14h et 16h30. Cette pollution est directement liée aux activités du centre-ville d'Argelès-Gazost qui est une zone de commerce active. Ensuite, le soir différents pics, dont le premier commençant à 19h, peuvent correspondre aux émissions issues des combustions de bois et des chaudières des habitants. Les observations effectuées avec la station de mesure sont tout à fait cohérentes avec les activités humaines. Tout au long de la journée, les concentrations mesurées sont faibles et les observations d'un SmS à l'autre peuvent varier en raison des problèmes de répétabilité évoqués dans l'état de l'art. Sur la figure 5.7 les observations du SmS 3 représentées en noir diffèrent des deux autres. L'utilisation d'un seul capteur de poussière dans ce cas d'utilisation peut mener à de mauvaises interprétations sur les pollutions et donc diminuer l'importance de l'exposition de la population aux polluants. C'est ainsi que la triple redondance permet de confirmer les pics de pollution par les observations communes des SmS et donc d'augmenter la fiabilité des mesures lorsqu'elles sont agrégées.

La singularité des capteurs de gaz

Les capteurs de gaz, lorsqu'ils sont déployés à l'extérieur, donnent des résultats très bruités, comme le montre la Fig. 5.8. Les concentrations sont exprimées en parties par milliard (ppb) et elles sont très faibles. Les limites de détection du capteur SGP30 sont atteintes, ce qui explique la variabilité d'une mesure à l'autre. Ainsi, ce type de comportement des capteurs conclut que la pollution de l'air en composés organiques volatils totaux (TVOC) est très faible pendant cette journée.

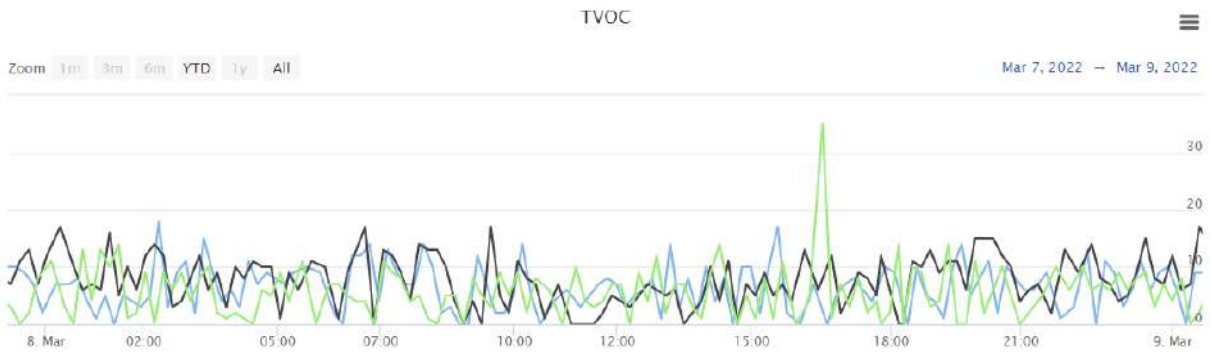


FIGURE 5.8 – Données brutes de TVOC.

Pour les capteurs de SO_2 (capteurs analogiques), les mesures obtenues après compensation de la température et de l'humidité montrent des résultats négatifs, conséquence de l'absence de pollution pour ce polluant. Sur la Fig. 5.9, les données issues des capteurs correspondent aux observations des températures avec un pic dans l'après-midi lorsque le soleil est au plus haut. Ces courbes correspondent aux formules de compensation appliquées aux observations du SO_2 et elles montrent la dépendance de la mesure avec la température.



FIGURE 5.9 – Données brutes de SO_2.

Pour le capteur MISC6814, qui est responsable des observation des grandeurs physiques correspondant aux polluants CO, NH_3 et NO_2, les concentrations sont très faibles et leur valeur analogique est difficile à interpréter par le microcontrôleur, ce qui conduit à un retour à la valeur zéro (Fig. 5.10).

Pour les capteurs de gaz, les limites techniques sont atteintes et il n'y a pas de mesures car les concentrations de pollution sont trop faibles. Lorsque les mesures sont disponibles, elles correspondent à des tests en intérieur (où les concentrations sont plus élevées qu'à l'extérieur). En effet, les capteurs de gaz LCS présentent des comportements singuliers à très faibles concentrations. Par conséquent, il n'est pas possible de détecter de fines concentrations, mais ces capteurs peuvent détecter des concentrations plus élevées de pollution, et dans certains cas contribuer à faire de la détection de fuites ou de pollutions importantes dans les zones sensibles.

Pour conclure sur ces premiers résultats, le choix du LCS est important pour mesurer la pollution et les capteurs de gaz utilisés pour cette station de mesure ne sont pas suffi-

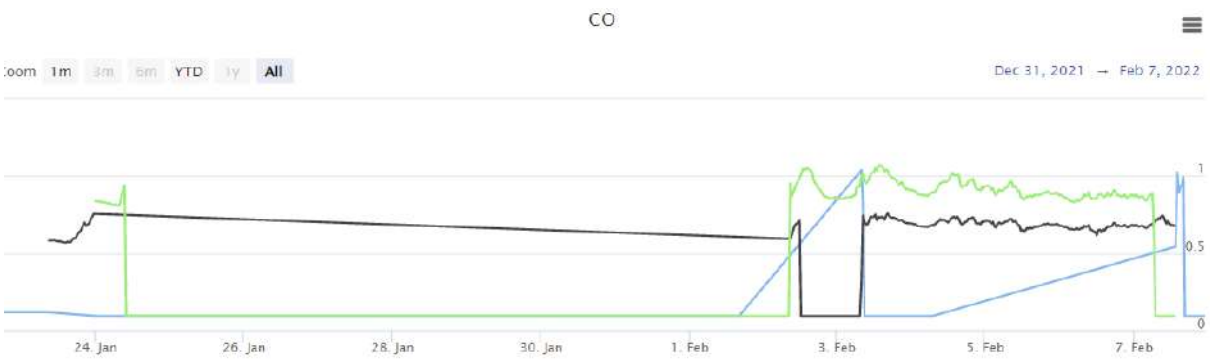


FIGURE 5.10 – Données brutes provenant du MISC6814.

samment sensibles malgré les efforts de calibration sur le terrain. Les mesures montrent que la qualité de l'air est excellente dans le centre-ville d'Argelès-Gazost pendant la période de déploiement des premiers SmS. De plus, d'autres mesures sont encore possibles en exploitant toutes les possibilités du LCS. Ainsi, pour améliorer les stations de mesure, deux stratégies peuvent être envisagées : la première consiste à choisir des capteurs plus sensibles et plus coûteux pour mesurer la pollution de manière plus fine, et la deuxième consiste à placer d'autres stations de mesure avec les capteurs actuels près des lieux identifiés comme sources de pollution (décharge, stations de traitement des eaux, etc.).

5.3.2 Fiabilisation des mesures : l'influence des hyperparamètres

La démarche de déploiement dans cette thèse est itérative, les premiers déploiements mettent en avant l'importance de la redondance pour élargir l'horizon de mesure mais d'autres efforts restent à faire pour fournir une synthèse fiable à partir des observations des LCS tout en excluant les valeurs aberrantes. Cette section étudie l'influence des deux hyperparamètres suivants :

— τ correspondant à la taille de la fenêtre d'observation de la station ;
— $R\varepsilon_i$, représentant l'erreur relative permise pour les LCS qui observent le paramètre φ_i pour la détection des valeurs incohérentes.

L'utilisation de la méthode va se concentrer sur les observations liées à φ_2 correspondant au paramètre physique $PM_{2.5}$. Cette focalisation s'explique par le fait que les LCS associés à $PM_{2.5}$ présentent plusieurs défaillances, faisant d'eux des cas d'étude pertinents.

Les synthèses pour une erreur relative autorisée de 30%

La méthode proposée est conçue pour être polyvalente, capable de fonctionner aussi bien avec des données préalablement collectées hors ligne qu'avec celles accessibles en ligne via une connexion. Pour montrer les influences des hyperparamètres, les données utilisées sont celles de la Figure 5.7 correspondant aux mesures de concentration de $PM_{2.5}$ sur une journée. Pour les SmS, leur période d'échantillonage est de cinq minutes ($Te_x = 5$ $minutes$) et la taille de la fenêtre d'observation est de dix minutes ($\tau = 10$ min). Le réglage de l'erreur relative autorisée $R\varepsilon_2$ est défini sur la base de l'erreur relative définie par le fabricant du LCS, elle est affinée au fur et à mesure des observations jusqu'à ce qu'une valeur seuil soit fixée. Ainsi, $R\varepsilon_2$ pour la détection des aberrations est fixée à 30% pour les LCS chargés des mesures de φ_2. En effet, la valeur de départ définie par le

fabricant est $\varepsilon r = \pm 15\%$, et les valeurs au-delà du double de l'erreur relative ne sont plus fiables et seront considérées comme des valeurs aberrantes.

Le résultat de la méthode de fiabilisation utilisant l'ensemble des données et se basant sur les paramètres précédents est présenté dans la Figure 5.11. La Figure 5.12 montre la sortie de l'algorithme lorsque $\tau = 20\ min$ et $R\varepsilon_2 = 30\%$. La taille des fenêtres glissantes a peu d'impact sur le nombre d'erreurs collectées par l'algorithme. En effet, plus la fenêtre d'observation est petite, plus les pics sont inclus dans la synthèse de la station. Un réglage de τ plus important aplatira la courbe de synthèse et augmentera le nombre de valeurs aberrantes. La comparaison des deux figures montre que lorsque τ augmente, les pics sont ignorés lors de la synthèse et le nombre d'erreurs sur les SmS 1 et 2 est approximativement identique. Cependant, il est possible de constater des groupements de valeurs aberrantes présent au niveau des pics sur les trois SmS lorsque τ est plus grand que la durée du pic sur la Figure 5.12. Cet évènement permet de conclure que c'est le paramètre mesuré qui atteint des valeurs anormales et n'est pas dû à la défaillance d'un LCS car toutes les observations sont identifiées comme étant des anomalies et sont regroupées dans un petit intervalle de temps. Les erreurs de SmS 3 sont plus fréquentes que celles des deux autres et leur nombre plus élevé indique des défaillances du composant en raison de plusieurs retour à la valeur zéro. L'indice de défaillance du LCS peut être calculé à partir du rapport entre les erreurs et le nombre total de mesures. La levée d'alertes sur les défauts matériels peut également être calculée en fonction du nombre d'observations comparées au autres SmS. En effet, lorsqu'un SmS n'envoie plus de données, la taille de sa série chronologique diminue par rapport aux autres SmS.

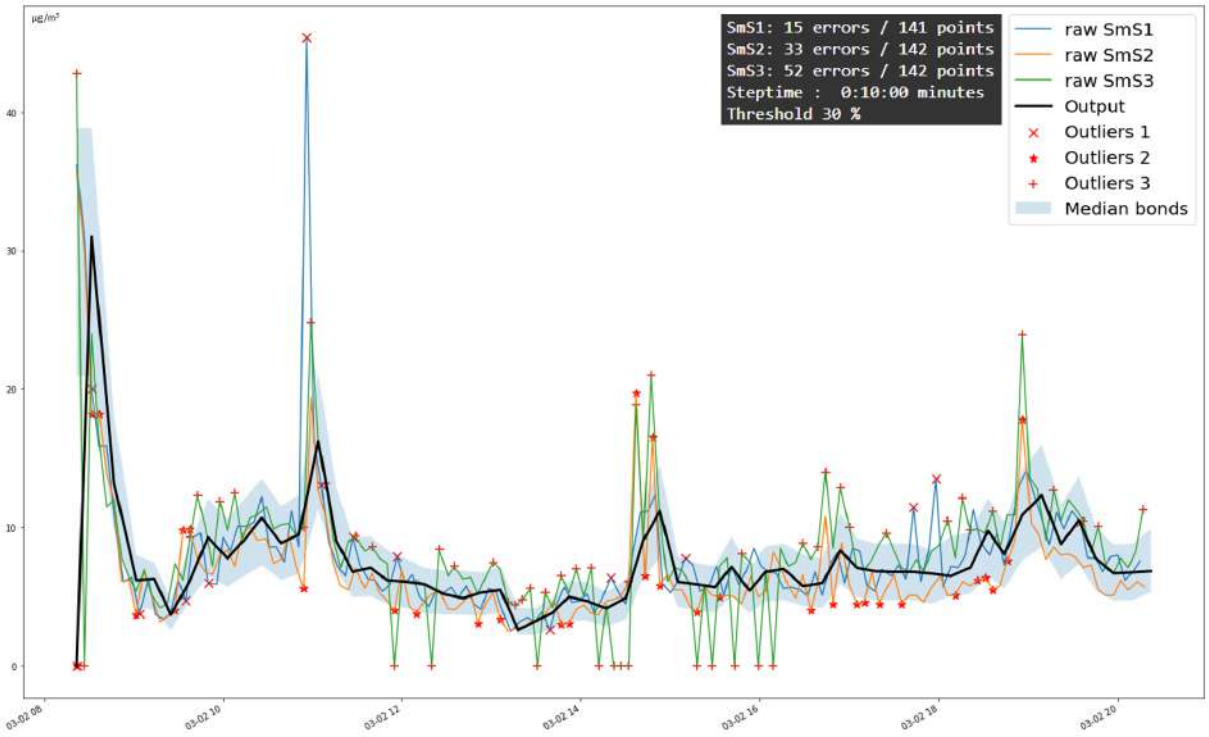


FIGURE 5.11 – Résultat de la méthode en utilisant la détection des valeurs aberrantes avec $\tau = 10\ min$ et $R\varepsilon_2 = 30\%$.

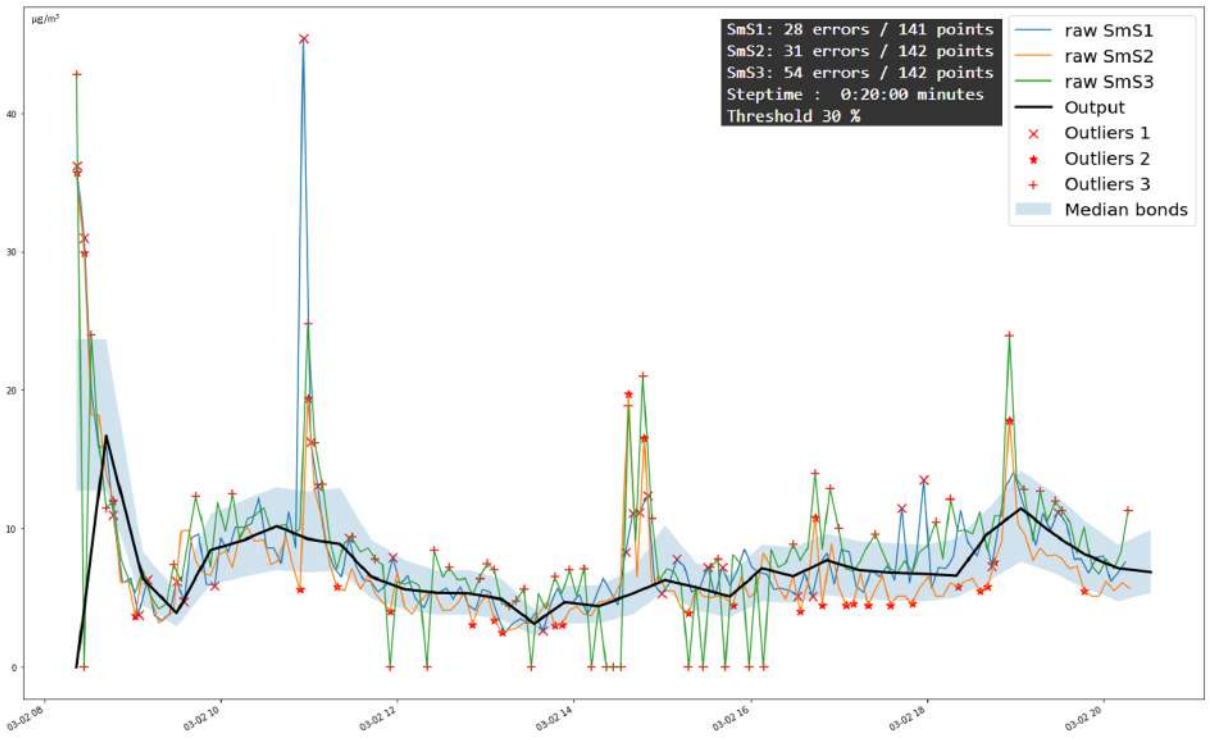


FIGURE 5.12 – Résultat de la méthode en utilisant la détection des valeurs aberrantes avec $\tau = 20\ min$ et $R\varepsilon_2 = 30\%$.

Les synthèses pour une erreur relative autorisée de 35%

Le paramètre $R\varepsilon_2$ a un impact significatif sur le nombre de valeurs aberrantes. Les comparaisons entre les figures 5.11 et 5.13 ainsi que les figures 5.12 et 5.14 confirment que plus le paramètre $R\varepsilon_2$ est élevé, plus les écarts entre les observations seront tolérés. Cependant, ce paramètre peut être décisif pour détecter une éventuelle dérive d'un LCS. En effet, en fonction du réglage de $R\varepsilon_2$, un grand nombre de mesures peut être détecté comme valeurs aberrantes par rapport à d'autres LCS, entraînant un décalage des valeurs observées par rapport à l'ensemble des observations collectées.

Pour la synthèse à l'échelle de la station de mesure, la taille de la fenêtre d'observation τ est un facteur important. Selon l'objectif de la mesure, elle sera adaptée soit pour suivre de près la pollution jusqu'à la détection des pics de courte durée, soit pour obtenir une synthèse sur une plage de temps plus large afin d'évaluer l'exposition à la pollution de fond d'un paramètre physique. Cette synthèse est peu influencée par τ en raison de la robustesse de la médiane sur les extrêmes. La combinaison de la synthèse et des valeurs aberrantes de la série chronologique permet d'observer les typologies de défaillance au niveau du matériel mais aussi au niveau de la qualité de l'air. Un indice de défaillance peut donc être calculé à partir du nombre d'erreurs de LCS et du nombre d'observations effectuées par le SmS.

5.3.3 Deuxième prototype avec fiabilisation des données

Lors du déploiement du premier prototype, il a été constaté que seulement les données brutes issues des SmS sont affichées dans l'interface graphique sans une synthèse facilitant la lecture des observations. Au niveau de la disponibilité de la station de mesure et des

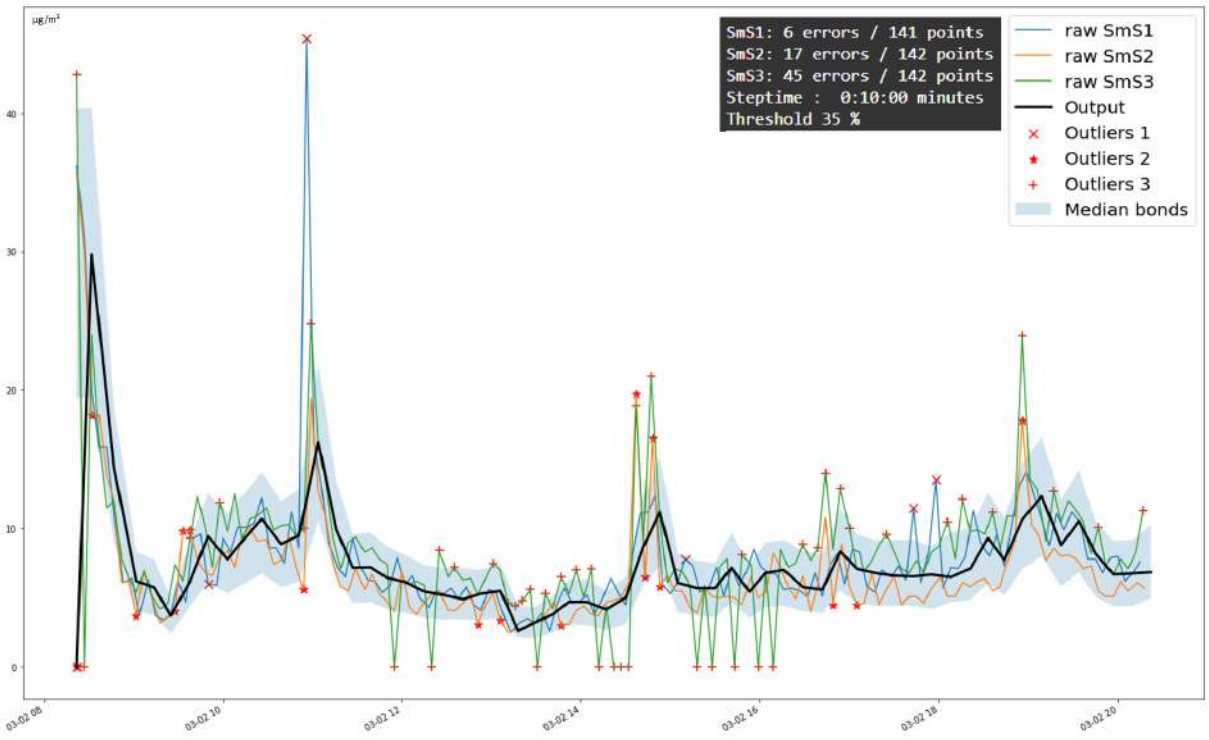


FIGURE 5.13 – Résultat de la méthode en utilisant la détection des valeurs aberrantes avec $\tau = 10\ min$ et $R\varepsilon_2 = 35\%$.

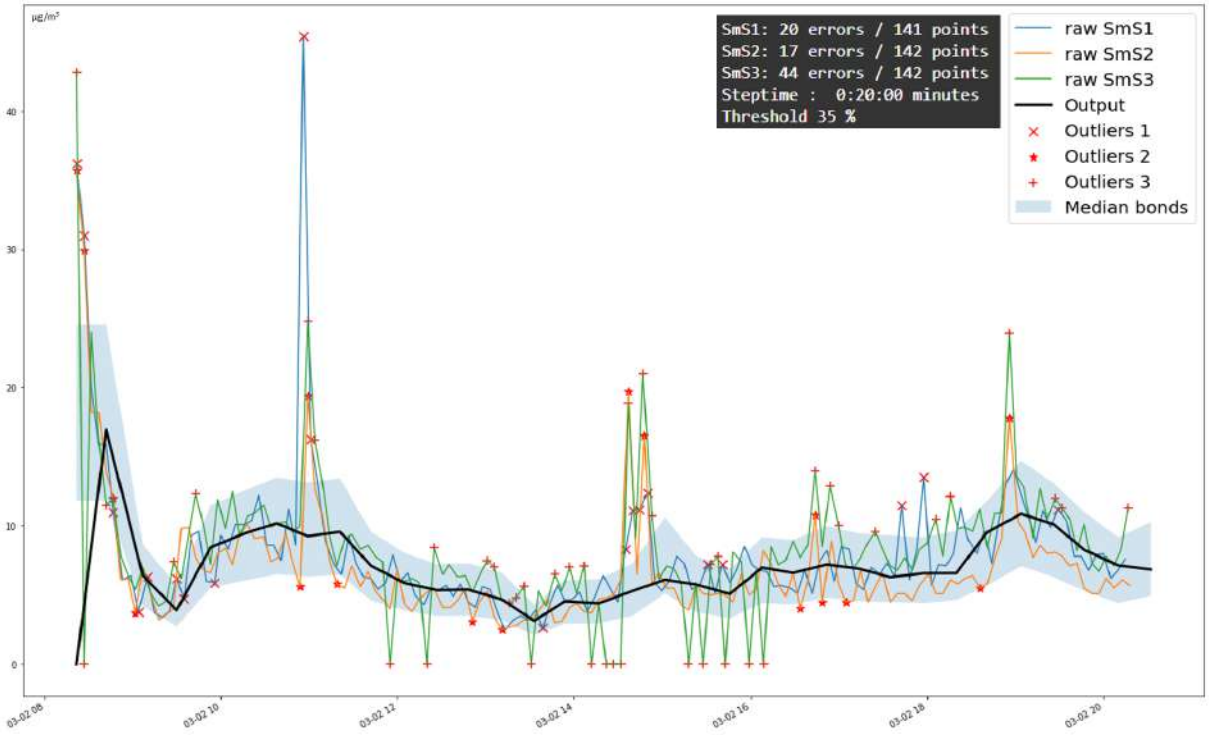


FIGURE 5.14 – Résultat de la méthode en utilisant la détection des valeurs aberrantes avec $\tau = 20\ min$ et $R\varepsilon_2 = 35\%$.

interventions sur les SmS, la détection des défaillances et de leur diagnostic était manuelle et nécessitait l'intervention d'un expert.

Pour la deuxième itération, au niveau matériel, des capteurs numériques ont été choisis. Cela permet un prototypage plus rapide et contourne les problèmes rencontrés avec les capteurs MISC6814 et SO_2 au niveau de l'interprétation du microcontroleur et peuvent donner des erreurs de conversion et/ou de compensation. Ces capteurs numériques intègrent directement d'autres capteurs d'humidité et de température et effectuent les conversions en micro-gramme par mètre cube lors de la mesure du gaz. Ils ont également l'avantage de faciliter la maintenance par leur connexion au circuit intégré sans soudure avec des fiches de connexion. Également, ils sont opérationnels dès le redémarrage de l'ESP32. La conception du circuit imprimé permet désormais plusieurs modifications pour évoluer en fonction des observations d'autres paramètres. Enfin, des mesures ont été prises pour prévenir une dégradation rapide des circuits électroniques dans un environnement extérieur, notamment en plaçant le PCB verticalement et en ajoutant une couche de vernis sur les soudures.

Au niveau logiciel, la fiabilisation des données issues des stations de mesures présentées dans cette thèse est implémentée au niveau de l'Agrégateur et du serveur local. Les Agrégateurs sont implémentés sur des micro-ordinateurs de type Raspberry Pi. Leur rôle consiste à agréger les observations SmS et à extraire le $S.I.D$ avant de le transmettre au serveur local. Ce dernier présente les résultats à travers l'interface graphique. A présent le SmS est robuste et peut supporter plusieurs modifications.

Les Figures présentées dans cette section illustrent les résultats générés par l'interface utilisateur du serveur local pour la station de mesure déployée à Argeles-Gazost. Les $S.I.D$ (Synthèse, Indices et Détection) concaténés sont représentés par trois graphiques au sein d'une même figure. Seuls les résultats les plus pertinents sont présentés, car il existe des similitudes entre les différents paramètres. Les hyperparamètres choisit sont les suivants : la fenêtre d'observation a une durée d'une heure ($\tau = 1h$) et un LCS est considéré fiable lorsqu'au moins les deux tiers ($Pc = 2/3$) de ses observations sont considérées fiables par la méthode. La collecte des données a été réalisée sur une période de quatre mois. Seules les périodes les plus représentatives, où les défaillances du SmS et du LCS étaient les plus présentes et critiques pour la mise en œuvre de l'algorithme, ont été sélectionnées dans cette sous-section. Il est essentiel de noter que cette sélection ciblée permet une analyse plus pointue des problèmes rencontrés et facilite l'identification des facteurs contribuant aux défaillances observées. La Figure 5.15 correspond au $S.I.D$ sur la période de minuit à midi du 8 février, pour les LCS mesurant le paramètre physique humidité, l'hyperparamètre $R\epsilon$ est de 20%. Le premier graphique affiche les données brutes de chaque LCS des trois SmS, les seuils supérieur et inférieur basés sur la matrice de seuils définissent la bande verte sur les observations brutes. La synthèse effectuée avec les valeurs du vecteur Reliable correspond à la courbe noire. Les détections de défaillance du matériel sont également représentées par des lignes verticales en rouge. Une icône en forme de coche verte située au-dessus d'une ligne verticale rouge indique que tous les SmS sont en état opérationnel et ont tous effectué des observations. L'icône en forme de croix rouge, annotée avec « N°3 », placée au-dessus d'une ligne verticale rouge, correspond à la détection d'une défaillance matérielle dans le SmS 3.

Le deuxième graphique représente l'indice de confiance global de la station. Il varie de 0 à 3 et correspond au nombre de SmS en fonctionnement. Dans le cas présent, il varie uniquement de 2 à 3 car un seul SmS était défectueux, ce qui donne un indice de confiance de 2 lorsque les observations d'un SmS étaient manquantes. Le troisième graphique affiche les indices de défaillance de chaque SmS pour le paramètre d'humidité. Une zone verte indique la plage de valeurs pour les indices de défaillance qui lorsque

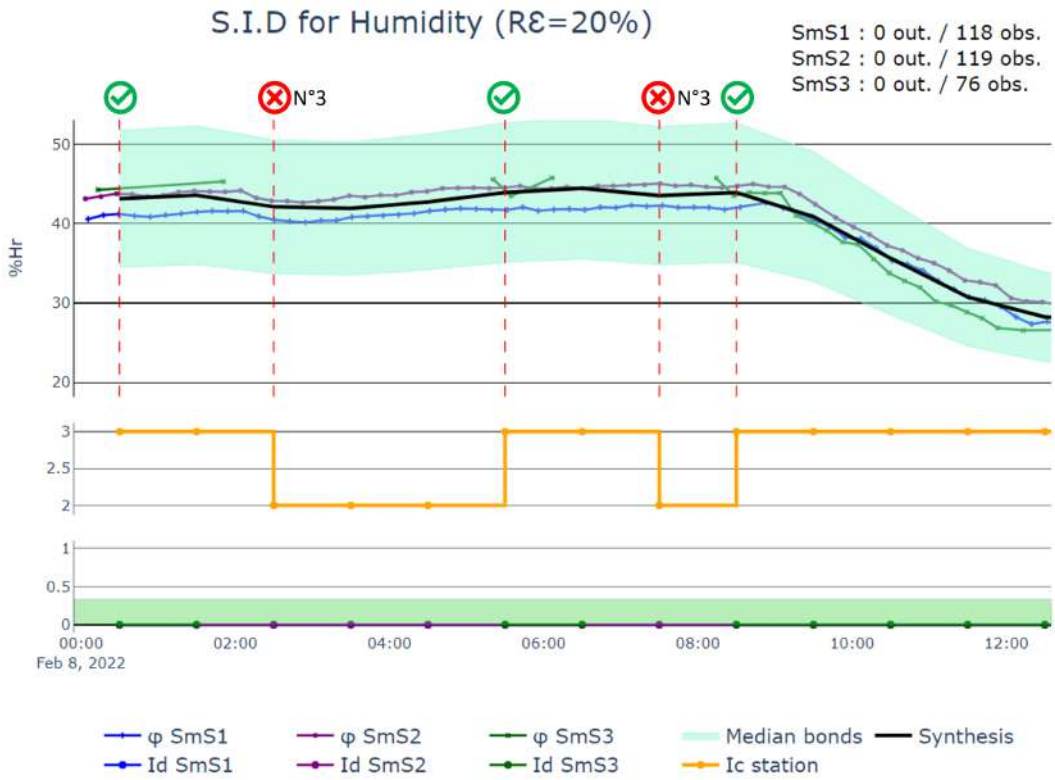


FIGURE 5.15 – Visualisation du *S.I.D* pour le paramètre Humidité lors de multiples déconnexions du SmS 3 sur une période d'une demi-journée.

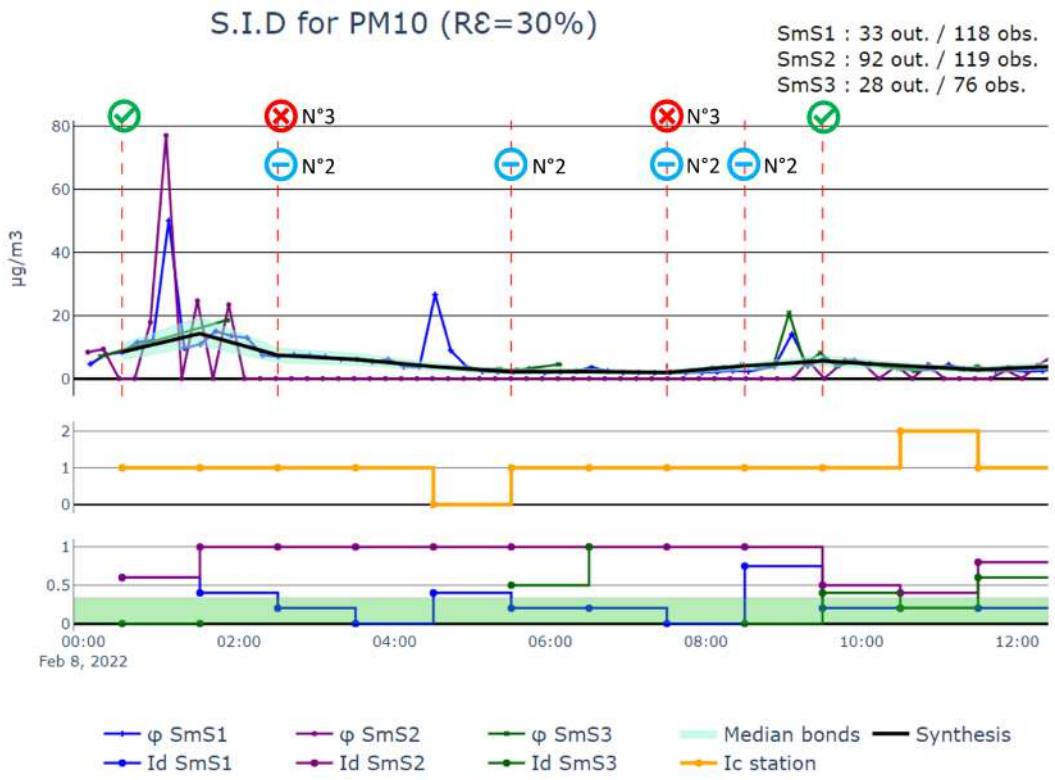


FIGURE 5.16 – Visualisation du *S.I.D* pour le paramètre PM_{10} lors de multiples déconnexions du SmS 3 et le blocage à la valeur zéro du LCS du SmS 2.

leur valeur est comprise entre 0 et $(1 - Pc)$ considère que les observations du LCS sont pertinentes pour la station de mesure. Sur ce graphique, les indices de défaillances sont tous égaux à zéro car aucune valeur aberrante n'a été détectée. Cette observation peut également être visualisée dans la synthèse, où les observations se situent dans la bande verte matérialisant les limites de la matrice de seuils. Outre la visualisation graphique du manque d'observations du SmS 3, une annotation globale correspondant aux erreurs cumulées par rapport aux observations totale par SmS sur la plage de donnée est située à côté du titre. Dans le cas du SmS 3, en comparaison des autres SmS, il apparaît qu'il y a un nombre moindre d'observations. Ces déficiences d'observation conduisent à une confirmation du diagnostic initial d'une défaillance matérielle du SmS 3. Des investigations ultérieures ont révélé que le SmS 3 avait été déconnecté du réseau local à deux reprises, entraînant une perte de données. Grâce aux indices de défaillances et de confiance, les premières détections automatiques peuvent être confirmées.

Les paramètres physiques de température et de pression présentent des comportements similaires et ne seront donc pas présentés. L'accent de ce déploiement est mis sur le comportement défectueux du LCS, où le paramètre PM_{10} implique un LCS basé sur un capteur de poussière sujet à diverses défaillances. L'avantage du capteur de particule PM_{10} est qu'il répond aux attentes en matière de mesure et constitue un excellent sujet de test pour la méthode. Le paramètre $PM_{2.5}$ ne sera pas présenté dans cette section car, dérivé du même capteur, il présente les mêmes caractéristiques techniques. Le $S.I.D$ du paramètre physique PM_{10} est présenté sur la Figure 5.16, où les graphiques sont dans le même ordre que sur la figure précédente et les observations sont prises sur la même période de temps. Cette période de temps a été choisie en raison des multiples défaillances du SmS 2. En effet, celui-ci présente une défaillance masquée et des écarts de mesure importants par rapport aux deux autres. Une icône en forme de trait bleu, annotée avec « N°2 » et positionnée au-dessus d'une ligne verticale rouge, correspond au LCS du SmS 2 qui est bloqué à la valeur zéro. L'indice de confiance reste à 1 dans la première période car le LCS du SmS 2 présente des valeurs aberrantes avant de se bloquer à la valeur zéro, ensuite, sans aucune intervention, il se remet à fonctionner correctement vers 9h00. Pendant la période où le SmS 2 est bloqué, seul le LCS du SmS 1 est suffisamment fiable et il sera le seul pris en compte dans l'indice de confiance. De manière similaire à ce qui est observé sur la Figure 5.15 pendant la même période, le SmS 3 est manquant. L'indice de confiance est réduit à zéro vers 4h00 en raison d'un pic de pollution, ce qui a augmenté le nombre d'erreurs dans le LCS du SmS 1. À partir de 9h00, lorsque le SmS est de nouveau en service, tous les LCS sont opérationnels, mais le SmS 2 continue à présenter des observations différentes des deux autres. À partir des indices de défaillance et de l'indice de confiance global, le comportement du LCS peut être mesuré et quantifié. Par exemple, lorsque l'indice de confiance est égal à zéro alors que tous les SmS n'ont pas montré des indices de défaillance élevés, on peut en déduire qu'un pic de pollution de courte durée (inférieur à τ) a été observé. Le résumé de la période montre le même nombre d'observations que sur la Figure 5.15, sauf que les valeurs aberrantes sont comptabilisées cette fois en raison de la nature du LCS présentant des problèmes de répétabilité et de blocage. Ainsi, la méthode permet de définir d'autres types de défaillances et montre que le LCS du SmS 1 est le plus fiable avec 33 erreurs sur 118 et que le SmS 2 a exclu certaines valeurs en raison de sa défaillance.

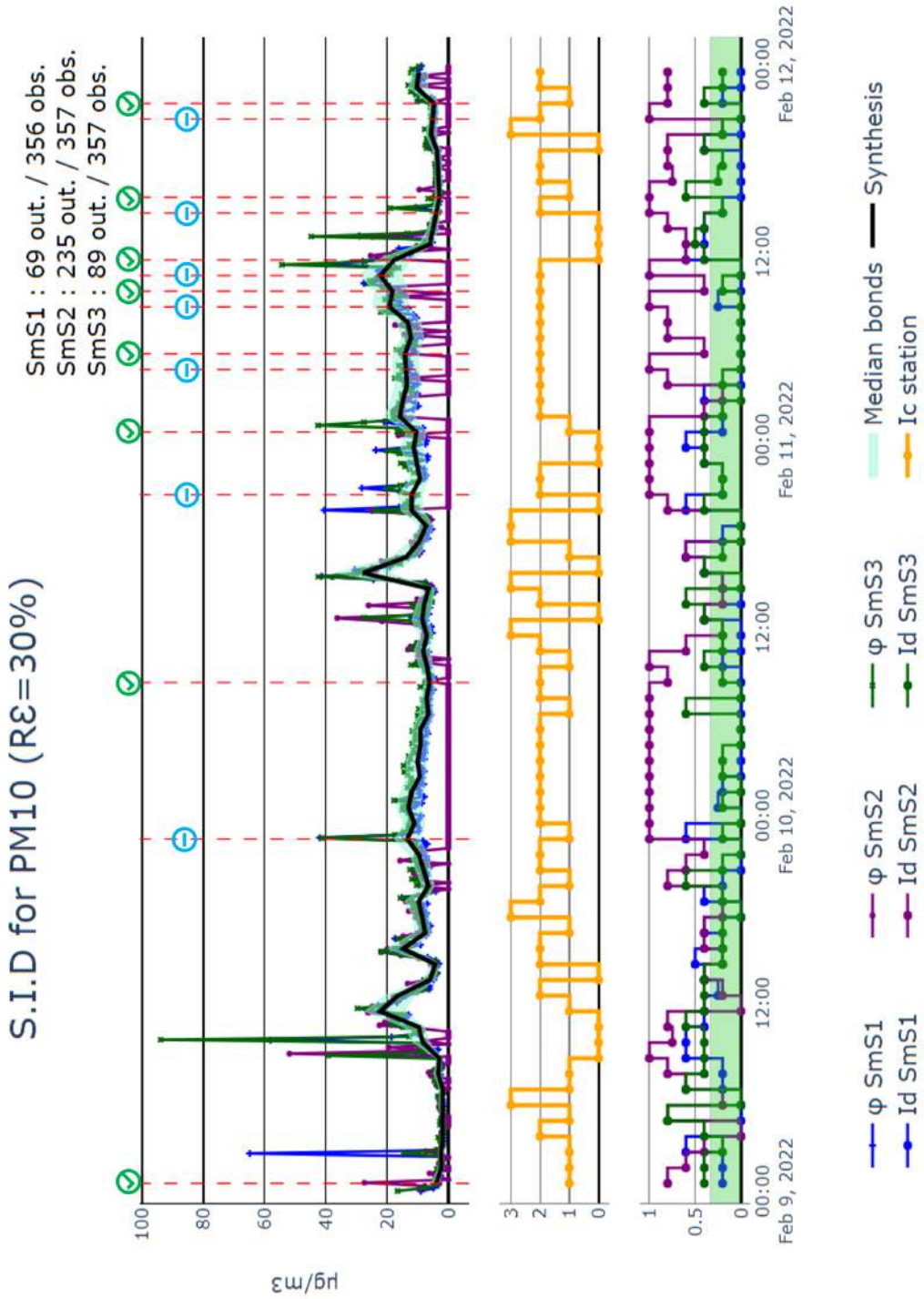


FIGURE 5.17 – Visualisation du $S.I.D$ pour le paramètre PM_{10} lors de multiples défaillances matérielles du LCS du SmS numéro 2 et des pics de pollution sur trois jours.

Les résultats dérivés de l'application de la méthodologie $S.I.D$ aux niveaux de PM_{10} sont présentés sur la Figure 5.17, où l'identification de différents schémas de comportement est facilitée par les indices de défaillance et de confiance. Sur cette figure, le rectangle vert sur le graphique de l'indice de défaillance correspond à la région dans laquelle les observations du LCS sont considérées comme fiables. Les pics de pollution peuvent être discernés en identifiant les moments où l'indice de confiance Ic est égal à zéro et le comportement défectueux du SmS 2 peut être évalué en calculant sa période de blocage constant à une seule valeur, ainsi que la fréquence à laquelle il est considéré comme peu fiable. On peut voir que le SmS 2 tombe rarement dans cette zone, ce qui réduit l'indice de confiance à 2. L'importance de cette figure réside dans sa capacité à présenter un aperçu complet et continu des différentes défaillances et une synthèse sur une période de temps étendue. Une telle représentation permet l'identification de problèmes potentiels et informe les décideurs dans le contexte de périodes de fonctionnement prolongées.

5.4 Conclusion du chapitre

Les résultats présentés dans ce chapitre sont issus d'un développement complet d'une station de mesure de la qualité de l'air. La démarche itérative permet de montrer les évolutions des prototypes par un affinage du co-design du matériel et du logiciel au sein des SmS, impliquant une évolution progressive des deux éléments pour améliorer les performances globales. En ce qui concerne le matériel, tous les logiciels ont été développés à l'aide d'outils et de bibliothèques standard et open source. La méthode d'agrégation a été formalisée et mise en œuvre sur un prototype entièrement développé dans notre laboratoire de recherche et par le doctorant.

Le premier prototype met en avant la possibilité de récolter des observations à partir de LCS en les rendant intelligents (SmS). Ainsi, les SmS sont des composants capables de fournir des observations de paramètres physiques à des moments et en des lieux spécifiques. Cependant, les remontées des observations brutes des SmS ne sont pas suffisantes pour les détections des anomalies autant de la qualité de l'air que celles dues aux SmS. Une amélioration du logiciel au sein de l'Agrégateur permet désormais de fiabiliser les Synthèses des paramètres physiques tout en excluant les valeurs aberrantes. L'expérience du premier déploiement a montré les limites des LCS spécialisés dans la mesure des gaz. C'est ainsi qu'il a été décidé, pour le deuxième prototype, de ne pas les intégrer. L'expérience a montré que les concentrations à mesurer sont trop faibles par rapport aux capacités physique des LCS, rendant les observations inutilisables.

En termes de conception matérielle, des ajustements peuvent être apportés au boîtier de protection des SmS. Lorsqu'ils sont enfermés dans une boîte, les composants chauffent et les mesures sont influencées par l'humidité de l'air. De plus, l'affirmation selon laquelle les trois SmS mesurent un seul point peut être améliorée grâce à une meilleure conception du boîtier.

Les périodes d'échantillonnage ont été sélectionnées en tenant compte des capacités des capteurs et du taux d'évolution modéré des phénomènes physiques. Pour la période choisie, τ, la détection des erreurs se produit après l'exécution de calculs dans une fenêtre d'observation d'une heure lors du déploiement en temps réel. Ce délai n'est pas critique pour la surveillance de la qualité de l'air car des actions de maintenance peuvent être planifiées dans un délai de réponse de 48 heures. Cependant, cela peut poser des défis pour la détection de polluants dans des environnements industriels à haut risque, par exemple.

À l'heure actuelle, une seule station de mesure est déployée avec trois SmS intégrés et un serveur opérationnel. Comme les données sont concentrées sur une seule station dans la ville d'Argelès-Gazost, l'Agrégateur a été intégré au serveur local, centralisant ainsi tout le traitement sur un seul nœud. L'inconvénient de cette concentration est le manque d'expérience dans le déploiement de l'Agrégateur sur le terrain, en particulier en ce qui concerne les problèmes de connexion ou les arrêts complets. Néanmoins, cette architecture est suffisante pour servir de preuve de concept et valider la conception des capteurs intelligents.

Enfin, les choix des hyper-paramètres τ, Pc et $R\varepsilon_S$ sont sujets à discussion, car ils influencent considérablement le $S.I.D.$ en sortie de l'Agrégateur. La caractérisation du PERV actuelle pour les erreurs relatives des mesure des capteurs nécessite l'utilisation des LCS du même fabricant. Des études futures devraient explorer la méthode avec des LCS provenant de différents fabricants. Dans le cadre de la surveillance de l'exposition de la population à la pollution de fond, τ est de l'ordre de l'heure. Au niveau de $R\epsilon_S$ des améliorations restent à faire car sur la Figure 5.15, les seuils sont trop importants en raison de la variabilité des observations entre les LCS mesurant l'humidité. Pc influence fortement le résultat de l'indicateur de confiance Ic, ce paramètre définit un seuil qui lorsqu'il est franchi permet de considérer que les observations du capteur comme étant fiables. Cet indice de confiance a pour but de comparer les synthèses des stations entre elles afin de prendre en compte uniquement celles ayant le score de confiance le plus élevé pour les heatmaps à l'échelle de la zone d'intérêt dans laquelle sont implantées les stations de mesure. Ces hyper-paramètres seront affinés à mesure que les stations seront déployées et exploitées.

La validation des observations est difficile au niveau de la synthèse des stations, car aucune station de mesure conventionnelle n'est disponible à proximité de la station de mesure pour servir de référence. Le déploiement d'une station de mesure à côté d'une station conventionnelle serait bénéfique pour valider les mesures dans un premier temps. Cependant, la topologie de la zone d'intérêt peut être à l'origine de grandes disparités entre les stations en fonction de leur emplacement géographique et de leur altitude. Des concentrations de pollution sont plus importantes au fond des vallées lors d'épisodes montagneux appelés « mer de nuages ».

La continuité du service peut être assurée tant qu'il y a au moins un SmS en service, et la détection des valeurs aberrantes permet de planifier des interventions pour le remplacement ou la réparation. En cas de panne de courant des SmS, les mesures ne sont pas effectuées et plusieurs jours d'observations peuvent être perdus si aucune intervention n'est réalisée. Cependant, ces premiers résultats permettent un post-traitement pour améliorer la qualité des mesures et tenir compte des variations de la qualité de l'air sur le terrain. Il est également envisageable d'intéger des batteries sur les SmS pour limiter les conséquences d'un tel évènement.

Chapitre 6

Conclusion générale

Sommaire

6.1 Résumé des chapitres 128
6.2 Rappel des objectifs principaux de la thèse 130
6.3 Contributions de la thèse 130
6.4 Implications et applications 131
6.5 Perspectives de recherche 132

6.1 Résumé des chapitres

La présente thèse aborde un défi complexe et multidisciplinaire dans le domaine de la surveillance de la qualité de l'air. Les enjeux liés à l'acquisition de données fiables, à la détection des défaillances et à la couverture territoriale sont examinés au travers des chapitres du document. Chaque chapitre contribue à une composante spécifique du problème global, menant à une compréhension plus profonde et à des solutions innovantes. Dans cette section de la conclusion, chaque chapitre est résumé pour récapituler les principaux développements et contributions.

Chapitre 1 : Introduction générale
Ce chapitre présente la problématique de la surveillance et de la gestion de la qualité de l'air dans le territoire de la CCPVG. Il détaille les particularités uniques du territoire, telles que sa topologie montagneuse et la présence persistante de la pollution de fond. Le chapitre souligne les limites des méthodes actuelles de mesure et propose une approche qui utilise la redondance des LCS de manière innovante. La nécessité d'une démarche scientifique rigoureuse pour protéger la santé des habitants et impliquer la communauté est clairement établie. Ce chapitre sert de point de départ pour l'exploration des questions de recherche clés et la réalisation des objectifs de la thèse.

Chapitre 2 : État de l'art
Ce chapitre éclaire la relation entre la pollution atmosphérique et la qualité de l'air. Il souligne l'importance de mesurer les concentrations de polluants, avec une attention particulière aux défis rencontrés dans les territoires à topologie particulière. Les diverses méthodes et technologies disponibles pour surveiller la qualité de l'air sont explorées, y compris les avantages et les préoccupations associés aux LCS, tels que leur facilité de mise

en œuvre, leur bas coût, et les questions de fiabilité. Une présentation des méthodes de surveillance conventionnelle selon une approche descendante est également proposée en parallèle avec les travaux menés avec des LCS.

Chapitre 3 : Proposition d'une architecture de monitoring territorial : une approche ascendante
Dans ce chapitre, l'architecture proposée pour une surveillance optimisée de la qualité de l'air a été détaillée, mettant en lumière sa dimension spatio-temporelle. La dimension spatiale présente des propres défis, notamment l'estimation des mesures en tout point du territoire. Les choix entre les approches descendante et ascendante, qui sont complémentaires, ont été explorés et la couverture territoriale rendue possible grâce aux ajouts de stations de mesure a été discutée. L'aspect temporel correspond aux observations en temps réel des stations de mesure sur les points de prélèvement et de l'ensemble des traitements de la chaîne d'acquisition tels que l'archivage et la prédiction. Des méthodes d'interpolation, tant dans la dimension spatiale que temporelle, ont été introduites. L'architecture prend en compte la couverture du territoire et la répartition géographique des zones d'intérêt. Elle englobe également l'ensemble de la chaîne d'acquisition, depuis les LCS utilisés pour concevoir les stations de mesure jusqu'aux serveurs locaux et central. Ces serveurs présentent des interfaces utilisateur adaptées aux différents décisionnaires.

Chapitre 4 : Données, Traitements et Décisions
Ce chapitre détaille un processus de traitement des données en huit blocs pour la surveillance de la qualité de l'air. Il se concentre sur la formalisation mathématique de la fiabilisation des données, décrivant une méthode en neuf étapes pour l'analyse précise et fiable des données, la gestion des données non synchronisées, la détection des anomalies et la génération des indicateurs fiables. Les matrices résultantes facilitent une évaluation rapide de la fiabilité des observations des stations de mesure, jetant les bases d'un système robuste pour la surveillance de la qualité de l'air.

Chapitre 5 : Mise en oeuvre et déploiement de l'architecture
Ce chapitre présente les travaux de développement d'une station de mesure de la qualité de l'air, en mettant en évidence la démarche itérative de la conception matérielle et logicielle appellé co-design. Les défis et les limites des LCS, les ajustements du boîtier de protection, les périodes d'échantillonnage et les choix des hyper-paramètres affectant la synthèse sont détaillés. Le déploiement d'une station dans la ville d'Argelès-Gazost est présenté, avec des réflexions sur la continuité du service, la validation des observations et la détection de polluants. Des considérations sur les défis futurs et les améliorations possibles sont également abordées.

Après avoir synthétisé les principaux développements et conclusions de chaque chapitre, il est essentiel de revenir aux fondements de cette recherche. Cette réexamination permet de mettre en lumière les objectifs principaux qui ont guidé ce travail et de mettre en perspective les défis et les questions de recherche abordés.

6.2 Rappel des objectifs principaux de la thèse

Les objectifs principaux de la thèse sont énumérés ci-après.

1. **Surveillance en temps réel de la qualité de l'air**
   L'objectif principal de ce travail était de déployer une instrumentation adaptée sur le territoire de la CCPVG pour obtenir une cartographie précise et en temps réel de la qualité de l'air. Cela incluait le développement ou l'adaptation de technologies de surveillance, tout en tenant compte des spécificités locales.
2. **Prédiction et proactivité**
   La thèse visait également à élaborer des modèles prédictifs fiables pour anticiper les variations de la qualité de l'air. Ces modèles étaient destinés à permettre des actions proactives, comme la régulation du trafic ou la limitation de certaines activités polluantes lors des pics de pollution prévus. Bien que étudié, cet objectif n'a pas été réalisé en raison du temps utilisé pour répondre au premier objectif.
3. **Engagement sociétal**
   L'objectif était aussi d'impliquer activement les acteurs locaux dans le processus de surveillance et de gestion. Cela comprenait la sensibilisation aux enjeux, la formation à l'utilisation des outils de surveillance, et la promotion de pratiques respectueuses de l'environnement. Cet objectif a été en partie atteint grâce à la sensibilisation du public lors de journées dédiées à la surveillance de la qualité de l'air. Des présentations ont été effectuées et des décisions prises en collaboration avec des représentants locaux.
4. **Développement d'une solution innovante et économique**
   Comparé à d'autres travaux dans ce domaine, cette thèse se distingue par son utilisation innovante de la redondance des LCS et la conception en partant de zéro d'une solution pour maîtriser le coût tout au long de la chaîne d'acquisition jusqu'aux outils d'aide à la décision. Le but était de fournir des outils plus accessibles et plus fiables pour la surveillance de la qualité de l'air, même dans les régions à ressources limitées.

En somme, ces objectifs étaient guidés par la nécessité de répondre aux défis spécifiques de la surveillance de la qualité de l'air sur le territoire de la CCPVG, en tenant compte des contraintes financières, logistiques, et des spécificités du territoire.

6.3 Contributions de la thèse

La présente thèse a apporté plusieurs contributions majeures dans le domaine de la surveillance de la qualité de l'air, notamment dans les territoires à topologie montagneuse. Ces contributions sont résumées ci-dessous.

— Proposition d'une architecture pour la surveillance de la qualité de l'air : une architecture optimisée a été proposée, incorporant la redondance des LCS. Cette architecture permet une surveillance en temps réel et une couverture territoriale étendue, en s'adaptant aux contraintes financières et aux besoins spécifiques du territoire de la CCPVG.
— Fiabilisation des données provenant des stations de mesure : un processus de traitement des données en huit blocs a été développé, permettant une fiabilisation des données. Une formalisation mathématique utilisant les LCS et la redondance active a été élaborée pour améliorer les synthèses provenant des données et la détection et

le diagnostic des défaillances. Cette méthodologie contribue à la détection des anomalies, à la gestion des données non synchronisées, et à la génération d'indicateurs fiables pour la surveillance de la qualité de l'air.

— Conception et mise en œuvre d'une station de mesure utilisant les LCS : ce projet a été initié afin d'élaborer une station de mesure s'appuyant sur la technologie des micro-capteurs. L'approche en co-design a prouvé son efficacité et a validé la possibilité d'utiliser les LCS en conditions réelles, tout en prenant en compte les spécificités locales du territoire de la CCPVG.

— Déploiement, validation sur le terrain et couverture territoriale : le déploiement d'une station de mesure dans la ville d'Argelès-Gazost a permis de valider les observations, de détecter les polluants et de faciliter la prise de décision concernant la maintenance des stations grâce aux comparaisons entre les stations de la zone d'intérêt. Cela a permis d'identifier les défis futurs notamment sur la continuité du service et la détection de polluants. L'approche présentée dans cette thèse peut être adaptée aux futures mises en œuvre des LCS, en vue de couvrir à terme l'ensemble du territoire de la collectivité.

— Contribution à l'engagement sociétal et difficultés rencontrées : les objectifs de la recherche incluaient également l'engagement sociétal et la fourniture d'éléments concrets pour la prise de décision. La thèse a mis en lumière la nécessité d'une démarche collaborative avec la communauté locale et a permis de proposer des solutions pour une action proactive sur la qualité de l'air. Les défis futurs et les améliorations possibles, ainsi que les lacunes du LCS, ont été identifiés, offrant une réflexion sur les obstacles à surmonter et les perspectives d'amélioration.

Ces contributions reflètent une tentative sérieuse d'aborder les défis de la surveillance de la qualité de l'air dans les territoires non couverts par une approche ascendante. Bien que la solution proposée s'inspire de méthodes existantes, elle offre une nouvelle perspective sur l'utilisation des LCS et tente d'améliorer la fiabilité des données et la couverture territoriale. La collaboration avec la collectivité locale a présenté des défis et n'a pas toujours abouti aux résultats escomptés. Cela souligne la complexité des enjeux rencontrés et la nécessité d'une réflexion continue pour rendre la surveillance de la qualité de l'air plus efficace et inclusive, tout en tenant compte des spécificités du territoire et des attentes des différents acteurs impliqués.

Après avoir souligné les contributions majeures de cette thèse, il est intéressant d'examiner les implications plus larges de ces travaux. La section suivante explore comment ces contributions peuvent être appliquées dans différents contextes et quel impact elles peuvent avoir sur la société, l'industrie et l'environnement.

6.4 Implications et applications

L'architecture développée dans cette thèse et les résultats obtenus ont des applications concrètes et des répercussions significatives dans divers domaines. Les implications potentielles et la mise en œuvre de l'architecture sur différents aspects de la société et de l'environnement sont explorés ci-dessous :

1. **Implications pour l'industrie et la politique publique :** la recherche présentée dans cette thèse, bien qu'expérimentale, démontre la possibilité de dupliquer les stations de mesure afin de couvrir l'entièreté du territoire. Une telle approche peut

non seulement bénéficier aux pouvoirs publics dans la surveillance de la qualité de l'air, mais aussi aux industries pour surveiller les zones à haut facteur de pollution. Les recommandations formulées soulignent l'importance d'une collaboration efficace entre les différentes parties, y compris la sensibilisation de la population à la surveillance de la qualité de l'air afin d'éviter les dégradations des stations.

2. **Applications potentielles dans différents contextes :** outre la surveillance de la qualité de l'air, l'architecture proposée peut être généralisée à d'autres applications, telles que la surveillance de la qualité de l'eau, la détection de vibrations dans les structures de génie civil, la surveillance de la température et de l'humidité dans l'agriculture, et le suivi des conditions météorologiques dans des zones éloignées.
3. **Impact sur la santé publique et l'environnement :** si généralisée et mise en application, la recherche de cette thèse pourrait avoir un impact significatif sur la santé publique. Une meilleure connaissance de l'exposition de la population aux activités humaines permettrait d'obtenir des preuves solides pour justifier des actions sur le territoire afin de réduire les émissions de pollution, contribuant ainsi à la protection de la santé et de l'environnement.
4. **Limitations et considérations éthiques :** bien que prometteuse, cette application est limitée par la technologie des LCS, et les résultats, assez grossiers, n'atteignent pas la précision des stations conventionnelles. Toutefois, cela peut être suffisant pour prévenir des pics de pollution. Sur le plan éthique, la surveillance de la qualité de l'air peut aussi être perçue comme une surveillance des activités humaines, avec des implications potentielles pour la vie privée. De plus, des conflits peuvent surgir avec les activités économiques du territoire, comme le tourisme ou les zones industrielles, où la pollution peut être tolérée pour des raisons économiques. La régulation des écobuages, pratique ancestrale, pourrait également être confrontée à des résistances.

Cette recherche ouvre la voie à de nouvelles possibilités en matière de surveillance environnementale tout en mettant en lumière des défis et des considérations éthiques. Les opportunités futures pour renforcer et élargir ces méthodes seront explorées dans la section suivante.

6.5 Perspectives de recherche

La recherche présentée dans cette thèse ouvre de nombreuses portes et soulève de nouvelles questions. Les travaux futurs peuvent se concentrer sur plusieurs aspects clés pour améliorer davantage l'architecture et étendre ses applications. Dans ce qui suit sont données quelques perspectives de recherche sélectionnées pour leur pertinence.

— **Amélioration de la couverture et intégration des interpolations :**
En prévision d'un futur déploiement de plusieurs stations et de l'ajout de points de prélèvement sur le territoire, une exploration des techniques d'interpolation spatiale et temporelle sera faite afin d'identifier les méthodes les plus adaptées au cas d'étude de la CCPVG. Une fois que des modèles auront été développés à partir de ces interpolations, couvrant à la fois les dimensions spatiales et temporelles de l'architecture proposée, des prévisions de la qualité de l'air de premier ordre seront établies. Ces prévisions seront ensuite soumises à validation, en s'appuyant sur les données d'ATMO. L'approche ascendante développée dans cette thèse prendra en

compte les conditions aux limites provenant de l'approche descendante d'ATMO, renforçant ainsi sa complémentarité pour la compréhension de la qualité de l'air et de ses variations.

— **Classification des défaillances avec un Autoencoder :**
Dans le cadre de travaux futurs, l'Agrégateur pourrait être amélioré avec des capacités d'apprentissage automatique pour apprendre les comportements défectueux des LCS et fournir des diagnostics plus précis et rapides. Une approche possible est l'utilisation d'un auto-encodeur, un réseau neuronal constitué d'un encodeur et d'un décodeur (voir Figure 6.1). L'encodeur a pour but de réduire la dimension des données d'entrée pour permettre de regrouper les données dans une dimension plus réduite mais bien plus caractéristique d'un fonctionnement normal. Le décodeur permet à partir de la dimension réduite de reconstruire les données d'entrée. L'erreur de reconstruction, la différence entre l'entrée et la sortie, permet de détecter une anomalie. Un processus d'étiquetage en fonction des défaillances pourra attribuer un diagnostic attaché à cette valeur d'erreur. Des travaux pour détecter des anomalies sur le taux de CO2 dans une pièce ont été menés en utilisant un autoencodeur [70], apprenant les dépendances à long terme entre les données issues de séquences de séries temporelles et identifiant le seuil optimal sur la base des erreurs de reconstruction.

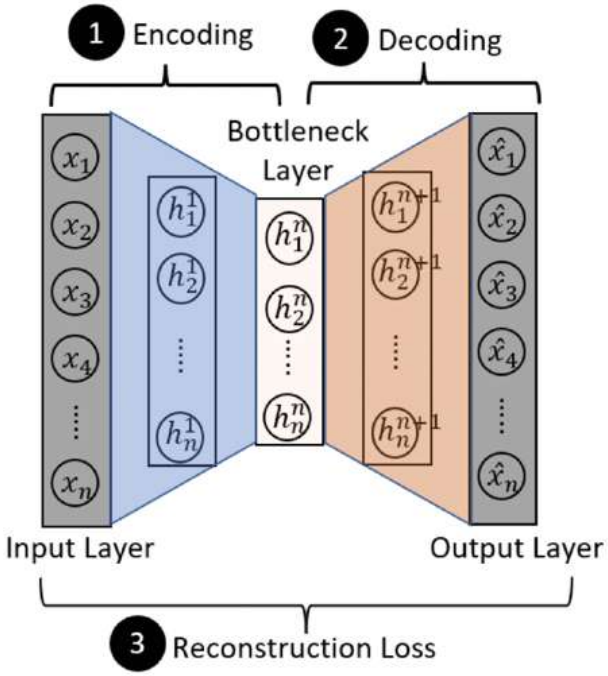


FIGURE 6.1 – Fonctionnement d'un autoencodeur [70]

— **Prédiction de la durée de vie des LCS :**
L'Agrégateur peut être amélioré pour communiquer les valeurs aberrantes en temps réel, plutôt qu'à la fin de la fenêtre d'observation, permettant ainsi une détection plus rapide des défaillances. En utilisant les observations en temps réel des LCS, l'Agrégateur peut apprendre à définir des indices de santé et à construire un modèle de défaillances pour chaque composant. Ceci permettrait de détecter les défaillances

et de prédire les durées de vie utiles restantes avant les arrêts complets pour maintenance corrective. Par exemple, un LCS poussiéreux pourrait être modélisé en fonction de sa tendance à présenter des valeurs anormales avant de se bloquer, se dégradant progressivement à mesure qu'il est saturé de poussière. Cette approche, bien que prometteuse, nécessiterait le déploiement de plusieurs stations pendant une année entière et un retour d'expérience considérable en matière de défaillances.

— **Amélioration du boîtier des SmS :**
Afin de converger vers l'hypothèse que les SmS mesurent les mêmes paramètres physiques en un point donné, et pour garantir des mesures plus précises, le boîtier des SmS doit être amélioré. La démarche consiste à modéliser les flux thermiques créés par les composants du LCS lors de son fonctionnement. Des mesures thermiques régulières permettront de créer un modèle thermique du PCB en fonctionnement, comme le montre la Figure 6.2.

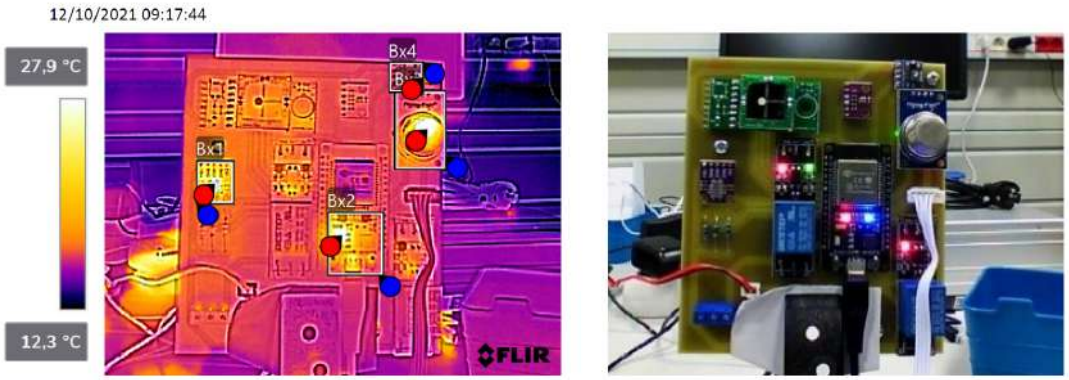


FIGURE 6.2 – Mesures thermiques du PCB pour modélisation des flux thermiques.

L'observation des mouvements des flux d'air à l'intérieur du boîtier sera essentielle pour expliquer les écarts de température observés par rapport aux mesures officielles, comme illustré dans la Figure 6.3. À partir du modèle thermique du PCB, différents boîtiers pourront être testés afin de sélectionner le concept optimal, permettant à la fois de mesurer la qualité de l'air et de protéger les équipements électroniques.

Un premier prototype est représenté dans la Figure 6.4, mais il doit encore être affiné par d'autres simulations et tests en conditions réelles lors du déploiement final du réseau de mesure sur l'ensemble du territoire.

— **La Prédiction de la qualité de l'air :**
Un autre défi scientifique à aborder après cette thèse concerne la prédiction de la qualité de l'air avec des LCS. La principale difficulté réside dans la différence de puissance de calcul entre une collectivité territoriale et une AASQA, qui utilise des supercalculateurs pour simuler des modèles complexes basés sur des données validées. La multiplication des points de prélèvement fiabilisés et la distribution du calcul sont des pistes prometteuses. L'utilisation de réseaux de neurones embarqués pourrait être une première étape vers la distribution du calcul. Après une année de collecte de données, les variations et les mouvements de pollution pourront être classés et identifiés. Cela permettra de reconnaître certains phénomènes et d'estimer leur durée. À l'instar des techniques utilisées en mathématiques financières,

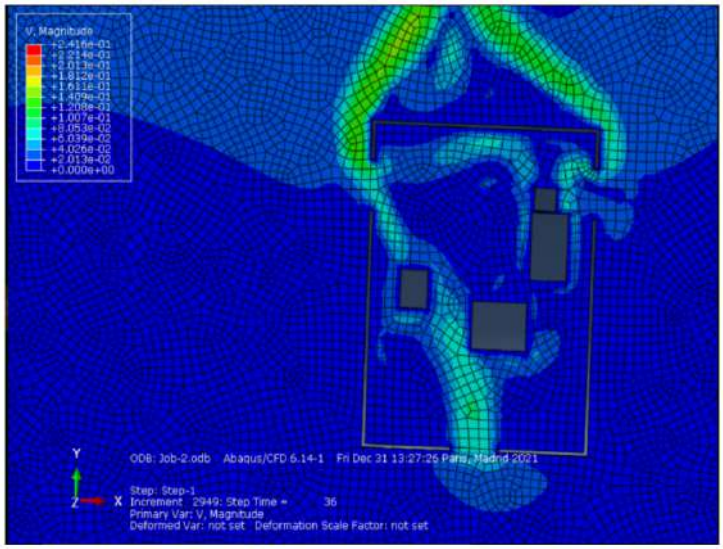


FIGURE 6.3 – Simulation des flux thermiques dans le boîtier initial modélisé en 2D.

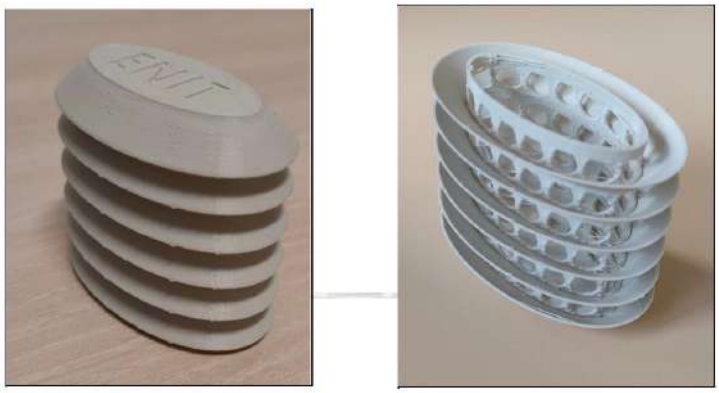


FIGURE 6.4 – Prototype de boîtier imprimé en 3D.

des croisements entre les moyennes mobiles sur plusieurs aspects temporels pourraient servir à émettre des signaux de fin d'épisode de pollution. Par exemple, le croisement d'une baisse de l'humidité atmosphérique avec une augmentation de la pression atmosphérique peut indiquer qu'il pleut. Cela signifie que la pollution due à la poussière est réduite par la capture des particules fines par les gouttes de pluie.

— **Le Système d'information du réseau de mesure :**
Pour conclure, des travaux supplémentaires peuvent être menés concernant le système d'information du réseau de mesure pour répondre aux défis liés à la gestion et au devenir des données. Ce système a pour objectif de surveiller la qualité de l'air à différentes échelles spatiales, allant de la zone d'intérêt jusqu'à l'échelle nationale. L'architecture du réseau de mesure, basée sur les LCS, se divise en trois dimensions principales : le territoire, la zone d'intérêt et la station de mesure. Chaque dimension a ses propres responsabilités en matière de confidentialité et d'infrastructures.

— *Stations de mesure :*
Chargées de transférer les données des SmS au serveur local responsable de la zone d'intérêt, cette dimension a été le principal axe de développement dans

cette thèse.

— *Serveurs locaux des zones d'intérêt :*
Ils agrègent et évaluent les données fiabilisées des stations, offrant deux types de support de décision. Le premier concerne la haute disponibilité du réseau de mesure, et le second l'évaluation des données relatives à la qualité de l'air, présentées des outils d'aides à la décision.

— *Liaisons entre le serveur central et les serveurs locaux :*
Le serveur central, responsable au niveau du territoire, récupère uniquement les synthèses de données liées à la qualité de l'air. Il doit travailler en collaboration avec les serveurs locaux responsables de la fiabilisation des données. Les développements futurs se concentreront sur la personnalisation des synthèses pour répondre aux besoins spécifiques des décideurs, en explorant différentes stratégies de déploiement et d'interpolation.

Ces dimensions, imbriquées entre elles, constituent un cadre complexe qui nécessite une exploration et une optimisation continues pour assurer une surveillance efficace et fiable de la qualité de l'air à différentes échelles.

Ces perspectives de recherche ouvrent des voies passionnantes pour l'avenir de la surveillance de la qualité de l'air et offrent de nombreuses possibilités d'innovation et d'amélioration.

Annexe A

Annexes

A.1 La pollution atmosphérique

La croissance de la population et de l'activité humaine a entraîné une augmentation des émissions de polluants depuis la révolution industrielle. Cette pollution atmosphérique est responsable de perturbations climatiques ainsi que d'une baisse de la qualité de l'air. L'atmosphère est majoritairement composée d'azote (78,1 %), de dioxygène (21 %), de vapeur d'eau (entre 1 % et 4 %), ainsi que d'autres gaz tels que l'argon (environ 0,93 %). L'hélium, le néon, et d'autres gaz inertes sont présents en quantités infimes et ne contribuent pas à la pollution atmosphérique comme le montre la Figure A.1. Les autres gaz se trouvent en concentrations infimes mesurées en partie par million (ppm). Cependant, l'augmentation de la concentration de certains gaz mineurs peut influencer de manière significative l'équilibre de l'atmosphère [15].

Les polluants sont des gaz et des composés chimiques ayant un effet néfaste sur l'environnement et la santé des êtres vivants, provenant soit de sources naturelles, soit de l'activité humaine. La dispersion des polluants dans l'air dépend de leur durée de vie et de leur réactivité, certains étant dégradés tandis que d'autres réagissent et deviennent des polluants secondaires. Bien que les polluants se trouvent principalement dans la troposphère, les plus stables peuvent migrer dans la couche supérieure de l'atmosphère. Les phénomènes météorologiques peuvent provoquer un brassage des concentrations de polluants, rendant la pollution locale un problème global. Enfin, certains polluants sont plus surveillés en raison de leur origine anthropique, qui peut être contrôlée, ainsi que de leurs impacts sanitaires à court et long termes.

A.1.1 Origine des polluants

Les polluants gazeux, tels que le dioxyde de soufre SO_2 et les oxydes d'azote NO_2, NO, sont classés selon leur espèce chimique. Bien que le dioxyde de carbone CO_2 n'ait pas d'effet direct sur la santé des êtres vivants, il est considéré comme un polluant en raison de sa contribution à l'effet de serre et au changement climatique, qui affectent indirectement les êtres vivants. Les particules en suspension, classées selon leur taille et appelées matière particulaire (PM), constituent le deuxième type de polluant. Les émissions de polluants sont considérées comme des sources, tandis que leur captation est appelée un puits (par exemple, les forêts sont des puits de CO_2). Les polluants primaires sont émis directement depuis la source, tandis que les polluants secondaires sont créés par des réactions chimiques dans l'atmosphère. La pollution peut avoir une origine naturelle ou anthropique, c'est-à-

Constituant	Formule chimique	Masse moléculaire (12C = 12)	Pourcentage en volume par rapport à l'air sec (10^{-2} v/v)	Masse totale dans l'atmosphère (g)	Temps de résidence moyen dans l'atmosphère
Atmosphère totale				$5{,}136.10^{21}$	
Air sec		28,9644	100,0	$5{,}119.10^{21}$	
Vapeur d'eau	H_2O	18,0153	Variable	$0{,}017.10^{21}$	6 à 15 jours
Diazote	N_2	28,0340	78,084	$3{,}866.10^{21}$	15.10^6 ans
Dioxygène	O_2	31,9988	20,948	$1{,}185.10^{21}$	8.10^3 ans
Argon	Ar	39,9480	0,934	$6{,}59.10^{19}$	infini
Dioxyde de carbone	CO_2	44,0099	0,0370	$2{,}72.10^{18}$	15 ans
Néon	Ne	20,1830	$1{,}818.10^{-3}$	$6{,}48.10^{16}$	infini
Hélium	He	4,0026	$5{,}2.10^{-4}$	$3{,}71.10^{185}$	infini
Méthane	CH_4	16,0430	$\approx 1{,}7.10^{-4}$	$\approx 4{,}3.10^{15}$	9 ans
Hydrogène	H_2	2,0159	$\approx 5{,}0.10^{-5}$	$\approx 1{,}8.10^{14}$	10 ans
Protoxyde d'azote	N_2O	44,0128	$\approx 3{,}1.10^{-5}$	$\approx 2{,}3.10^{15}$	150 ans
Monoxyde de carbone	CO	28,0106	$\approx 1{,}2.10^{-5}$	$\approx 5{,}9.10^{14}$	2 mois
Ozone	O_3	47,9982	$2\text{-}200.10^{-6}$	$\approx 3{,}3.10^{15}$	1 à 2 mois
Ammoniac	NH_3	17,0306	$\approx 0{,}1$ à 1.10^{-6}	$\approx 3{,}0.10^{13}$	20 jours
Dioxyde d'azote	NO_2	46,0055	$\approx 1.10^{-7}$	$\approx 8.10^{12}$	1 jour
COV (composés organiques volatils)	$C_xH_yO_z$	variable	$\approx 0{,}1$ à 1.10^{-6}	$\approx 10^{14}$	heures-jours
Dioxyde de soufre	SO_2	64,063	$\approx 2.10^{-8}$	$\approx 2{,}3.10^{12}$	1 jour
Sulfure d'hydrogène	H_2S	34,080	$\approx 2.10^{-8}$	$\approx 1{,}2.10^{12}$	1 jour

FIGURE A.1 – Gaz constituant l'atmosphère [15].

dire créée par les activités humaines.

Les sources naturelles

Les sources de pollution naturelle proviennent généralement des phénomènes suivants [17] :

- Les volcans émettent du soufre S, des particules en suspension et du CO_2 ;
- La faune et la flore émettent des composés organiques volatils COV et du CO_2 ;
- Les vents sont à l'origine des particules en suspension (poussières) et des aérosols marins (gaz + particules) ;
- Les éclairs émettent des oxydes d'azote NO et de l'ozone O_3 par combinaison du N_2 et du NO_2 ;
- Le soleil, par rayonnement, crée des polluants photochimiques, dont l'ozone, c'est une pollution géogénique.

Les sources antropiques

Les sources anthropiques de pollution sont diverses et dépendent des secteurs d'activité. L'industrie et les transports utilisent la combustion d'énergie fossile, qui produit des polluants tels que le CO, le CO_2, les COV, les NO_x, les SO_2 et les métaux lourds, ainsi que des particules fines et ultrafines (PM2,5 et PM1). D'autres secteurs industriels émettent

des COV et des particules en suspension, selon leur secteur d'activité et leur consommation d'énergie. Les incendies, considérés comme des activités humaines, émettent également des COV, du CO, du CO_2 et des particules. L'agriculture est également responsable d'une part importante des émissions, notamment l'épandage et la dispersion en aérosols de pesticides et d'engrais NH_3, le lisier et les boues d'épuration, responsables des COV et du NH_3, ainsi que l'élevage, qui contribue au méthane CH_4 et aux gaz à effet de serre. Ces polluants sont transportés par les vents ou les phénomènes dynamiques des masses d'air, et sont finalement absorbés par des puits ou continuent d'augmenter leur concentration dans l'atmosphère.

A.1.2 Cycle de vie des émissions

Les polluants primaires proviennent principalement de sources situées à la surface du sol. Leur dispersion s'effectue de manière horizontale et verticale. En fonction de leur avancée verticale, ils subissent une transformation secondaire. Les polluants sont transportés sur une distance qui est limitée par leur durée de vie, appelée temps caractéristique du transport atmosphérique. Le schéma présenté dans la Figure A.2 illustre le cycle de vie des polluants.

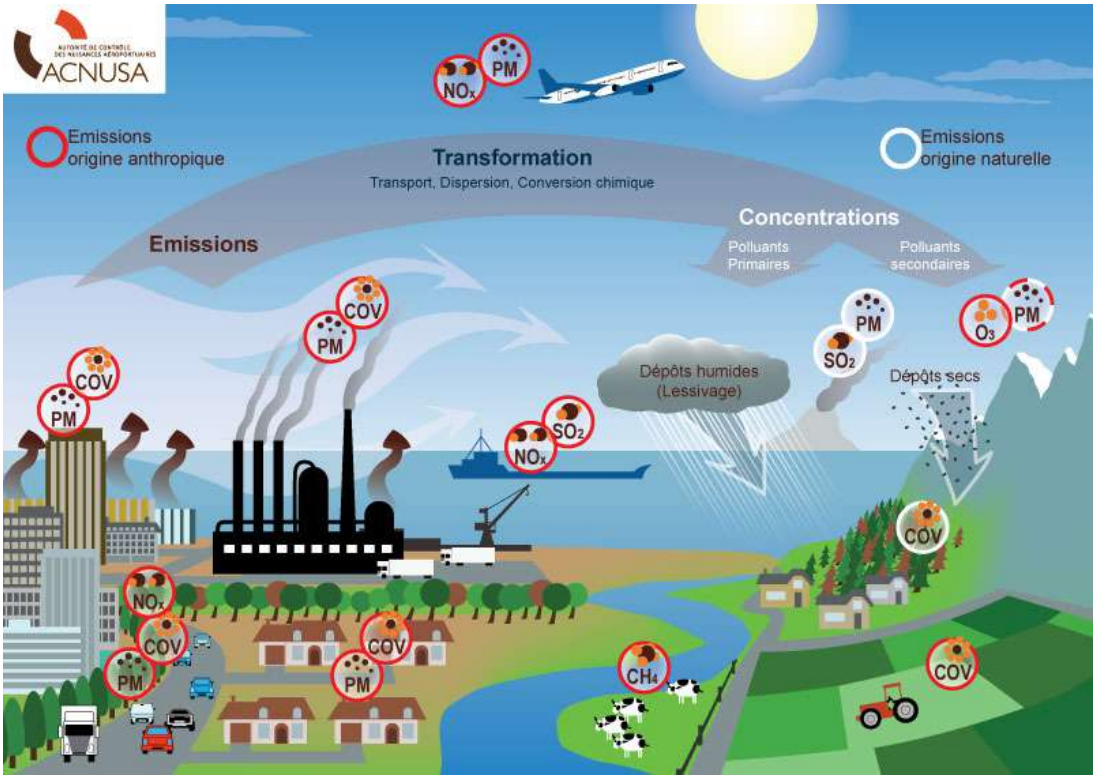


FIGURE A.2 – Cycle de vie des émissions [1].

Les impacts de la pollution sont évalués à trois échelles géographiques distinctes : locale, régionale et planétaire. Les polluants ayant une durée de vie courte ont un impact local, tandis que les polluants persistants peuvent être transportés sur de longues distances, même à haute altitude, et avoir un impact planétaire en affectant le climat ou la couche d'ozone [65].

Pour mesurer la qualité de l'air, les concentrations de fond et les polluants transportés sont additionnés pour la prise en compte de la pollution totale afin de mesurer la qualité

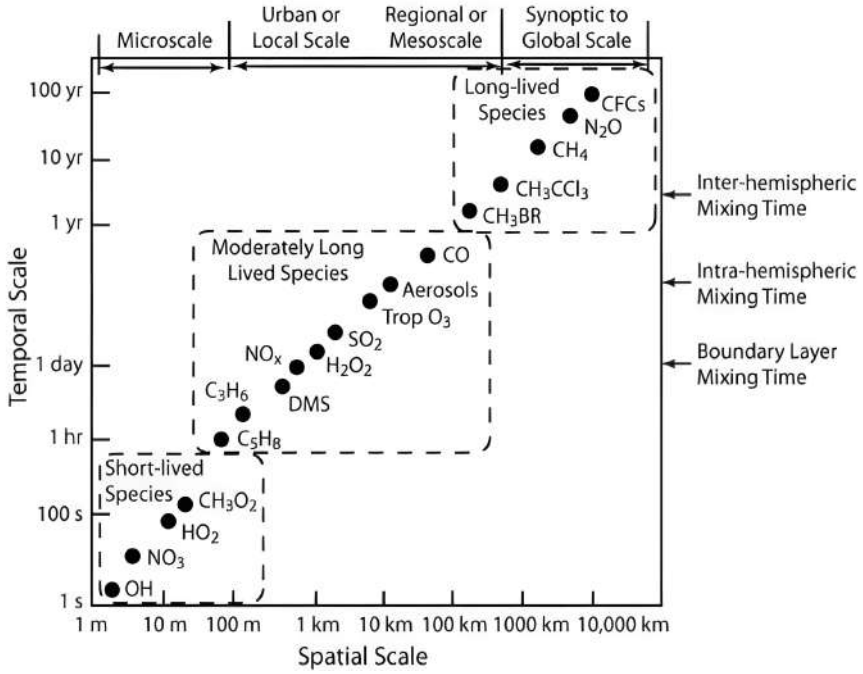


FIGURE A.3 – Variabilité spatiale et temporelle des constituants clés de l'atmosphère [13].

de l'air. La durée de vie d'un polluant permet d'estimer son origine potentielle. La figure A.3 illustre la durée de vie des polluants et leur résidence dans les couches atmosphériques, permettant ainsi de différencier les sources de pollution locales des événements globaux.

A.1.3 La dégradation des polluants

Les polluants atmosphériques sont en grande partie dégradés par le rayonnement solaire (photochimie), produisant des radicaux tels que OH^-, H_2O^-, NO_3^-, et O_3 qui réagissent avec les polluants pour former d'autres composés [7]. Les composés organiques se dégradent successivement jusqu'à se transformer en CO_2. La durée de vie des particules en suspension dépend de leur forme et de leur nature, leur solubilité ou leur sédimentation (regroupement en couche) étant influencée par leur emprisonnement dans les puits.

Les puits de polluants

Il existe deux types de puits pour la pollution : les dépôts secs et les dépôts humides. Pour les dépôts secs, le polluant se dépose par contact. L'efficacité du puits est caractérisée par la nature de la surface de contact (sol, étendue d'eau, végétation, etc.). Les dépôts humides permettent le lessivage des polluants par la pluie. Leur efficacité dépend de l'humidité, de la solubilité des polluants et de la quantité de pluie. Les phénomènes météorologiques vont soit améliorer l'efficacité d'absorption des puits soit permettre la stagnation des polluants dans l'atmosphère.

A.1.4 La dispersion des polluants

Pour commencer, il est essentiel de comprendre la structure de l'atmosphère afin de mieux appréhender les mouvements des polluants ainsi que de leurs trajectoires.

La structure de l'atmosphère

L'atmosphère est composée de plusieurs couches :

— La Troposphère est la couche la plus proche du sol, où nous vivons. Elle est chauffée par les rayonnements du sol et présente un gradient de température négatif : la température diminue avec l'altitude ;

— La Stratosphère est située juste au-dessus de la Troposphère et contient la couche d'ozone, qui absorbe les rayonnements solaires. Elle présente un gradient de température positif : la température augmente avec l'altitude ;

— La Mésosphère est une zone de transition entre la terre et l'espace. Son gradient de température est identique à celui de la Troposphère ;

— La Thermosphère contient du dioxygène, qui absorbe les rayonnements et explique son gradient de température positif. Elle peut atteindre des températures de l'ordre de 2 000°C.

La Figure A.4 présente les couches de l'atmosphère en fonction de l'altitude.

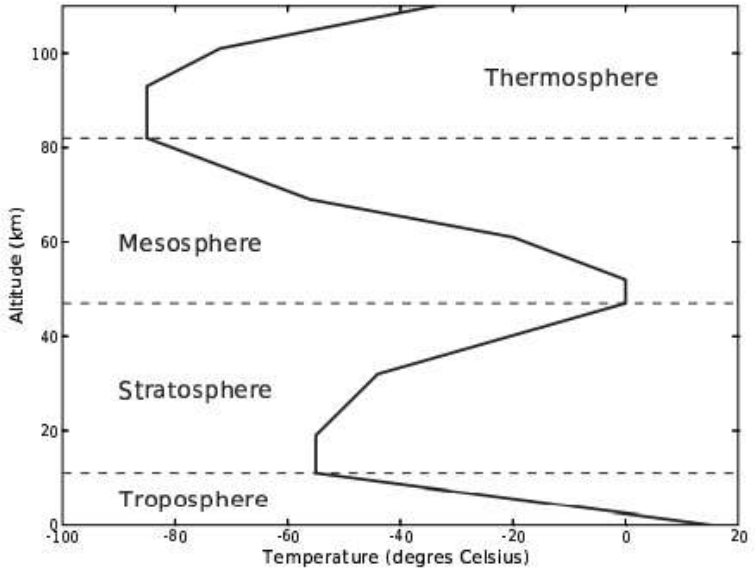


FIGURE A.4 – Profil vertical de température et principales couches de l'atmosphère [66].

Au-delà de la Thermosphère, se trouve l'Exosphère, la température continue d'augmenter mais les gaz sont en quantité infime. La Troposphère, la couche où nous vivons, permet grâce au gradient de température négatif, de disperser les polluants vers la Stratosphère. Comme le montre la Figure A.5, la troposphère est divisée en deux couches, la couche limite atmosphérique (CLA) et la troposphère libre. Cela est dû au fait que la qualité de l'air est observée dans les premiers mètres au niveau du sol. Selon Stull [66], la CLA désigne la partie de la troposphère qui subit des variations de l'ordre de heure au minimum. Cette couche est très réactive aux phénomènes naturels tels que l'évaporation, le réchauffement et la pollution. Au-dessus de la CLA, se trouve la troposphère libre où se forment les nuages. Ces derniers ont un impact sur le transport des polluants via leur effet sur le rayonnement (température du sol, pluie, réactions photochimiques, etc.). Ces couches ne sont pas étanches entres elles, les effets météorologiques vont influencer sur la dispersion des polluants.

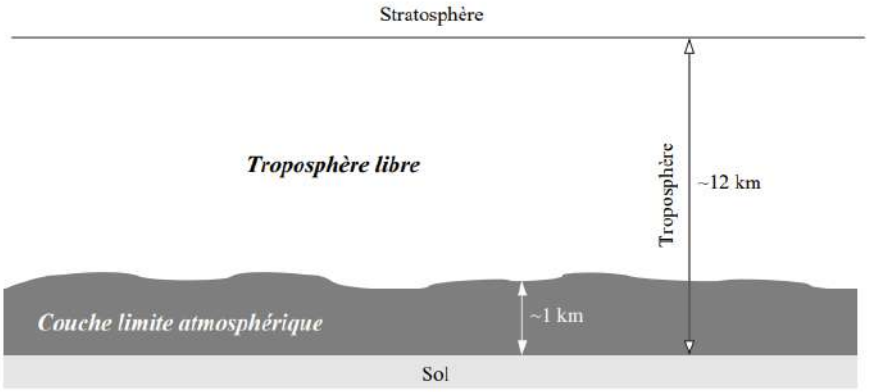


FIGURE A.5 – Structure de la troposphère [66].

A.1.5 Les phénomènes météorologiques particuliers

Pour comprendre la qualité de l'air, il ne suffit pas de considérer uniquement les polluants, les phénomènes météorologiques jouent également un rôle important. Ces derniers sont responsables de la dispersion ou du transport des polluants. La convection, les vents et les précipitations, par exemple, contribuent à disperser les polluants dans l'air, ce qui peut améliorer la qualité de l'air, ou au contraire augmenter leur concentration. Ainsi, les phénomènes météorologiques sont un facteur clé à prendre en compte pour évaluer la qualité de l'air.

Les phénomènes dispersifs

La dispersion des polluants est influencée par plusieurs facteurs, notamment la topologie du territoire et les paramètres météorologiques. La couche limite atmosphérique (CLA) est la partie de la troposphère qui est influencée par les limites du sol et de la couche d'atmosphère libre. La dynamique des masses d'air dans cette couche est complexe et désordonnée, avec une épaisseur moyenne de 1 km qui varie de jusqu'à 3 km en pleine journée et descend à 10 m la nuit. Le vent et la pression atmosphérique sont des facteurs clés de la dispersion des polluants, avec la vitesse et la direction du vent qui déterminent la rapidité et la direction de la dispersion. Cependant, la vitesse horizontale du vent peut varier en fonction de la pression atmosphérique et de la topologie du terrain, et les phénomènes de convection peuvent influencer la dispersion des polluants en renforçant les vents de surface pendant la journée et les vents d'altitude pendant la nuit. Le rayonnement joue un rôle important dans la formation des polluants secondaires par photochimie, tandis que l'instabilité de l'atmosphère peut entraîner un courant ascendant qui transporte la pollution vers le haut. Les obstacles naturels peuvent également modifier la circulation des vents et concentrer l'écoulement des polluants dans certaines directions, tels que les brises de terre et de mer, les obstacles naturels et les constructions.

Les phénomènes concentratifs

Les effets des phénomènes météorologiques peuvent être observables à l'œil nu, notamment par l'apparition d'un voile dans le paysage. L'humidité, la température et le rayonnement ont un impact sur la concentration des polluants. L'humidité favorise la concentration des particules de pollution en les captant avec des gouttelettes en suspen-

sion, qui seront ensuite relâchées lors de pluies ou brouillards. Une faible humidité accélère les réactions photochimiques et augmente la concentration de polluants. La température élevée favorise les réactions, tandis qu'une température basse peut concentrer les polluants en réduisant leur volatilité.

Une atmosphère stable, caractérisée par une couche d'air chaud au-dessus d'une couche plus froide près du sol, peut favoriser la concentration de polluants en rendant l'air froid plus dense et plus lourd que la couche du dessus. Ce phénomène se produit fréquemment la nuit par vent faible. Les configurations urbaines, telles que les parkings souterrains, les tunnels, les bâtiments encaissés et les rues canyons, combinées à un vent faible, contribuent à une augmentation de la concentration des polluants en agissant comme des obstacles concentrateurs, augmentant ainsi leur temps de présence et réduisant leur dégradation.

Les phénomènes spécifiques en montagne

Les phénomènes spécifiques en montagne sont étudiés pour mieux comprendre les mouvements de pollution. Pour expliquer et prévenir les pics de pollution, il est crucial d'étudier le comportement des masses d'air pollué, également appelé dynamique atmosphérique [69].

Dans cette optique, les brises d'aval (brises montantes) et les brises d'amont (brises descendantes) sont des phénomènes importants à prendre en compte. Les brises d'aval se déclenchent en fonction de l'intensité du rayonnement solaire et se produisent de la mimatinée jusqu'au coucher du soleil. Plus il fait chaud, plus l'effet est important, atteignant une vitesse de 15 à 20 km/h. Le réchauffement des versants de la montagne permet à l'air près du sol de s'alléger et de créer une brise montante vers le sommet de la montagne, créant ainsi un flux d'air provenant des plaines et finissant vers les sommets. Si une vallée basse contient une mer de nuage, les brises d'aval peuvent transporter les nuages jusqu'à une vallée plus haute. Lorsque le soleil se couche, un reflux est créé et les nuages retournent à leur emplacement d'origine. Cet effet apporte de la chaleur et de l'humidité vers les sommets, aidant à la formation de nuages sur le relief.

Les brises d'amont se produisent après la fin des brises d'aval et sont accompagnées d'un « renversement de brise« . Le sol rayonne et se refroidit ainsi que la couche d'air proche, qui devient plus dense. Un phénomène de glissement le long des pentes de cet air dense se crée, et l'air froid se concentre en fond de vallée, créant une couche épaisse d'air froid (100 à 200 m) qui s'écoule vers les plaines. La brise d'amont est une brise nocturne et est plus forte lorsque le ciel est dégagé. Elle peut être arrêtée si le ciel est couvert, car la déperdition de chaleur du sol est stoppée par la couche de nuage.

Ces phénomènes de brise dépendent de la configuration des vallées, de l'exposition au soleil et de la présence de glaciers ou de neige sur les versants. Ils sont essentiels au climat montagnard. En été, lors de ciel dégagé, la brise d'aval fait circuler l'air et transporte l'humidité, formant des nuages au niveau du sommet. Elle est plus forte que la brise d'amont, ce qui donne l'impression que la montagne aspire l'air de la plaine. En hiver, lors d'épisodes anticycloniques, la brise d'amont transporte un air sec (libéré de son humidité au contact de la neige sur les versants) vers la plaine en traversant les vallées.

En outre, les effets dynamiques des vallées face aux vents peuvent entraîner plusieurs phénomènes, en fonction de la direction du vent par rapport à l'axe de la vallée. Lorsque le vent est perpendiculaire, la vallée devient le centre de la turbulence et une couche d'air dense peut être chassée pour laisser place à l'air froid provenant de l'extérieur de la vallée qui descend vers le fond. Dans le cas où le vent souffle dans l'axe de la vallée, des turbulences se créent au niveau des versants, le vent est accéléré dans les vallées étroites

et crée des remous en sortie. Si le vent souffle avec un angle inférieur à 90°, une partie de l'air s'engouffre dans la vallée, l'effet de canalisation est accentué lorsque le vent est proche de l'axe. Les effets météorologiques spécifiques à la montagne sont dus aux brises locales et à la canalisation des vents extérieurs dans les vallées.

A.1.6 Les principaux polluants surveillés

Cette section énumère les principaux polluants prioritaires à surveiller par l'union européenne. Pour chacun d'eux, l'origine, les effets sur la santé et les conséquences sur l'environnement sont présentés. Les polluants sont présentés dans l'ordre suivant : particules en suspension, dioxyde de soufre, oxydes d'azote, ozone, ammoniac et composés organiques volatils.

Les particules en suspensions

Le polluant PM_{10}, mesuré par son diamètre de moins de 10 micromètres, provient de phénomènes d'érosion, de pollens et d'émissions industrielles. Les particules fines $PM2,5$, quant à elles, pénètrent plus profondément dans l'appareil respiratoire et sont émises par la combustion, le trafic, les carrières et l'industrie du batiment [17]. Les deux types de particules peuvent être de nature différente, allant des pollens aux particules minuscules de métaux ou d'hydrocarbures aromatiques polycycliques (HAP). Les PM_{10} sont généralement retenues par le système de défense du corps humain, mais leur absorption par voie orale peut entraîner des cancers de la vessie. Les $PM2,5$, en revanche, peuvent être évacuées par la toux ou le mucus, mais provoquent des troubles respiratoires, cardiovasculaires et peuvent entraîner des cancers du poumon.

Le Dioxyde de soufre SO_2

Le dioxyde de soufre SO_2 provient de la combustion des énergies fossiles contenant du soufre, telles que le pétrole, le fioul et le charbon, émis par des activités telles que l'industrie métallurgique, l'industrie chimique, les centrales thermiques et le chauffage au fioul. Ce gaz irritant peut entraîner des troubles respiratoires et cardiaques à long terme. Il est également à l'origine des pluies acides par sa transformation secondaire (particule de sulfate), il devient au contact de l'eau de l'acide sulfurique, causant ainsi des dommages directs aux végétaux et modifiant l'équilibre des sols. Il contribue notamment au noircissement des façades des bâtiments.

Les Oxydes d'Azote NO et NO_2

Les oxydes d'azote NO_x, quant à eux, proviennent principalement de la combustion de diesel et de bois énergie et sont émis par le trafic routier diesel, le chauffage des serres et l'industrie manufacturière. Ils contribuent à la formation des pluies acides et à la formation d'Ozone dans sa transformation secondaire. Irritant, ils pénètrent dans le système respiratoire plus profondément que le SO_2 ce qui, à long terme, provoque des infections pulmonaires, des problèmes respiratoires chroniques et la broncho-pneumopathie chronique obstructive (BPCO). Il perturbe également la croissance et les défenses naturelles des végétaux.

L'Ozone O_3

L'ozone O_3 est un composant normal de l'atmosphère, mais devient un polluant secondaire lors de fortes concentrations, provoqué par les NO_x, les COV et le CO. Ce polluant photochimique perturbe les poumons, les reins, le cerveau et les yeux par oxydation et peut causer des œdèmes pulmonaires à forte concentration. Il perturbe également les processus physiologiques tels que la photosynthèse et la respiration nocturne des végétaux et dégrade certains polymères et caoutchoucs. Pour terminer, il contribue à l'effet de serre.

L'ammoniac NH_3

L'ammoniac NH_3 provient principalement de l'engrais et est produit à 90 % par l'agriculture, notamment lors d'épisodes d'épandage et de fertilisation des sols. Il est irritant et possède une forte odeur. Il contribue à la formation de particules de nitrate d'ammonium. Ces composés azotés acidifient l'environnement car ils sont injectés dans la terre.

Les Composés Organiques Volatils COV

Les composés organiques volatils COV sont des gaz provenant de sources telles que l'utilisation de solvants organiques, le bois énergie, le trafic routier essence, les industries chimiques et agroalimentaires. Les COV comprennent tous les gaz contenant au moins un atome de carbone et étant sous forme gazeuse à partir de 0°C. Ils peuvent être divisés en deux catégories : les $COVNM$ et les COV. Les $COVNM$ excluent le méthane (CH4), qui est présent en grande quantité et qui provient de la dégradation de la matière organique. Ils ont un impact plus important sur la santé que les autres COV. ces derniers contribuent à la formation de l'ozone dans la troposphère et sont des réactifs photochimiques. Ils peuvent être irritants et réduire la capacité respiratoire à court terme. À long terme, ils peuvent devenir mutagènes et cancérigènes.

Autres polluants

Il existe deux autres polluants atmosphériques : les Hydrocarbures Aromatiques Polycycliques (HAP) et le monoxyde de carbone CO. Ces derniers sont produits lors de combustions réalisées dans des conditions défavorables telles que la combustion de bois humide, une chaudière mal réglée, une cheminée ouverte ou un moteur mal réglé. Le monoxyde de carbone est un gaz inodore, incolore et très toxique qui peut être difficile à détecter. Le HAP, quant à lui, peut affecter le système immunitaire tandis que le monoxyde de carbone peut provoquer des troubles respiratoires, des maux de tête et même entraîner la mort par asphyxie.

A.1.7 Conclusion sur la pollution atmosphérique

Cette section a été consacrée à la compréhension de l'origine, du cycle de vie et de la dispersion des polluants dans l'atmosphère. La structure de l'atmosphère joue un rôle important dans le transport des polluants et leur devenir. Les spécificités topologiques des montagnes influencent également la pollution dans les vallées.

Comprendre les phénomènes de diffusion des polluants est essentiel pour évaluer la qualité de l'air. Il est important de suivre l'évolution de la pollution sur le terrain et à des

fréquences régulières, voire en temps réel, pour évaluer les risques pour la santé publique. Ce sujet concerne tout le monde, y compris la CCPVG.

A.2 Le PHM dans l'industrie

Les industriels tendent à renforcer leur capacité à anticiper les défaillances afin de recourir à des actions préventives les plus justes possibles dans un objectif de réduction des coûts et des risques. Aussi, la mise en œuvre de solutions de PHM prend une part grandissante dans les activités de maintenance et le processus de pronostic est aujourd'hui considéré comme un des principaux leviers d'action dans la recherche d'une performance globale [22].

Le PHM est une discipline qui vise à fiabiliser un système, améliorer sa disponibilité, sa sécurité, mais aussi de réduire les coûts d'entretiens. Son implémentation au sein d'un système industriel se fait au travers de sept étapes illustrées dans la figure A.6 :

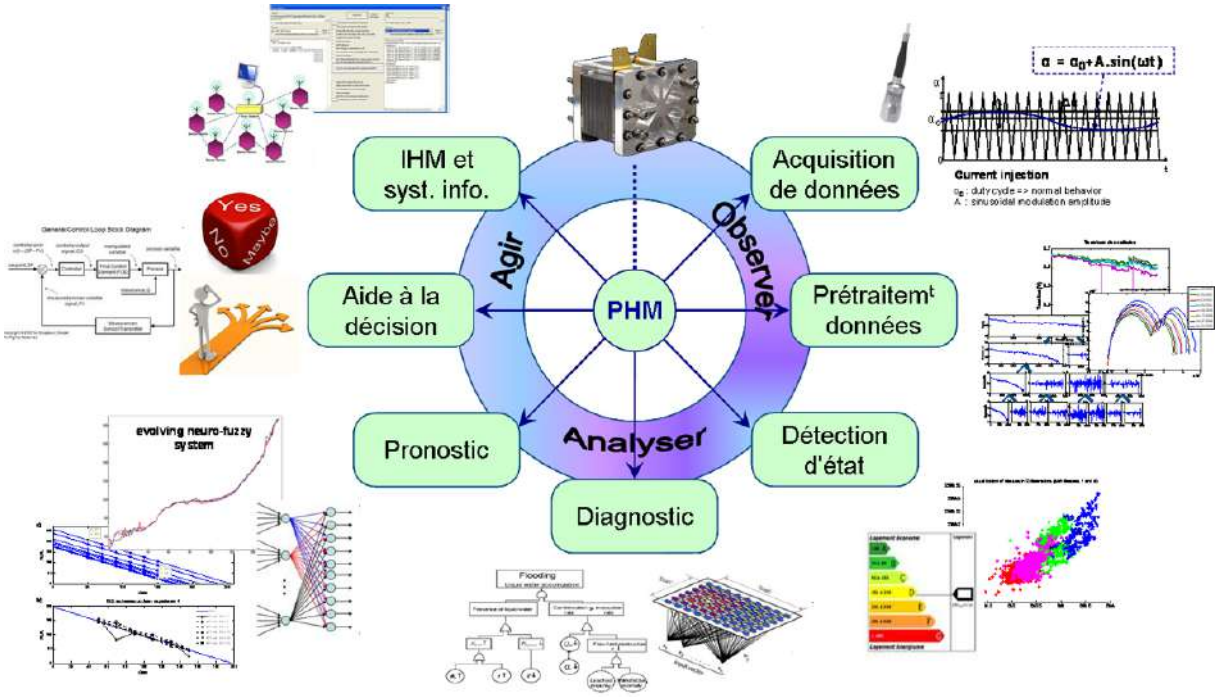


FIGURE A.6 – Étapes d'implémentation du PHM [22].

- **Acquisition des données**
  La mise en place des capteurs ne doit pas influencer le système physique, ni modifier son comportement. L'acquisition des données est l'étape la plus importante du PHM.
- **Post-traitement des données**
  Il permet de nettoyer le signal des bruits permet d'exploiter les données bruts sans erreurs.
- **Détection**
  C'est la différence entre le comportement attendu et le comportement actuel permet de lever une détection à partir d'analyse par consistance.
- **Diagnostic**

La levée de la détection et la dégradation de l'état du système observé pouvant amener à une défaillance, (perte d'une fonction du système) est traduite par un diagnostic qui identifie le ou les composants en défaut.

- **Pronostic**
  Ce processus prévoit le temps de vie restant du composant en défaut (survie).
- **Injonction de décision**
  Les analyses postérieures vont amener à proposer plusieurs alternatives possibles en fonction du résultat attendu.
- **Interface**
  Les décisions sont présentées à l'interface, deux cas se présentent :
  — l'homme reçoit la décision et va agir sur le système (ex : maintenance) dans le cas où l'action est physiquement possible par l'homme ;
  — l'interface utilisera un actionneur qui agira sur le système physique.

Le PHM est une discipline proche de la surveillance de l'environnement mais au niveau principalement industriel : il observe un système physique, traite les données (détection, diagnostic et pronostic), présente des décisions et a une interface pour interagir avec un humain.

A.3 Vue d'ensemble de l'architecture avec les logiciels utilisés

L'architecture globale est représentée dans la figure A.3. Cet article traite principalement de la phase d'Observation (1) du projet global BOLDAIR. Tout d'abord, les données sont collectées par le SmS, puis traitées et triées par l'Aggregator, comme précédemment défini. Ensuite, les données décisionnelles temporelles agrégées sont envoyées à une base de données (Influxdb) via des services web mis en œuvre dans le serveur web, qui est l'un des deux sous-systèmes le composant.

La phase de traitement (2) et la phase de décision (3) sont mises en œuvre dans les sous-systèmes terminaux. Enfin, les autorités ont accès au support décisionnel via des interfaces web (4) pour prendre des décisions politiques et convenir d'actions (5) visant à améliorer la qualité de l'air et/ou à anticiper les épisodes de pollution. Il est important de noter que les serveurs locaux et le serveur central sont presque identiques. L'idée de base est de pouvoir déployer une architecture fédérée les couvrant pour de vastes zones géographiques. Ainsi, leur principale différence réside dans la zone géographique qu'ils couvrent par rapport aux autorités locales. Dans le cas du CCPVG, les serveurs locaux collectent les données des stations dans une vallée et proposent des décisions au niveau de la ville, tandis que le serveur central collecte des données des autres et propose des décisions pour l'ensemble du territoire. Leur interface varie en fonction des utilisateurs et de leurs responsabilités.

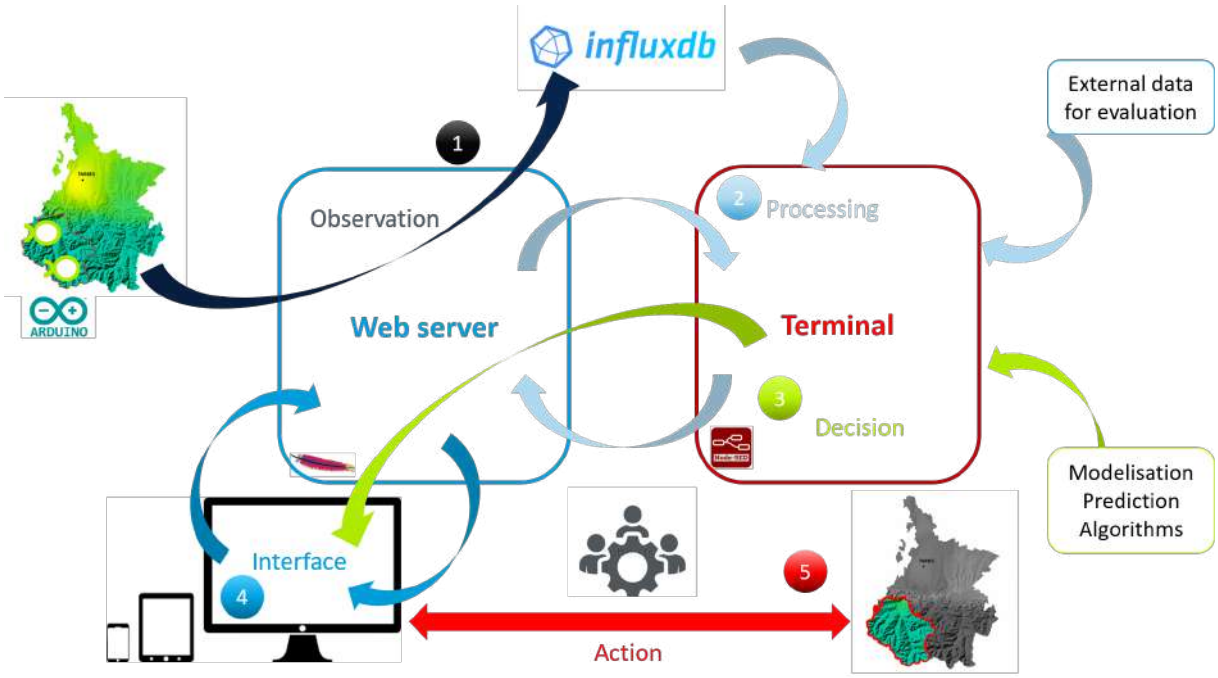


FIGURE A.7 – Vue d'ensemble de l'architecture avec les logiciels utilisés.

A.4 Modes de fonctionnement du capteur intelligent

Dans l'architecture proposée, le SmS est le composant central et est dupliqué pour rendre le réseau de mesure hautement réparable et évolutif. Étant de conception matérielle identique, il réduit le temps de mise en service. Sa conception logicielle est conçue pour avoir deux modes de fonctionnement : configuration et mesure. Le passage d'un mode à l'autre se fait automatiquement. La sous-section suivante décrit chaque mode de fonctionnement et le basculement avec un diagramme algorithmique.

Le mode de configuration

Une fois le programme chargé et enregistré dans l'ensemble de l'architecture, le SmS est considéré comme sortant de l'usine. Pour sa première mise en service, toutes les variables sont initialisées pour éviter les erreurs d'attribution de mémoire. Ainsi, le mode de configuration correspond au mode par défaut lorsqu'il est mis sous tension. Il permet au SmS de créer un point d'accès WiFi pour se connecter à une page d'accueil qui offre les fonctionnalités suivantes :

1. **Configuration WiFi et des paramètres :** par mesure de sécurité, le SSID et le mot de passe du réseau WiFi fourni par le routeur local sont enregistrés dans la mémoire du microcontrôleur et non dans le code. La période d'échantillonage Te_x, les adresses de l'Aggregator et du Datamart local sont également enregistrés dans les variables ;
2. **Test LCS :** un protocole de test est effectué pour chaque LCS afin de s'assurer qu'ils fonctionnent correctement ;
3. **Lecture des diagnostics :** après la sélection de la fonction de test LCS, un rapport détaillé des événements est présenté au client web ;

4. **Réinitialisation d'usine :** la mémoire du microcontrôleur est effacée, réinitialisant ainsi tous les paramètres ;
5. **Mode de mesure :** Le capteur intelligent prendra des mesures des paramètres sélectionnés et les transmettra au serveur local à une période te_x. Même en cas de panne de courant, il restera dans ce mode.

Ces fonctionnalités constituent le point de départ de la configuration du réseau de mesure, facilitant ainsi l'intégration et le diagnostic rapides en cas de défaillance sans accès physique. Le mode de mesure est choisi sur la page d'accueil pour mettre en service le SmS.

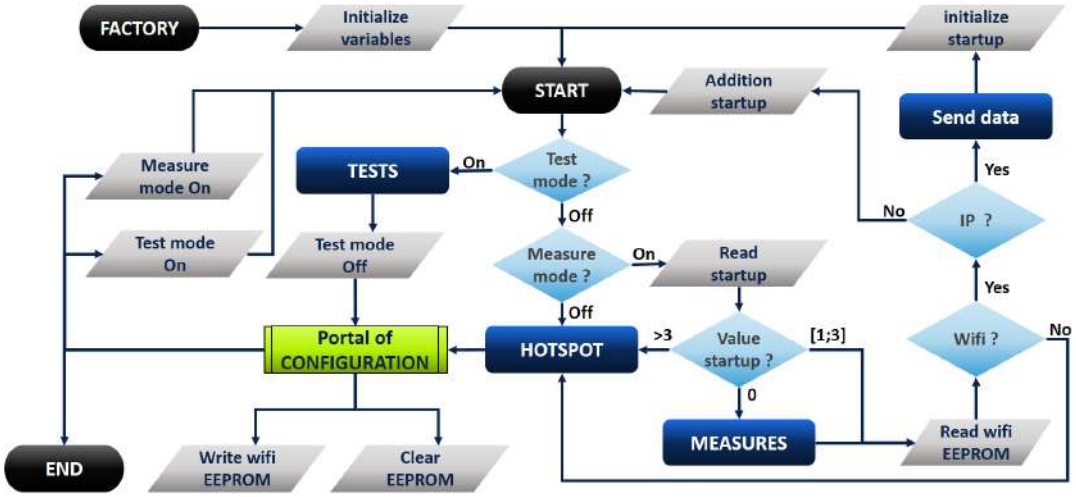


FIGURE A.8 – Diagramme algorithmique du SmS.

Le mode de mesure

Activé depuis le mode de configuration, le mode de mesure met le SmS en mode veille profonde pour économiser de l'énergie entre deux mesures. Lorsqu'il se réveille, il veille à ce que le protocole de mesure décrit dans la section 5.2.5 soit correctement suivi en suivant les directives des fabricants de LCS respectifs. Une attention particulière a été accordée à une circulation forcée de l'air qui renouvelle l'air à l'intérieur de la boîte du capteur intelligent pour que les polluants des mesures précédentes ne perturbent pas les nouvelles mesures. En effet, les capteurs de gaz ont tendance à accumuler des molécules dans leur partie sensible. De plus, en tenant compte de la température et de l'humidité à différentes étapes du protocole de mesure, les mesures sont ajustées en compensant leur influence sur le paramètre cible.

Le passage d'un mode à l'autre est décrit par un diagramme algorithmique illustré à la figure A.8. Les modes de configuration et de mesure sont mis en œuvre en fonction de l'expérience sur le terrain acquise au fil du temps. Ce développement itératif est au cœur de notre stratégie de développement, concernant la conception matérielle et les plans, le microcontrôleur et son micrologiciel, l'architecture du serveur local et le logiciel décisionnel. L'objectif est d'améliorer et de converger de manière itérative grâce aux retours d'expérience en conditions réelles. De plus, tous les développements sont réalisés conformément aux normes open source, de sorte que l'évolution matérielle et logicielle reste accessible.

Description du diagramme algorithmique

Après l'assemblage de tous les composants, le flashage de la mémoire et la mise sous tension, le capteur intelligent est à l'état "USINE". Avant l'état "DÉMARRER", toutes les variables sont initialisées. À partir de cet état, les valeurs des variables *test_mode* et *measure_mode* sont lues, et actuellement, elles ont la valeur "désactivé". Ainsi, un point d'accès WiFi est créé pour accéder au "portail de CONFIGURATION". À partir de ce portail, le mainteneur peut prendre certaines décisions : l'état "FIN" est atteint s'il décide de le retirer du réseau pour réparation, de modifier la variable *test_mode* pour tester les mesures des LCS ou de modifier la variable *measure_mode* pour mettre le capteur intelligent en service. Les accès à la mémoire EEPROM permettent de sauvegarder le SSID et le mot de passe du réseau WiFi local ou de réinitialiser complètement pour revenir à l'état "USINE".

Lorsque la variable *test_mode* est activée, le protocole de "TESTS" sera activé et fournira un rapport de diagnostic au portail de configuration après avoir recréé le point d'accès WiFi. Lorsque la variable *measure_mode* est activée, la variable *startup* est alors lue. Le but de cette variable est de compter le nombre de tentatives de connexion et de détecter la perte du réseau WiFi pour l'envoi des données. Lorsque la variable *startup* est nulle, l'état "MESURE" est atteint, et le protocole de mesure est activé. Les mesures prises sont enregistrées dans la mémoire du microcontrôleur. Si le nombre de redémarrages est inférieur à trois fois (valeur de *startup* entre 1 et 3), les mesures ne sont pas refaites pour envoyer les données le plus rapidement possible. Lorsque le nombre de redémarrages dépasse trois fois (la variable *startup* est supérieure à 3), l'état "POINT D'ACCÈS" est atteint, permettant l'accès au portail de configuration pour des diagnostics. Le déploiement sur le terrain montre que le personnel n'a pas pu intervenir et le SmS est resté en mode point d'accès pendant de longues périodes, tandis que la connexion était restaurée entre-temps. Par conséquent, le point d'accès (lorsqu'il est activé à partir du mode de configuration) offre une fenêtre de connexion limitée avant de redémarrer et de revenir à l'état "DÉMARRER". Cette modification permet au SmS d'être autonome en cas de défaillance de ce type de connexion.

En cas de panne de courant, le SmS se réinitialise, atteint l'état "DÉMARRER" et renouvelle les mesures. Lorsque la connexion au réseau WiFi local est réussie, l'état "ENVOI DES DONNÉES" est atteint, et les données (au format JSON) sont envoyées à l'Aggregator ou au serveur local en fonction de leur disponibilité.

A.5 Détails techniques du SmS

Dans cette section sont présentés les éléments techniques concernant la partie matérielle du SmS.

A.5.1 Le PCB en détail

Le schémas électrique présenté dans la Figure A.5.1 a été conçu avec le logiciel libre Kicad.

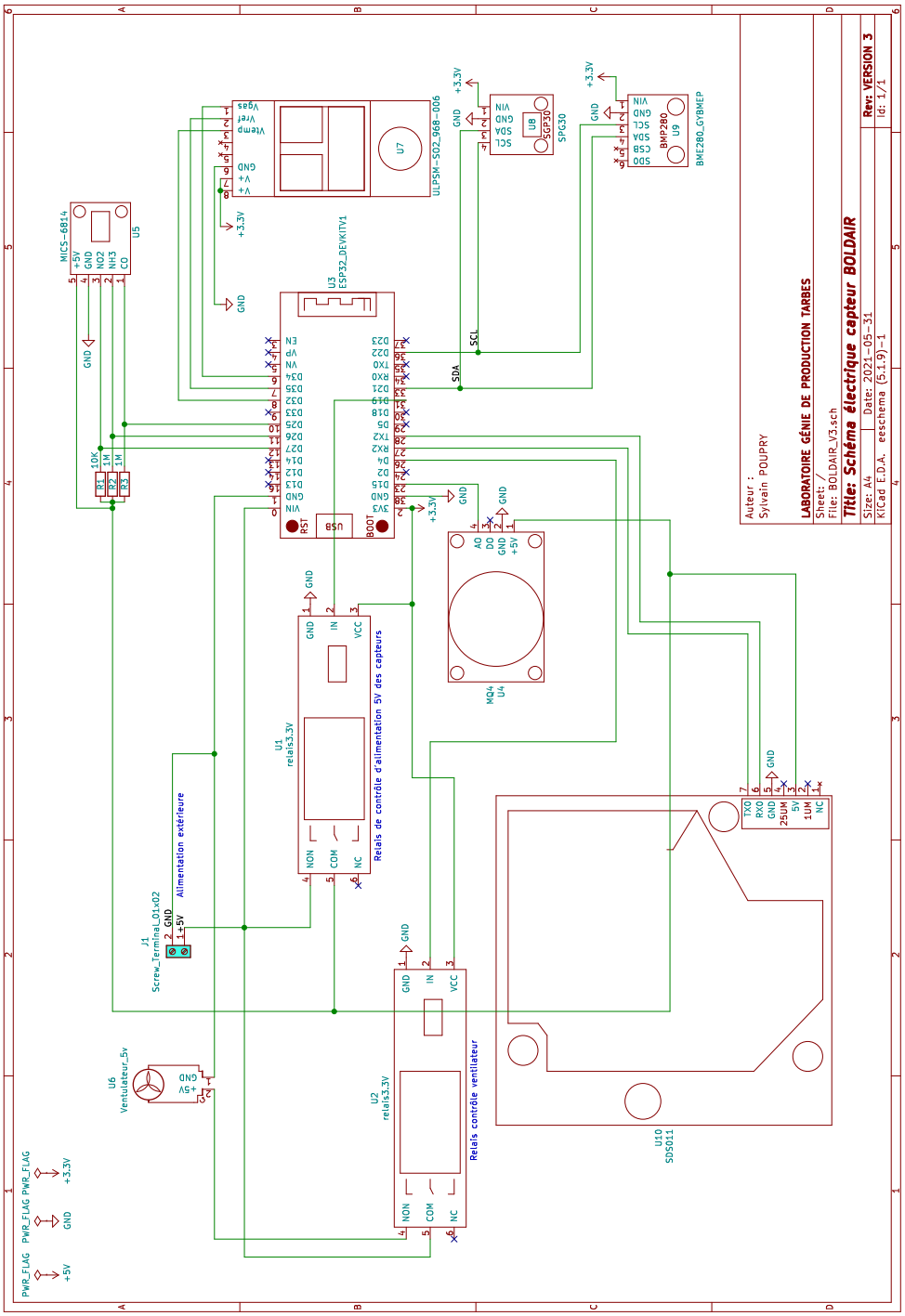


FIGURE A.9 – Schémas électrique complet réalisé avec le logiciel Kicad.

A.5.2 Les schémas électriques détaillés pour chaque composants

Les schémas électriques suivant détaillent les connexions électriques avec l'ESP32 permettant leur utilisation.

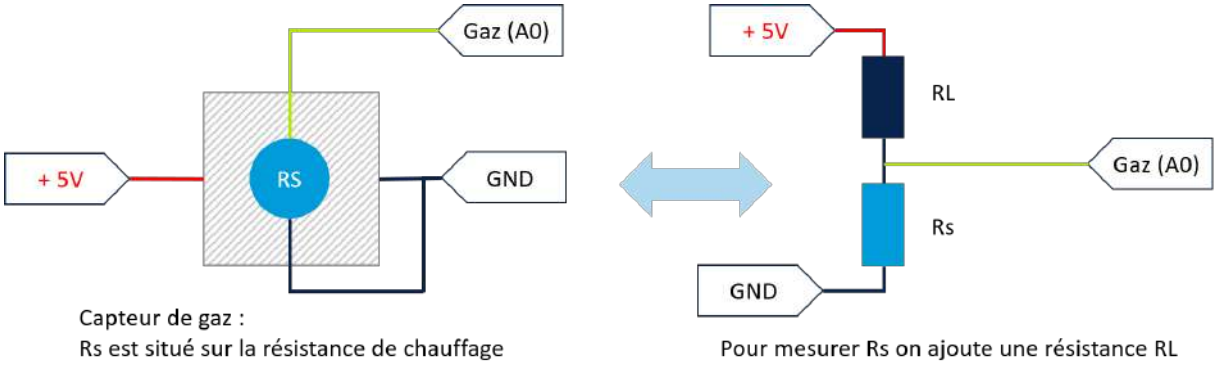


FIGURE A.10 – La spécificité des capteurs de gaz analogiques par l'ajout de résistance de charge pour évaluer la résistance interne du capteur dont la valeur évolue en fonction des concentrations.

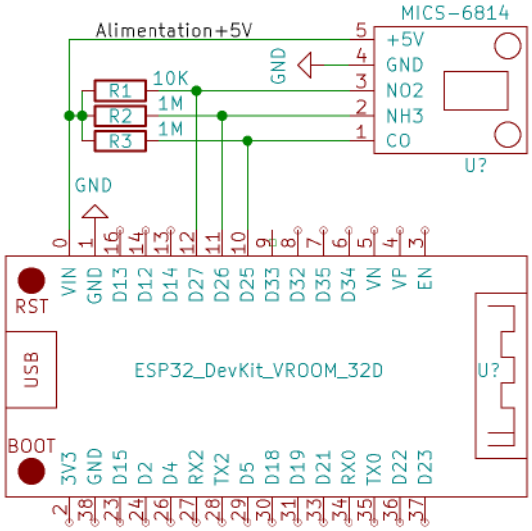


FIGURE A.11 – MICS 6814 en connexion analogique avec un pont diviseur de tension.

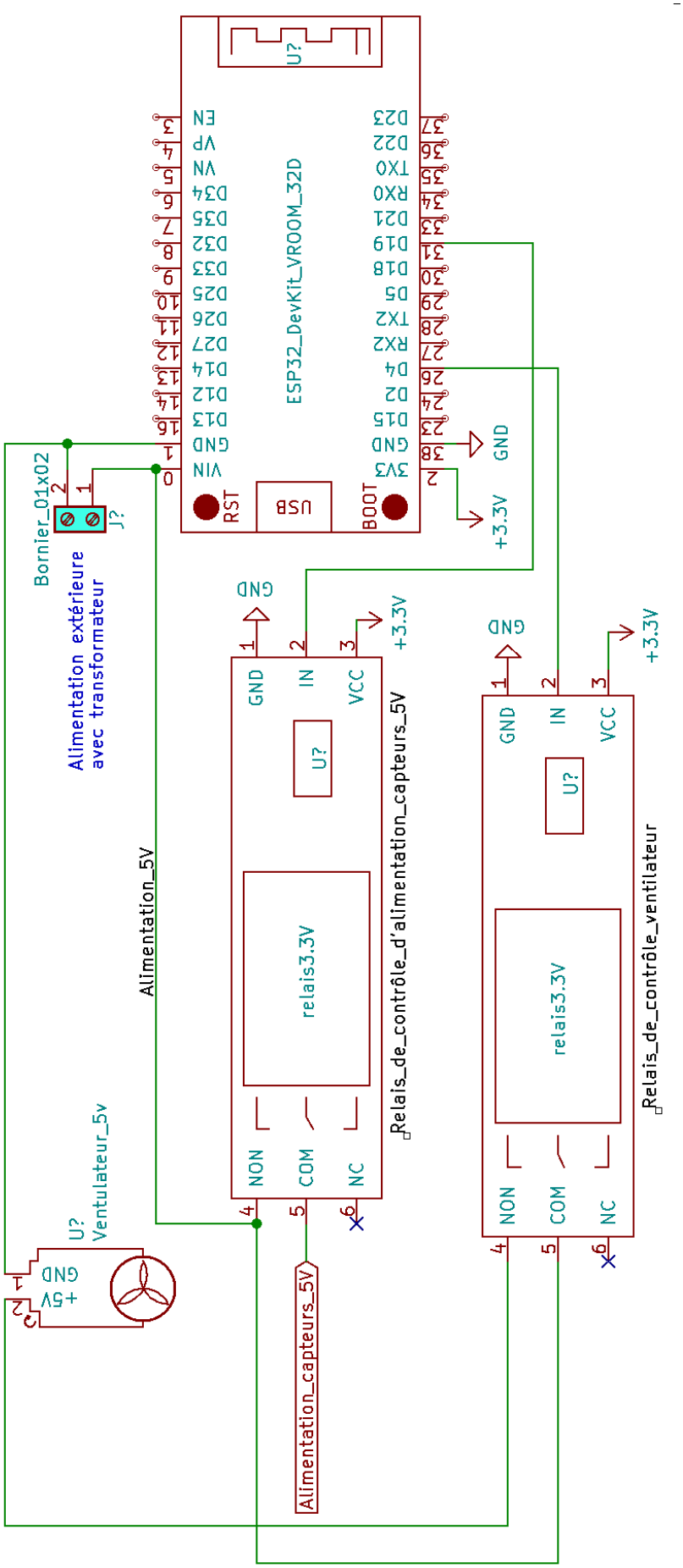


FIGURE A.12 – Contrôle de l'alimentation des capteurs de gaz et du ventilateur en 5V.

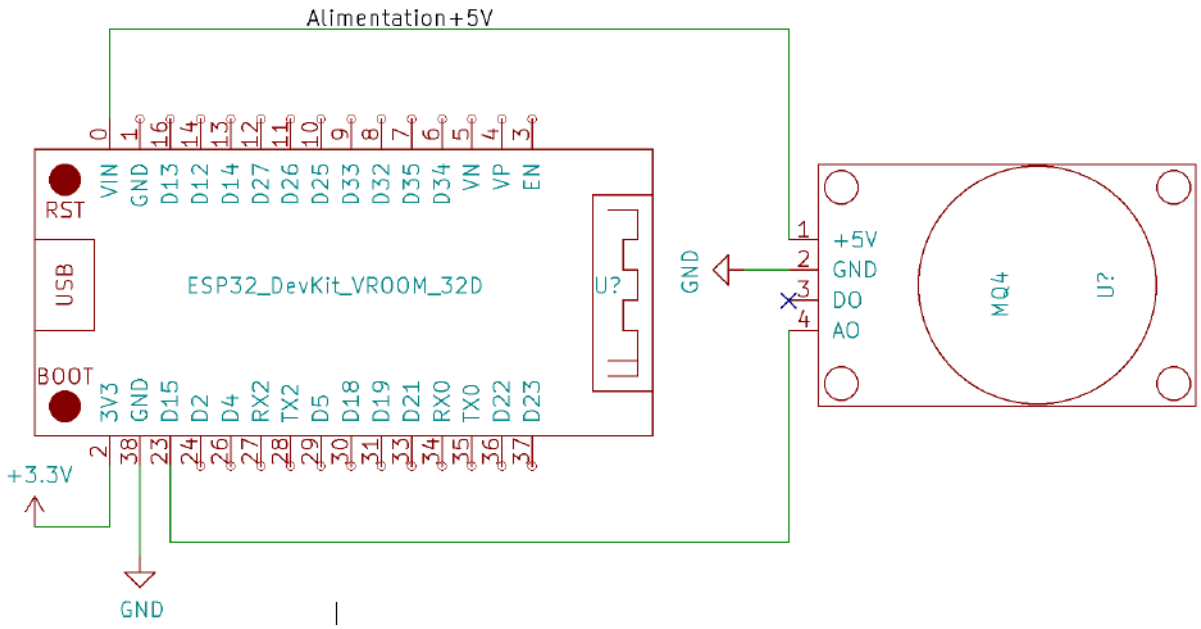


FIGURE A.13 – Le capteur de gaz MQ4 connecté en analogique et alimenté en 5 V.

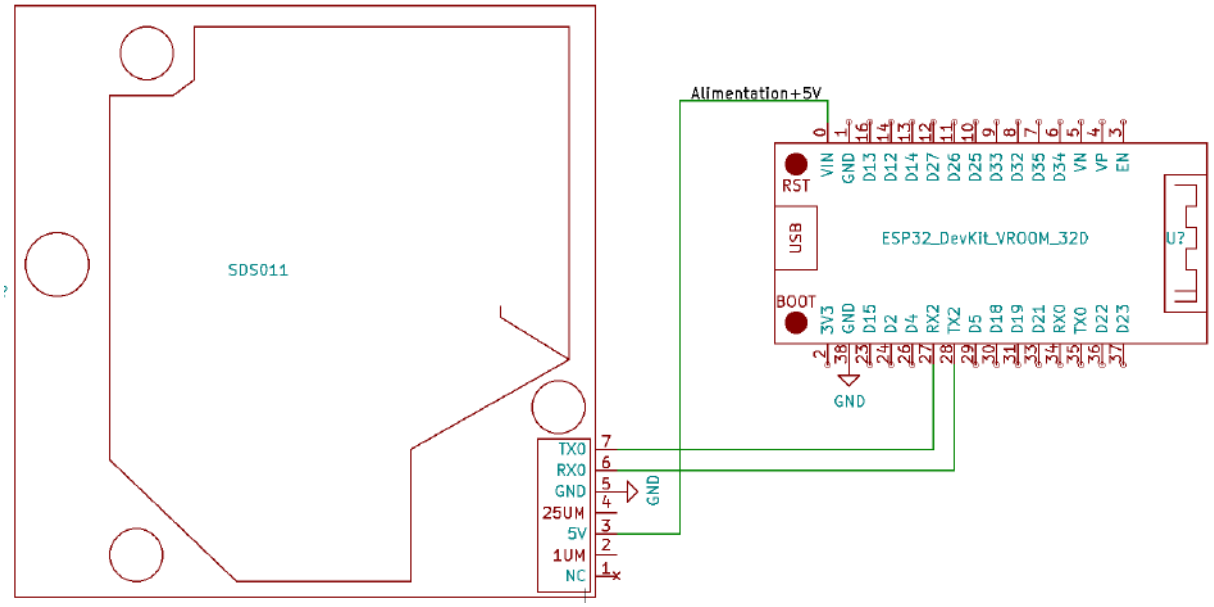


FIGURE A.14 – Le capteur de poussière SDS011 en connexion UART avec l'ESP32.

Ces schémas électriques ont été nécessaires pour l'intégration complète et la conception du PCB. Pour des raisons budgétaires, il a été décidé que les circuits imprimés seraient gravés *à l'anglaise* sur une seule face. La Figure A.5.2 est le résultat de deux ans de co-design et d'expérience sur le terrain.

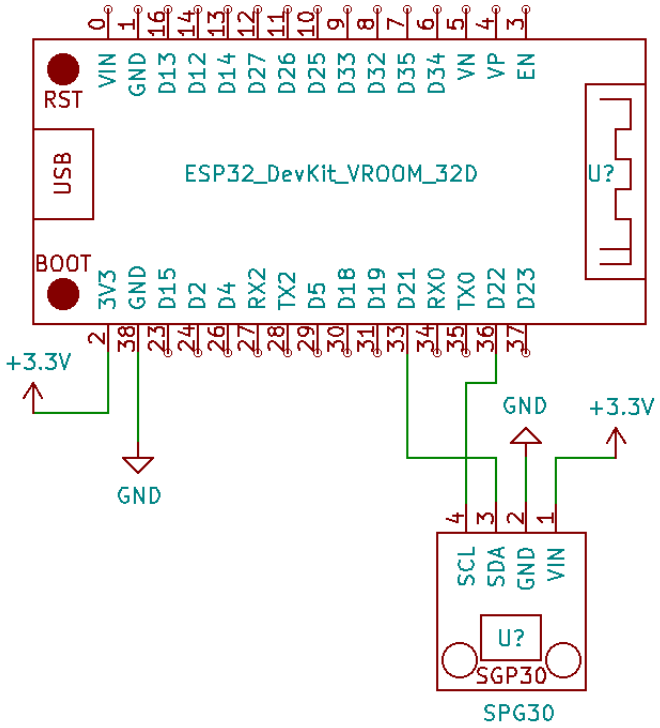


FIGURE A.15 – Capteur de gaz SPG30 en connexion I²C.

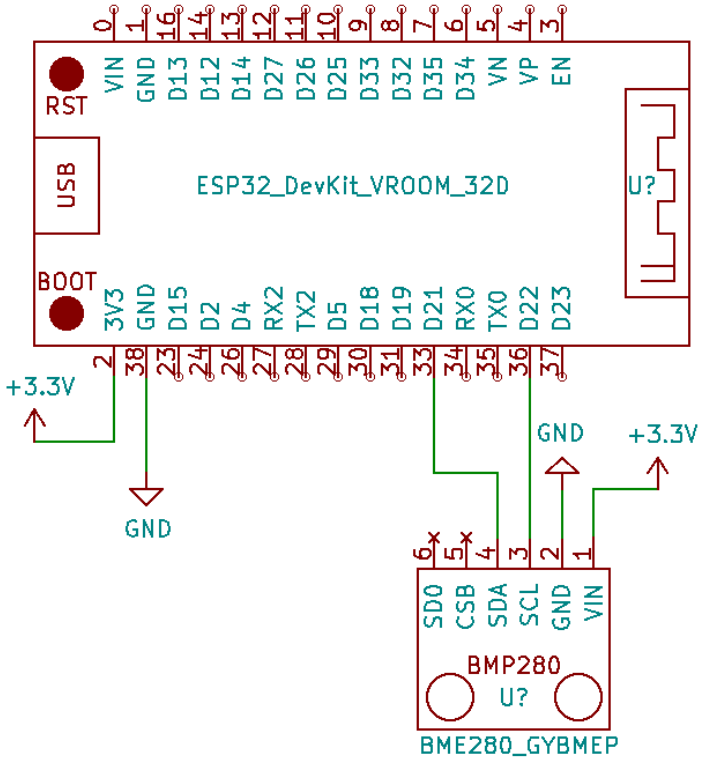


FIGURE A.16 – Le BMP 280 en connexion I²C avec l'ESP 32.

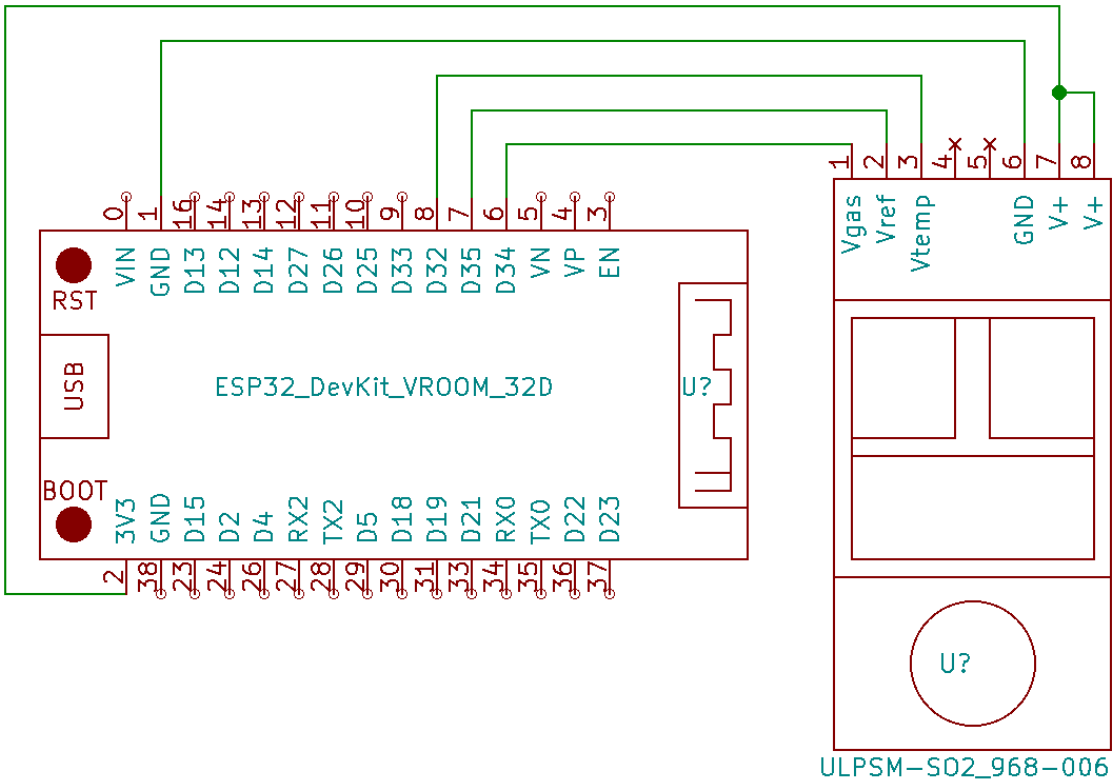


FIGURE A.17 – Le capteur de SO_2 en connexion analogique.

A.5.3 Les réalisations : du procept au prototype final

Pour parvenir au déploiement du premier prototype, plusieurs processus ont été créés afin d'assurer l'intégration de tous les composants.

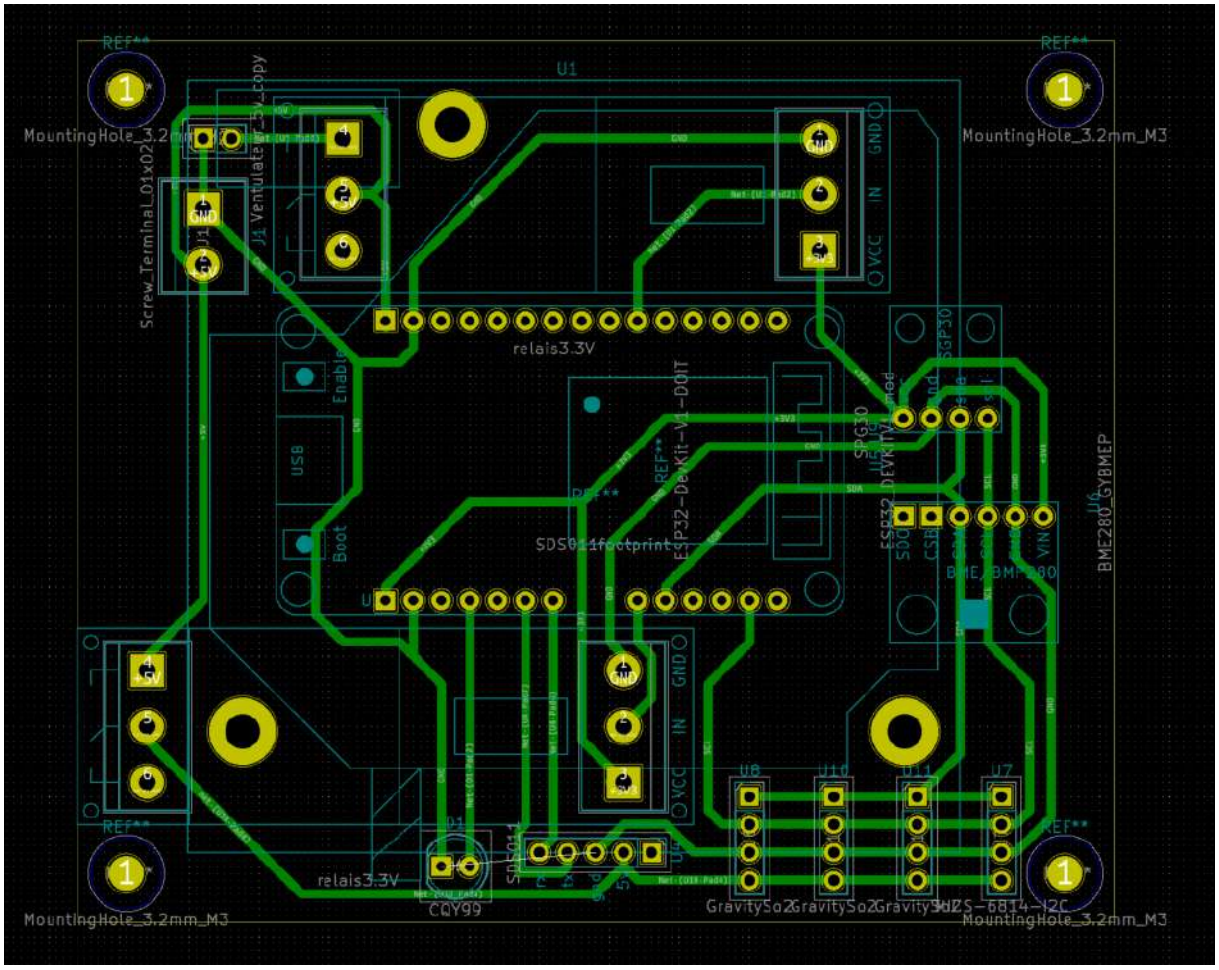


FIGURE A.18 – Le design du PCB final pour le deuxième prototype.

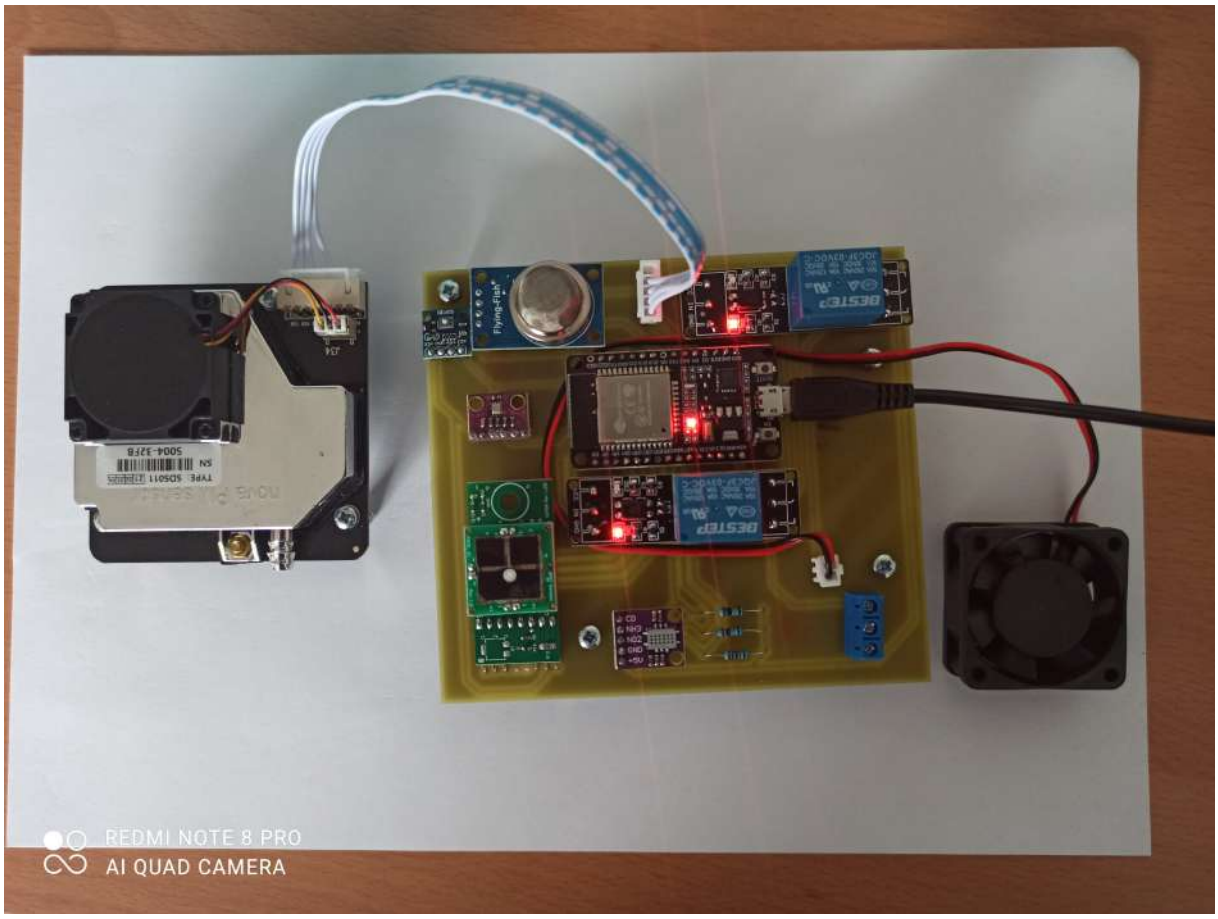


FIGURE A.19 – La preuve du concept avant mise en fabrication du 1er prototype.

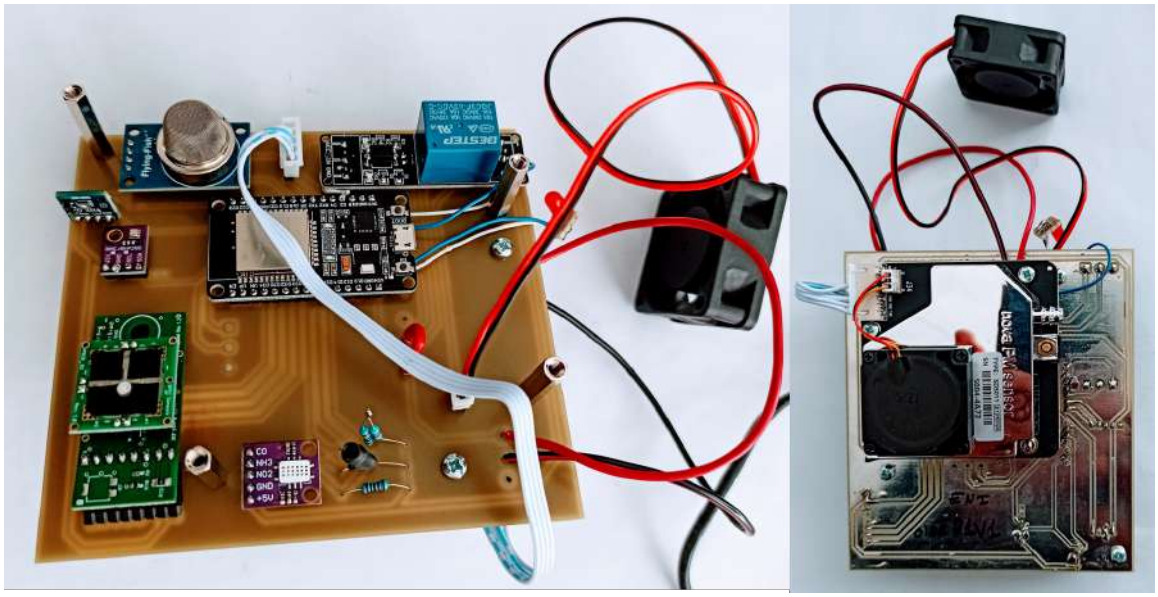


FIGURE A.20 – 1er prototype déployé.

Bibliographie

[1] ACNUSA. *Pollution aéroport | Pollution atmosphérique et qualité de l'air aéroport*. fr. 2021. URL : https://www.acnusa.fr/fr/la-pollution-de-lair/definitions/16 (visité le 10/02/2021).

[2] Matthias AICHINGER-ROSENBERGER et al. « MPG-NET : A low-cost, multi-purpose GNSS co-location station network for environmental monitoring ». In : *Measurement* 216 (2023), p. 112981. ISSN : 0263-2241. DOI : https://doi.org/10.1016/j.measurement.2023.112981. URL : https://www.sciencedirect.com/science/article/pii/S0263224123005456.

[3] *Atmo France - Réseau national des Associations agréées de surveillance de la qualité de l'air*. fr-FR. URL : https://atmo-france.org/ (visité le 15/01/2021).

[4] Zaher Al BARAKEH et al. « Development of a normalized multi-sensors system for low cost on-line atmospheric pollution detection ». en. In : *Sensors and Actuators B : Chemical* 241 (mars 2017), p. 1235-1243. ISSN : 0925-4005. DOI : 10.1016/j.snb.2016.10.006. URL : http://www.sciencedirect.com/science/article/pii/S092540051631632X (visité le 20/01/2021).

[5] Julian BASS, Gholam Reza LATIF-SHABGAHI et Stuart BENNETT. *Experimental comparison of voting algorithms in cases of disagreement*. Pages : 523. Oct. 1997. ISBN : 978-0-8186-8129-5. DOI : 10.1109/EURMIC.1997.617368.

[6] Petra BAUEROVÁ et al. « Low-Cost Air Quality Sensors : One-Year Field Comparative Measurement of Different Gas Sensors and Particle Counters with Reference Monitors at Tušimice Observatory ». en. In : *Atmosphere* 11.5 (mai 2020). Number : 5 Publisher : Multidisciplinary Digital Publishing Institute, p. 492. DOI : 10.3390/atmos11050492. URL : https://www.mdpi.com/2073-4433/11/5/492 (visité le 06/10/2021).

[7] Marie CAMREDON et Aumont BERNARD. « I – L'ozone troposphérique : production/consommation et régimes chimiques ». In : *Pollution atmosphérique* (jan. 2007). DOI : 10.4267/pollution-atmospherique.1404.

[8] *Capteur électrochimique : les caractéristiques*. URL : https://www.ineris.fr/sites/ineris.fr/files/contribution/Documents/DRA_76_OMEGA_22_B2_web.pdf. (visité le 22/01/2021).

[9] Nuria CASTELL et al. « Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates ? » en. In : *Environment International* 99 (fév. 2017), p. 293-302. ISSN : 0160-4120. DOI : 10.1016/j.envint.2016.12.007. URL : http://www.sciencedirect.com/science/article/pii/S0160412016309989 (visité le 20/01/2021).

[10] Meo Vincent C. Caya et al. « Air Quality Monitoring Platform with the Integration of Dual Sensor Redundancy Mechanism through Internet of Things ». In : *2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology,Communication and Control, Environment and Management (HNICEM)*. Nov. 2018, p. 1-6. DOI : 10.1109/HNICEM.2018.8666428.

[11] I Christakis et al. « Barometric Pressure as a correction factor for low-cost particulate matter sensors ». In : *IOP Conference Series : Earth and Environmental Science*. T. 1123. 1. IOP Publishing. 2022, p. 012068. DOI : 10.1088/1755-1315/1123/1/012068.

[12] *contribution ATMO assise tranistion-energie 2018*. fr-FR. URL : https://atmo-france.org/wp-content/uploads/2018/09/contribution_ATMO_assise_tranistion_energie_2018.pdf (visité le 18/01/2021).

[13] National Research Council. *The Atmospheric Sciences : Entering the Twenty-First Century*. en. Déc. 1969. ISBN : 978-0-309-06415-6. DOI : 10.17226/6021. URL : https://www.nap.edu/catalog/6021/the-atmospheric-sciences-entering-the-twenty-first-century (visité le 05/02/2021).

[14] Louis De Mesnard. « Pollution models and inverse distance weighting : Some critical remarks ». In : *Computers & Geosciences* 52 (2013), p. 459-469.

[15] Robert Delmas et Gérard Mégie. *Physique et chimie de l'atmosphère : 2e édition*. fr. Google-Books-ID : 6ROWDgAAQBAJ. Humensis, juill. 2015. ISBN : 978-2-7011-8555-2.

[16] *Electrochemical Gas Sensors – Membrapor*. URL : https://www.membrapor.ch/electrochemical-gas-sensors/ (visité le 22/01/2021).

[17] C. Elichegaray et al. « État et évolution de la pollution atmosphérique ». en. In : *Revue des Maladies Respiratoires* 26.2 (fév. 2009), p. 191-206. ISSN : 0761-8425. DOI : 10.1016/S0761-8425(09)71596-7. URL : http://www.sciencedirect.com/science/article/pii/S0761842509715967 (visité le 03/02/2021).

[18] *Figure.2 : Constitution du capteur de gaz MOX*. en. URL : https://www.researchgate.net/figure/Constitution-du-capteur-de-gaz-MOX_fig3_45636412 (visité le 22/01/2021).

[19] Walter Fuertes et al. « Distributed System as Internet of Things for a New Low-Cost, Air Pollution Wireless Monitoring on Real Time ». In : *2015 IEEE/ACM 19th International Symposium on Distributed Simulation and Real Time Applications (DS-RT)*. ISSN : 1550-6525. Oct. 2015, p. 58-67. DOI : 10.1109/DS-RT.2015.28.

[20] Rosa Amalia González Rivero et al. « A Low-Cost Calibration Method for Temperature, Relative Humidity, and Carbon Dioxide Sensors Used in Air Quality Monitoring Systems ». In : *Atmosphere* 14.2 (2023), p. 191. DOI : https://doi.org/10.3390/atmos14020191.

[21] Rosa Amalia González Rivero et al. « Relevance and Reliability of Outdoor SO2 Monitoring in Low-Income Countries Using Low-Cost Sensors ». In : *Atmosphere* 14.6 (2023), p. 912. DOI : https://doi.org/10.3390/atmos14060912.

[22] Rafael Gouriveau et al. « PHM – Prognostics and health management - De la surveillance au pronostic de défaillances de systèmes complexes ». fr. In : (2013), p. 21.

[23] Hanna Yang et al. *Air pollution and NCDs*. English. Avr. 2019. URL : https://www.euro.who.int/__data/assets/pdf_file/0005/397787/Air-Pollution-and-NCDs.pdf (visité le 01/10/2021).

[24] Robert L Harder et Robert N Desmarais. « Interpolation using surface splines. » In : *Journal of aircraft* 9.2 (1972), p. 189-191.

[25] IQAir. *Live Animated Air Quality Map (AQI, PM2.5...) | IQAir*. en. URL : https://www.iqair.com/air-quality-map (visité le 18/05/2022).

[26] M. Kassab et al. « A Novel Voting Algorithm Based on Weighted Average Voting with a Classification Technique ». en. In : *International Conference on Aerospace Sciences and Aviation Technology* 15.AEROSPACE SCIENCES (mai 2013), p. 1-8. ISSN : 2636-364X. DOI : 10.21608/asat.2013.22266. URL : http://asat.journals.ekb.eg/article_22266.html (visité le 20/01/2022).

[27] Anwar Ulla Khan et al. « An Efficient Wireless Sensor Network Based on the ESP-MESH Protocol for Indoor and Outdoor Air Quality Monitoring ». In : *Sustainability* 14.24 (2022), p. 16630. DOI : https://doi.org/10.3390/su142416630.

[28] P. Kucera et al. « Realibility model of TMR system with fault detection ». en. In : *IFAC Proceedings Volumes*. IFAC Workshop on Programmable Devices and Embedded Systems PDeS 2006, Brno, Czech Republic, February 14-16, 2006 39.21 (fév. 2006), p. 468-472. ISSN : 1474-6670. DOI : 10.1016/S1474-6670(17)30233-1. URL : https://www.sciencedirect.com/science/article/pii/S1474667017302331 (visité le 03/03/2022).

[29] Prashant Kumar et al. « The rise of low-cost sensing for managing air pollution in cities ». en. In : *Environment International* 75 (fév. 2015), p. 199-205. ISSN : 0160-4120. DOI : 10.1016/j.envint.2014.11.019. URL : http://www.sciencedirect.com/science/article/pii/S0160412014003547 (visité le 22/01/2021).

[30] G. R Latif-Shabgahi. « A novel algorithm for weighted average voting used in fault tolerant computing systems ». en. In : *Microprocessors and Microsystems* 28.7 (sept. 2004), p. 357-361. ISSN : 0141-9331. DOI : 10.1016/j.micpro.2004.02.006. URL : https://www.sciencedirect.com/science/article/pii/S0141933104000134 (visité le 20/01/2022).

[31] Jean Legras. « Méthodes et techniques de l'analyse numérique ». In : *(No Title)* (1971).

[32] Alastair C. Lewis et al. « Evaluating the performance of low cost chemical sensors for air pollution research ». en. In : *Faraday Discuss.* 189.0 (juill. 2016). Publisher : The Royal Society of Chemistry, p. 85-103. ISSN : 1364-5498. DOI : 10.1039/C5FD00201J. URL : https://pubs.rsc.org/en/content/articlelanding/2016/fd/c5fd00201j (visité le 20/01/2021).

[33] Christophe Leys et al. « Detecting outliers : Do not use standard deviation around the mean, use absolute deviation around the median ». en. In : *Journal of Experimental Social Psychology* 49.4 (juill. 2013), p. 764-766. ISSN : 0022-1031. DOI : 10.1016/j.jesp.2013.03.013. URL : https://www.sciencedirect.com/science/article/pii/S0022103113000668 (visité le 21/03/2022).

[34] Wenhua Li, Lili Yi et Xiang Yin. « Real Time Air Monitoring, Analysis and Prediction System Based on Internet of Things and LSTM ». In : *2020 International Conference on Wireless Communications and Signal Processing (WCSP)*. ISSN : 2472-7628. Oct. 2020, p. 188-194. DOI : 10.1109/WCSP49889.2020.9299738.

[35] *L'indice de la qualité de l'air ATMO*. fr-FR. Déc. 2020. URL : https://atmo-france.org/lindice-atmo/ (visité le 17/02/2021).

[36] *Liste des stations*. URL : https://www.atmo-occitanie.org/sites/default/files/publications/2020-10/Liste\%20des\%20stations.pdf (visité le 16/02/2021).

[37] *Loi de transition énergétique pour la croissance verte*. fr. URL : https://www.ecologie.gouv.fr/loi-transition-energetique-croissance-verte (visité le 18/01/2021).

[38] P.R. Lorczak, A.K. Caglayan et D.E. Eckhardt. « A theoretical investigation of generalized voters for redundant systems ». In : *[1989] The Nineteenth International Symposium on Fault-Tolerant Computing. Digest of Papers*. Juin 1989, p. 444-451. DOI : 10.1109/FTCS.1989.105617.

[39] Tatiana Macé, Eva Leoz et François Mathé. *Garantir la qualité des mesures en air ambiant de polluants atmosphériques réglementés : le modèle français*. fr. text. Publisher : Association pour la prévention de la pollution atmosphérique. Avr. 2013. URL : http://lodel.irevues.inist.fr/pollution-atmospherique/index.php?id=991 (visité le 15/01/2021).

[40] Tatiana Macé, François Mathé et Eva Léoz-Garziandia. « Surveillance de la qualité de l'air : Description du système français d'assurance qualité ». fr. In : (2010), p. 97.

[41] Sumukh Marathe et al. « Fault diagnosis system for low-cost air pollution sensors : demo abstract ». In : *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*. 2020, p. 611-612. DOI : https://doi.org/10.1145/3384419.3431191.

[42] M. I. Mead et al. « The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks ». en. In : *Atmospheric Environment* 70 (mai 2013), p. 186-203. ISSN : 1352-2310. DOI : 10.1016/j.atmosenv.2012.11.060. URL : http://www.sciencedirect.com/science/article/pii/S1352231012011284 (visité le 20/01/2021).

[43] Erik Meijering. « A chronology of interpolation : from ancient astronomy to modern signal and image processing ». In : *Proceedings of the IEEE* 90.3 (2002), p. 319-342.

[44] Lidia Morawska et al. « Applications of low-cost sensing technologies for air quality monitoring and exposure assessment : How far have they gone ? » en. In : *Environment International* 116 (juill. 2018), p. 286-299. ISSN : 0160-4120. DOI : 10.1016/j.envint.2018.04.018. URL : http://www.sciencedirect.com/science/article/pii/S0160412018302460 (visité le 10/11/2020).

[45] Thaha MUHAMMED et Riaz Ahmed SHAIKH. « An analysis of fault detection strategies in wireless sensor networks ». en. In : *Journal of Network and Computer Applications* 78 (jan. 2017), p. 267-287. ISSN : 1084-8045. DOI : 10.1016/j.jnca.2016.10.019. URL : https://www.sciencedirect.com/science/article/pii/S1084804516302545 (visité le 21/03/2022).

[46] Chi Vuong NGUYEN. « Assimilation de données et couplage d'échelles pour la simulation de la dispersion atmosphérique en milieu urbain ». Thèse de doct. Université de Lyon, 2017.

[47] MJ OATES et al. « Automatic fault detection in a low cost frequency domain (capacitance based) soil moisture sensor ». In : *Agricultural Water Management* 183 (2017), p. 41-48. DOI : https://doi.org/10.1016/j.agwat.2016.12.002.

[48] Margaret A OLIVER et Richard WEBSTER. « Kriging : a method of interpolation for geographical information systems ». In : *International Journal of Geographical Information System* 4.3 (1990), p. 313-332.

[49] Meenakshi PANDA et P. M. KHILAR. « Distributed self fault diagnosis algorithm for large scale wireless sensor networks using modified three sigma edit test ». en. In : *Ad Hoc Networks* 25 (fév. 2015), p. 170-184. ISSN : 1570-8705. DOI : 10.1016/j.adhoc.2014.10.006. URL : https://www.sciencedirect.com/science/article/pii/S1570870514002224 (visité le 23/03/2022).

[50] *Photoionization Detector - an overview | ScienceDirect Topics*. URL : https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/photoionization-detector (visité le 22/01/2021).

[51] Sylvain POUPRY. *Data and algorithm on Github*. fr. URL : https://github.com/spoupry/AQdata (visité le 10/10/2023).

[52] Sylvain POUPRY, Cédrick BÉLER et Kamal MEDJAHER. « Development of a reliable measurement station for air quality monitoring based on low-cost sensors and active redundancy ». In : *IFAC-PapersOnLine*. 2nd IFAC Workshop on Integrated Assessment Modelling for Environmental Systems IAMES 2022 55.5 (2022), p. 7-12. ISSN : 2405-8963. DOI : 10.1016/j.ifacol.2022.07.631. URL : https://www.sciencedirect.com/science/article/pii/S2405896322010527 (visité le 26/05/2023).

[53] Sylvain POUPRY, Cédrick BÉLER et Kamal MEDJAHER. « Towards Data Reliability Based on Triple Redundancy and Online Outlier Detection ». In : *PHM Society European Conference* 7.1 (2022). Number : 1, p. 410-420. ISSN : 2325-016X. DOI : 10.36001/phme.2022.v7i1.3366. URL : http://papers.phmsociety.org/index.php/phme/article/view/3366 (visité le 26/05/2023).

[54] Sylvain POUPRY, Kamal MEDJAHER et Cédrick BÉLER. « Automatic detection of hardware failures in an air quality measuring station with low cost sensors ». In : *PHM Society Asia-Pacific Conference*. T. 4. 1. 2023. DOI : https://doi.org/10.36001/phmap.2023.v4i1.3706.

[55] Sylvain POUPRY, Kamal MEDJAHER et Cédrick BÉLER. « Data reliability and fault diagnostic for air quality monitoring station based on low cost sensors and active redundancy ». In : *Measurement* 223 (2023), p. 113800. ISSN : 0263-2241. DOI : https://doi.org/10.1016/j.measurement.2023.113800. URL : https://www.sciencedirect.com/science/article/pii/S0263224123013647.

[56] The World Air Quality Index PROJECT. *World's Air Pollution : Real-time Air Quality Index*. en. URL : https://waqi.info/ (visité le 18/05/2022).

[57] *Protocole de détermination des caractéristiques de performance métrologique des micro-capteurs - étude comparative des performances en laboratoire de micro-capteurs de NO2 | LCSQA*. URL : https://www.lcsqa.org/fr/rapport/2016/mines-douai/protocole-determination-caracteristiques (visité le 22/01/2021).

[58] Aakash C. RAI et al. « End-user perspective of low-cost sensors for outdoor air pollution monitoring ». en. In : *Science of The Total Environment* 607-608 (déc. 2017), p. 691-705. ISSN : 0048-9697. DOI : 10.1016/j.scitotenv.2017.06.266. URL : https://www.sciencedirect.com/science/article/pii/S0048969717316935 (visité le 01/10/2021).

[59] Federica ROLLO, Chiara BACHECHI et Laura PO. « Anomaly Detection and Repairing for Improving Air Quality Monitoring ». In : *Sensors* 23.2 (2023), p. 640. DOI : https://doi.org/10.3390/s23020640.

[60] *Réglementation et Normes QA version finale*. fr. URL : https://www.ecologie.gouv.fr/sites/default/files/01_Tableau-Normes-Seuils\%20r\%C3\%A9glementaires.pdf (visité le 18/01/2021).

[61] Noor SAAD BAQER et al. « Development of the Internet of Things sensory technology for ensuring proper indoor air quality in hospital facilities : Taxonomy analysis, challenges, motivations, open issues and recommended solution ». In : *Measurement* 192 (2022), p. 110920. ISSN : 0263-2241. DOI : https://doi.org/10.1016/j.measurement.2022.110920. URL : https://www.sciencedirect.com/science/article/pii/S0263224122001993.

[62] Robin SIBSON. « A brief description of natural neighbour interpolation ». In : *Interpreting multivariate data* (1981), p. 21-36.

[63] Vaclav SKALA. « RBF interpolation with CSRBF of large data sets ». In : *Procedia Computer Science* 108 (2017), p. 2433-2437.

[64] Kate R. SMITH et al. *An improved low power measurement of ambient NO_2 and O_3 combining electrochemical sensor clusters and machine learning*. en. preprint. Gases/In Situ Measurement/Validation et Intercomparisons, sept. 2018. DOI : 10.5194/amt-2018-285. URL : https://amt.copernicus.org/preprints/amt-2018-285/amt-2018-285.pdf (visité le 20/01/2021).

[65] B. SPORTISSE. « Atmospheric pollution. From processes to modelling ; Pollution atmospherique. Des processus a la modelisation ». French. In : (juill. 2008). URL : https://www.osti.gov/etdeweb/biblio/21048157 (visité le 05/02/2021).

[66] Roland B. STULL. « Mean Boundary Layer Characteristics ». en. In : *An Introduction to Boundary Layer Meteorology*. Sous la dir. de Roland B. STULL. Atmospheric Sciences Library. Dordrecht : Springer Netherlands, 1988, p. 1-27. ISBN : 978-94-009-3027-8. DOI : 10.1007/978-94-009-3027-8_1. URL : https://doi.org/10.1007/978-94-009-3027-8_1 (visité le 08/02/2021).

[67] Olivier TALAGRAND. « Assimilation of observations, an introduction (gtspecial issueltdata assimilation in meteology and oceanography : Theory and practice) ». In : *Journal of the Meteorological Society of Japan. Ser. II* 75.1B (1997), p. 191-209.

[68] Mehmet TAŞTAN. « A low-cost air quality monitoring system based on Internet of Things for smart homes ». In : *Journal of Ambient Intelligence and Smart Environments* 1.Preprint (2022), p. 1-24. DOI : 10.3233/AIS-210458.

[69] Jean-Jacques THILLET. *La météo de montagne - Jean-Jacques Thillet*. fr. guides du club alpin français. EAN 9782020205788. Seuil, jan. 1997. ISBN : 2-02-020578-5. URL : https://www.decitre.fr/livres/la-meteo-de-montagne-9782020205788.html (visité le 10/02/2021).

[70] Yuanyuan WEI et al. « LSTM-Autoencoder-Based Anomaly Detection for Indoor Air Quality Time-Series Data ». In : *IEEE Sensors Journal* 23.4 (2023), p. 3787-3800. DOI : 10.1109/JSEN.2022.3230361.

[71] WHO. *Ambient (outdoor) air pollution*. en. organisation. Juill. 2021. URL : https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health (visité le 01/10/2021).

[72] WHO OCCUPATIONAL AND ENVIRONMENTAL HEALTH TEAM. *WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide*. ar. Rapp. tech. WHO/SDE/PHE/OEH/06.02. number-of-pages : 20. World Health Organization, 2006. URL : https://apps.who.int/iris/handle/10665/69477 (visité le 18/05/2022).

[73] Marek WOJCIKOWSKI et al. « A surrogate-assisted measurement correction method for accurate and low-cost monitoring of particulate matter pollutants ». In : *Measurement* 200 (2022), p. 111601. ISSN : 0263-2241. DOI : https://doi.org/10.1016/j.measurement.2022.111601. URL : https://www.sciencedirect.com/science/article/pii/S0263224122008120.

[74] WORLD HEALTH ORGANIZATION AND OCCUPATIONAL AND ENVIRONMENTAL HEALTH TEAM. *New WHO Global Air Quality Guidelines aim to save millions of lives from air pollution*. en. 2021. URL : https://www.who.int/news/item/22-09-2021-new-who-global-air-quality-guidelines-aim-to-save-millions-of-lives-from-air-pollution (visité le 18/05/2022).

[75] Chen XIAOJUN, Liu XIANPENG et Xu PENG. « IOT-based air pollution monitoring and forecasting system ». In : *2015 international conference on computer and computational sciences (ICCCS)*. IEEE. 2015, p. 257-260.

[76] Pamela YOBOUE. « Etude technologique pour l'amélioration des performances d'un capteur de gaz à oxyde métallique : développement d'une plateforme chauffante haute température et intégration de couche sensible nanostructurée ». In : (mai 2010).