



**HAL**  
open science

# Étude d'un système à base de micro-capteurs de gaz pour le suivi et la cartographie de la pollution atmosphérique

Rachid Laref

► **To cite this version:**

Rachid Laref. Étude d'un système à base de micro-capteurs de gaz pour le suivi et la cartographie de la pollution atmosphérique. Micro et nanotechnologies/Microélectronique. Université de Lorraine, 2020. Français. NNT : 2020LORR0048 . tel-02940704

**HAL Id: tel-02940704**

**<https://hal.univ-lorraine.fr/tel-02940704v1>**

Submitted on 16 Sep 2020

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

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



## AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : [ddoc-theses-contact@univ-lorraine.fr](mailto:ddoc-theses-contact@univ-lorraine.fr)

## LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

[http://www.cfcopies.com/V2/leg/leg\\_droi.php](http://www.cfcopies.com/V2/leg/leg_droi.php)

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

# THESE

Présentée à l'Université de Lorraine pour l'obtention du diplôme de

DOCTEUR DE L'UNIVERSITE DE LORRAINE

Spécialité : **Systèmes Electroniques**

Ecole doctorale : **IAEM**

## **ETUDE D'UN SYSTEME A BASE DE MICRO- CAPTEURS DE GAZ POUR LE SUIVI ET LA CARTOGRAPHIE DE LA POLLUTION ATMOSPHERIQUE**

Par

**Rachid Laref**

Soutenue le 18 juin 2020 devant le jury :

Rapporteur	Ph. Ménini	Professeur à l'Université Toulouse III - Paul Sabatier
Rapporteur	P. Siarry	Professeur à l'Université Paris-Est Créteil
Examineur	N. Zerhouni	Professeur à l'Université Bourgogne Franche-Comté
Directeur de thèse	M. Siadat	Maître de conférences à l'Université de Lorraine
Co-directeur	A. Sava	Maître de conférences à l'Université de Lorraine
Co-encadrant	E. Losson	Maître de conférences à l'Université de Lorraine
Membre invité	D. Durant	Responsable de l'unité métrologie d'ATMO Grand Est

## **Résumé**

L'objectif de cette thèse est de concevoir un dispositif multi-capteurs de gaz à faible coût pour la surveillance des gaz polluants en environnement externe et d'élaborer des méthodes de calibrage automatique et d'analyse de données adaptées à son exploitation. Ce dispositif vise à contribuer à la densification du réseau de surveillance de la pollution atmosphérique actuel afin de permettre une meilleure résolution spatiale et temporelle.

Premièrement, nous avons sélectionné des micro-capteurs électrochimiques et des micro-capteurs à oxyde métallique capables de détecter et de quantifier des faibles concentrations de dioxyde d'azote et d'ozone dans l'atmosphère. Nous avons caractérisé ces micro-capteurs en laboratoire en termes de sensibilité et de reproductibilité, et les avons testés sur le terrain (à proximité d'un axe autoroutier) en suivant le protocole d'évaluation de micro-capteurs de gaz proposé par le Centre Commun de Recherche Européen (C.C.R.).

L'analyse individuelle des données obtenues à partir de chaque capteur a mis en évidence la nécessité de fusionner les réponses de tous les capteurs ainsi que les valeurs de la température et de l'humidité de l'air, à l'aide des méthodes multi-variables de classification et de quantification. Les performances de plusieurs méthodes de classification, utilisées généralement dans le domaine des dispositifs multi-capteurs de gaz, ont été étudiées. Leur comparaison nous a conduit à choisir la régression SVM (Séparateur à Vaste Marge) grâce à sa précision et à sa robustesse. Les hyper-paramètres de cette méthode ont été optimisés à partir de l'algorithme de Recherche par Motifs Généralisés en raison de sa simplicité et de sa fiabilité. La méthode d'étalonnage proposée permet à notre dispositif de quantifier les deux polluants atmosphériques visés dans la gamme des précisions recommandées par les directives européennes concernant les mesures de type « indicative ».

Ensuite, nous avons abordé le problème des dérives des capteurs qui engendrent des réétalonnages périodiques. Nous avons proposé une méthode de standardisation basée sur la régression SVM pour permettre de réduire le coût et l'effort d'un étalonnage complet. Cette même méthode de standardisation a été utilisée avec succès pour le transfert d'étalonnage entre plusieurs systèmes identiques afin d'éviter des étalonnages individuels.

Nos travaux de recherche ont été valorisés par la publication de trois articles dans les journaux référencés JCR et deux communications dans des conférences internationales IEEE.

### **Mots clés :**

Micro-capteurs de gaz ; polluants atmosphériques ; étalonnage, estimation et optimisation ; méthodes multi-variables ; transfert d'étalonnage.

## **Abstract**

The objective of this thesis is to design a low-cost gas sensor array for outdoor gas pollutants monitoring and to develop adapted standardization and data analysis methods. The device aim is to contribute to the densification of the current atmospheric pollution-monitoring network for better spatial and temporal resolution.

First, we select electrochemical and metal oxide gas sensors capable to detecting and quantifying low concentrations of nitrogen dioxide and ozone in ambient air near a highway. We have characterized these micro-sensors in the laboratory using a diffusing controller system and tested them in the field by following procedures inspired from the evaluation protocol of low cost gas sensors proposed by the European Joint Research Center (JRC).

The individual analysis of the data from each sensor highlights the need to merge the responses of all the sensors as well as the values of the temperature and the humidity of the atmosphere using multivariate methods of classification. The performance of several classification methods generally used in the field of multi-sensor gas systems was studied. This comparison leads us to choose Support Vector Machine regression (SVM) thanks to its performance in terms of precision and robustness. We also conduct a study on the optimization of the SVM hyper-parameters using optimization algorithms. The Generalized Pattern Search algorithm (GPS) was chosen among other algorithms due to its simplicity and reliability.

Then, we highlight the problem of sensor drift over time and the need for periodic calibration. We proposed a standardization method based on SVM regression, thereby reducing the cost and the effort of full recalibration. This standardization method was also used for calibration transfer between several identical systems in order to avoid individual calibration of each system.

The research work elaborated in this PhD was published in three journal papers and two IEEE conference papers.

### **Key words:**

Low cost gas sensors; atmospheric air pollution; calibration, estimation, optimization; multivariate methods; calibration transfer.

*À la mémoire de mon père*

## Remerciements

Je voudrais remercier tout d'abord Madame Maryam Siadat Maître de conférences HDR à l'Université de Lorraine pour avoir dirigé mes recherches et pour sa confiance en mes compétences pour mener à bien ces travaux. Elle a toujours été disponible, à l'écoute de mes nombreuses questions, et s'est toujours intéressée à l'avancée de mes travaux. Je suis très reconnaissant pour tout ce qu'elle m'a apporté au cours de ces années.

Je voudrais remercier grandement Monsieur Alexandre Sava Maître de conférences HDR à l'Université de Lorraine qui m'a encadré durant ces années de thèse, je le remercie également pour sa gentillesse, sa disponibilité permanente et pour les nombreux encouragements qu'il m'a prodigués. Sa compétence, sa rigueur scientifique et sa clairvoyance m'ont beaucoup appris.

Mes vifs remerciements vont à Monsieur Etienne Losson, Maître de conférences à l'Université de Lorraine pour son implication dans ce travail de thèse. Je suis très ravi d'avoir travaillé en sa compagnie depuis mon stage de master. Il a toujours été là pour me soutenir et m'encourager. Ses conseils et ses remarques constructives m'ont permis d'améliorer grandement la qualité de mes travaux et de ce mémoire.

J'exprime tous mes remerciements à l'ensemble des membres de mon jury : Messieurs Patrick Siarry Professeur à l'Université Paris-Est Créteil, Philippe Ménini Professeur à l'Université Toulouse III - Paul Sabatier, et Nouredine Zerhouni Professeur à l'Université de Bourgogne Franche-Comté, pour l'honneur qu'ils m'ont fait en participant à ce jury.

Je remercie également Monsieur Damien Durant Responsable de l'unité métrologie d'ATMO Grand Est, d'avoir accepté de mettre à notre disposition des équipements et des données propres d'ATMO Grand Est. Je le remercie également pour son aide et ses conseils techniques précieux.

Mes remerciements vont aussi au professeur Imed KACEM Directeur du Laboratoire LCOMS de m'avoir accueilli au sien de son laboratoire et de m'avoir permis de travailler dans d'aussi bonnes conditions. Je tiens à remercier tous les membres du laboratoire LCOMS et particulièrement les thésards pour les nombreux bons moments passés ensemble.

Enfin, un énorme merci à ma famille, à ma mère, à mes frères et sœurs et à mes amis pour leurs encouragements et leur soutien indéfectible.

# Sommaire

<b>Introduction générale.....</b>	<b>10</b>
<b>Chapitre I : La pollution environnementale et les méthodes de mesure.....</b>	<b>13</b>
<b>I.1 Introduction .....</b>	<b>14</b>
<b>I.2 Les polluants, leurs sources et leurs impacts sur la santé et l'environnement .....</b>	<b>14</b>
I.2.1 Les oxydes d'azote (NO <sub>2</sub> ) .....	15
I.2.2 Le dioxyde de soufre (SO <sub>2</sub> ) .....	16
I.2.3 Les composés organiques volatils (COV) .....	16
I.2.4 Le monoxyde de carbone (CO) .....	16
I.2.5 L'ozone (O <sub>3</sub> ) .....	17
I.2.6 L'ammoniac (NH <sub>3</sub> ).....	17
I.2.7 Les métaux lourds.....	17
I.2.8 Les particules ou poussières en suspension (PM) .....	18
<b>I.3 Réglementation de la qualité de l'air .....</b>	<b>18</b>
<b>I.4 Définition des normes de la qualité de l'air.....</b>	<b>20</b>
<b>I.5 Les objectifs environnementaux par polluant .....</b>	<b>20</b>
<b>I.6 Les méthodes de référence utilisées.....</b>	<b>21</b>
I.6.1 Mesure des oxydes d'azote (NO et NO <sub>x</sub> ) par chimiluminescence.....	22
I.6.2 Mesure d'ozone par absorption de rayonnement UV.....	23
I.6.3 Conclusion .....	24
<b>I.7 Stations de mesure de la pollution .....</b>	<b>24</b>
I.7.1 Les stations « de fond » .....	24
I.7.2 Les stations « de proximité » .....	24
<b>I.8 Types de mesure et incertitudes.....</b>	<b>25</b>
<b>I.9 Evolution de la pollution en France (2000-2017) .....</b>	<b>27</b>
<b>I.10 Les micro-capteurs de gaz à faible coût .....</b>	<b>28</b>
I.10.1 Les capteurs électrochimiques (EC).....	29
I.10.2 Les capteurs à oxyde métallique (MOS) .....	31
I.10.3 Les capteurs infrarouges.....	32
I.10.4 Les capteurs à photo-ionisation .....	32
I.10.5 Les micro-capteurs de gaz commerciaux.....	33
<b>I.11 Surveillance de la pollution atmosphérique avec des micro-capteurs à faible coût.....</b>	<b>34</b>
<b>I.12 Conclusion.....</b>	<b>42</b>
<b>Chapitre II : Protocole d'évaluation des micro-capteurs de gaz.....</b>	<b>44</b>
<b>II.1 Introduction .....</b>	<b>45</b>
<b>II.2 Protocole d'évaluation et d'étalonnage des micro-capteurs recommandé par Centre Commun de Recherche Européen.....</b>	<b>45</b>
II.2.1 Définition du module « mesure ».....	46



II.2.2	Les mesures en laboratoire .....	48
II.2.3	Chambre d'exposition .....	48
II.2.4	Génération de mélange de gaz.....	48
II.2.5	Temps de réponse .....	49
II.2.6	Pré-calibration .....	49
II.2.7	Dérives des capteurs.....	49
II.2.8	Facteurs d'influence (interférents).....	51
<b>II.3</b>	<b>Etalonnage et validation en laboratoire.....</b>	<b>52</b>
<b>II.4</b>	<b>Validation sur le terrain .....</b>	<b>52</b>
<b>II.5</b>	<b>Conclusion.....</b>	<b>53</b>
<b>II.6</b>	<b>Caractérisation et test des micro-capteurs.....</b>	<b>53</b>
<b>II.7</b>	<b>Sélection de micro-capteurs de gaz .....</b>	<b>54</b>
<b>II.8</b>	<b>Mesures en laboratoire.....</b>	<b>55</b>
II.8.1	Temps de réponse .....	57
II.8.2	Répétabilité .....	59
II.8.3	Dérives à court terme .....	60
<b>II.9</b>	<b>Etalonnage en laboratoire.....</b>	<b>60</b>
II.9.1	Etalonnage du capteur NO <sub>2</sub> -B43F .....	60
II.9.2	L'étalonnage des capteurs MiCS-2714 et de GGS 7530 .....	61
<b>II.10</b>	<b>Mesures sur le terrain .....</b>	<b>63</b>
<b>II.11</b>	<b>Étalonnage sur le terrain.....</b>	<b>65</b>
II.11.1	Mesure de concentrations de dioxyde d'azote .....	65
II.11.2	Mesure de l'ozone .....	66
II.11.3	Influence de la température et de l'humidité .....	68
II.11.4	Incertitudes des mesures.....	69
<b>II.12</b>	<b>Conclusion.....</b>	<b>70</b>

**Chapitre III : Méthodes multi-variables pour l'estimation de la concentration de gaz à l'aide d'un réseau de capteurs .....** **72**

<b>III.1</b>	<b>Introduction .....</b>	<b>73</b>
<b>III.2</b>	<b>Régression SVM.....</b>	<b>74</b>
III.2.1	Cas linéaire.....	74
III.2.2	Cas non linéaire .....	78
<b>III.3</b>	<b>Optimisation des hyper-paramètres de la régression SVM .....</b>	<b>80</b>
III.3.1	Influence de chaque hyper-paramètre sur la performance de la régression SVM .....	80
III.3.2	Relation entre C et $\sigma$ .....	81
<b>III.4</b>	<b>Présentation des algorithmes d'optimisation .....</b>	<b>82</b>
III.4.1	Algorithme génétique.....	83
III.4.2	Recuit simulé.....	85
III.4.3	Optimisation par essais particuliers .....	86
III.4.4	Recherche directe GPS.....	88
<b>III.5</b>	<b>Application et comparaison des algorithmes d'optimisation .....</b>	<b>89</b>
III.5.1	Application de l'algorithme GPS.....	89

III.5.2	Applications d'autres algorithmes d'optimisation et comparaison .....	92
<b>III.6</b>	<b>Application des modèles d'étalonnage et comparaison.....</b>	<b>93</b>
III.6.1	Extreme Learning Machine (ELM) .....	93
III.6.2	Régression des moindres carrés partiels (PLS) .....	95
III.6.3	Résultats de la comparaison .....	96
<b>III.7</b>	<b>Etude de l'influence de la température.....</b>	<b>98</b>
<b>III.8</b>	<b>Comparaison entre l'utilisation individuelle des capteurs et la fusion des données des capteurs par les méthodes multi-variables .....</b>	<b>99</b>
<b>III.9</b>	<b>Conclusion.....</b>	<b>100</b>
<b>Chapitre IV : Transfert d'étalonnage pour une utilisation à long terme et sur des instruments identiques.....</b>		<b>102</b>
<b>IV.1</b>	<b>Introduction .....</b>	<b>103</b>
<b>IV.2</b>	<b>Les causes d'invalidité d'un modèle d'étalonnage.....</b>	<b>103</b>
<b>IV.3</b>	<b>Les stratégies d'adaptation d'un modèle d'étalonnage .....</b>	<b>105</b>
<b>IV.4</b>	<b>Méthodes de standardisation .....</b>	<b>105</b>
IV.4.1	Standardisation des coefficients du modèle.....	106
IV.4.2	Standardisation des valeurs prédites .....	107
IV.4.3	Standardisation des réponses des capteurs .....	107
IV.4.4	Conclusion.....	109
<b>IV.5</b>	<b>Algorithmes de sélection d'échantillons .....</b>	<b>109</b>
IV.5.1	L'algorithme Kennard-Stone (K-S) .....	109
IV.5.2	L'algorithme SPXY .....	110
IV.5.3	L'algorithme kernel SPXY .....	110
<b>IV.6</b>	<b>Algorithme proposé pour la standardisation.....</b>	<b>111</b>
<b>IV.7</b>	<b>Transfert d'étalonnage entre deux instruments identiques.....</b>	<b>112</b>
IV.7.1	Nécessité de la standardisation .....	113
IV.7.2	Application de la standardisation SVM .....	114
IV.7.3	Comparaison avec la standardisation directe classique.....	115
<b>IV.8</b>	<b>Standardisation pour éviter un réétalonnage complet .....</b>	<b>116</b>
<b>IV.9</b>	<b>Transfert d'étalonnage pour un seul dispositif entre deux périodes .....</b>	<b>118</b>
IV.9.1	Nécessité d'un étalonnage fréquent ou d'une standardisation pour corriger les dérives .....	118
IV.9.2	Application de la standardisation.....	119
<b>IV.10</b>	<b>Conclusion .....</b>	<b>120</b>
<b>Conclusion générale .....</b>		<b>123</b>
<b>Annexe: Normes de surveillance pour la pollution de l'air.....</b>		<b>128</b>
<b>Références bibliographiques .....</b>		<b>134</b>

# **Introduction générale**

# Introduction générale

La pollution atmosphérique, selon les derniers rapports de l'Organisation Mondiale de la Santé (OMS), est la cause d'environ sept millions de décès prématurés dans le monde chaque année [Lelieveld et al. 2015]. Ces décès sont liés principalement aux accidents vasculaires cérébraux, aux maladies cardiaques, aux maladies pulmonaires obstructives chroniques, aux cancers du poumon et aux infections respiratoires aiguës [Pascal 2009; Sun et Zhu 2019; Myrto 2009; Mabahwi et al. 2014]. L'OMS estime que 9 personnes sur 10 respirent de l'air contenant des niveaux élevés de polluants, et que 80% des personnes vivant dans des zones urbaines sont exposées à des niveaux de polluants de l'air qui dépassent les limites fixées par l'OMS. De plus, la qualité de l'air est fortement liée aux enjeux climatiques et aux écosystèmes dans le monde [Babich et Stotzky 1972; Vitousek et al. 1997].

L'une des causes essentielles de la pollution atmosphérique est due aux processus de combustion [Kampa et Castanas 2008] liés au trafic routier, au chauffage domestique, aux centrales électriques et aux sources industrielles. Dans les zones urbaines, le trafic routier constitue la source majeure d'émission de différents polluants. Les polluants émis sont de nature gazeuse comme les oxydes d'azote ( $\text{NO}_x$ ), le monoxyde de carbone (CO), l'ozone ( $\text{O}_3$ ) ou de nature solide comme les particules en suspension [Curtis et al. 2006].

Les réglementations et les moyens de surveillance de la pollution ont évolué au fil des années. De nos jours, des directives détaillées décrivent les seuils et les limites d'émission pour chaque polluant ainsi que les méthodes de surveillance adoptées (Arrêté du 19 avril 2017 relatif au dispositif national de surveillance de la qualité de l'air ambiant). La pollution atmosphérique est continuellement surveillée par un réseau de stations de mesure fixes, parfois complété par une modélisation de la propagation de la pollution [Clements et al. 2017; Kamionka et al. 2006]. Le coût élevé des instruments de mesure dits « de référence », comme les analyseurs utilisés par les agences de la qualité de l'air en France, est responsable du nombre limité de stations déployées. Ce nombre limité de stations de mesure demeure insuffisant pour bien caractériser la pollution et sa dispersion spatiale et temporelle à l'échelle urbaine [Wilson 2012; McKercher et al. 2017; Cavaliere et al. 2018; Sayago et al. 2019].

Aujourd'hui avec l'évolution technologique des micro-capteurs à faible coût disponibles sur le marché et montés dans des dispositifs adéquats sont capables de fournir une estimation des concentrations de divers polluants atmosphériques [White et al. 2012; McKercher et al. 2017; Sun et al. 2016]. Bien que ces capteurs de gaz ne puissent pas fournir les mêmes sensibilité, stabilité et précision que les instruments de référence [Williams et al. 2019], ils peuvent néanmoins constituer un complément précieux dans les programmes de surveillance de la qualité de l'air pour une meilleure couverture spatiale et temporelle [Laref et al. 2019; Trincavelli et al. 2008; Mukherjee et al. 2017; Spinelle et al. 2014; De Vito et al. 2009; Gillooly et al. 2019].

Plusieurs projets internationaux ont été créés pour évaluer des systèmes à prix modéré pour la surveillance des principaux polluants atmosphériques [Castell et al. 2017; Spinelle et al. 2017; Jiao et al. 2016]. Ces projets sont basés essentiellement sur des systèmes commercialisés sans proposer des algorithmes de traitement de données innovants permettant de pallier aux problèmes rencontrés dans l'exploitation de ces systèmes. Dans ce cadre, le groupe « nez électronique » du laboratoire LCOMS a monté un projet de recherche pour concevoir des dispositifs innovants à prix modéré en se basant sur les méthodes de traitement de données pertinentes pour la mesure du dioxyde d'azote et de l'ozone. Le choix de ces deux polluants est basé sur le fait qu'ils sont régulièrement en dépassement dans plusieurs agglomérations françaises [Le Moullec 2018]. Ce groupe travaille sur la conception et la valorisation des systèmes multi-capteurs intelligents de gaz dédiés appliqués en particulier aux mesures d'odeurs et de la pollution atmosphérique [Ahmadou 2015; Fuchs 2008].

Notre travail concerne l'étude d'un système à base de micro-capteurs chimiques pour la surveillance de la concentration de deux gaz polluants : le dioxyde d'azote et l'ozone. Plus précisément, nous portons notre attention sur les performances de notre système par rapport aux seuils et aux limites fixés par l'OMS et les directives européennes. Ces types de systèmes, que l'on appelle des nez électroniques, sont constitués de deux parties principales : la partie matérielle contenant les capteurs ainsi que l'électronique associée et la partie logicielle qui concerne l'analyse des données des capteurs à l'aide des méthodes de traitement de données et de reconnaissance de formes.

L'essentiel de notre travail s'est porté sur le test des capteurs et l'élaboration des algorithmes d'analyse de données basés sur des méthodes multi-variables, dont nous avons optimisé les hyper-paramètres afin d'obtenir la meilleure estimation possible de la concentration des gaz polluants ciblés. Par la suite, nous avons étudié des méthodes de transfert

d'étalonnage pour permettre d'une part de surmonter les problèmes liés aux dérives des capteurs et d'autre part d'adapter un modèle d'étalonnage pour plusieurs nez électroniques identiques.

Notre étude est organisée selon quatre chapitres.

Dans le premier chapitre, nous proposons un état de l'art sur la pollution de l'air ambiant ainsi que sur les systèmes et les méthodes déployés pour la surveillance de la qualité de l'air. Nous présentons brièvement les principaux polluants, leurs sources et leurs impacts sur la santé et la nature. Nous présentons également les principales directives concernant la pollution de l'air, les analyseurs actuels de la qualité de l'air et les micro-capteurs de gaz à faible coût. Nous décrivons par la suite, quelques travaux de recherche effectués sur ces types de capteurs de gaz afin de justifier notre choix de capteurs.

Dans le chapitre II, nous présentons la démarche que nous avons adoptée pour caractériser et tester les capteurs choisis en suivant un protocole allégé de la version proposée dans le cadre de la directive européenne par le Centre Commun de Recherche européen C.C.R (ou JCR, Joint Research Center) [Spinelle et al. 2013]. Dans un premier temps nous réalisons des expérimentations en laboratoire afin de déterminer les caractéristiques et l'adéquation des capteurs choisis à notre application. Ensuite, nous conduisons des mesures et des tests sur le terrain afin d'étudier l'influence des facteurs environnementaux tels que les gaz interférents et la température sur la performance ou encore la dégradation des capteurs.

Le chapitre III est consacré aux méthodes d'étalonnage des systèmes multi-capteurs. Nous évoquons les méthodes d'apprentissage statistique utilisées pour les nez électroniques. Nous montrons l'intérêt de la méthode *support vector machine* (ou Séparateur à Vaste Marge, SVM) pour notre application. Le choix des hyper-paramètres d'une méthode d'apprentissage statistique étant primordial pour obtenir une bonne performance, nous étudions également des algorithmes d'optimisation et nous proposons un algorithme d'optimisation adapté en termes d'effort de calcul, de rapidité et de précision.

Le chapitre IV aborde le problème du transfert du modèle d'étalonnage. Nous étudions les différentes techniques de transfert d'étalonnage afin de généraliser ou d'adapter un modèle prédéfini. Nous exploitons cette approche pour corriger les dérives à long terme ainsi que la non reproductibilité dans la fabrication de capteurs de gaz identiques. Notre contribution porte sur l'élaboration d'une méthode de standardisation basée sur la méthode SVM.

**Chapitre I : La pollution environnementale  
et les méthodes de mesure**

### I.1 Introduction

La pollution de l'air est définie par la présence des gaz et des particules qui ont des effets nocifs sur tous les vivants sans discrimination (l'homme, les animaux, les plantes) et sur l'environnement naturel. Cette pollution a été longtemps considérée comme un problème local, mais aujourd'hui elle est devenue un problème mondial puisqu'elle provoque en plus des dérèglements climatiques. Ceci explique la résolution de nombreux pays à agir contre ce problème en établissant des réglementations précises. Cependant, la lutte contre la pollution atmosphérique passe avant tout par une caractérisation des différents polluants, leurs origines, ainsi que l'évolution de leur concentration dans le temps et dans l'espace. C'est pourquoi, nous dédions ce chapitre à un état de l'art sur les polluants atmosphériques, les directives européennes en ce qui concerne les objectifs de qualité d'air, les méthodes conventionnelles de mesures et les nouvelles approches.

Dans la première partie, nous abordons la problématique de la pollution atmosphérique, et décrivons les différents polluants, leurs sources et leurs effets sur la santé et la nature. Ensuite, nous détaillons les réglementations, les objectifs européens de la qualité de l'air concernant les principaux polluants, ainsi que les méthodes de mesure adoptées selon les normes européennes pour les deux gaz polluants étudiés dans ce travail (le dioxyde d'azote et l'ozone).

La deuxième partie est consacrée aux micro-capteurs de gaz à faible coût. Ces capteurs présentent une alternative très prometteuse pour la mesure de certains polluants de l'air. Après avoir décrit les principes de fonctionnement de différents types de capteurs, nous présenterons quelques travaux réalisés sur ces micro-capteurs de gaz et appliqués à la détection de la pollution de l'air. Ces travaux ont été choisis afin de mettre en évidence les principaux aspects à prendre en compte lors de l'utilisation de ces micro-capteurs, qui concernent le choix de types de capteurs, l'influence de différents facteurs environnementaux, les méthodes d'étalonnage, la différence observée entre un étalonnage en laboratoire et un étalonnage sur le terrain.

### I.2 Les polluants, leurs sources et leurs impacts sur la santé et l'environnement

Les sources de polluants sont multiples et peuvent être regroupées dans deux catégories : les sources d'origine naturelle et les sources d'origine anthropique.

- Les sources d'origine naturelle, appelées biogéniques, sont liées aux phénomènes naturels tels que les feux de forêts, les éruptions volcaniques, les brumes des sables ou encore



les embruns marins. Ces phénomènes produisent plusieurs polluants comme les composés organiques volatils (COV), le dioxyde d'azote ( $\text{NO}_2$ ), l'ammoniac ( $\text{NH}_3$ ), le méthane ( $\text{CH}_4$ ), le dioxyde de carbone ( $\text{CO}_2$ ), des composés soufrés et des aérosols (Fenger 2009). La photochimie provoque également plusieurs réactions dans l'air ambiant, y compris la transformation de polluants primaires en polluants secondaires, par exemple, les oxydes d'azote ( $\text{NO}_x$ ) en ozone ( $\text{O}_3$ ).

- Les sources de pollution d'origine anthropique sont liées à l'activité humaine [Mayer 1999] et appartiennent aux différents secteurs tels que : agricole (97% des émissions d'ammoniac et 28% des particules fines en suspension ( $\text{PM}_x$ )), industriel (83% du dioxyde de soufre émis et 40% des COV), résidentiel (31% des  $\text{PM}_{10}$  et 49% des  $\text{PM}_{2.5}$ ) et aux transports (61% des émissions d'oxyde d'azote). La pollution atmosphérique anthropique est principalement due aux processus de combustion dans les moteurs (pour la propulsion de véhicules), les poêles à bois et les centrales électriques. En milieu urbain, ces moteurs émettent des gaz tels que : des oxydes de carbone ( $\text{CO}_x$ ), des hydrocarbures non brûlés (HC), des oxydes d'azote ( $\text{NO}_x$ ), des dioxydes de soufre ( $\text{SO}_x$ ), des particules fines en suspension ( $\text{PM}_x$ ) et des métaux lourds [Fenger 2009].

### I.2.1 Les oxydes d'azote ( $\text{NO}_2$ )

Les principaux oxydes d'azote sont le monoxyde d'azote (NO) et le dioxyde d'azote ( $\text{NO}_2$ ). Ces oxydes résultent des processus de combustion (chauffage, centrales électriques, moteurs thermiques des véhicules...) [Charpin et Raheison 2007]. Ils sont également émis lors des processus de chimie de l'azote (fabrication de nitrate d'ammonium) ou de l'utilisation de produits nitrés dans les procédés industriels. Dans le secteur agricole, l'utilisation des engrais azotés entraîne des rejets de  $\text{NO}_x$ . Des sources naturelles comme les volcans et les sols naturels secs sont aussi responsables de la formation d'oxydes d'azote.

Les oxydes d'azote présents dans l'atmosphère sont très nocifs. Ils peuvent causer des crises d'asthmes et des infections pulmonaires [Pénard-Morand et Annesi-Maesano 2008]. Ils réagissent en particulier avec les composés organiques volatils pour conduire à la formation d'ozone ou d'ammoniac. Ils provoquent l'acidification et l'eutrophisation des sols, des eaux souterraines et des eaux de surface, entraînant la réduction de la vitalité des écosystèmes et de la biodiversité. **La moyenne annuelle des concentrations de  $\text{NO}_2$  est plus élevée en milieu urbain, en particulier à proximité des principales routes, qu'en milieu rural.**

### **I.2.2 Le dioxyde de soufre (SO<sub>2</sub>)**

Les sources principales de dioxyde de soufre sont les processus de combustion des énergies fossiles (fioul, charbon, lignite, gazole, etc.) ainsi que les procédés industriels (production d'acide sulfurique, production de pâte à papier, raffinage du pétrole). Les volcans sont aussi des sources naturelles d'émission de dioxyde de soufre [Godish et al. 2014].

Le dioxyde de soufre provoque des troubles respiratoires : toux, gêne respiratoire, troubles asthmatiques [Charpin et Raheison 2007]. Il peut aussi causer une irritation des muqueuses et de la peau. Sur le plan naturel il rentre dans la formation des pluies acides. Il est aussi responsable de la formation des particules secondaires en réaction avec les NO<sub>x</sub>.

### **I.2.3 Les composés organiques volatils (COV)**

La Directive Européenne 2008 a défini les composés organiques volatils comme étant « *des composés organiques provenant de sources anthropiques et biogènes, autres que le méthane, capables de produire des oxydants photochimiques par réaction avec des oxydes d'azote sous l'effet du rayonnement solaire* ». Parmi ces composés, on trouve le benzène, l'acétone et le perchloroéthylène.

Les COV conduisent à une diminution de la capacité respiratoire et à des irritations. Ils sont aussi des sources de nuisances olfactives. Certains composés, comme le benzène, peuvent aussi être cancérogènes.

La nature est responsable de 90% des émissions de COV. Pour les sources liées aux activités humaines, le secteur résidentiel et tertiaire, en raison de l'utilisation de solvants à usage domestique (peintures, colles, etc.), est le principal secteur d'émission des COV. Il est suivi par l'industrie manufacturière (peintures) et le transport routier.

### **I.2.4 Le monoxyde de carbone (CO)**

Le monoxyde de carbone est un polluant atmosphérique inodore et omniprésent qui résulte de l'oxydation incomplète du carbone au cours des processus de combustion. Les centrales électriques, les chaudières, les incendies de forêt, les transports et le tabagisme sont les sources les plus courantes des émissions du CO dans l'environnement [Godish et al. 2014]. Le CO est le polluant principal de la qualité de l'air intérieur.

L'exposition au CO constitue un risque majeur pour les êtres humains [Estrella et al. 2005]. En fonction de sa concentration il peut causer des problèmes de santé allant jusqu'à des asphyxies mortelles. Des expositions à des concentrations de CO supérieure à 100µg/m<sup>3</sup> provoque une déficience du système cardiovasculaire [Elichegaray et al. 2009].

### **I.2.5 L'ozone (O<sub>3</sub>)**

L'ozone forme une couche de 12 à 50km au-dessus du sol. Cette couche permet d'intercepter des rayons ultraviolets et de protéger ainsi la terre. En revanche sa présence dans les basses couches de l'atmosphère (de 0 à 12km au-dessus du sol) est très nocive pour la santé humaine, les animaux et les végétaux.

L'ozone est produit par les transformations photochimiques complexes entre certains polluants comme les oxydes d'azote, le monoxyde de carbone et les composés organiques volatils [Awang et al. 2000].

Les effets toxiques de l'ozone pour la santé humaine comprennent les douleurs thoraciques, la toux, l'irritation de la gorge et la congestion, ainsi que l'aggravation de la bronchite, de l'emphysème et de l'asthme. Il peut réduire la fonction pulmonaire et enflammer les parois des poumons. Les effets de l'O<sub>3</sub> sur l'environnement naturel se manifestent principalement par la dégradation de la végétation à la suite de l'endommagement des feuilles, la réduction de la photosynthèse et la diminution du rendement des cultures. Les dommages causés aux plantes peuvent altérer la structure de l'écosystème, réduire la biodiversité et la croissance des plantes. Généralement, c'est en été lors des périodes de canicule que surviennent les épisodes de pollution à l'ozone.

### **I.2.6 L'ammoniac (NH<sub>3</sub>)**

Les activités agricoles comme la volatilisation lors des épandages et le stockage des effluents d'élevage sont les sources principales du gaz d'ammoniac. Ce gaz toxique est caractérisé par son odeur piquante, qui brûle les yeux et les poumons. C'est l'une des causes de la pollution olfactive à l'origine de nombreuses plaintes de voisinage [Fuchs 2008].

Son inhalation est toxique, voire mortelle à des fortes concentrations. Sur le plan écologique, il provoque une acidification des eaux et des sols. Il peut aussi être à l'origine des particules fines secondaires par une réaction chimique avec d'autres substances.

### **I.2.7 Les métaux lourds**

Les métaux lourds sont des constituants de la croûte terrestre. Parmi les métaux lourds on retrouve : Plomb, Cadmium, Mercure, Fer, Zinc, Nickel, Arsenic, Chrome. Ces métaux peuvent se trouver en suspension à cause d'une érosion, une activité volcanique ou des feux de forêts. L'homme peut aussi participer à l'apparition de ces métaux dans l'air grâce aux activités métallurgiques, l'extraction minière, la production énergétique, l'incinération de déchets et le trafic routier [Godish et al. 2014].

Bien que certains éléments métalliques dont le Fer, le Zinc, le Nickel, l'Arsenic et le Chrome soient indispensables à la nature à faibles doses, leur existence à des doses importantes est très nocive. Sur le plan de la santé humaine, les métaux lourds peuvent être très dangereux sur les fonctions rénales, hépatiques et respiratoires. Ils provoquent aussi certains cancers [Elichegaray et al. 2009].

### **I.2.8 Les particules ou poussières en suspension (PM)**

Selon leur mécanisme de formation ou d'émission, on peut regrouper les PM en deux types : les particules primaires et les particules secondaires.

- Les particules primaires sont majoritairement émises par des combustions incomplètes. Elles sont aussi générées par des processus naturels comme l'érosion des sols, les pollens et les feux de biomasse.
- Les particules secondaires sont principalement dues aux réactions physico-chimiques entre certains polluants comme le dioxyde de soufre (SO<sub>2</sub>), les oxydes d'azote (NO<sub>x</sub>), les composés organiques volatils (COV) ou même les particules primaires.

Ces particules sont classées en fonction de leur diamètre. Les particules de diamètre inférieur à 10 micromètres sont appelées PM<sub>10</sub>. Vu leur taille elles restent au niveau du nez et des voies aériennes supérieures. En revanche les PM<sub>2,5</sub>, avec un diamètre inférieur à 2.5 micromètres, peuvent pénétrer profondément jusqu'aux alvéoles pulmonaires et même circuler dans le sang [Brunekreef et Forsberg 2005].

Sur le plan sanitaire, la pénétration des particules provoque des irritations et des problèmes respiratoires chez les personnes sensibles. Elles peuvent aussi être une cause des maladies cardiovasculaires et des cancers [Simon et Charpin 2010; Brook et al. 2002].

## **I.3 Réglementation de la qualité de l'air**

La qualité de l'air est un sujet de préoccupation de l'autorité française depuis le 19<sup>ème</sup> siècle. En octobre 1810, un décret impérial est apparu pour réglementer les émissions d'odeur par les usines. En avril 1932 en France, la loi Morizet a été adoptée pour limiter les gaz toxiques, les fumées et les poussières émis par les manufactures. Une autre loi apparue en août 1961 vise non seulement les industries mais aussi les ateliers artisanaux, les établissements agricoles et même les immeubles pour lutter contre la pollution de l'air et la nuisance olfactive.

A partir des années 1980, ce sont les directives européennes qui organisent les normes et les réglementations de la qualité de l'air. L'évolution de ces directives concernant la qualité

de l'air est résumée dans le tableau I-1 [Al Barakeh 2012]. A ce jour, la directive européenne 2008/50/CE fixe les normes de la qualité de l'air pour les polluants gazeux, les particules fines et le plomb.

Directive	Date de parution	Polluants
80/779/EEC (C.E., 1980)	15-juillet-1980	SO <sub>2</sub> , PM <sub>10</sub>
82/884/EEC (C.E., 1982)	3-December-1982	NO <sub>x</sub>
85/283/EEC (C.E., 1985)	7-Mars-1985	
92/72/EEC (C.E., 1992)	21-Septembre-1992	O <sub>3</sub>
1999/30/CE (C.E., 1999)	22-Avril-1999	SO <sub>2</sub> , NO <sub>2</sub> & NO <sub>x</sub> , PM <sub>10</sub> , Pb
2000/69/CE (C.E., 2000)	16-Novembre-2000	C <sub>6</sub> H <sub>6</sub> , CO
2002/3/CE (C.E., 2002)	12-Février-2002	O <sub>3</sub>
2004/107/CE (C.E., 2004)	15-décembre-2004	As, Cd, Hg, Ni, HAP
2008/50/CE (C.E., 2008)	21-mai-2008	SO <sub>2</sub> , NO <sub>2</sub> & NO <sub>x</sub> , PM <sub>10</sub> , PM <sub>2,5</sub> , Pb, C <sub>6</sub> H <sub>6</sub> , CO, O <sub>3</sub> .

**Tableau I- 1 Evolution de la directive européenne pour la qualité de l'air.**

Sur ce tableau nous pouvons constater que le nombre de polluants concernés par la réglementation européenne s'est accru considérablement depuis les années 80. Ceci est principalement dû au fait que les sources d'émission de polluants deviennent de plus en plus importantes, mais également au fait que de nos jours l'évolution de la technologie permet de réaliser des mesures de nombreux polluants même à de très faibles concentrations. Ces directives imposent des surveillances en continu et/ou ponctuelles de la qualité de l'air à partir des appareillages complexes mais précis.

### I.4 Définition des normes de la qualité de l'air

Dans cette partie, nous définissons tout d'abord la terminologie utilisée pour les objectifs de la qualité de l'air, puis nous donnons les seuils fixés pour la concentration des différents polluants. Nous terminons cette partie par une présentation des méthodes de mesure de référence utilisées dans cette étude pour les oxydes d'azote et l'ozone.

Conformément à l'article R. 221-1 du code de l'environnement en France, les termes employés pour caractériser la qualité de l'air sont les suivants :

- **Objectif de qualité** : *« un niveau de concentration de substances polluantes dans l'atmosphère à atteindre à long terme, sauf lorsque cela n'est pas réalisable par des mesures proportionnées, afin d'assurer une protection efficace de la santé humaine et de l'environnement dans son ensemble ».*
- **Valeur cible** : *« un niveau de concentration de substances polluantes dans l'atmosphère fixé dans le but d'éviter, de prévenir ou de réduire les effets nocifs sur la santé humaine ou sur l'environnement dans son ensemble, à atteindre, dans la mesure du possible, dans un délai donné ».*
- **Valeur limite** : *« un niveau de concentration de substances polluantes dans l'atmosphère fixé sur la base des connaissances scientifiques à ne pas dépasser dans le but d'éviter, de prévenir ou de réduire les effets nocifs de ces substances sur la santé humaine ou sur l'environnement dans son ensemble ».*
- **Seuil d'information et de recommandation** : *« un niveau de concentration de substances polluantes dans l'atmosphère au-delà duquel une exposition de courte durée présente un risque pour la santé humaine des groupes particulièrement sensibles de la population rendant nécessaires des informations immédiates et adéquates ».*
- **Seuil d'alerte** : *« un niveau de concentration de substances polluantes dans l'atmosphère au-delà duquel une exposition de courte durée présente un risque pour la santé de l'ensemble de la population ou de dégradation de l'environnement justifiant l'intervention de mesures d'urgence ».*

### I.5 Les objectifs environnementaux par polluant

Le laboratoire central de la surveillance de la qualité de l'air (LCSQA) en France a établi des recommandations sur chaque polluant atmosphérique. Les tableaux suivants présentent ces

recommandations pour le dioxyde d'azote et l'ozone, les deux gaz polluants concernés par cette étude.

<b>Dioxyde d'azote (NO<sub>2</sub>)</b>	
Objectif de qualité	moyenne annuelle : 40µg/m <sup>3</sup>
Valeurs limites pour la protection de la santé humaine	moyenne horaire à ne pas dépasser 18h/an : 200µg/m <sup>3</sup>
Seuil d'information et de recommandation	moyenne horaire : 200µg/m <sup>3</sup>
Seuil d'alerte	moyenne horaire pendant 3h consécutives : 400µg/m <sup>3</sup>

**Tableau I- 2 Les seuils et les normes de la qualité de l'air pour le dioxyde d'azote.**

<b>OZONE (O<sub>3</sub>)</b>	
Objectif de qualité pour la protection de la santé humaine	maximum journalier : 120µg/m <sup>3</sup> moyenne sur 8h/an
Valeur cible pour la protection de la santé humaine	maximum journalier de la moyenne sur 8h à ne pas dépasser plus de 25 jours par an (en moyenne sur 3 ans) 120µg/m <sup>3</sup>
Seuil d'information et de recommandation	moyenne horaire : 180µg/m <sup>3</sup>
Seuil d'alerte pour une protection sanitaire pour toute la population	moyenne horaire : 240µg/m <sup>3</sup>
Seuils d'alerte nécessitant la mise en œuvre progressive de mesures d'urgence	1 <sup>er</sup> seuil : 240 µg/m <sup>3</sup> (horaire moyenne) pendant 3h consécutives
	2 <sup>ème</sup> seuil : 300 µg/m <sup>3</sup> (horaire moyenne) pendant 3h consécutives
	3 <sup>ème</sup> seuil : 360 µg/m <sup>3</sup> (horaire moyenne)

**Tableau I- 3 Les seuils et les normes de la qualité de l'air pour l'ozone.**

Ces tableaux montrent l'intérêt des mesures de la concentration de ces gaz avec une bonne résolution temporelle et spatiale, afin de pouvoir alerter la population, mais également de chercher à maîtriser les taux de ces polluants.

### I.6 Les méthodes de référence utilisées

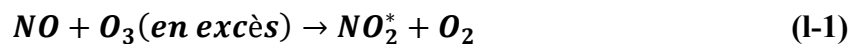
Pour la mesure des polluants, la directive européenne autorise l'utilisation des principes de mesure appelés méthodes de référence reconnues comme moyens de mesure sûrs et précis (Annexe). Le tableau I-4 résume les normes et le principe de mesure de certains polluants. Nous présentons par la suite le principe de fonctionnement des méthodes de référence utilisées pour le dioxyde d'azote et l'ozone.

Polluant	Normes	Type et principe
Dioxyde d'azote et monoxyde d'azote	NF EN 14211	Chimiluminescence
Dioxyde de soufre	NF EN 14212	Fluorescence UV
Ozone	NF EN 14625	Photométrie UV
Monoxyde de carbone	NF EN 14626	Spectroscopie à rayonnement infrarouge non dispersif
Benzène	NF EN 14662-1	Désorption thermique + Chromatographie en phase gazeuse
	NF EN 14662-2	Désorption au solvant + Chromatographie en phase gazeuse

**Tableau I- 4 Méthodes de référence adoptées pour l'évaluation des concentrations des polluants.**

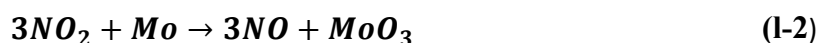
### **I.6.1 Mesure des oxydes d'azote (NO et NO<sub>x</sub>) par chimiluminescence**

La méthode est basée sur une réaction de chimiluminescence entre le NO et l'O<sub>3</sub>, ce dernier doit être en haute concentration [Mathé et Houdret 2009] :



Cette réaction génère des molécules de NO<sub>2</sub><sup>\*</sup> à l'état excité. Celles-ci émettent un rayonnement à une longueur d'onde proche de 1100nm en revenant à l'état fondamental. La concentration de NO est alors proportionnelle à l'intensité de la raie d'émission.

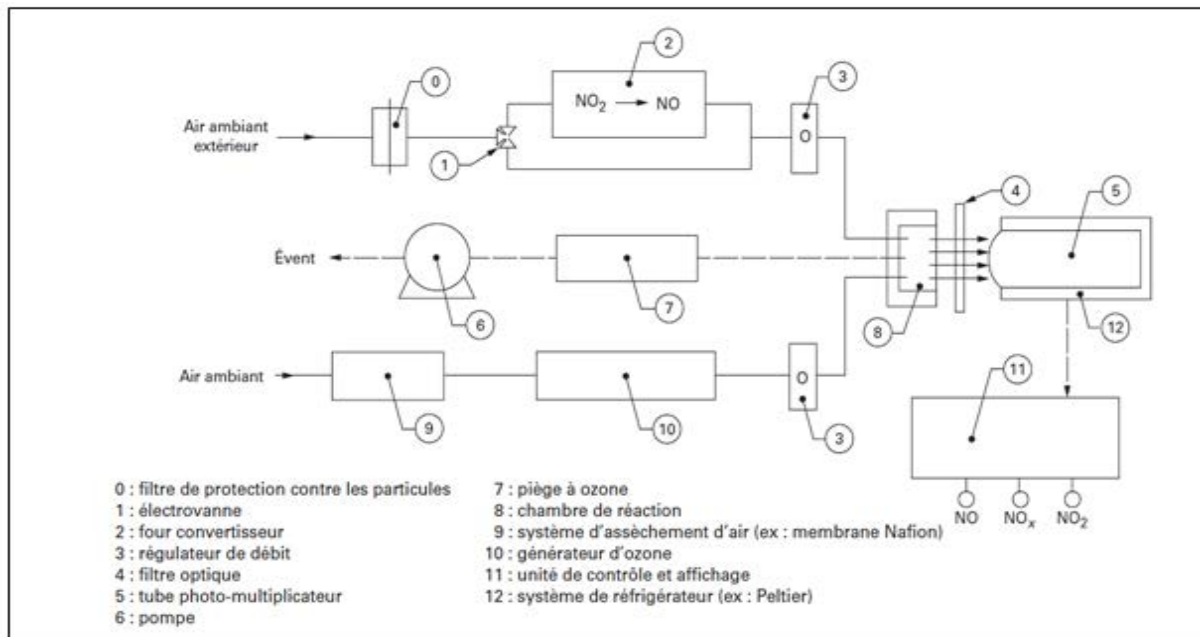
La mesure du NO<sub>2</sub> se fait indirectement. D'abord le NO<sub>2</sub> doit être réduit chimiquement en NO à l'aide de molybdène chauffé (four convertisseur) selon la réaction suivante :



Par la suite, pour obtenir la concentration de NO<sub>2</sub>, on soustrait la concentration de l'échantillon passé directement dans la cellule de mesure, à la concentration de l'échantillon



passé par le four convertisseur (qui est situé en amont de la cellule de mesure), comme l'illustre la figure I-1.



**Figure I-1 Principe de fonctionnement de l'analyseur d'oxydes d'azote [Mathé et Houdret 2009].**

Nous avons utilisé ce type d'analyseur comme dispositif de référence pour valider notre système à base de micro-capteurs de gaz pour la quantification du  $\text{NO}_2$ . Le flux d'air ambiant utilisé étant fixé à 500ml/min.

### I.6.2 Mesure d'ozone par absorption de rayonnement UV

Cette méthode est basée sur le phénomène d'absorption du rayonnement monochromatique par l'ozone qui suit la loi de Beer-Lambert [Daudé 2012]. L'instrument de mesure est doté d'une cellule irradiée par un rayonnement UV monochromatique de longueur d'onde 253.7nm pour assurer une absorption maximale de l'ozone. L'intensité du rayonnement  $I$  à la sortie de la cellule est décrite par l'équation :

$$I = I_0 \exp(-\alpha LC) \quad (\text{I-3})$$

$I_0$  : l'intensité du rayonnement à la sortie de la cellule en l'absence d'ozone ;

$\alpha$  : coefficient d'absorption de l'ozone ;

$C$  : concentration en ozone ;

$L$  : longueur de la cellule.

Le calcul de la concentration en ozone est réalisé à l'aide de l'équation suivante :

$$C = \left( \frac{-10^9}{\alpha L} \ln \left( \frac{I}{I_0} \right) \right) \frac{T p_0}{T_0 p} \quad (I-4)$$

où  $T$  et  $T_0$  sont, respectivement, la température et la température de référence ;  
 $p$  et  $p_0$  sont la pression et la pression de référence respectivement.

### I.6.3 Conclusion

Le principe de fonctionnement de ces méthodes de référence montre leur complexité et par conséquent le coût élevé des analyseurs conventionnels que les agences de surveillance de la qualité de l'air utilisent. De plus, ces appareils certes de grande précision, nécessitent une maintenance régulière et un réétalonnage périodique afin d'assurer la fiabilité des mesures. Ils sont également encombrants et, en général, placés dans des armoires fixes appelées stations de surveillance.

## I.7 Stations de mesure de la pollution

En France, les stations de mesure sont gérées par les associations agréées de surveillance de la qualité de l'air (AASQA) . Les stations de mesure sont implantées de telle sorte qu'elles donnent la meilleure représentativité possible. Selon leur implantation (ville ou campagne), et leurs proximités à des sources de pollution (proche ou lointaine), elles sont classées en deux types : **les stations de fond et les stations de proximité** [Mathé et Malherbe 2017].

### I.7.1 Les stations « de fond »

Elles sont destinées à surveiller l'exposition moyenne des personnes et de l'environnement. Ces stations sont classées en plusieurs types selon les problèmes de pollution atmosphérique abordés et la zone géographique où elles sont implantées. On parle alors de stations urbaines, périurbaines, régionales et nationales. Par exemple, à Metz qui est une ville de taille moyenne, ATMO Grand Est dispose de trois stations de fond et deux stations de proximité dont une station de trafic et une station industrielle.

### I.7.2 Les stations « de proximité »

Elles ont pour objet de mesurer la concentration de certains polluants et le niveau d'exposition de la population qui se trouve à proximité de la source d'émission. On distingue deux types de stations selon la source de pollution : **station industrielle et station trafic**.

Le tableau I-5 résume les différents types de stations, leurs objectifs et les polluants ciblés.

Type de station	Rayon de l'aire de représentation	Polluants	Objectifs
Urbaine	De 100m à 2km	NO <sub>x</sub> , O <sub>3</sub> , SO <sub>2</sub> , PM <sub>10</sub> , COV	Suivi de l'exposition moyenne de la population aux phénomènes de pollution atmosphérique dits de « fond » dans les centres urbains.
Périurbaine	De 1km à 5km	NO <sub>x</sub> , O <sub>3</sub>	Suivi de la pollution photochimique (notamment O <sub>3</sub> et ses précurseurs) et éventuellement les polluants primaires. Suivi du niveau d'exposition moyen de la population aux phénomènes de pollution atmosphérique dits de « fond » à la périphérie du centre urbain.
Régionale rurale	De 5km à 25km	NO <sub>x</sub> , O <sub>3</sub> , NH <sub>3</sub>	Surveillance de l'exposition de la population et des écosystèmes à la pollution atmosphérique de « fond » (notamment photochimique) à l'échelle régionale, participation à la surveillance de la qualité de l'air sur l'ensemble du territoire, notamment dans les zones rurales.
Nationale rurale	De 30km à 500km	NO <sub>x</sub> , O <sub>3</sub>	Surveillance dans les zones rurales de la pollution atmosphérique dite de « fond » issue des transports de masses d'air à longue distance, notamment transfrontaliers.
Industrielle	De 100m à 5km	SO <sub>2</sub> , COV, HAP, NO <sub>x</sub> , HP, O <sub>3</sub> et CO (sont optionnels)	Sources d'informations représentatives du niveau maximum auquel la population riveraine d'une source fixe est susceptible d'être exposée.
Trafic	De 100m à 50m	No <sub>x</sub> , CO, SO <sub>2</sub> , Bb et HAP	Sources d'informations représentatives du niveau maximum auquel la population située en proximité d'une infrastructure routière est susceptible d'être exposée.
Station d'observation spécifique	Selon les besoins	Selon les besoins	Réponse à des besoins spécifiques tels que l'aide à la modélisation, la prévision...

**Tableau I- 5 Types de stations, les polluants ciblés et les objectifs.**

### I.8 Types de mesure et incertitudes

Les mesures de la concentration des polluants sont classées selon les méthodes utilisées, l'incertitude et la période minimale de mesure, aussi bien dans le cas des mesures en continu que dans le cas des campagnes de mesures périodiques. Trois types de mesure sont définis : mesures fixes, mesures indicatives et estimations de concentrations par modèle.

- **Les mesures fixes** sont des mesures prélevées en continu dans des stations de surveillance fixes en utilisant des méthodes de mesure de grande précision. Dans ce cas, **les incertitudes de mesures, réalisées par les analyseurs de référence, ne dépassent pas les 15% pour le dioxyde d'azote et pour l'ozone.**

Mesures fixes			
	SO <sub>2</sub> , NO <sub>2</sub> , NO <sub>x</sub> , CO	PM <sub>10</sub> , PM <sub>2,5</sub> , PB	O <sub>3</sub> , NO, NO <sub>2</sub>
Incertitude	15%	25%	15%
Saisie minimale de données	90%	90%	90% en été 75% en hiver

**Tableau I- 6 Incertitudes des mesures fixes effectuées en continu.**

- **Les mesures indicatives** sont des mesures moins strictes généralement prises par échantillonnage de l'air à des périodes aléatoires réparties uniformément sur l'année. Ces mesures sont réalisées soit dans des camionnettes mobiles équipées, soit dans des laboratoires. **Les incertitudes des mesures indicatives ne doivent pas dépasser les 25% pour le dioxyde d'azote et les 30% pour l'ozone.**

Mesures indicatives			
	SO <sub>2</sub> , NO <sub>2</sub> , NO <sub>x</sub> , CO	PM <sub>10</sub> , PM <sub>2,5</sub> , Pb	O <sub>3</sub> , NO, NO <sub>2</sub>
Incertitude	25%	50%	30%
Saisie minimale de données	90%	90%	90%
Période minimale	14% (une journée par semaine)	14% (une journée par semaine)	> 10% en été

**Tableau I- 7 Incertitudes des mesures indicatives.**

- **L'estimation par modèle** est une technique basée sur des algorithmes théoriques qui modélisent les différents phénomènes afin de simuler la dispersion de la pollution atmosphérique. Elle permet de prédire les concentrations des polluants à l'échelle de la rue ou d'un quartier en se basant sur les données des stations fixes. Ces algorithmes utilisent aussi les caractéristiques géographiques de la zone étudiée (bâtiments, arbres, etc.) et les facteurs climatiques comme la direction du vent ou la température. Le modèle d'estimation est considéré valide si l'incertitude ne dépasse pas 50%.

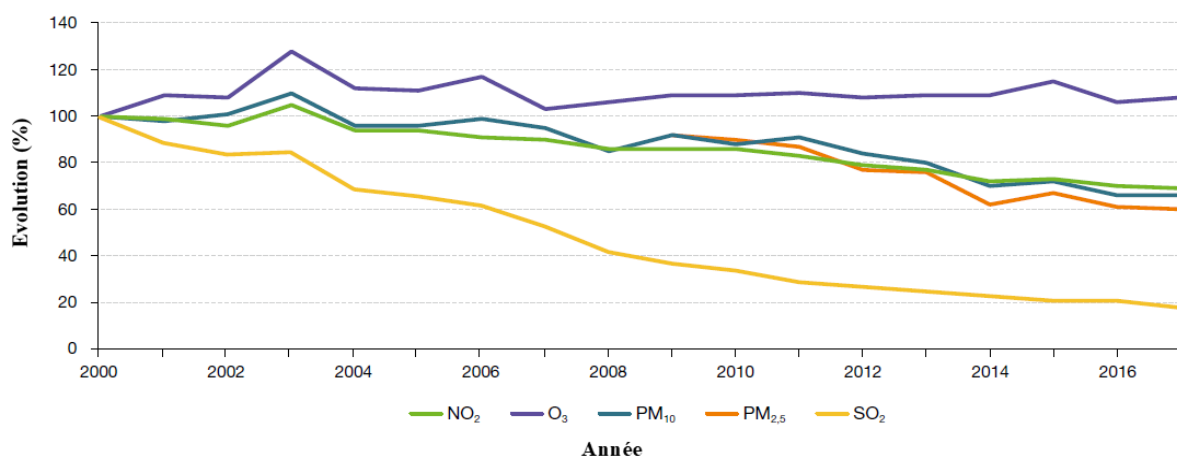
Ces logiciels sont très coûteux, nécessitent des mises à jour régulières, et enfin ne donnent qu'une estimation très globale de la qualité de l'air dans l'espace et dans le temps.

Incertitude du modèle			
	SO <sub>2</sub> , NO <sub>2</sub> , NO <sub>x</sub> , CO	PM <sub>10</sub> , PM <sub>2.5</sub> , PB	O <sub>3</sub> , NO, NO <sub>2</sub>
Incertitude par heure, par 8h ou journalière	50%	-	50%
Incertitude par moyenne annuelle	30%	50%	-

**Tableau I- 8 Incertitude de l'estimation de polluants par modèle.**

### I.9 Evolution de la pollution en France (2000-2017)

En France, les taux de pollution ont nettement baissé durant ces vingt dernières années. Ceci est dû aux nombreuses dispositions prises par les autorités, mais également grâce aux évolutions technologiques (usines propres, voitures moins polluantes). En 2017, la France a enregistré par rapport à l'année 2000, une baisse de 35% d'émission de dioxydes d'azote, de 38% pour les particules PM<sub>10</sub>. Les émissions de dioxyde de soufre (SO<sub>2</sub>) ont baissé de 80%. En revanche, les émissions d'ozone n'ont pas connu de baisse. L'évolution des émissions des différents polluants entre 2000 et 2017 [Le Moullec 2018] est présentée sur la figure I-2.














**Figure I-2 L'évolution de la concentration moyenne des polluants en France [Le Moullec 2018].**

En ce qui concerne les objectifs de la qualité de l'air, d'après le tableau I-9, on constate qu'en 2017 les taux de la plupart des polluants restent dans les normes, à l'exception de PM<sub>10</sub>, Ni et B[a]P dont les dépassements ne touchent que quelques agglomérations. **En revanche 12 à 28 agglomérations en France ont dépassé les normes en ce qui concerne le dioxyde d'azote et l'ozone.**

## Les micro-capteurs de gaz à faible coût

En effet, les données présentées dans ce tableau correspondent aux valeurs moyennes annuelles sur tout le territoire, et par conséquent ne peuvent pas représenter le dépassement d'un seuil en un lieu et un temps donné. Par exemple, le trafic routier qui ne cesse d'augmenter même en milieu urbain et dans des quartiers résidentiels, peut entraîner un dépassement certain des taux de dioxyde d'azote.

Polluants	Principales sources primaires	Respect de la réglementation en 2018	Nombre d'agglomérations en dépassement en 2018
SO <sub>2</sub>		✓	0
NO <sub>2</sub>		✗	12
O <sub>3</sub>	Aucune source primaire	✗	28
PM <sub>10</sub>		✗	3
PM <sub>2.5</sub>		✓	0
CO		✓	0
C <sub>6</sub> H <sub>6</sub>		✓	0
As		✓	0
Cd		✓	0
Ni		✗	2
Pb		✓	0
B[a]P		✗	2

**Tableau I- 9 Dépassements des normes réglementaires de qualité de l'air en France (2017) [Le Moullec 2018].**

Ces remarques, qui sont également généralisables sur les états membres de l'Union Européenne, nous confortent dans l'intérêt de ce travail de thèse portant sur la validation d'une méthode de mesure de ces deux polluants (NO<sub>2</sub>, O<sub>3</sub>) à l'aide de systèmes basés sur les micro-capteurs de gaz à bas coût. Ces dispositifs, du fait de leur facilité d'utilisation, vont permettre de réaliser une cartographie de la pollution autour d'une zone émissive.

### I.10 Les micro-capteurs de gaz à faible coût

Les progrès récents dans le domaine des capteurs, de l'électronique numérique, des méthodes de traitement et d'analyse de données ont conduit à l'émergence des micro-capteurs de gaz pour la surveillance de la pollution atmosphérique [Kumar et al. 2015; Morawska et al. 2018]. Bien que ces capteurs ne puissent pas offrir la même sensibilité, stabilité et précision

que les instruments de référence traditionnels, ils peuvent potentiellement être un complément précieux dans les programmes de surveillance. **L'utilisation de capteurs à faible coût peut apporter une meilleure résolution spatiale et temporelle des cartes de pollution, améliorer les connaissances sur la dynamique de la pollution et mieux identifier les points chauds** [Munir et al. 2019]. Ils seront utiles dans la gestion des interventions d'urgence, de détection des fuites dangereuses et de surveillance de la conformité des sources. Ces micro-capteurs peuvent aussi être utilisés pour accroître la sensibilisation et l'engagement de la communauté à l'égard des problèmes de qualité de l'air.

Une définition claire et précise des micro-capteurs à bas coût de gaz a été donnée dans un rapport établi en 2016 par N. Redon et al. [Redon et al. 2016]. Tout d'abord le mot « bas coût » signifie que le prix d'achat d'un capteur et les équipements nécessaires pour sa mise en œuvre soient beaucoup moins chers que pour les instruments de référence utilisés habituellement. Deuxièmement, le mot « micro » signifie que les systèmes à base de ces capteurs doivent avoir un poids et un volume qui permettent à un individu de les porter sans gêne, soit un poids maximal de 2kg et un volume de 30×30×30cm. L'expression « de gaz », signifie que le capteur doit livrer une réponse pouvant se transformer directement en concentration de gaz cible en ppb (parties par billion), en  $\mu\text{g}/\text{m}^3$  ou en  $\text{mg}/\text{m}^3$ .

La plupart des micro-capteurs à faible coût servant dans la détection de la pollution d'air sont de types électrochimique, résistif (à base d'oxyde métallique), infrarouge ou à photo ionisation.

### **I.10.1 Les capteurs électrochimiques (EC)**

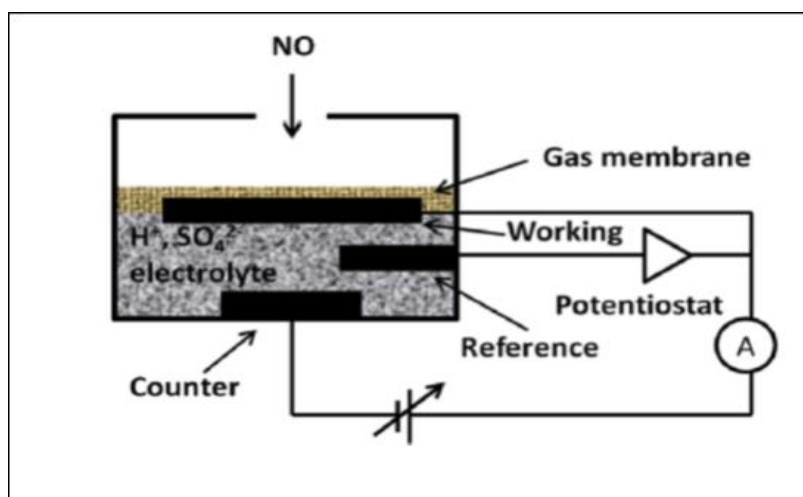
Les capteurs de gaz électrochimiques représentent l'une des technologies les plus anciennes. Ils sont largement utilisés pour les mesures de concentration des gaz [Mead et al. 2013]. Selon le signal de sortie il existe deux types des capteurs électrochimiques. Dans le cas d'un capteur électrochimique ampérométrique on mesure le courant généré par une réaction d'oxydo-réduction, alors que dans le cas des capteurs potentiométriques la mesure est basée sur une différence de potentiel [Baron et Saffell 2017].

Le principe de base de ces capteurs est une réaction électrochimique due au transfert des charges électriques de l'électrode vers la contre-électrode à travers l'électrolyte. Cet électrolyte peut être solide, gel, liquide ou gazeux. Le gaz cible diffuse à travers les membranes du capteur jusqu'à l'électrode de mesure. Il se produit alors un transfert d'électrons entre l'électrode de travail et la contre l'électrode induisant ainsi un courant électrique proportionnel aux concentrations de gaz (figure I-3). Aujourd'hui la majorité des capteurs électrochimiques sont

constitués d'une électrode de mesure, d'une contre-électrode ainsi qu'une électrode de référence supplémentaire. Certains ont une quatrième électrode pour la surveillance des modifications physiques et de la dérive dans l'architecture du capteur.

Les cellules électrochimiques sont des capteurs à large bande, et montrent ainsi une faible sélectivité. Leur sélectivité peut être améliorée pour un gaz cible de plusieurs manières. L'un des moyens le plus utilisé est de cibler une réaction chimique spécifique pour contrôler la sélectivité en ajustant les caractéristiques physiques de la barrière de diffusion à l'aide de molécules particulières. D'autres facteurs ont également un impact sur la sélectivité comme la tension de polarisation entre l'électrode de référence et la contre-électrode, le type d'électrolyte ou encore le matériau de l'électrode de mesure.

Généralement ces capteurs sont compacts, économiques et à faible coût. Leur taille est d'environ 20mm. Le temps de réponse est d'environ 120 secondes et varie en fonction de la température de l'air. Leur consommation électrique est très basse en raison du faible courant électrique généré. Par contre, une amplification du signal est nécessaire pour lire la mesure. Pour un bon fonctionnement, la plupart des capteurs électrochimiques ont besoin d'humidité, puisqu'un très faible taux d'humidité peut endommager certains électrolytes et entraîner un décalage dans les mesures. La température a également une influence sur la réponse de ces capteurs, mais cette interférence peut être modélisée et compensée. La vitesse du vent, en particulier dans les applications à l'air ambiant, peut également influencer sur la diffusion à travers la membrane de ces capteurs, et donc sa réponse. Ces capteurs montrent une stabilité à long terme avec des valeurs de dérive comprises entre 2% et 15% par an. Ils peuvent être étalonnés pour les gaz à mesurer en utilisant une fonction linéaire ou logarithmique.



**Figure I-3 Schéma d'un capteur électrochimique montrant les éléments de base d'une cellule ampérométrique [Popoola et al. 2016].**



### I.10.2 Les capteurs à oxyde métallique (MOS)

Les capteurs à oxyde métallique sont généralement plus petits que les capteurs électrochimiques, avec une taille de quelques millimètres et un poids de quelques grammes. Leur mécanisme de transduction est basé sur la modification des propriétés électriques (la résistance ou la conductivité) d'une fine couche d'oxyde métallique lorsqu'ils sont exposés aux différents gaz (figure I-4) [Wang et al. 2010]. L'oxyde métallique le plus utilisé est l'oxyde d'étain ( $\text{SnO}_2$ ) car il présente une réactivité étendue et de fortes variations de la résistance [Wilson et Baietto 2009]. Les molécules de gaz cible réagissent avec l'oxyde utilisé piégeant les électrons sur la surface sensible, ce qui permet de créer un espace de charge positif constituant une barrière à la conductivité [Liu et al. 2012]. Ces capteurs ont besoin d'une température élevée au niveau de leurs couches sensibles pour que les réactions se déroulent plus rapidement. Une résistance de chauffage est généralement intégrée au capteur à cet effet [Korotcenkov 2007].

Les capteurs à oxyde métallique répondent à une large gamme de concentrations de gaz, qui s'étend de quelques ppb pour des gaz comme le  $\text{NO}_2$  et le  $\text{O}_3$  [Faleh et al. 2016], à plusieurs centaines de ppm pour d'autres gaz. Leur facilité d'utilisation et leur durée de vie de plusieurs années complètent les avantages des capteurs MOS par rapport à d'autres micro-capteurs. A ce jour les caractéristiques de ces capteurs en termes de sélectivité, de seuil de détection et de consommation énergétique ont été largement améliorées. Ce qui permet d'envisager de les utiliser dans des applications plus sensibles telles que la mesure des concentrations de polluants ou en médecine pour dépister certaines maladies pulmonaires.

Les capteurs à oxyde métallique sont généralement compacts, peu coûteux mais leur consommation d'énergie supérieure par rapport à d'autres types de capteurs (de l'ordre de 30mW et de 280mW pour les deux capteurs choisis dans cette étude). Leur temps de réponse est relativement long à cause de la lenteur du processus de désorption des gaz. Le manque de sélectivité de ces capteurs reste le facteur limitant de leur commercialisation et de leur utilisation. Pour améliorer leur sélectivité à un gaz particulier, les fabricants utilisent différents dopants dans la couche d'oxyde métallique (l'élément sensible) ou bien des filtres physiques ou chimiques. La température et l'humidité sont également des interférents importants du signal et doivent être contrôlés ou mesurés avec précision afin de pouvoir corriger le signal de mesure. Un autre obstacle à l'utilisation de ce type de capteurs est leur stabilité dans le temps. En effet, ces capteurs présentent des dérives à court et à long termes, ce qui implique que les systèmes à base de ces capteurs doivent être étalonnés régulièrement.

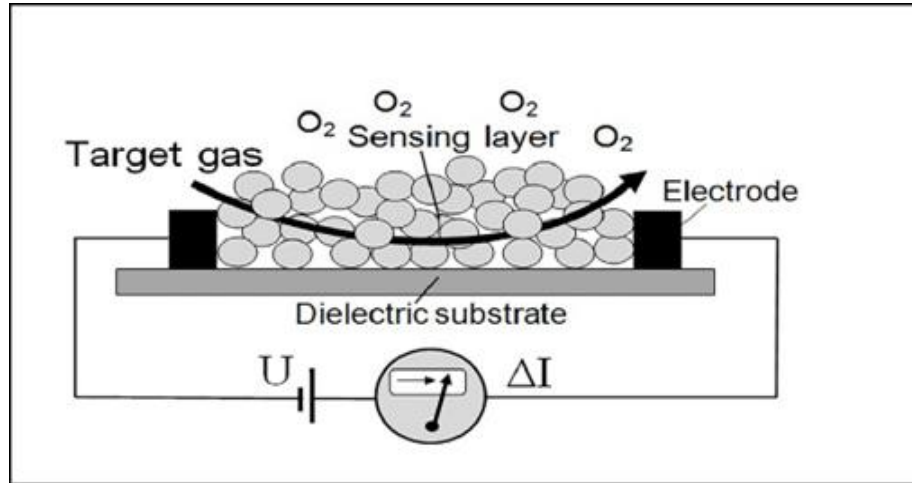


Figure I-4 Schéma de principe de fonctionnement des capteurs MO [Korotcenkov 2007].

### I.10.3 Les capteurs infrarouges

Ces capteurs sont constitués de transducteurs optiques basés sur les spectres d'absorption infrarouge des gaz [Aleixandre et Gerboles 2012]. La lumière infrarouge illumine l'atmosphère à étudier et les molécules présentes dans un mélange de gaz absorbent le rayonnement sur une bande étroite d'adsorption bien déterminée pour chaque type de molécule. L'intensité de cette absorption suit l'équation de Beer-Lambert et elle est proportionnelle à la concentration du gaz cible.

Ces capteurs ont une taille de quelques millimètres et une consommation électrique de quelques centaines de mW. Leur temps de réponse est très court, de l'ordre du dixième de seconde. Ces capteurs sont très sélectifs au CO<sub>2</sub> puisqu'ils ont une bande d'absorption très caractéristique, mais ils souffrent d'un manque de sélectivité dans le cas des hydrocarbures car ils partagent une bande d'absorption similaire. Un autre inconvénient de l'utilisation de ce type de capteurs est que le rapport signal/bruit est très faible.

### I.10.4 Les capteurs à photo-ionisation

Ce type de capteurs est basé sur la photo-ionisation [Aleixandre et Gerboles 2012]. Les molécules du gaz cible sont éclairées par une lumière UV à haute énergie. L'absorption d'un photon par une molécule peut briser cette molécule et générer des ions. Ces derniers sont exposés à un champ électrique externe qui permet de générer un courant électrique. Ce courant électrique est proportionnel à la concentration du gaz cible [Spinelle et al. 2017].

La taille des capteurs à photo-ionisation est généralement d'environ 20x17mm pour un poids d'environ 10g avec une consommation d'énergie de 110mW. Ces capteurs sont rapides, mais leur sélectivité varie selon le polluant à mesurer. La dépendance à la température de ces

capteurs est faible ainsi que leur sensibilité à l'humidité. Cependant, ils nécessitent un étalonnage fréquent. Les fabricants recommandent un recalibrage une fois par mois.

### I.10.5 Les micro-capteurs de gaz commerciaux

Avant de présenter les micro-capteurs de gaz disponibles sur le marché il faut distinguer entre un capteur et un système de capteurs. Ce dernier est un dispositif complet, composé de capteurs avec un boîtier de protection, d'un système d'échantillonnage d'air, d'un module d'alimentation électrique et d'une carte électronique pour l'acquisition et le traitement de données [Karagulian et al. 2019]. Du point de vue de l'utilisateur, les systèmes de capteurs sont prêts à l'utilisation et peuvent être déployés même par des non-spécialistes qui souhaitent surveiller la qualité de l'air dans un environnement choisi. Par contre, les utilisateurs de capteurs doivent ajouter des composants matériels et des logiciels de traitement de données pour leur mise en œuvre.

Un certain nombre de micro-capteurs et de systèmes à base de micro-capteurs commerciaux existants sont résumés dans les tableaux I-10 et I-11. Ces tableaux illustrent également les types de polluants cibles et le prix indicatif de chacun d'eux.

modèle	polluant cible	type	prix en euros
MiCS-2714	NO <sub>2</sub>	MOS	10
NO2-B43F	NO <sub>2</sub>	électrochimique	50
NO2-A1	NO <sub>2</sub>	électrochimique	50
MiCS-4514	CO, NO <sub>2</sub>	MOS	20
GG5 7530	NO <sub>2</sub>	MOS	20
NO2-3E50	NO <sub>2</sub>	électrochimique	100
FIS SP-61	O <sub>3</sub>	MOS	50
OX3-B4	O <sub>3</sub>	électrochimique	55
O3-3E1F	O <sub>3</sub>	électrochimique	500
SM50	O <sub>3</sub>	MOS	500
TGS-5042	CO	MOS	40
CO-B4	CO	électrochimique	50
CO 3E300	CO	électrochimique	100

**Tableau I- 10 Micro-capteurs de gaz commerciaux [Karagulian et al. 2019].**

## Surveillance de la pollution atmosphérique avec des micro-capteurs à faible coût

modèle	polluants	type	prix en euros
AirNut	PM2.5	néphélométrie	150
MAS	CO, NO2, O3, PM2.5	électrochimique, UV, néphélométrie	5500
PA-II	PM1, MP2.5, PM10	optique	200
Egg v1	CO, NO2, O3	MOS	200
SENS-IT	O3, CO, NO2	MOS	2200
S-500	O3, NO2	MOS	500
CairClip O3/NO2	PM1	électrochimique	600
Portable AS-LUNG	PM1, MP2.5, PM10	optique	1000
AirCube	NO2, O3, NO	électrochimique	3538
AirSensEUR (v.2)	NO2, O3, CO, NO	électrochimique	1600
Air Quality Station	PM1, PM2.5	optique	2000
AQY v0.5	PM2.5, NO2, O3	optique, MOS, électrochimique	3000
Vaisala AQT410 v.1.15	CO, NO2	électrochimique	3700
2B Tech. (POM)	O3	UV	4500
AQMesh v.4.0	NO2, CO, NO, O3	électrochimique	10000

**Tableau I- 11** Systèmes de capteurs existants [Karagulian et al. 2019].

### **I.11 Surveillance de la pollution atmosphérique avec des micro-capteurs à faible coût**

Ce paragraphe présente les travaux récents qui portent sur la surveillance de la pollution atmosphérique à partir des micro-capteurs de gaz. Nous avons choisi un ensemble de travaux

qui mettent en évidence les nombreux aspects à prendre en compte lors de l'exploitation de ces capteurs.

- **Castell et al. 2017**

Les auteurs ont mené une évaluation exhaustive de 24 systèmes identiques de type AQMESH, basés sur des capteurs électrochimiques du fabricant Alphasense (U.K), pour la mesure des principaux polluants gazeux (CO, NO, NO<sub>2</sub>, O<sub>3</sub>). Une évaluation est réalisée par rapport aux analyseurs de référence. Durant les tests, les systèmes AQMESH sont soumis à une variation importante des conditions environnementales. Les auteurs ont conclu que la performance de ces systèmes varie à la fois dans le temps et dans l'espace, dans la mesure où elle dépend de la composition de l'atmosphère et des conditions météorologiques. Leurs résultats montrent également que les performances de ces systèmes varient d'une unité à l'autre, impliquant la nécessité d'examiner la qualité des données de chaque système avant son utilisation. Les auteurs signalent qu'il est nécessaire de bien définir la manière de tester ces systèmes afin de s'assurer qu'ils répondent aux performances souhaitées avant de les commercialiser.

Dans cette étude plusieurs tests sont menés afin d'évaluer si le système peut être utilisé pour des applications nécessitant : 1) une grande précision en accord avec les objectifs de qualité des données définis dans la législation sur la qualité de l'air, ou 2) une précision adéquate à des fins tels que la sensibilisation de la population. Les résultats de ces tests montrent que les plateformes AQMESH présentent une faible précision par rapport aux objectifs de qualité de l'air fixés par les directives européennes, mais qu'elles peuvent néanmoins fournir des informations indicatives sur la qualité de l'air observée.

- **Aleixandre et Gerboles 2012**

Cet article examine plus de 100 systèmes à base de capteurs à faible coût de différentes technologies (MOS, électrochimiques, infrarouges, capteurs à photo-ionisation) en suivant les protocoles de tests et de validation exigés par la directive européenne 2008 sur la qualité de l'air. Dans cette étude les auteurs concluent que les capteurs résistifs MOS ont généralement la sensibilité la plus élevée, mais avec une reproductibilité et une stabilité médiocre. Les capteurs électrochimiques ont la deuxième meilleure sensibilité avec une meilleure stabilité. Les capteurs infrarouges sont les moins sensibles, mais avec une bonne stabilité. Cependant ils ne sont disponibles que pour certains gaz (par exemple CO<sub>2</sub>). Les capteurs à photo-ionisation sont sélectifs avec une bonne sensibilité et reproductibilité, mais nécessitent une maintenance

importante, et ils ne sont disponibles que pour un nombre restreint de polluants (par exemple les COV).

En ce qui concerne les polluants NO<sub>2</sub> et O<sub>3</sub>, la plupart des capteurs à oxyde métallique et électrochimiques ont une bonne sensibilité, mais seuls quelques-uns peuvent mesurer des concentrations de l'ordre du ppb. Dans cette catégorie, les auteurs citent les capteurs électrochimiques Alphasense NO<sub>2</sub>-D4, -A1 et -B1, et certains capteurs de Citytech.

- **Peterson et al. 2017**

Cet article propose une étude en laboratoire sur les capteurs à oxyde métallique de type MiCS-4514, MiCS-2614, MiCS-5914, et SGX de Sensortech, pour la mesure des concentrations de NO<sub>2</sub> et d'O<sub>3</sub>. Les auteurs ont utilisé des modèles analytiques linéaires multiples permettant d'accroître la précision à long terme en incluant dans le modèle tous les facteurs environnementaux. Des tests étendus sur 18 mois sont faits sur des données collectées à partir d'un mois de la période d'étalonnage. Ces méthodes permettent de mesurer les polluants avec un intervalle de confiance de 95% et une précision de 20µg/m<sup>3</sup> pour le dioxyde d'azote et de 26,8µg/m<sup>3</sup> pour l'ozone. Au-delà de quatre mois après l'étalonnage, la dérive des capteurs devient significative et la précision est considérablement réduite.

Les variations de fabrication dans les caractéristiques de capteurs individuels ont été également examinées car l'instrument utilisé est équipé de paires de capteurs du même type. Les résultats montrent une bonne répétabilité avec un indice de corrélation entre les éléments de même type supérieur à 0.7. Ce qui est quand même faible.

Les auteurs recommandent l'utilisation des ventilateurs pour favoriser la circulation d'air vers l'élément sensible du capteur. La comparaison entre un montage avec un flux passif et un montage avec un flux dynamique par aspiration montre clairement l'amélioration de la sensibilité des capteurs ventilés.

A la fin de cette étude, des pistes possibles sont suggérées pour améliorer la fiabilité de cette technologie de capteurs pour la mesure de la pollution en milieu urbain.

- **Brienza et al. 2015**

Les auteurs ont élaboré un système de détection à faible coût (uSense) pour la surveillance de la qualité de l'air dans les zones urbaines. Il contient des capteurs à oxyde métallique de type TGS2442, MiCS-2714 et MiCS-2614. Ce système permet aux utilisateurs de surveiller le niveau de la pollution de l'air à l'intérieur de leurs propriétés, en installant simplement une ou plusieurs unités. Le système proposé suit une approche de détection

coopérative où les données collectées par chaque unité sont téléchargées sur un serveur distant et partagées entre les utilisateurs du système, de sorte que chacun puisse obtenir des informations plus complètes et fiables sur la pollution dans leur zone d'habitation. Les données collectées par uSense peuvent être consultées de différentes manières, via une interface Web, une application mobile ou des services Web.

Le système uSense a été testé dans trois zones urbaines différentes (Toscane, Italie). Les mesures expérimentales, comparées aux valeurs fournies par l'agence de la qualité de l'air locale, montrent que même si les mesures obtenues à partir de capteurs à faible coût ne sont pas aussi précises que les données officielles, elles peuvent néanmoins fournir des indications utiles sur la qualité de l'air à un endroit donné. La qualité des données fournies par uSense dépend en grande partie de la précision de ses capteurs de gaz. À mesure que la technologie évoluera, des capteurs de gaz plus précis seront rendus disponibles et la qualité des données collectées s'améliorera en conséquence.

- **Penza et al. 2012**

Penza et al. ont réalisé une expérience combinée en laboratoire et sur le terrain en utilisant des capteurs commerciaux, des capteurs électrochimiques Alphasense, des capteurs à oxyde métallique Figaro et des nanotubes de carbone modifiés ENEA IT. Les capteurs sont étalonnés en laboratoire et validés sur le terrain par comparaison avec un instrument de référence. Les auteurs ont conclu que ces capteurs à faible coût sont très prometteurs pour le développement d'un système de capteurs portable pour la surveillance en temps réel de la pollution atmosphérique. Les auteurs insistent également sur la nécessité de la validation et de l'étalonnage sur terrain pour une campagne expérimentale à long terme lors du développement futur de ce type de systèmes de mesure.

- **Jiao et al. 2016**

Les travaux publiés dans cet article se situent dans le cadre du projet de réseau de capteurs de la qualité de l'air communautaire (CAIRSENSE) aux Etats Unis. Un réseau de capteurs de pollution atmosphérique à faible coût a été déployé dans une zone de 2km<sup>2</sup> de surface. Dans cette étude plusieurs systèmes de capteurs disponibles sur le marché {AQMesh G.3 (AQMesh Corp, UK), Air Quality Egg (Wicked Device, USA), Cairclip (Cairpol, France) SM50(Aeroqual, New Zealand), Aercet 831(MetOne, USA), PMS-SYS-1 (Shinyei, Japan), Dylos Particle Counter (USA), Airbeam (HabitatMap, USA)} ont été utilisés pour mesurer les concentrations d'oxyde d'azote, d'ozone, de monoxyde de carbone, de dioxyde de soufre et de

particules fines. De multiples capteurs identiques d'O<sub>3</sub>, de NO<sub>2</sub> et de CO ont révélé une corrélation très variable avec les instruments de référence, avec un coefficient de corrélation allant respectivement de 0.39 à 0.97, de -0.25 à 0.76 et de -0.40 à 0.82. Une régression linéaire multiple par étapes a été effectuée pour sélectionner les capteurs les plus pertinents afin d'améliorer les performances. L'amélioration la plus importante en accord avec les instruments de référence est obtenue en intégrant dans la méthode de traitement de données tous les paramètres environnementaux. Cette amélioration a été observée pour la quantification du NO<sub>2</sub>, avec un coefficient de corrélation qui augmente de 0.57 à 0.81. Par contre, la quantification d'autres gaz n'a montré aucune amélioration apparente.

Cette étude montre aussi qu'un réseau de capteurs a pu être utilisé avec succès pendant 8 mois pour la surveillance de l'ozone avec deux capteurs Cairclip et SM50, et la surveillance des particules avec des capteurs PMS-SYS-1, Dylos Particle Counter, Airbeam, et Aercet 831.

- **Williams et al. 2013**

Williams et al. ont élaboré un nouvel instrument économique à base d'un capteur à oxyde métallique. Sa couche conductrice est de type WO<sub>3</sub> capable de mesurer avec précision les concentrations ambiantes d'ozone. Une procédure d'étalonnage en deux étapes est présentée, dans laquelle une transformation non linéaire convertit la résistance du capteur en un signal proportionnel à la concentration d'O<sub>3</sub>. Ensuite, une régression linéaire est utilisée pour l'étalonnage avec un instrument de référence. Les systèmes de contrôle du débit d'air et de la température du capteur, sont logés avec le capteur dans un ensemble compact, simple à échanger. Cet instrument peut fonctionner à l'énergie solaire et utilise la technologie du téléphone cellulaire pour permettre la surveillance dans des endroits éloignés.

Les données de tests sur le terrain ont montré que la précision et la stabilité de l'instrument sur des périodes de plusieurs mois sont de l'ordre de quelques parties par billion (ppb). Les auteurs ont aussi montré que les modes de défaillance courants peuvent être détectés grâce à l'enregistrement et l'étude des signaux disponibles à partir de l'instrument. Finalement, les auteurs ont conclu que la combinaison d'une stabilité à long terme, d'un autodiagnostic et d'une réparation simple et peu coûteuse signifie que le coût de fonctionnement et d'étalonnage de ces instruments est considérablement réduit par rapport à l'instrumentation de référence traditionnelle.



- **Mead et al. 2013**

Le but de cette étude est de montrer que les capteurs de gaz électrochimiques du fabricant Alphasense, généralement utilisés pour la détection des mélanges de gaz à des concentrations de l'ordre de la partie par million (ppm), peuvent être utilisés à l'échelle du ppb, s'ils sont configurés et utilisés de manière appropriée. Les auteurs ont effectué des expériences en laboratoire et sur le terrain à l'aide des capteurs électrochimiques de type NO<sub>2</sub>-A1, NO-A1 et CO-AF56 de ce fabricant pour mesurer respectivement la concentration de NO<sub>2</sub>, NO et CO.

Plusieurs remarques intéressantes ont découlé de cette étude. Les auteurs ont montré que les changements de la température ambiante et de l'humidité relative affectent la sensibilité et la ligne de base (réponse du capteur à de l'air pur) des capteurs. Pour les mesures à l'échelle du ppb, la correction de la ligne de base des capteurs pour les effets de température ou d'humidité relative est plus importante à prendre en compte que la correction de la sensibilité (gain). Les auteurs proposent donc une méthodologie de post-traitement des données permettant d'éliminer les effets de la température et de l'humidité relative et de corriger l'interférence de l'ozone sur la quantification du NO<sub>2</sub>. Les réponses des capteurs sont nettement améliorées et deviennent plus linéaires par rapport à la concentration des gaz cibles à l'échelle du ppb après le post-traitement proposé. Certains résultats montrent également que les capteurs peuvent être utilisés sans atténuation significative du gain sur de longues périodes. Ainsi pendant onze mois de mesures, ils indiquent une déviation de  $\pm 13\%$ .

- **Piedrahita et al. 2014**

Les auteurs ont développé un système appelé « M-Pods » conçu avec des capteurs commerciaux à oxyde métallique SGX de Sensortech pour les mesures de CO, O<sub>3</sub>, NO<sub>2</sub> et des COV et un capteur infrarouge pour le CO<sub>2</sub>. L'étude combinée, réalisée en laboratoire et sur le terrain a révélé qu'une meilleure estimation de la quantité de polluants est obtenue si l'étalonnage du système est réalisé sur le terrain. Par contre, dans le cas d'un étalonnage en laboratoire, les estimations sur le terrain sont nettement moins bonnes. Lors de l'étalonnage sur le terrain, la précision était de 4.0 à 6.1 ppb pour l'O<sub>3</sub> et de 6.4 à 8.4 ppb pour le NO<sub>2</sub>. Cependant, le rapport signal/bruit était élevé. Au cours d'une période d'essai de quatre semaines, des dérives de -2.6 à 2.0 ppb par jour pour l'O<sub>3</sub> et de -1.56 à 0.51 ppb par jour pour le NO<sub>2</sub> sont observées.

Les auteurs confirment que ce système permet d'accéder à la surveillance de la qualité de l'air pour un public plus large de scientifiques et de citoyens. Ils signalent qu'un système d'étalonnage en laboratoire peut coûter des milliers, voire des dizaines de milliers de dollars, alors qu'un système d'étalonnage en collocation avec des systèmes de références ne nécessite

qu'un boîtier convenable pour les instruments. La collocation et l'étalonnage collaboratifs seront un outil précieux dans la prochaine génération de surveillance de la qualité de l'air. Les systèmes de micro-capteurs couplés aux dispositifs des agences de surveillance, et aidés par des scientifiques permettront de construire des cartes détaillées de la pollution.

- **Lösch et al. 2008**

Ces travaux de recherche présentent une étude en laboratoire concernant la mesure de l'ozone à l'échelle du ppb en utilisant un capteur de gaz à oxyde métallique commercial de type MiCS-2611, SGX SENSORTECH. Les auteurs ont utilisé la modulation de température de la couche sensible du capteur comme facteur permettant d'améliorer les performances de ce capteur. En effet, la modification de la température de chauffe permet d'extraire des paramètres pertinents car les caractéristiques du capteur changent de manière remarquable aux différentes températures utilisées. Les résultats montrent qu'il est ainsi possible de prédire des faibles concentrations d'ozone.

- **Spinelle et al. 2015; Spinelle et al. 2017**

Dans ce travail, les performances de plusieurs méthodes d'étalonnage sur le terrain pour les capteurs à faible coût sont comparées. Cette étude utilise un réseau de capteurs à oxyde métallique et de capteurs électrochimiques pour le dioxyde d'azote, l'ozone et le monoxyde de carbone, ainsi que des capteurs infrarouges miniaturisés pour le dioxyde de carbone. L'étalonnage a été effectué au cours des deux premières semaines d'évaluation à l'aide des instruments de référence sur un site semi-rural. Les données de validation sont collectées sur cinq mois. Les capteurs de CO peuvent atteindre l'objectif de qualité des données avec 25% d'incertitude, définie par la directive européenne sur la qualité de l'air pour des méthodes indicatives. En revanche l'incertitude prévue par cette directive pour le NO<sub>2</sub> n'était pas atteinte. Ceci est probablement dû au faible taux de NO<sub>2</sub> associé à un niveau élevé d'O<sub>3</sub> expliqué par la localisation de l'étude sur un site semi-rural. Pour l'estimation de la concentration d'un gaz à partir des données des capteurs, les résultats de cette étude montrent que les techniques d'apprentissage supervisées sont plus performantes par rapport aux méthodes de régression linéaire ou multilinéaire.

- **Lerner et al. 2015**

Ces travaux portent sur l'évaluation de l'effet du vent sur les capteurs lors des mesures en laboratoire, dans une soufflerie et sur le terrain avec des capteurs montés sur des voitures. Les capteurs utilisés sont des capteurs résistifs à oxyde métallique, disponibles dans le

commerce sous le nom Elm du fabricant PerkinElmer (USA). Les auteurs concluent qu'il existe une influence évidente de la vitesse du véhicule et du vent sur les mesures des concentrations de NO<sub>2</sub> et d'O<sub>3</sub>. Il est donc important de prendre en compte les variations du débit d'air, en envisageant soit une conception qui élimine cette influence, soit une correction des signaux de réponse.

- **Moltchanov et al. 2015**

Moltchanov et al. ont mené une étude sur le terrain à l'aide de six unités de capteurs commerciaux d'Airbase Systems (ISR), Aeroqual LTD, NZL et IAQ-100 AppliedSensors pour mesurer des concentrations de NO<sub>2</sub>, O<sub>3</sub>, COV, CO<sub>2</sub>, et des particules en suspension. Ces systèmes mesurent aussi le niveau de bruit, la température et l'humidité relative. Les capteurs sont testés en parallèle avec des instruments de référence dans trois sites urbains de Haïfa en Israël pendant une période de 71 jours. Les données des capteurs sont fournies toutes les 20 secondes pour le NO<sub>2</sub> et 60 secondes pour l'O<sub>3</sub>. Une bonne corrélation a été mise en évidence entre des valeurs moyennes sur 30 minutes avec un coefficient de détermination R<sup>2</sup> compris entre 0.92 et 0.99 pour l'O<sub>3</sub> et entre 0.78 et 0.98 pour le NO<sub>2</sub>.

Les auteurs utilisent un schéma de rotation intéressant dans lequel trois capteurs sont fixes et trois autres en rotation entre les sites, afin de fournir une certaine inter comparabilité et un contrôle minutieux de la qualité des données. Les auteurs soulignent la nécessité d'un étalonnage fréquent sur le terrain pour maintenir les performances de tels systèmes.

- **De Vito et al. 2018**

Ce travail évalue et compare les performances de cinq méthodes d'apprentissage statistique dans un ensemble de scénarios d'échantillonnage continu de l'air. Les tests ont été réalisés en utilisant trois jeux de données différents couvrant une grande variété de conditions et conçus pour remettre en cause les limites technologiques des micro-capteurs chimiques de gaz, pour finalement déterminer leurs performances absolues. Ces jeux de données ont été générés à l'aide de dispositifs basés sur différentes technologies de capteurs (Sensor Network for Air Quality SNAQ développé à l'Université de Cambridge), et exposés à des polluants intérieurs ou extérieurs typiques, à des mesures non contrôlées avec des interférents inconnus ou connus.

Les données sont recueillies et moyennées à différentes résolutions temporelles. Les résultats obtenus ont montré que les approches d'apprentissage automatique dynamique qui reposent également sur des réponses instantanées des capteurs dépassent leurs homologues

statiques. En ce qui concernent les techniques d'apprentissage, il a été démontré que la méthode *support vector machine* (SVM) donne les meilleures performances dans la plupart des scénarios, quels que soient le nombre d'échantillons et la technologie du capteur. Bien qu'ils ne présentent généralement pas les meilleures performances dans de nombreux cas, les réseaux de neurones peu profonds (NN) fournissent des modèles les plus adaptés en termes de temps de calcul et de mémoire nécessaire pour l'implémentation de l'algorithme sur un microcontrôleur. Les auteurs ont aussi conclu que les techniques non linéaires ont montré un avantage significatif en termes de performance par rapport à la régression linéaire classique en utilisant moins d'échantillons d'apprentissage.

### I.12 Conclusion

Bien que les émissions des polluants aient nettement baissé, le dioxyde d'azote, l'ozone et les particules fines sont toujours en dépassement dans plusieurs agglomérations en France. C'est pourquoi il est nécessaire de compléter et de densifier les réseaux de surveillance de la qualité de l'air en utilisant des instruments basés sur des micro-capteurs de gaz à faible coût pour aboutir à une meilleure résolution aussi bien spatiale que temporelle.

Les travaux réalisés à ce jour montrent le potentiel de ces capteurs pour la surveillance de la pollution atmosphérique. Cependant, de nombreux facteurs, difficiles à contrôler, entraînent une baisse de précision des mesures en les rendant inadéquats à de telles utilisations où les concentrations de gaz et leur évolution sont très faibles. Parmi ces facteurs on note les dérives propres des capteurs, leur manque plus ou moins important de sélectivité, de même que l'influence des paramètres environnementaux tels que la température, l'humidité relative, la vitesse du vent et la présence d'autres gaz. Ceci implique la nécessité d'un travail de recherche important au niveau de la conception du système de mesure, mais également au niveau du traitement de données et des méthodes d'étalonnage ainsi qu'au niveau des procédures de contrôle du vieillissement des capteurs.

Les travaux de recherche développés dans ce mémoire se situent dans ce contexte. La première étape a pour but de sélectionner les capteurs les plus prometteurs en se basant sur les travaux de recherche présentés dans ce chapitre, suivis de la caractérisation et de tests de ces capteurs en laboratoire et sur site. Cette démarche nous permet de mettre en évidence leurs performances et leurs faiblesses. Dans la seconde étape nous utilisons des méthodes de traitement des données et plus particulièrement des méthodes d'apprentissage statistique afin

## **Chapitre I : la pollution environnementale et les méthodes de mesure**

---

de proposer des solutions aux problèmes d'analyse des données liés à l'utilisation des micro-capteurs de gaz appliqués au domaine de la surveillance de la qualité de l'air.

## **Chapitre II : Protocole d'évaluation des micro-capteurs de gaz**

### II.1 Introduction

Les caractéristiques de micro-capteurs de gaz ont bénéficié d'énormes améliorations de sorte qu'aujourd'hui il est possible d'envisager de les utiliser dans des domaines plus pointus comme la détection et l'estimation de la concentration des polluants atmosphériques. La directive européenne autorise l'utilisation sans restriction de ces micro-capteurs à faible coût pour les mesures indicatives dans les zones où le polluant à mesurer ne dépasse généralement pas le seuil limite [Spinelle et al. 2013]. Par ailleurs, dans les zones où le seuil limite est souvent dépassé, ces capteurs peuvent être un complément aux mesures de référence afin de réduire ces dernières à 50%. Cependant la directive européenne ne spécifie aucune méthode indicative en particulier, mais exige **de démontrer que les dispositifs à base de micro-capteurs peuvent répondre à un objectif de qualité basé sur une incertitude de mesure inférieure au double de l'incertitude qui caractérise les méthodes de référence.**

Dans ce cadre, un protocole d'évaluation et de tests a été proposé en 2013 par le Centre Commun de Recherche Européen (C.C.R) pour l'évaluation des micro-capteurs en tant que méthodes de mesures indicatives [Spinelle et al. 2013]. La mise en œuvre d'un protocole d'évaluation et de tests en laboratoire conformes aux recommandations du C.C.R étant au-delà de la portée de notre projet, nous avons suivi une démarche d'évaluation allégée inspirée de ce protocole, afin de les adapter aux moyens et aux dispositifs présents au sein de notre équipe. Dans cette étude, notre objectif est de montrer la capacité des micro-capteurs de gaz choisis pour évaluer la concentration de dioxyde d'azote et d'ozone avec une incertitude de mesure de 25% et 30% respectivement, fixée dans le cas des mesures indicatives par la directive européenne.

Dans la première partie de ce chapitre nous résumons le protocole d'évaluation des systèmes à base de micro-capteurs de gaz, spécifié par le C.C.R. Dans la deuxième partie, nous présentons notre démarche d'évaluation en commençant par une caractérisation en laboratoire des capteurs choisis suivie par des tests et des évaluations sur le terrain.

### II.2 Protocole d'évaluation et d'étalonnage des micro-capteurs recommandé par Centre Commun de Recherche Européen

Dans le cadre de l'évaluation et de l'étalonnage des micro-capteurs pour la prise de mesures indicatives, la démarche globale du protocole proposé par le C.C.R recommande que les capteurs soient à la fois étalonnés en laboratoire et testés sur le terrain. Si l'incertitude

estimée à partir des tests sur le terrain est conforme aux critères de qualité des données de la directive européenne, le capteur est alors accepté pour des mesures indicatives. Sinon, il est rejeté et des expériences supplémentaires en laboratoire doivent être effectuées afin de modifier le modèle de calcul et de réduire les incertitudes. S'il n'est toujours pas possible de réduire l'incertitude à l'aide d'un modèle plus étendu, le capteur peut toutefois être testé sur le terrain. Au cas où l'incertitude du nouveau modèle sur le terrain est conforme à l'objectif de qualité des données, le capteur est utilisable pour des mesures indicatives mais seulement sur le site où l'étalonnage et le test ont été menés. En effet, les micro-capteurs de gaz sont trop sensibles aux conditions météorologiques et aux interférents gazeux qui varient d'un site à l'autre.

La procédure complète de test et d'évaluation proposée par le C.C.R est présentée par l'organigramme de la figure II-1.

### **II.2.1 Définition du module « mesure »**

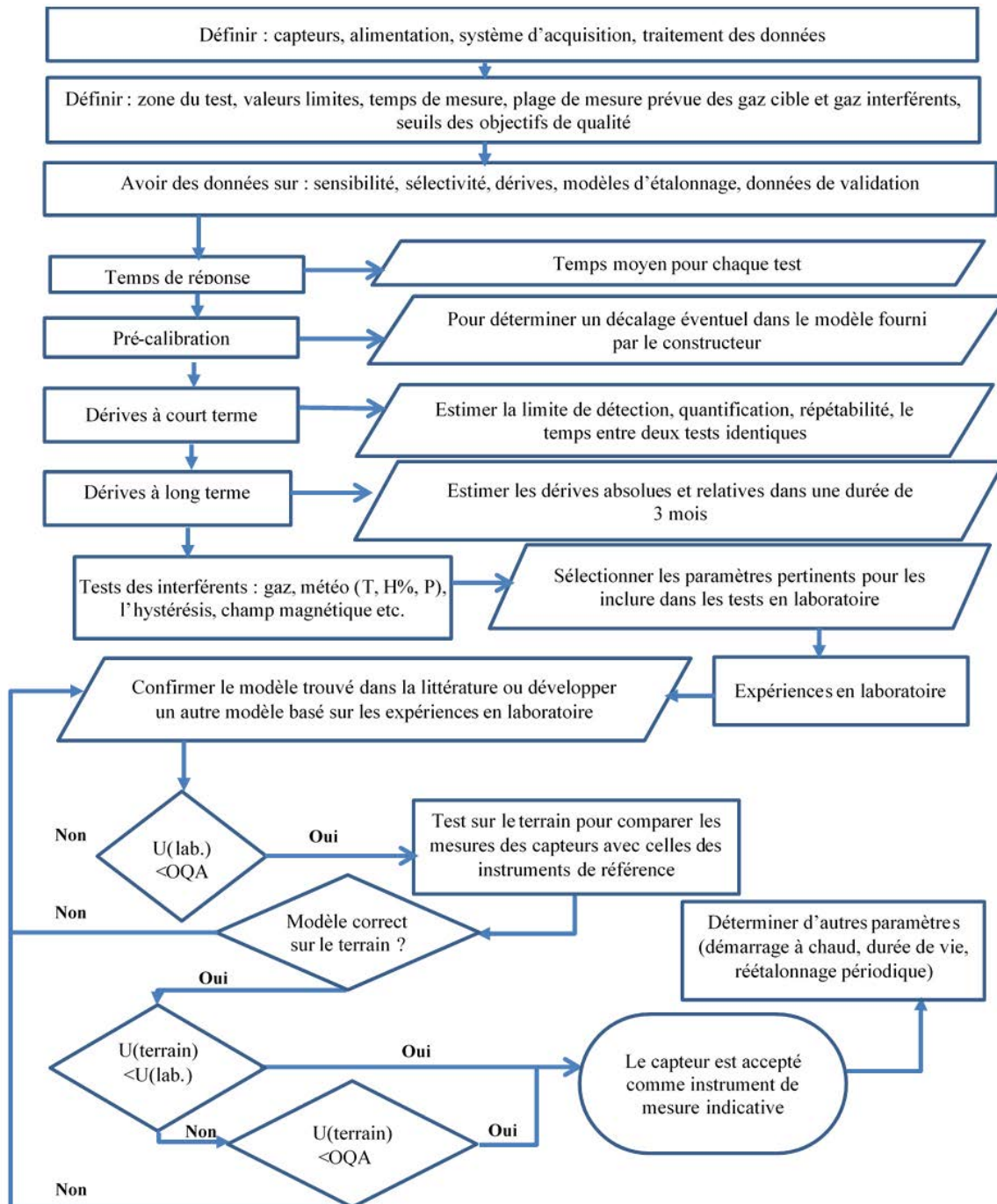
Avant de réaliser l'unité de prise de mesure, il faut tout d'abord définir les points suivants :

- La cellule de mesure, qui est le module contenant les capteurs, le circuit de conditionnement associé, ainsi que la partie acquisition des données.
- Le fabricant et le fournisseur des capteurs.
- Le modèle et le numéro de série de chaque capteur.
- L'alimentation et la carte d'acquisition des données.
- Les détails sur le boîtier de protection et sur le support des capteurs ainsi que le matériau utilisé pour leur construction.
- Il faut également fournir toutes les informations de sorte que l'utilisateur puisse répéter l'expérience dans les mêmes conditions. Par exemple, dans le cas de l'utilisation des capteurs à oxyde métallique, il faut spécifier la température de l'élément chauffant et préciser si la température de chauffage est constante ou si le capteur est chauffé selon un cycle de températures.
- Le microenvironnement (zones urbaines, rurales, industrielles ou de trafics, etc.), dans lequel il est prévu d'utiliser le système, doit être défini. Ce qui permet d'estimer la plage attendue de polluants à surveiller et les interférences gazeuses probables ainsi que la température moyenne et l'humidité relative de l'air.



## Chapitre II : Protocole d'évaluation des micro-capteurs de gaz

- Une étude bibliographique sur les capteurs sera également utile afin d'avoir au préalable des informations sur leur sensibilité, leur sélectivité, leur temps de réponse, et même sur leur modèle d'équation d'étalonnage s'il existe.



**Figure II- 1 Organigramme de la procédure d'évaluation des micro-capteurs pour les mesures indicatives selon le C.C.R (U: incertitude, DQO : objectif qualité des données) [Spinelle et al. 2013].**

## **II.2.2 Les mesures en laboratoire**

Les recommandations adoptées dans le protocole préconisé par le C.C.R pour les mesures en laboratoire concernent la procédure de génération de mélanges gazeux, les caractéristiques de la chambre d'exposition et les méthodes de calcul des caractéristiques du capteur telles que le temps de réponse, la répétabilité, les dérives à court et à long terme. Comme nous avons précisé précédemment, ces recommandations étant très complexes et trop coûteuses à réaliser dans notre laboratoire, nous les avons adaptés tout en cherchant à satisfaire et à conserver la rigueur demandée. Nous verrons, par la suite, dans la partie spécifiquement consacrée au dispositif expérimental, la démarche adoptée dans notre cas.

## **II.2.3 Chambre d'exposition**

La chambre d'exposition pour les tests des capteurs doit avoir une capacité suffisante pour accueillir simultanément le nombre de capteurs prévu. Elle doit être fabriquée en utilisant de la matière inerte telle que le verre, le polytétrafluoroéthylène (PTFE) ou l'acier inoxydable. Elle doit être équipée d'un système de génération d'atmosphère gazeuse pouvant mesurer, contrôler et modifier le débit d'air, la concentration gazeuse dans la chambre, ainsi que la température et l'humidité relative et tous les autres interférents. La mesure de mélanges de gaz se fait en utilisant un instrument de référence puisque les gaz risquent d'être absorbés sur la paroi de la chambre d'exposition ou bien d'autres réactions peuvent avoir lieu. Cela peut modifier de manière significative les niveaux de concentration réelle auxquels les capteurs sont exposés.

Dans le cadre de cette étude, la chambre de mesure contenant les capteurs a été réalisée en acier inoxydable, inerte aux gaz.

## **II.2.4 Génération de mélange de gaz**

La procédure pour générer un mélange de gaz doit être conforme à la norme ISO-6145. Les concentrations des gaz doivent être vérifiées par des méthodes de référence. La dilution doit être effectuée avec de l'air synthétique ou de l'air filtré afin d'éliminer les interférents gazeux.

Dans notre cas, les concentrations de gaz générées n'ont pas été vérifiées par des méthodes de référence mais la rampe de gaz que nous avons utilisée pour la génération des mélanges gazeux est équipée de contrôleurs de débit massique à haute précision ( $\pm 0.1\%$ ).

### **II.2.5 Temps de réponse**

Selon le protocole préconisé par le C.C.R, le temps de réponse des capteurs est défini par le temps nécessaire pour atteindre 80% de la valeur stable finale. Le temps initial est le moment de l'injection du gaz. Le temps de réponse permet d'évaluer la durée nécessaire pour tous les tests du protocole. Il permet également de vérifier si le capteur parvient à atteindre la stabilité dans un certain délai, puisque ce temps de réponse doit être inférieur à un quart du temps requis pour l'établissement des mesures. Par exemple, la société ATMO Grand Est fournit les mesures de concentrations avec une périodicité de 15 minutes.

Généralement trois répétitions d'introduction (montée) et d'arrêt (descente) de gaz sont suffisantes pour calculer le temps de réponse sachant qu'il est fastidieux d'attendre une réponse stable du capteur à chaque mesure. Nous notons ici que les capteurs ayant un temps de réponse long peuvent être utilisés pour la surveillance dans les zones où les changements de niveaux de polluants gazeux sont lents. Ceci est le cas des stations rurales fixes.

### **II.2.6 Pré-calibration**

Pour certains capteurs, on dispose d'un modèle de transformation des réponses de capteurs en concentration de gaz fourni par le constructeur et établi dans des conditions environnementales de température et d'humidité relative optimales. Dans de tels cas, l'étape de pré-calibration permet de s'assurer que l'utilisation de l'équation de transformation est justifiée sans entraîner de décalage avec la valeur de la concentration réelle.

Dans le cas où la relation entre les réponses des capteurs et la concentration du gaz polluant n'est pas connue, un modèle d'étalonnage doit être établi par une régression entre les réponses du capteur et les données de référence. Au moins trois concentrations de gaz doivent être utilisées pour cette étape de pré-calibration, y compris la valeur limite. Si la fonction d'étalonnage n'est pas linéaire il faut utiliser plusieurs concentrations avec un ordre aléatoire (par exemple 80, 40, 0, 60, 20, 95% de la valeur limite) afin de considérer également l'hystérésis éventuel du capteur.

### **II.2.7 Dérives des capteurs**

Un facteur important dans l'utilisation des micro-capteurs est la dérive de leur réponse, et en particulier celle des capteurs à oxyde métallique. D'une façon générale, la dérive représente la variation de la réponse du capteur obtenue pour des conditions identiques. Cette dérive peut être de deux types : la dérive à court terme et la dérive à long terme. La notion de stabilité à court terme est utilisée pour définir la durée maximale entre des tests similaires et

contribue à la détermination de l'incertitude globale du capteur. Tandis que la stabilité à long terme est utilisée pour définir la périodicité du réétalonnage. Ces phénomènes de dérives peuvent être observés dans plusieurs cas :

- lorsque le capteur a été exposé à un mélange gazeux durant une certaine période;
- lorsque les conditions environnementales, en particulier la température et l'humidité relative, varient.

De manière générale, en fonction de la nature des gaz, de la durée d'exposition et des conditions d'environnement, la dérive des micro-capteurs de gaz peut être plus ou moins importante que ce soit à court ou à long terme. La dérive des micro-capteurs de gaz provoque à la fois un décalage de leur ligne de base et une modification de leur sensibilité. En général, trois caractéristiques sont définies en rapport avec la dérive des capteurs de gaz : la répétabilité, les dérives à court terme et les dérives à long terme.

- **Répétabilité**

Les tests de répétabilité, tel qu'ils sont définis par le protocole du C.C.R, permettent d'estimer non seulement le facteur « répétabilité du capteur », mais aussi des caractéristiques telles que la limite de détection et la limite de quantification, pour un gaz donné.

L'écart type de ces mesures  $s_r$  est calculé comme suit :

$$s_r = \sqrt{\frac{\sum_{i=1}^N (R_i - \bar{R})^2}{N - 1}} \quad \text{(II-1)}$$

où  $R_i$  dénote des réponses du capteur à 0, 50 et 80% de la valeur limite. Chaque concentration devant être répétée trois fois de manière consécutive,  $\bar{R}$  est la valeur moyenne de la réponse du capteur et  $N$  est le nombre de mesures ( $N = 3$  concentrations\*3 répétitions).

Les mesures doivent être effectuées à une température et une humidité constante avec un écart type relatif inférieur à 2%. Le capteur doit être exposé uniquement au gaz cible sans interférent.

Selon les recommandations du C.C.R, la répétabilité est alors définie par  $2\sqrt{2}s_r$ . La limite de détection est estimée à  $3s_r$  et la limite de quantification à  $10s_r$ .

- **Dérives à court terme**

Pour estimer les dérives à court terme, il est nécessaire d'effectuer des mesures à trois concentrations différentes : 0, 50% et 80% de la valeur limite. Ces mesures sont à réaliser chaque jour pendant trois jours consécutifs. Au cours de l'expérience, il est important de maîtriser tous les paramètres susceptibles d'influencer la réponse du capteur (température, humidité et autres facteurs influents possibles) avec un écart type relatif d'environ 2%. Il faut également que l'ordre des concentrations soit croissant, ou décroissant pour éviter les effets d'hystérésis.

La stabilité moyenne à court terme se calcule à partir de la différence entre les réponses des capteurs à la même concentration avec un délai de 24 heures sur les trois jours :

$$D_{ss} = \frac{\sum_1^{N-3} |R_{s,avant} - R_{s,après}|}{N - 3} \quad (\text{II-2})$$

$R_{s,avant}$  et  $R_{s,après}$  étant les réponses du capteur à 0, 50% et 80% de la valeur limite à  $t_0$  (avant 24 heures) et à  $t_{24h}$  (après 24 heures) respectivement.

- **Dérives à long terme**

Les dérives à long terme sont estimées de la même manière que les dérives à court terme mais sur une durée plus longue soit de 3 mois au lieu de 3 jours, et en réalisant des mesures tous les 15 jours mais sans répétition. Trois mesures sont effectuées à 0, 50% et 80% de la valeur limite. Ainsi six mesures pour chacune des trois concentrations sont nécessaires et on utilise la même formule que dans le cas des dérives à court terme.

### II.2.8 Facteurs d'influence (interférents)

Des tests sur l'influence de chaque interférent sont nécessaires pour les inclure dans le modèle d'étalonnage. Chaque interférent doit être estimé séparément. Par exemple, lorsqu'on fait des tests de température, les autres facteurs doivent rester constants et aucun gaz interférent ne doit être présent dans le mélange de gaz. Ces tests incluent l'influence de la température, de l'humidité, de la vitesse du vent, de la pression, de l'effet hystérésis et des gaz interférents.

Les tests de facteurs d'influence étant très lourds, coûteux et chronophages, nous n'avons pas réalisé ces tests dans le cadre de cette étude, de plus nous avons estimé que notre étalonnage sur le terrain remplacerait favorablement ces tests en laboratoire.

### II.3 Etalonnage et validation en laboratoire

Pour la phase d'étalonnage et de validation en laboratoire, les mesures doivent être effectuées à trois températures différentes, sous trois taux d'humidité différents et des interférents jugés significatifs, à condition qu'il soit possible de les mesurer et de corriger leurs effets pendant les essais sur le terrain. Si aucun modèle physique n'est disponible, un modèle de régression doit être adopté pour compenser l'influence des variables testées.

### II.4 Validation sur le terrain

Les tests sur le terrain sont conçus pour identifier d'autres sources d'influence qui ne sont pas couvertes de manière adéquate par les tests en laboratoire, puisqu'elles proviennent de l'utilisation du capteur en environnement réel. Un test d'au moins trois mois doit être effectué sur un site de test où le capteur est prévu d'être exploité. La durée des tests doit couvrir des périodes représentatives afin de considérer toutes les variations des conditions environnementales extrêmes et des épisodes de pollution.

Un accès aux mesures de référence est indispensable pour établir une relation linéaire entre la réponse du capteur étalonné et les mesures de référence à l'aide d'une régression linéaire de type :

$$y_i = a + bz_i \quad (\text{II-3})$$

avec :

$y_i$  : la concentration obtenue par le modèle d'étalonnage,

$z_i$  : la concentration obtenue par un instrument de référence,

$a$  : le décalage à zéro (biais de la régression),

$b$  : la pente de la régression.

La régression linéaire se calcule en utilisant la méthode des moindres carrés. Le calcul de  $a$  et  $b$  se fait selon les équations suivantes :

$$S_{ZZ} = \sum_{i=1}^N (Z_i - \bar{Z})^2 \quad (\text{II-4})$$

$$S_{YY} = \sum_{i=1}^N (Y_i - \bar{Y})^2 \quad (\text{II-5})$$

$$S_{YZ} = \sum_{i=1}^N (Z_i - \bar{Z})(Y_i - \bar{Y}) \quad (\text{II-6})$$

$$\text{où} \quad \bar{Z} = \frac{\sum_{i=1}^N Z_i}{N} \text{ et } \bar{Y} = \frac{\sum_{i=1}^N Y_i}{N} \quad (\text{II-7})$$

$$b = \frac{S_{YY} - S_{ZZ} + \sqrt{(S_{YY} - S_{ZZ})^2 + 4S_{YZ}^2}}{2S_{YZ}} \quad (\text{II-8})$$

$$a = \bar{Y} - b\bar{Z} \quad (\text{II-9})$$

L'erreur quadratique moyenne (RMSE), définie par l'équation II-10 est généralement utilisée pour quantifier l'erreur entre la mesure du capteur et la référence :

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Z_i - Y_i)^2}{N}} \quad (\text{II-10})$$

où  $N$  est le nombre de mesures.

### II.5 Conclusion

Nous venons de présenter en détail les recommandations du C.C.R pour la caractérisation des micro-capteurs appliqués à la mesure des polluants atmosphériques. Des protocoles d'expérimentation très précis, ainsi que des modèles de calcul pour la détermination des caractéristiques des capteurs sont fournis. Concernant la validation en laboratoire, les expériences exigées sont trop lourdes et coûteuses à réaliser dans le cadre de notre thèse. En effet, selon les recommandations, il serait nécessaire de reproduire en laboratoire une atmosphère quasi réelle en générant des modifications aussi bien gazeuses que climatiques.

Dans le cadre de ce travail, pour la caractérisation de nos capteurs en laboratoire, nous avons étudié la sensibilité, la répétabilité, la limite de détection, ainsi que la dérive à court terme des capteurs utilisés.

Nous verrons dans le paragraphe consacré au système expérimental et à l'étalonnage en laboratoire quelles sont les précautions adoptées dans notre cas. L'essentiel de notre travail a porté sur la caractérisation et la validation sur site réel où notre dispositif a été en fonctionnement durant environ un an.

### II.6 Caractérisation et test des micro-capteurs

Les moyens et les équipements de notre laboratoire ne permettent pas de réaliser une évaluation complète des capteurs à faible coût en respectant le protocole proposé par le C.C.R dans sa globalité. Par exemple, les tests qui concernent la quantification de l'impact des interférents gazeux sur la qualité de la mesure reposent sur l'utilisation des équipements

spécialisés pour générer les gaz interférents et des instruments de référence pour leur quantification de manière exacte.

Pour cette raison, les tests en laboratoire que nous avons effectués portent uniquement sur la caractérisation des capteurs avec notre principal gaz cible, soit le dioxyde d'azote. L'évaluation sur le terrain est effectuée dans l'une des stations de mesure d'ATMO Grands Est située sur un site « trafic » où les différents instruments de référence sont disponibles et où les facteurs environnementaux sont mesurables et/ou contrôlables.

### II.7 Sélection de micro-capteurs de gaz

D'après les études bibliographiques que nous avons menées, les capteurs électrochimiques et les capteurs à oxyde métallique sont les plus adaptés pour la surveillance des gaz polluants comme le dioxyde d'azote et l'ozone.

La figure II-2 résume une étude comparative de la limite de détection pour différentes familles de capteurs à faible coût et concerne les concentrations des polluants typiques trouvés dans l'union européenne (Concentration in EU). Les polluants sont mentionnés en rectangle bleu sur la figure et les valeurs limites des polluants spécifiées dans la directive européenne (EU specified limit) en traits rouges. Le graphique montre que les capteurs électrochimiques (EC sensors, mentionnés par des croix vertes) et les oxydes métalliques (MOS mentionnés par des triangles noirs) sont adéquats pour mesurer la concentration d'ozone et de dioxyde d'azote dans l'air ambiant [Rai et al. 2017]. On peut noter que les capteurs optiques sont surtout utilisés pour la détection des particules fines.

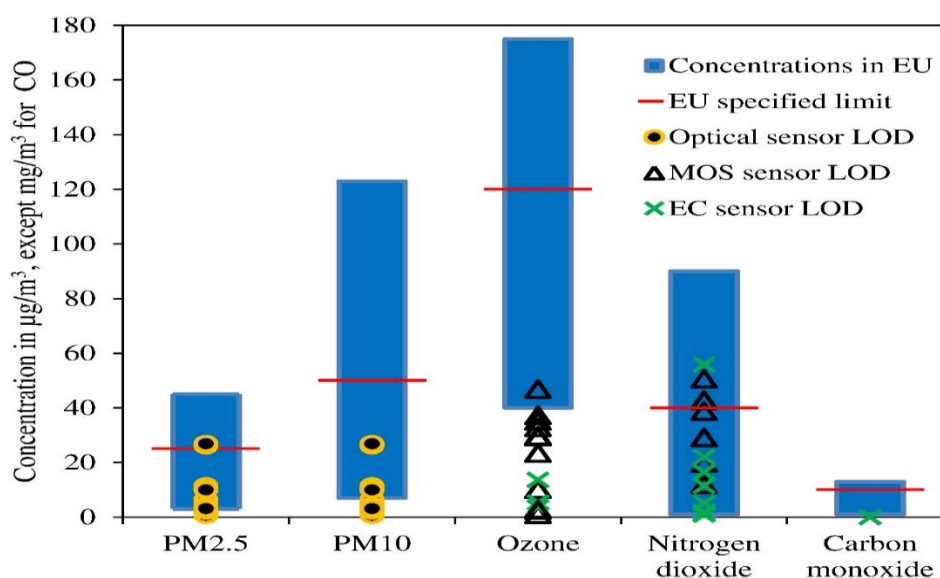


Figure II- 2 Evaluation de différents types de capteurs à faible coût pour différents polluants [Rai et al. 2017].



Parmi les micro-capteurs électrochimiques commercialement disponibles, notre choix s'est porté sur les capteurs électrochimiques NO<sub>2</sub>-B43F (NO<sub>2</sub>), OX-B431 (O<sub>3</sub>) fabriqués par le constructeur Britannique Alphasense en raison de leurs caractéristiques et de leur fiabilité. En ce qui concerne les capteurs à oxyde métallique, nous avons choisi le capteur MiCS-2714 (NO<sub>2</sub>) de la société Suisse SGX SENSORTECH et le capteur GGS 7530 (NO<sub>2</sub>+O<sub>3</sub>) de la société Allemande UST Umweltsensortechnik.



**Figure II- 3 Capteurs choisis pour la mesure de dioxyde d'azote et d'ozone.**

Le choix de ces capteurs se justifie au regard des résultats de la littérature et de leurs caractéristiques en termes de précision et gamme de concentration. Dans ce travail, nous nous focalisons plus particulièrement sur la valorisation d'un nez électronique en utilisant des algorithmes de modélisation.

### II.8 Mesures en laboratoire

En laboratoire, nous avons réalisé la caractérisation du capteur électrochimique NO<sub>2</sub>-B43F, et des deux capteurs à oxyde métallique MiCS-2714, GGS 7530 destinés à mesurer le dioxyde d'azote. En effet, le capteur d'ozone (OX-B431) a été rajouté à notre réseau de capteurs suite à des observations sur les mesures de terrain montrant une interaction entre le dioxyde d'azote et l'ozone, entraînant la modification de la concentration de NO<sub>2</sub>.

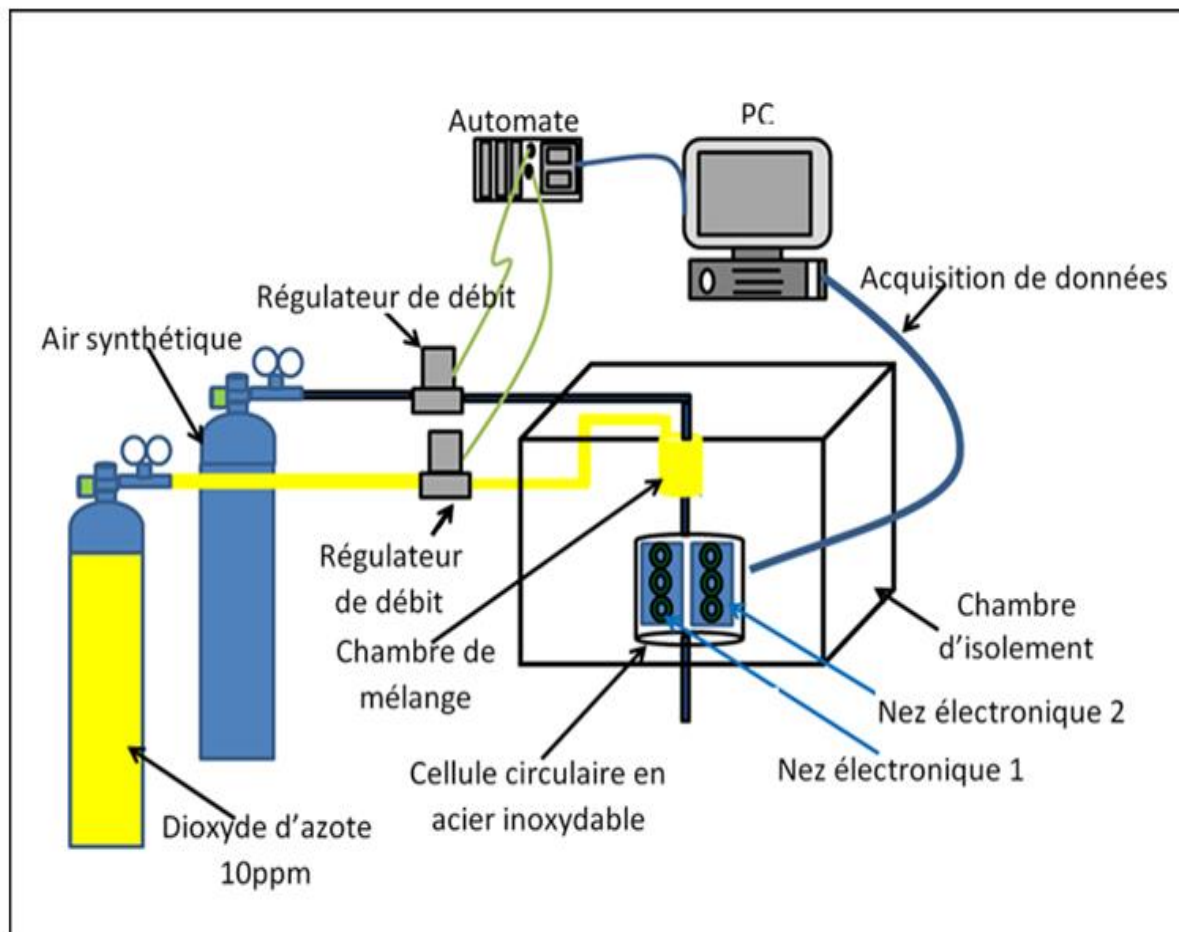
Nous avons utilisé ces capteurs en deux exemplaires afin d'étudier par la suite des techniques de transfert d'étalonnage entre deux dispositifs identiques, présentées dans le chapitre IV. Les capteurs électrochimiques ont leurs propres cellules de diffusion tandis que les

capteurs MOS sont placés dans une cellule circulaire en acier inoxydable étudiée et usinée à cet effet.

Les mesures en laboratoire consistent à exposer les capteurs au dioxyde d'azote à des faibles concentrations, de l'ordre de la partie par billion (ppb) en accord avec les concentrations trouvées dans les atmosphères polluées. On dispose d'une bouteille de NO<sub>2</sub> à 10ppm diluée dans de l'azote et une bouteille d'air synthétique. Les différentes concentrations de NO<sub>2</sub> sont obtenues en diluant le dioxyde d'azote de la bouteille avec de l'air synthétique à l'aide d'une rampe de gaz conçue dans notre laboratoire et adaptée pour nos mesures [Ahmadou 2015]. Ce système de génération de gaz contient deux régulateurs de débit massiques pour contrôler le débit à la sortie des bouteilles d'air synthétique et de dioxyde d'azote. Le flux total du gaz généré a été fixé à 400ml/min. Le régulateur de débit relié à la bouteille de NO<sub>2</sub> est calibré à un débit maximal de 10ml/min, pour permettre d'obtenir avec une bonne précision les dilutions souhaitées. En faisant varier le pourcentage du flux du dioxyde d'azote par rapport au flux total, nous pouvons générer les concentrations souhaitées. Ainsi, pour obtenir 250ppb de NO<sub>2</sub>, il faut mélanger un flux de 10ml/min de dioxyde d'azote avec un flux de 390ml/min d'air synthétique (équation II-11). Nous notons ici, que dans le cadre des mesures en laboratoire, nous utilisons l'unité ppb pour définir les concentrations de NO<sub>2</sub>, contrairement à l'unité utilisée habituellement pour les concentrations de polluants gazeux qui est le µg/m<sup>3</sup>. La conversion de ppb en µg/m<sup>3</sup> se fait à l'aide d'un facteur de 1.88 pour le NO<sub>2</sub>.

L'acquisition des données de caractérisation se fait via deux cartes d'acquisition pour pouvoir recueillir tous les signaux (DAQ usb 6009, National instruments). La gestion de la commande des mesures de caractérisation est réalisée à l'aide d'un automate programmable et d'une interface utilisateur que nous avons développée sous Matlab. Cette interface permet d'effectuer des mesures automatiques, répétitives et précises. Tous les paramètres de mesure tels que les débits de gaz, la fréquence d'échantillonnage des signaux, les temps d'exposition ou de régénération des capteurs ainsi que la création de mélanges gazeux sont configurables. Cette interface permet également d'observer l'évolution de la mesure et les signaux des capteurs en temps réel.

$$\text{concentration générée (ppb)} = \frac{\text{débit NO}_2 * \text{concentration de la bouteille NO}_2 * 1000}{\text{débit total (400ml/min)}} \quad (\text{II-11})$$



**Figure II- 4 Schéma du dispositif pour les expérimentations en laboratoire.**

### II.8.1 Temps de réponse

Pour l'évaluation du temps de réponse des capteurs, nous les avons tout d'abord soumis à une concentration fixe de 250ppb de NO<sub>2</sub> pendant une durée de 20min, afin d'assurer le remplissage complet de la chambre de mesure et un temps de réaction « capteur-gaz » suffisant pour atteindre une réponse stable de tous les capteurs.

D'après les figures II-5a) à II-7a), les temps de réponse des capteurs utilisés sont tous inférieurs à 5 minutes ( $t = 0$  correspond à l'introduction du gaz). Les deux capteurs électrochimiques NO<sub>2</sub>-B43F mettent un temps de 83 secondes pour atteindre 80% de la réponse stable à 250ppb. Leur réponse en descente (après l'arrêt du flux de NO<sub>2</sub>) est aussi rapide, avec un temps d'environ 38 secondes (figures II-5b) à II-7b). Nous observons le même comportement pour les deux capteurs MiCS-2714 où les temps de réponse en montée et en descente sont de 62s et 71s respectivement. Les deux capteurs GGS 7530 ont le temps de réponse le plus long, avec une réponse en montée de 188s et une réponse en descente de 443s.

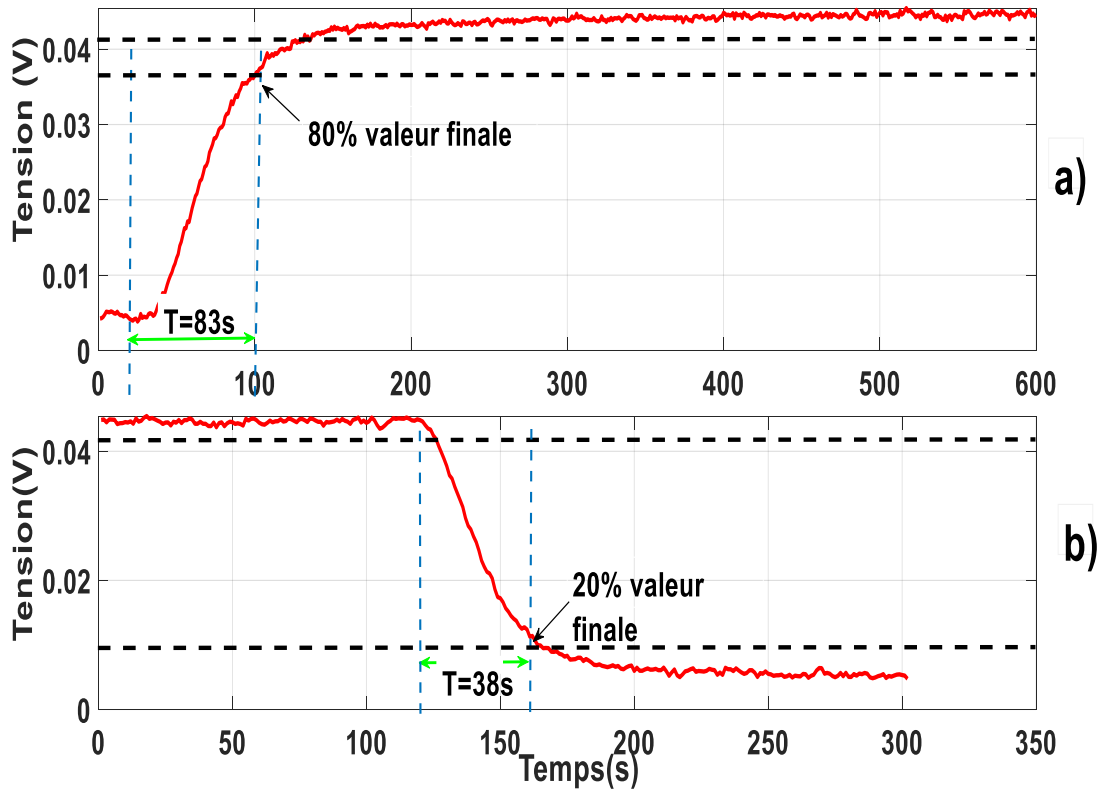


Figure II- 5 Etude du temps de réponse pour le capteur Alphasense NO2-B43F  
 a: montée (introduction du gaz) ; b : descente (introduction d'air pur).

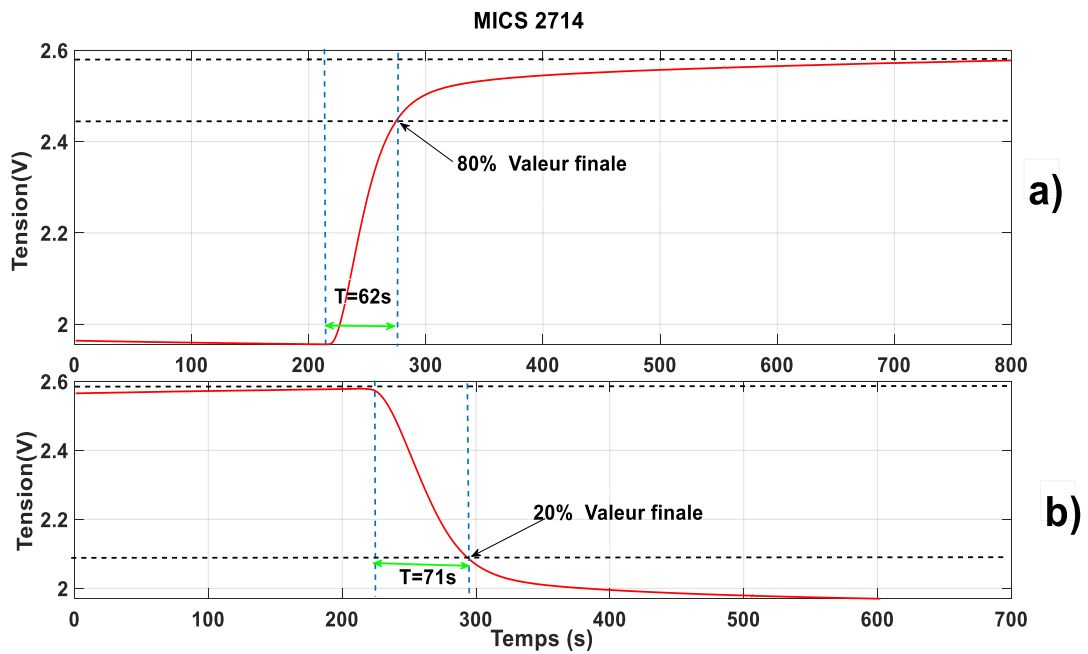
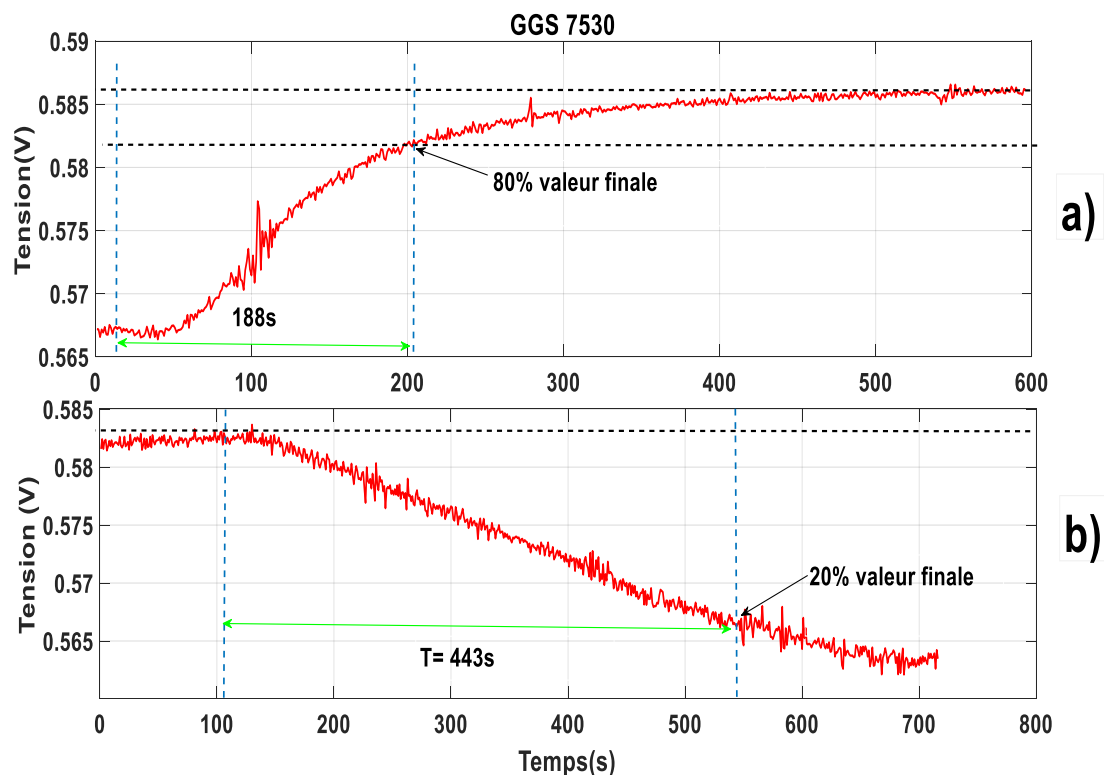


Figure II- 6 Temps de réponse pour le capteur MiCS-2714  
 a : montée (introduction du gaz) ; b : descente (introduction d'air pur).



**Figure II- 7 Temps de réponse pour le capteur GGS 7530  
a : montée (introduction du gaz) ; b : descente (introduction d'air pur).**

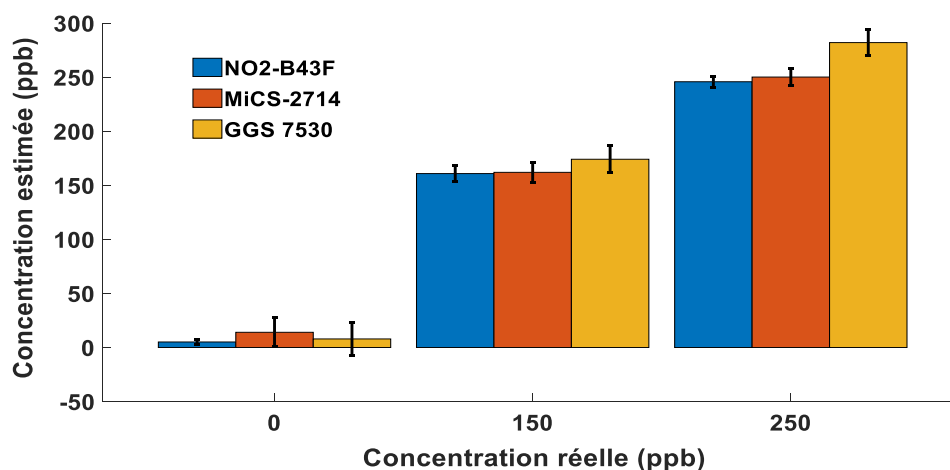
### II.8.2 Répétabilité

Pour l'étude de la répétabilité, nous utilisons le protocole fixé par le C.C.R. Nous avons ainsi réalisé trois mesures à 0, à 150 et 250ppb, et nous répétons ce cycle trois fois consécutives pour évaluer la répétabilité, la limite de détection et la limite de quantification.

Le tableau II-1 résume les résultats obtenus pour chacun des trois types de capteurs. D'après ces résultats, on peut conclure que le capteur Alphasense NO2-B43F a une meilleure répétabilité que les deux capteurs MOS. Ceci est illustré également par la figure II-8. C'est un résultat attendu, du fait que d'après la littérature la dérive à court terme des capteurs électrochimiques est beaucoup moins importante que celle des capteurs MOS.

	NO2-B43F	GGs 7530	MiCS-2714
Répétabilité (ppb)	5.32	8.17	7.04
Limite de détection (ppb)	5.65	8.69	7.48
Limite de quantification (ppb)	18.86	28.96	24.96

**Tableau II- 1 Répétabilité des capteurs pour les mesures en laboratoire calculée à partir des sensibilités obtenues par un modèle linéaire d'étalonnage.**



**Figure II- 8 Répétabilité des trois capteurs pour les trois mesures à 0, 150 et 250ppb.**

### II.8.3 Dérives à court terme

Pour calculer les dérives à court terme nous avons réalisé trois mesures à 0, 150 et 250ppb. Ces mesures ont été prises une fois par jour pendant 3 jours, selon les recommandations du C.C.R.

Le tableau II-2 résume les dérives à court terme pour les trois capteurs utilisés. Les dérives obtenues sont calculées selon la formule préconisée par le C.C.R. Nous notons, conformément à la littérature, une dérive plus faible pour le capteur électrochimique.

	NO2-B43F	GGs 7530	MiCS-2714
Dérives à court terme (ppb)	10.27	20.61	15.34

**Tableau II- 2 Dérives à court terme obtenues par un modèle linéaire d'étalonnage.**

## II.9 Étalonnage en laboratoire

D'après le C.C.R, l'étalonnage en laboratoire consiste à établir un modèle linéaire entre les concentrations de dioxyde d'azote et les réponses de chaque capteur individuellement. Il est important de noter que ces modèles ne sont valables que dans des conditions de laboratoire car ils sont établis à partir des données de capteurs exposés uniquement à un seul gaz, soit le NO<sub>2</sub> dans notre cas.

### II.9.1 Étalonnage du capteur NO2-B43F

Le constructeur Alphasense fournit le modèle suivant pour le calcul de la concentration de NO<sub>2</sub> :

$$[NO_2] = [(WE - WE_0) - (AE - AE_0)]/S_T \quad \text{(II-12)}$$

où :

$WE$  est la tension fournie par l'électrode de travail ;

$AE$  est la tension fournie par l'électrode auxiliaire ;

$S_T$  est la sensibilité du capteur ;

$WE_0$  et  $AE_0$  désignent respectivement, les tensions de l'électrode de travail et celle de l'électrode auxiliaire exposées à l'air pur.

L'utilisation de l'équation fournie par le constructeur et des valeurs fournies de  $WE_0$ ,  $AE_0$  et  $S_T$  ne donnent pas des résultats corrects car les conditions de mesure au sein de notre laboratoire diffèrent des conditions du constructeur. C'est pourquoi nous avons établi notre propre équation en utilisant la méthode des moindres carrés ordinaires. L'équation d'étalonnage obtenue est la suivante :

$$[NO_2]_{NO2-B43F} = signal_{NO2-B43F} * a_{NO2-B43F} + b_{NO2-B43F} \quad (II-13)$$

où :

$signal_{NO2-B43F}$  est le signal de la différence de potentiel entre l'électrode de travail et l'électrode auxiliaire ;

$a_{NO2-B43F}$  et  $b_{NO2-B43F}$  désignent respectivement la pente et le biais déterminés par la régression linéaire.

Les données utilisées en laboratoire correspondent aux 508 données réparties sur 10 concentrations allant de 25ppb à 250ppb avec un pas de 25ppb, soit environ 50 mesures par concentration.

### II.9.2 L'étalonnage des capteurs MiCS-2714 et de GGS 7530

La même procédure d'étalonnage est appliquée pour chacun des deux types de capteurs à oxyde métallique qui ont également un comportement linéaire, selon l'équation d'étalonnage suivante :

$$[NO_2] = signal_{capteur} * a + b \quad (II-14)$$

où :

$signal_{capteur}$  est la différence entre la tension mesurée aux bornes de la résistance du capteur en présence du gaz cible et celle mesurée quand le capteur est exposé à l'air pur ;

$a$  et  $b$  correspondent respectivement à la pente et au biais, déterminés par une régression linéaire.

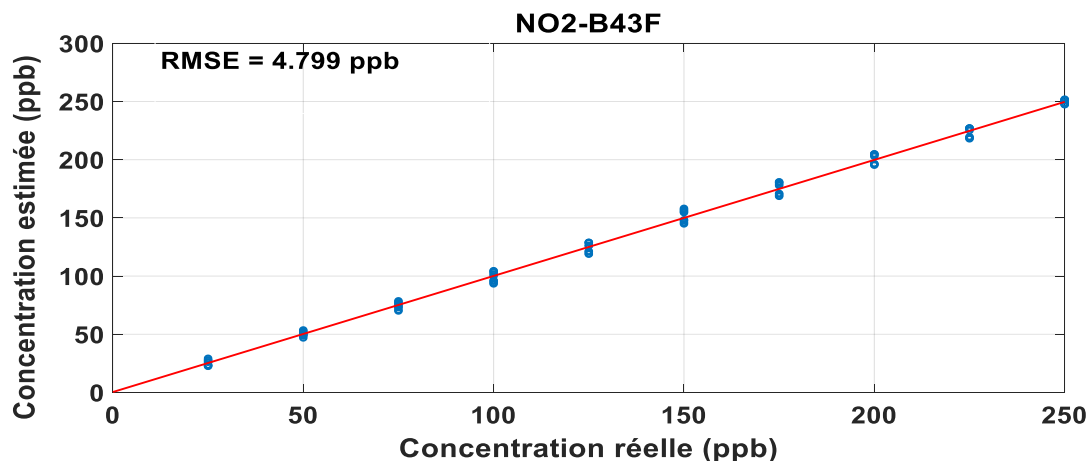


Figure II- 9 Corrélation entre les concentrations réelles et celles estimées en laboratoire par le capteur NO2-B43F, en utilisant un modèle linéaire.

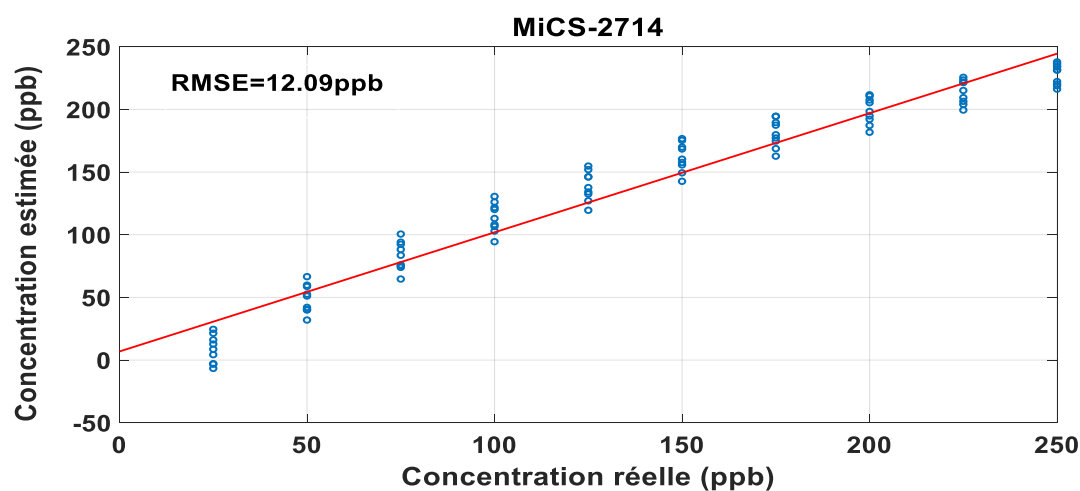


Figure II- 10 Corrélation entre les concentrations réelles et celles estimées en laboratoire par le capteur MiCS-2714, en utilisant un modèle linéaire.

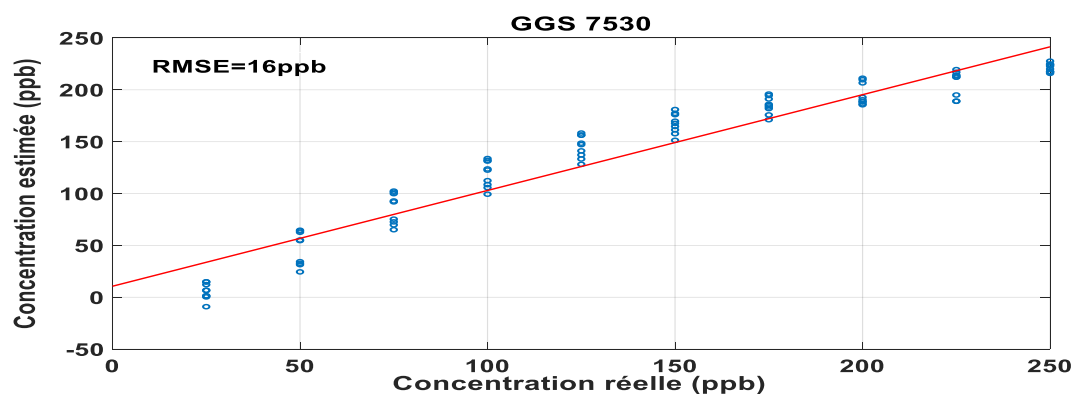


Figure II- 11 Corrélation entre les concentrations réelles et celles estimées en laboratoire par le capteur GGS 7530, en utilisant un modèle linéaire.

D'après les figures II-9 à II-11, nous pouvons conclure que les trois capteurs sont capables de quantifier les différentes concentrations de NO<sub>2</sub> dans les conditions de laboratoire



et exposés uniquement au NO<sub>2</sub>. Le capteur électrochimique Alphasense NO<sub>2</sub>-B43F fournit la meilleure estimation avec une erreur quadratique moyenne RMSE = 4.79ppb contre 12.09ppb pour le capteur MOS MiCS-2714. Le capteur MOS GGS 7530 donne des performances les moins bonnes parmi ces trois capteurs avec un RMSE de 16ppb et un signal très bruité.

### II.10 Mesures sur le terrain

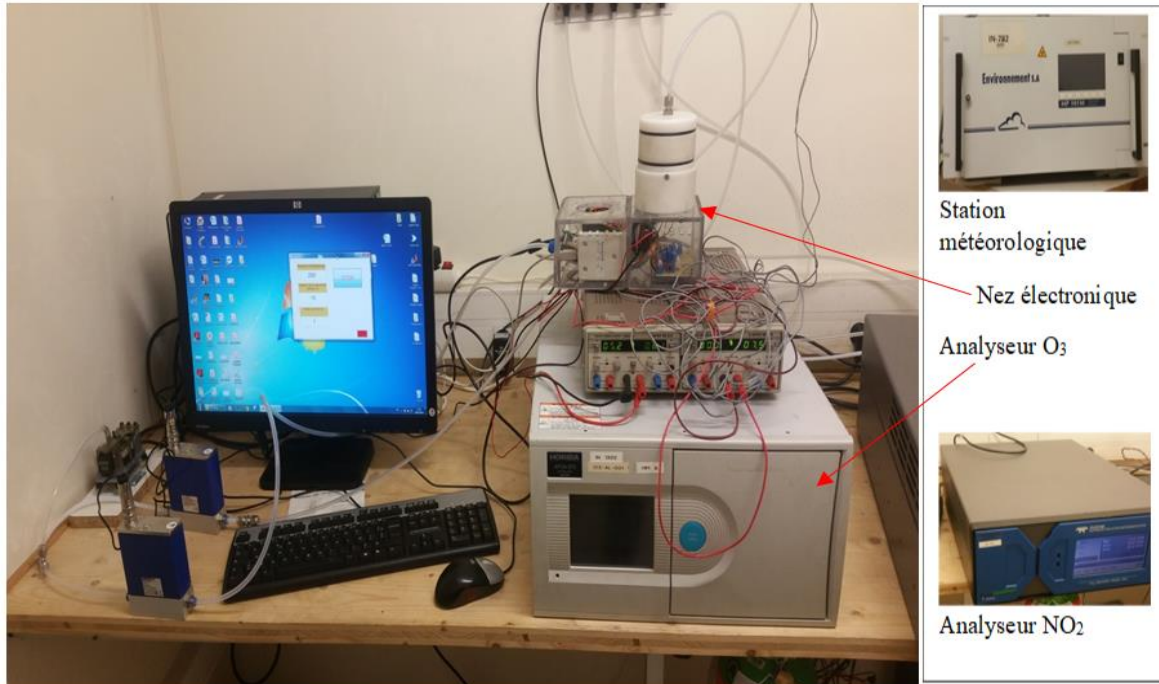
Pour les mesures sur le terrain, nous avons rajouté un second type de capteur électrochimique pour la mesure de l'ozone et nous avons retiré le capteur GGS 7530 car le bruit sur son signal amplifié était très important et, par conséquent, l'incertitude de mesure trop élevée. Nous avons conçu deux dispositifs identiques. L'un des dispositifs est placé à l'intérieur de la station de mesure d'ATMO Grand Est situé au bord de l'autoroute A31 traversant la ville de Metz, et le second à l'extérieur, sur le toit de la même station, comme illustrés sur les figures II-12 et II-13.

La figure II-12 montre les différents équipements utilisés pour la collecte des données. Les données de référence sont obtenues à partir de deux analyseurs d'ATMO Grand Est, considérés comme instruments de référence. Le premier permet de mesurer la concentration de dioxyde d'azote par la méthode de chimiluminescence (norme NF EN 14211) et le deuxième mesure le taux d'ozone par la spectroscopie UV (norme NF EN 14625). Nous avons cherché à reproduire sur le terrain un dispositif identique à celui utilisé en laboratoire.

- L'unité de mesure contenant les capteurs, leurs circuits de conditionnement et la chambre d'exposition aux gaz sont les mêmes.
- Ensuite, en ce qui concerne l'exposition des capteurs à l'air à analyser (l'échantillonnage d'air), nous avons souhaité travailler en mode dynamique, contrairement aux systèmes existants qui fonctionnent en mode statique. Ainsi, une pompe et un régulateur de débit massique sont placés à la sortie de la chambre d'exposition en aspiration, afin de générer un flux d'air constant et continu. A l'aide de ce régulateur de débit, nous avons fixé le débit d'air à 500ml/min, afin d'avoir un flux d'air identique à celui utilisé par les analyseurs d'ATMO Grand Est.
- Les réponses des capteurs sont collectées avec une fréquence d'échantillonnage de données de 200Hz. Ces réponses sont ensuite moyennées sur une période de 10s et enregistrées sur un ordinateur à l'aide d'un programme et une interface créés sous Matlab.
- Nous avons collecté les données en continu sur plusieurs mois. Le système à l'intérieur de la station a été mis en fonctionnement depuis juillet 2018, tandis que le système à l'extérieur a été installé en février 2019.

## Mesures sur le terrain

Nous précisons qu'à l'intérieur de la station, la température est contrôlée et fixée à 22°C ainsi la variation de la température du gaz reste très faible. Par contre, le système placé sur le toit de la station, n'est pas contrôlé en température et nous avons donc deux facteurs interférents importants, soient la température et l'humidité relative qui peuvent varier fortement.



**Figure II- 12 Dispositif micro-capteurs de gaz placé à l'intérieur de la station de mesure d'ATMO Grand Est.**



**Figure II- 13 Dispositif micro-capteurs de gaz placé à l'extérieur sur le toit de la station d'ATMO Grand Est.**

### II.11 Étalonnage sur le terrain

#### II.11.1 Mesure de concentrations de dioxyde d'azote

Pour l'étalonnage sur le terrain de nos dispositifs, nous avons repris la procédure d'étalonnage que nous avons réalisé en laboratoire, la seule différence étant au niveau des données de référence qui sont fournies par les analyseurs d'ATMO Grand Est.

Pour étudier la corrélation de nos mesures avec celles d'ATMO Grand Est, nous utilisons nos données obtenues à partir du dispositif placé à l'intérieur de la station de surveillance. Les résultats d'estimation de la concentration de dioxyde d'azote à partir d'un modèle de régression, montrent que le capteur électrochimique NO<sub>2</sub>-B43F donne une meilleure estimation par rapport au capteur à oxyde métallique MiCS-2714 avec un RMSE = 9.16µg/m<sup>3</sup> (figure II-14). Le capteur à oxyde métallique MiCS-2714 a un RMSE = 21µg/m<sup>3</sup> (figure II-15). Cette différence de performance peut être expliquée par l'influence des autres gaz contenus dans l'air, tels que l'ozone sur le capteur MiCS-2714 alors que le capteur électrochimique NO<sub>2</sub>-B43F est plus sélectif car il est équipé d'un filtre évitant le contact des gaz interférents avec l'électrode de mesure.

Il est important de préciser que pour le dispositif situé à l'intérieur de la station de surveillance, les facteurs influents tels que la température et l'humidité de l'air sont maintenus constants.

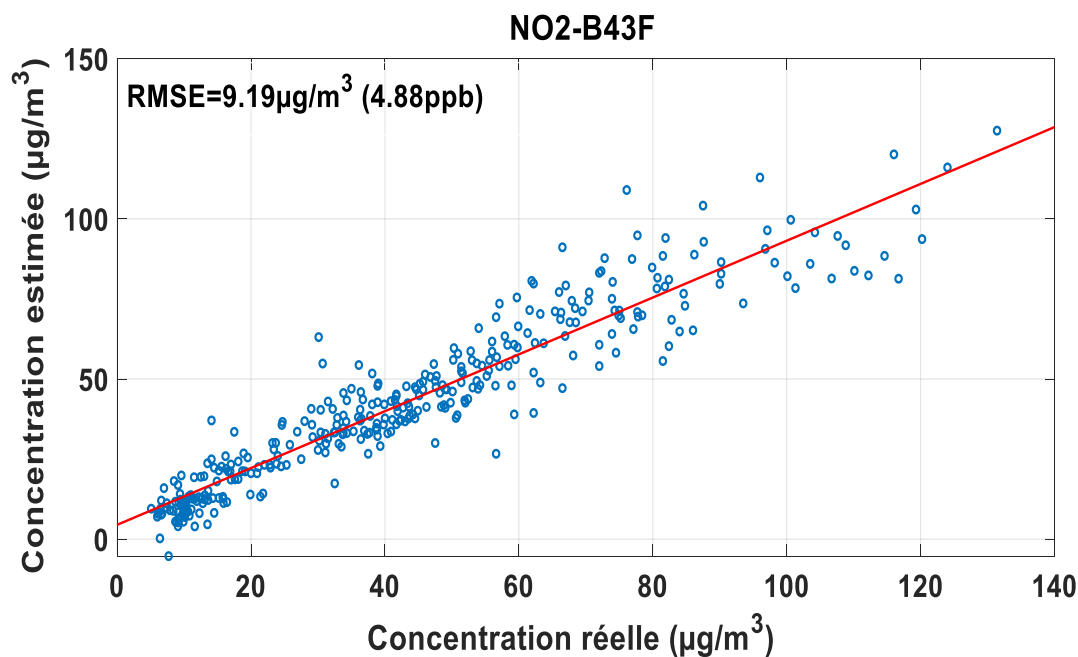
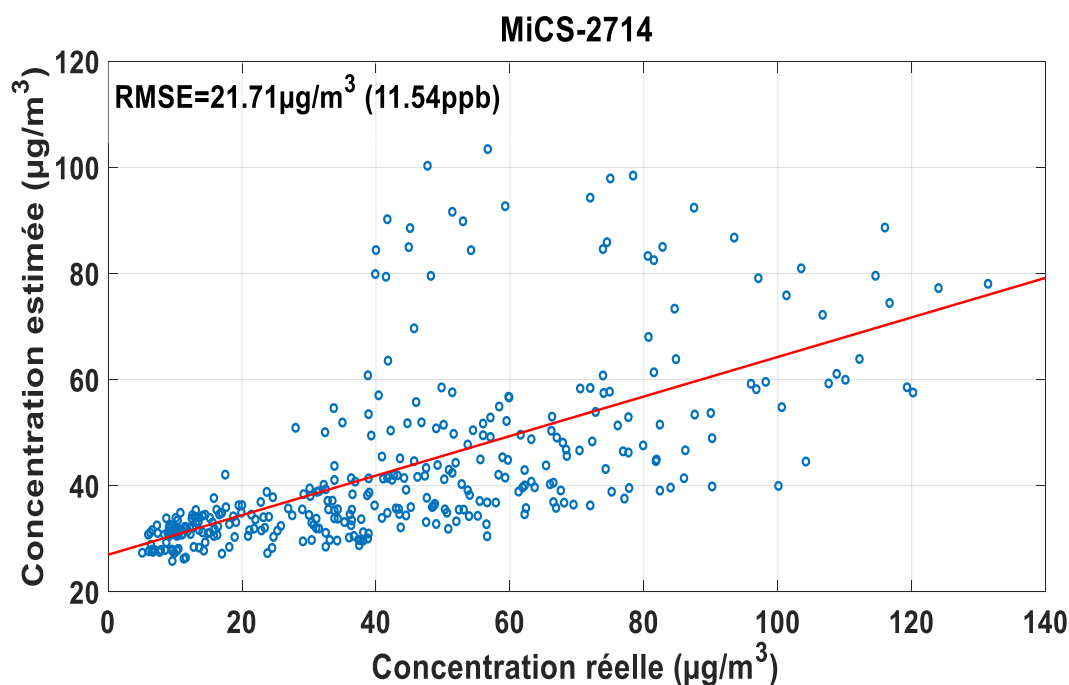


Figure II- 14 Corrélation entre les concentrations réelles et estimées de dioxyde d'azote par le capteur NO<sub>2</sub>-B43F en utilisant un modèle linéaire.



**Figure II- 15** Corrélation entre les concentrations réelles et estimées de dioxyde d’azote par le capteur MiCS-2714, en utilisant un modèle linéaire.

### II.11.2 Mesure de l’ozone

Les capteurs électrochimiques commerciaux existants pour mesurer l’ozone répondent simultanément à l’O<sub>3</sub> et au NO<sub>2</sub>, sans discrimination, du fait que le NO<sub>2</sub> et l’O<sub>3</sub> sont réductibles à des potentiels similaires sur les électrodes en carbone ou en or. Par conséquent, les réponses de ces capteurs sont proportionnelles à la concentration combinée d’O<sub>3</sub> et de NO<sub>2</sub>. Cette non-sélectivité des capteurs devient un obstacle majeur pour les applications de la surveillance de la pollution de l’air où le NO<sub>2</sub> et l’O<sub>3</sub> sont présents simultanément avec des concentrations de même ordre de grandeur.

Pour répondre à la quantification simultanée des concentrations en O<sub>3</sub> et en NO<sub>2</sub>, le constructeur Alphasense recommande l’utilisation d’une paire de capteurs électrochimiques : un capteur modèle OX-B431 qui répond aux deux gaz (O<sub>3</sub>+NO<sub>2</sub>) et un capteur modèle NO<sub>2</sub>-B43F qui répond uniquement au NO<sub>2</sub> [Zuidema et al. 2019]. Le capteur NO<sub>2</sub>-B43F est équipé d’un filtre au dioxyde de manganèse qui catalyse l’O<sub>3</sub> en oxygène, empêchant ainsi la réponse du capteur à l’ozone présent dans l’environnement (Figure II-16).

Pour déterminer la concentration en ozone, la contribution du NO<sub>2</sub> à la réponse du capteur OX-B431 doit être retirée. Il nous faut d’abord calculer la concentration en NO<sub>2</sub> avec le capteur NO<sub>2</sub>-B43F et ensuite la soustraire à la concentration fournie par le capteur OX-B431 [Hossain et al. 2016].

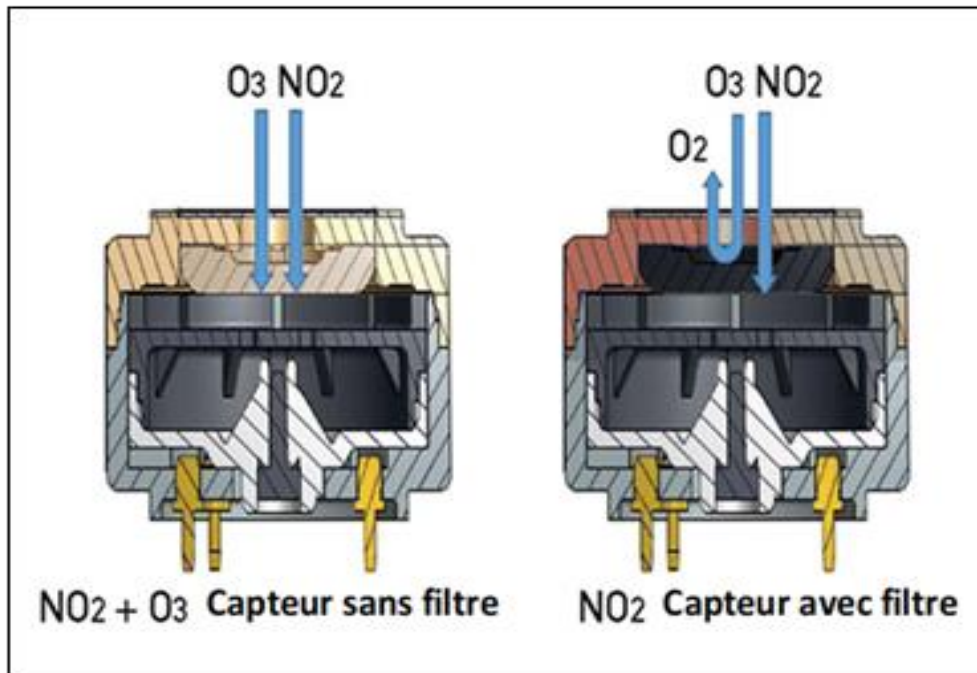


Figure II- 16 Schéma représentatif de la mesure de l'ozone à l'aide d'une paire de capteurs, l'un sans filtre et l'autre équipé d'un filtre à l'ozone [Hossain et al. 2016].

L'étalonnage de cette paire de capteurs se fait comme suit.

- Etalonner le capteur NO<sub>2</sub>-B43F pour la mesure du NO<sub>2</sub> (selon l'équation II-13).
- Etalonner le capteur OX-B431 pour la mesure du mélange (NO<sub>2</sub> + O<sub>3</sub>) également selon une droite de régression (équation II-15).

$$[NO_2 + O_3]_{OX-B431} = signal_{OX-B431} * a_{OX-B431} + b_{OX-B431} \quad (II-15)$$

- La concentration de O<sub>3</sub> sera la différence entre la concentration obtenue par le capteur OX-B431 et la concentration obtenue par le capteur NO<sub>2</sub>-B43F (équation II-16).

$$[O_3] = [NO_2 + O_3]_{OX-B431} - [NO_2]_{NO2-B43F} \quad (II-16)$$

Sur la figure II-17, nous montrons la corrélation entre les concentrations d'ozone estimées et les concentrations réelles fournies par ATMO Grand Est. Les performances d'estimation de l'ozone sont moins bonnes par rapport à l'estimation du dioxyde d'azote, ce qui peut être justifié par l'utilisation de deux capteurs et l'utilisation d'un modèle d'étalonnage spécifique pour chacun des capteurs.

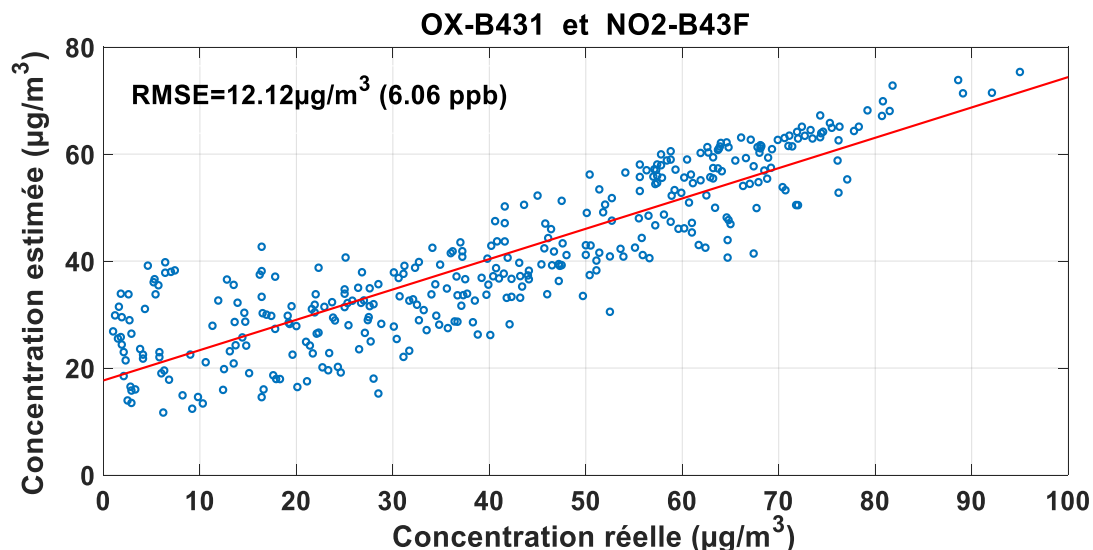


Figure II- 17 Corrélation entre les concentrations réelles et estimées d’ozone par le capteur OX-B431 en utilisant un modèle linéaire.

### II.11.3 Influence de la température et de l’humidité

Les capteurs installés sur le toit, à l’extérieur de la station de mesure, subissent des variations importantes de la température et de l’humidité entre jour et nuit ou d’une saison à une autre. Afin de mettre en évidence l’effet de la variation de la température nous avons estimé les concentrations de dioxyde d’azote en utilisant le capteur NO2-B43F installé à l’extérieur avec un modèle d’étalonnage sans prendre en considération la variation de la température. La figure II-18 montre que le capteur installé à l’extérieur donne une précision moins bonne que celui installé à l’intérieur de la station de mesure (figure II-14), soit un RMSE de 12.08µg/m<sup>3</sup> contre 9.19 µg/m<sup>3</sup>.

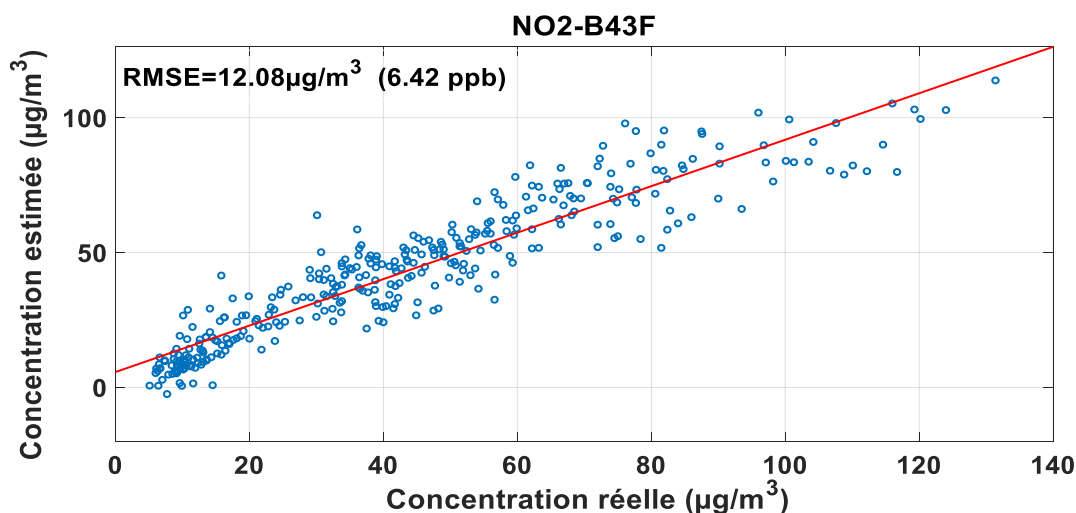
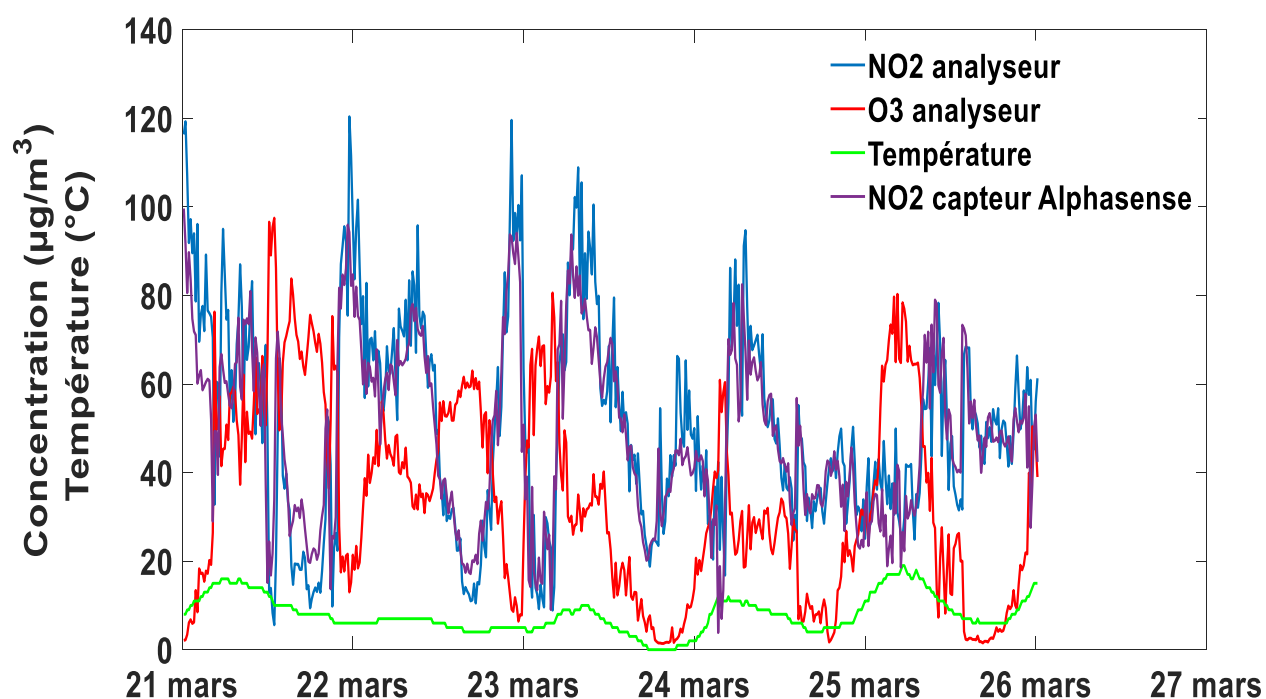


Figure II- 18 Corrélation entre les concentrations réelles et estimées de dioxyde d’azote sur le terrain à l’extérieur de la station par le capteur Alphasense NO2-B43F.

En effet la variation de la température influe sur la réponse du capteur ainsi que sur la variation de la concentration d'ozone et de dioxyde d'azote. La figure II-19 montre cette relation, lorsque la température augmente avec l'ensoleillement. Ce dernier participe à la transformation du dioxyde d'azote en ozone d'où l'augmentation de la concentration d'ozone et la diminution de la concentration de dioxyde d'azote (la journée du 25 mars) ; à l'inverse lorsque la température est faible (la nuit du 25 mars), le taux d'ozone décroît alors que celui de NO<sub>2</sub> augmente.

La température influe également sur la sensibilité des micro-capteurs, bien que cette influence ne soit pas directement identifiable sur les réponses des micro-capteurs.



**Figure II- 19 Corrélation entre la variation de la température et les concentrations d'ozone et de dioxyde d'azote.**

Ces remarques seront utilisées dans le chapitre III où nous montrons que la prise en compte de la variation de la température dans le modèle d'étalonnage va considérablement améliorer la précision du modèle.

### II.11.4 Incertitudes des mesures

Nous avons calculé les incertitudes relatives de nos mesures afin de les comparer avec celles exigées par la directive européenne concernant les mesures indicatives selon l'équation II-17.



$$\text{Incertitude } \% = 100 * \sqrt{\frac{\sum_{i=1}^N (Z_i - Y_i)^2}{N}} / \frac{\sum_{i=1}^N (Z_i)}{N} \quad (\text{II-17})$$

où :

$Z_i$  : concentration réelle du polluant fournie par ATMO Grand Est

$Y_i$  : concentration du polluant estimée par l'un des capteurs individuellement

$N$  : nombre de mesures.

Les résultats obtenus à partir de nos mesures sont regroupés dans le tableau II-3 suivant.

Capteur	Gaz à mesurer	Incertitude de mesure	Incertitude exigée par la directive européenne
NO2-B43F	NO <sub>2</sub>	23.12%	25%
MiCS-2714	NO <sub>2</sub>	49.77%	25%
OX-B431 + NO2-B43F	O <sub>3</sub>	29.75%	30%

**Tableau II- 3 L'incertitude relative par capteur, calculée à partir des données collectées sur le terrain et des données de référence obtenues par les analyseurs.**

D'après le tableau II-3 les résultats fournis par les deux capteurs électrochimiques NO2-B43F et OX-B431 installés à l'intérieur peuvent être utilisés en tant que mesures indicatives puisqu'ils respectent les objectifs exigés par la directive européenne. En revanche, l'incertitude du capteur à oxyde métallique MiCS-2714 dépasse de manière importante (deux fois plus) le seuil des incertitudes recommandé pour les mesures indicatives.

## II.12 Conclusion

Dans ce chapitre nous avons présenté la démarche que nous avons suivie pour caractériser des micro-capteurs de gaz en laboratoire, et puis, pour les tester en conditions réelles sur le terrain. Leurs caractéristiques telles que le temps de réponse, les limites de détection et de quantification ont permis d'envisager de les utiliser dans un dispositif de surveillance de la pollution de l'air pour le dioxyde d'azote. Nous avons constaté qu'en laboratoire les deux types de capteurs électrochimiques et à oxyde métallique sont capables de détecter et de quantifier des concentrations faibles de dioxyde d'azote. Cependant, dans des conditions réelles, la présence d'autres gaz interférents dans l'air ambiant empêche les capteurs



## **Chapitre II : Protocole d'évaluation des micro-capteurs de gaz**

---

de gaz à oxyde métallique de quantifier les concentrations de dioxyde d'azote avec une incertitude acceptable. Alors que les capteurs électrochimiques actuels pour le  $\text{NO}_2$  sont équipés d'un filtre qui les rend insensibles aux autres gaz interférents tels que  $\text{O}_3$ , et permettent une estimation correcte du  $\text{NO}_2$  avec une incertitude  $\approx 23\%$ .

En plaçant notre dispositif de micro-capteurs sur le toit de la station, nous avons également mis en évidence l'influence de la température et de l'humidité relative sur les réponses des capteurs. Nous en avons conclu qu'il était donc impératif d'introduire et d'exploiter toute information utile dans le modèle d'étalonnage, en particulier la température et l'humidité.

L'une des solutions possibles est de combiner les différentes informations provenant de différents capteurs ainsi que les valeurs de la température et de l'humidité de l'air à l'aide des méthodes multi-variables d'apprentissage statistique. Nous consacrons donc le chapitre suivant à l'étude et à l'application de ces méthodes sur les données recueillies sur le terrain.

**Chapitre III : Méthodes multi-variables pour  
l'estimation de la concentration de gaz à  
l'aide d'un réseau de capteurs**

### III.1 Introduction

Un nez électronique se compose de deux parties essentielles : la partie matérielle qui représente le réseau de capteurs de gaz et la partie logicielle qui applique les méthodes de traitement des données et de reconnaissance de formes [Gardner et Bartlett 2000; Gardner et Bartlett 1994]. Les méthodes de traitement de données et de reconnaissance de formes permettent d'exploiter les données des capteurs afin de classifier et/ou de quantifier un mélange de gaz. La méthode multi-variables adoptée doit pouvoir fournir une grande précision, une bonne généralisation, et permettre un paramétrage à partir d'un ensemble réduit d'échantillons d'apprentissage afin de faciliter le calibrage de ces appareils. Dans la littérature, de nombreuses méthodes de classification et de régression ont été mentionnées pour exploiter les données d'un nez électronique [Cordero et al. 2018; Fonollosa et al. 2014; Zhang et al. 2012; Cevoli et al. 2011; Harrou et al. 2018; Lentka et al. 2015]. Elles peuvent être regroupées en deux catégories principales : méthodes linéaires et méthodes non linéaires [Hines et al. 1999].

Parmi les méthodes linéaires les plus utilisées on trouve la Régression en Composantes Principales (PCR) [Zellers et al. 1993; Blanco et al. 1999] et les Moindres Carrés Partiels (PLS) [Apetrei et al. 2010; Song et al. 2010; Blanco-Rodríguez et al. 2018; Lozano et al. 2007]. La méthode PLS est très répandue dans les applications de nez électronique ainsi que dans le domaine de la spectroscopie grâce à sa simplicité et son efficacité.

Les méthodes non linéaires telles que les différents types de réseaux de neurones (NN) ont été largement testées dans des applications de nez électronique [Hadjiiski et al. 1999; Martín et al. 2001; Esposito et al. 2016; Zuppa et al. 2004; Gorji-Chakespari et al. 2016; Li et al. 2007]. Des travaux montrent que les NN souffrent d'une mauvaise généralisation surtout dans le cas où le nombre d'échantillons destinées à l'apprentissage est relativement faible [Song et al. 2011]. La méthode des Séparateurs à Vaste Marge (*support vector machine* SVM) a également été utilisée pour la classification et la quantification des gaz avec de bonnes performances en termes de généralisation et de précision [Laref et al. 2018; Bigi et al. 2018; Men et al. 2018]. D'autres méthodes sont aussi exploitées pour les réseaux de capteurs de gaz comme *random forests* [Zimmerman et al. 2018] ou encore *extreme learning machine* [Qiu et al. 2015; Men et al. 2018].

Dans ce chapitre nous allons d'abord présenter la théorie de la régression SVM, choisie dans le cadre de nos travaux, pour mieux comprendre son fonctionnement. Ensuite, nous proposons un algorithme basé sur la méthode de recherche par motifs généralisée (*Generalized*

*Pattern Search* GPS) pour sélectionner les valeurs de ses hyper-paramètres. Ce choix a été fait après une comparaison par rapport à d'autres algorithmes d'optimisation (par exemple, les algorithmes génétiques, le recuit simulé, etc.). La deuxième partie de ce chapitre est consacrée à une comparaison de la méthode SVM par rapport aux méthodes PLS et ELM (*extreme learning machine*). Cette étude comparative sera basée sur la précision de la quantification gazeuse souhaitée et le volume de données utilisé pour la phase d'apprentissage.

### III.2 Régression SVM

La famille des méthodes *support vector machines* (SVM), souvent traduits par Séparateurs à Vaste Marge, fait partie des méthodes d'apprentissage statistique (*machine learning*) initialement définies pour la classification binaire. Ces méthodes ont été ensuite généralisées pour une classification multi-classes ainsi que pour les tâches de quantification (régression SVM) [Basak et al. 2007], qui est préférentiellement utilisée dans notre étude.

Nous commençons par décrire la théorie de base pour la méthode de la régression SVM [Smola et Schölkopf 2004].

Supposons que nous avons les données suivantes :  $\{(x_1, y_1), \dots, (x_i, y_i), \dots, (x_n, y_n)\} \subset \mathfrak{R}^d \times \mathfrak{R}$

telles que :

$x_i$  sont les «  $d$  » variables explicatives correspondant, dans notre cas, aux tensions mesurées aux bornes des capteurs de la  $i^{\text{ème}}$  mesure.

$y_i$  est la sortie ou la variable à expliquer qui représente la concentration réelle de la  $i^{\text{ème}}$  mesure.

L'objectif de la régression SVM est de trouver une fonction  $f(x)$  capable de prédire la valeur de  $y_i$  pour chacune des mesures d'apprentissage avec une erreur maximale  $\varepsilon$ , et que  $f(x)$  soit en même temps la plus simple possible. En d'autres termes, les erreurs sont acceptées tant qu'elles sont inférieures à  $\varepsilon$ . Cette méthode de régression est appelée  $\varepsilon$ -SVM.

Afin de simplifier la description de la SVM nous abordons d'abord le cas où la fonction  $f(x)$  est linéaire.

#### III.2.1 Cas linéaire

Dans ce cas, la fonction  $f(x)$  est définie par :

$$f(x) = \langle x, w \rangle + b \quad (\text{III-1})$$

où  $w \in \mathfrak{R}^d$ ,  $b \in \mathfrak{R}$  et  $\langle . \rangle$  est le produit scalaire dans l'espace  $\mathfrak{R}^d$ .

### Chapitre III : Méthodes multi-variables pour l'estimation de la concentration de gaz à l'aide d'un réseau de capteurs

Pour avoir une fonction  $f(x)$  la plus plate (la plus simple) possible, il faut minimiser la norme de  $w$  :  $\|w\|^2$ .

On peut alors écrire le problème sous forme d'un problème d'optimisation, soit :

$$\text{Minimiser } \frac{1}{2} \|w\|^2 \quad (\text{III-2})$$

$$\text{tel que } \begin{cases} y_i - \langle x_i, w \rangle - b \leq \varepsilon \\ \langle x_i, w \rangle + b - y_i \leq \varepsilon \end{cases}$$

Dans certains cas on ne peut pas satisfaire la condition que toutes les données d'entraînement soient prédites avec une précision de  $\varepsilon$ . Pour permettre de tolérer certaines erreurs supérieures à  $\varepsilon$ , on peut introduire deux variables de tolérance  $\xi_i, \xi_i^*$  dans le but de faire face aux contraintes impossibles du problème d'optimisation (équation III-2). Le problème d'optimisation devient comme suit :

$$\text{Minimiser } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (\text{III-3})$$

$$\text{tel que } \begin{cases} y_i - f(x_i, w) - b \leq \varepsilon + \xi_i^* \\ f(x_i, w) + b - y_i \leq \varepsilon + \xi_i \\ \xi_i^*, \xi_i \geq 0 \end{cases}$$

La constante  $C$  qui doit être positive, détermine le compromis entre la simplicité de  $f(x)$  et la valeur tolérée de la somme des écarts supérieurs à  $\varepsilon$ . Cela correspond à une fonction de tolérance  $|\xi|_\varepsilon$  ou de perte dite  $\varepsilon$ -insensitive, décrite par l'équation III-4 et schématisée sur la figure III-1 :

$$|\xi|_\varepsilon = \begin{cases} 0 & \text{si } |\xi| \leq \varepsilon \\ |\xi| - \varepsilon & \text{sinon} \end{cases} \quad (\text{III-4})$$

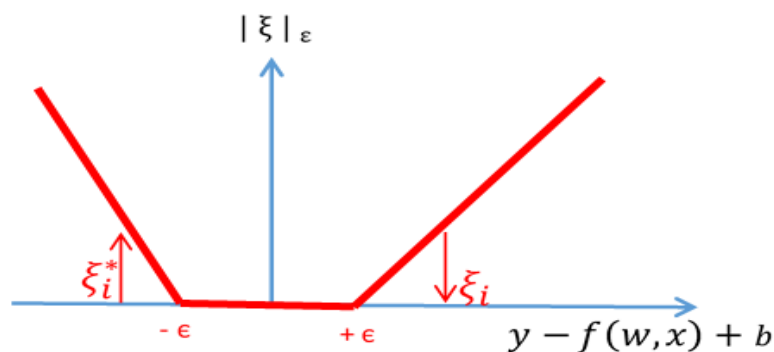
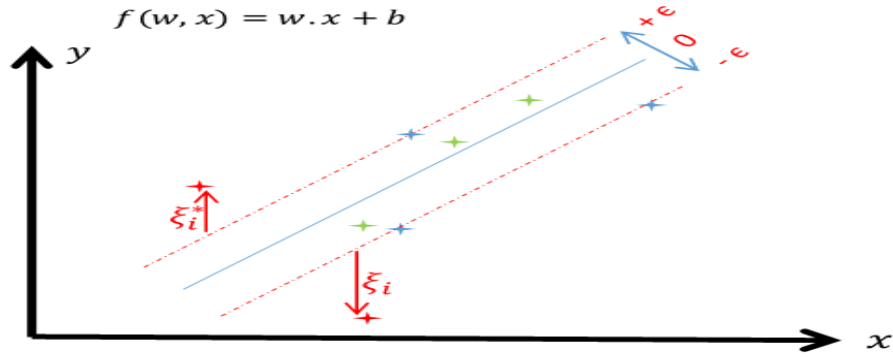


Figure III- 1 Schématisation de la fonction de perte  $|\xi|_\varepsilon$ .



**Figure III- 2 Schéma d'un modèle SVM à une dimension ; les points à l'extérieur des traits (traits en rouge pointillés) sont pris en considération pour calculer l'erreur.**

La figure III-2 illustre graphiquement la situation de ces écarts tolérés par la fonction  $f(x)$ . Seuls les points situés en dehors du tube contribuent au coût et les écarts sont pénalisés de manière linéaire.

Il s'avère que dans la plupart des cas, le problème d'optimisation (équation III-3) peut être résolu plus facilement dans sa forme duale en utilisant les multiplicateurs de Lagrange. L'idée est de construire une fonction de Lagrange à partir de la fonction à minimiser et des contraintes correspondantes comme suit :

$$\begin{aligned}
 L = & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) - \sum_{i=1}^n (\eta_i \xi_i + \eta_i^* \xi_i^*) \\
 & - \sum_{i=1}^n \alpha_i (\varepsilon + \xi_i - y_i + \langle x_i, w \rangle + b) \\
 & - \sum_{i=1}^n \alpha_i^* (\varepsilon + \xi_i^* + y_i - \langle x_i, w \rangle - b)
 \end{aligned} \tag{III-5}$$

$L$  étant le lagrangien et  $\eta_i, \eta_i^*, \alpha_i, \alpha_i^*$  les multiplicateurs de Lagrange.

Les multiplicateurs de Lagrange doivent satisfaire les contraintes de positivité, c'est-à-dire :

$$\eta_i, \eta_i^*, \alpha_i, \alpha_i^* \geq 0 \tag{III-6}$$

Les dérivées partielles de  $L$  par rapport aux variables primales  $w, b, \xi_i, \xi_i^*$  doivent s'annuler pour être optimales, ce qui se traduit par les quatre équations suivantes :

$$\partial_b L = \sum_{i=1}^n (\alpha_i^* - \alpha_i) = 0 \quad (\text{III-7})$$

$$\partial_w L = w - \sum_{i=1}^n (\alpha_i - \alpha_i^*) x_i = 0 \quad (\text{III-8})$$

$$\partial_{\xi_i} L = C - (\alpha_i - \eta_i) = 0 \quad (\text{III-9})$$

$$\partial_{\xi_i^*} L = C - (\alpha_i^* - \eta_i^*) = 0 \quad (\text{III-10})$$

En remplaçant les équations III-7, III-8, III-9 et III-10 dans l'équation III-5, le problème d'optimisation devient comme suit :

$$\text{maximiser} \left\{ -\frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle x_i, x_j \rangle - \varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) + \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*) \right\} \quad (\text{III-11})$$

$$\text{tel que :} \quad \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \quad \text{et } \alpha_i, \alpha_i^* \in [0, C]$$

Or d'après l'équation III-8 nous avons :

$$w = \sum_{i=1}^n (\alpha_i - \alpha_i^*) x_i \quad (\text{III-12})$$

et la fonction  $f(x)$  peut alors s'exprimer de la manière suivante :

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \langle x_i, x \rangle + b \quad (\text{III-13})$$

D'après l'équation III-12,  $w$  peut être complètement décrit par une combinaison linéaire des données d'apprentissage  $x_i$  où  $(\alpha_i - \alpha_i^*) \neq 0$ . Ces données sont appelées vecteurs supports. On peut aussi remarquer que la complexité de la représentation d'une fonction par les vecteurs supports est indépendante de la dimensionnalité de l'espace de saisie  $X$  et dépend uniquement du nombre de vecteurs supports. Par conséquent, l'algorithme complet peut être décrit en termes

de produits scalaires entre les données d'apprentissage. Il est à noter que lors de l'évaluation de  $f(x)$ , il n'est pas nécessaire de calculer  $w$  explicitement. Ces observations seront utiles pour la formulation d'une extension non linéaire.

Le calcul du paramètre  $b$  de la fonction  $f(x) = \langle x, w \rangle + b$  peut être réalisé en exploitant les conditions de Karush – Kuhn –Tucker (KKT). Celles-ci indiquent qu'au moment de la recherche de la solution, le produit entre variables duales et contraintes doit s'annuler.

$$\begin{cases} \alpha_i(\varepsilon + \xi_i - y_i + \langle x, w \rangle + b) = 0 \\ \alpha_i^*(\varepsilon + \xi_i^* + y_i - \langle x, w \rangle - b) = 0 \end{cases} \quad (\text{III-14})$$

et

$$\begin{cases} (C - \alpha_i) \xi_i = 0 \\ (C - \alpha_i^*) \xi_i^* = 0 \end{cases} \quad (\text{III-15})$$

Il suffit de prendre une donnée où  $\alpha_i \in ]0, C]$  et de la remplacer dans l'équation III-13 pour calculer la valeur de  $b$  [Lauer et Bloch 2006].

### III.2.2 Cas non linéaire

Dans le cas non linéaire, les données sont projetées par une projection  $\Phi$  dans un espace de haute dimension  $\mathcal{F}$  dans lequel le problème devient linéaire.

$$\begin{cases} \Phi: \mathbb{R}^d \rightarrow \mathcal{F} \\ x_i \rightarrow \Phi(x_i) \end{cases} \quad (\text{III-16})$$

$\Phi(x_i)$  étant la projection de  $x_i$  dans l'espace  $\mathcal{F}$

Comme il a été évoqué auparavant, l'algorithme SVM dépend uniquement des produits scalaires des données  $x_i$ . Cela permet de ne pas calculer la projection des données mais uniquement de calculer le produit scalaire des données projetées  $\langle \Phi(x_i), \Phi(x) \rangle$ .

Ce produit scalaire des données projetées dans un espace de haute dimension est égale à une fonction noyau (*Kernel function*)  $k(x_i, x)$ .

Ainsi le problème d'optimisation correspond à la recherche de la fonction la plus plate dans l'espace des fonctions noyau et non pas dans l'espace d'entrée. La question qui se pose maintenant est de savoir quelle fonction  $k(x_i, x)$  correspond à un produit scalaire des données projetées ( $k(x_i, x) = \langle \Phi(x_i), \Phi(x) \rangle$ ). D'après le théorème de Mercer [Sun 2005], toute fonction continue, symétrique et semi-définie positive est une fonction noyau.

L'algorithme SVM, dans le cas non linéaire, devient ainsi :



$$\text{maximiser} \left\{ -\frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)k(x_i, x_j) - \varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) + \sum_{i=1}^n y_i(\alpha_i - \alpha_i^*) \right\} \quad (\text{III-17})$$

$$\text{tel que :} \quad \begin{cases} \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, C] \end{cases}$$

La solution à l'équation III-17 nous donne la fonction  $f(x)$

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x_i, x) + b \quad (\text{III-18})$$

Parmi les fonctions noyau les plus connues on trouve :

Noyau linéaire

$$K(x_i, x_j) = x_i^t x_j \quad (\text{III-19})$$

Noyau exponentiel

$$K(x_i, x_j) = \exp(-\gamma \|x_i, x_j\|) \quad (\text{III-20})$$

Noyau gaussien

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i, x_j\|^2}{2\sigma^2}\right) \quad (\text{III-21})$$

Noyau hyperbolique

$$K(x_i, x_j) = \frac{1}{\epsilon + \gamma \|x_i, x_j\|} \quad (\text{III-22})$$

Noyau polynomial

$$K(x_i, x_j) = (x_i^t x_j + c)^d \quad (\text{III-23})$$

Dans ce travail nous avons choisi le noyau gaussien comme fonction noyau. Notre choix a été basé tout d'abord sur une étude bibliographique et ensuite à partir des comparaisons des résultats obtenus entre les différents types de noyaux.

### III.3 Optimisation des hyper-paramètres de la régression SVM

Le choix des hyper-paramètres de la méthode SVM est primordial pour assurer une bonne performance. Dans la majorité des travaux de recherche sur les réseaux de capteurs de gaz où la régression SVM a été utilisée, la sélection de ses hyper-paramètres optimaux est effectuée à l'aide d'une recherche par 'grille' [Qiu et al. 2015; Qiu et Wang 2017; Men et al. 2018]. Cette technique de recherche exhaustive balaye un sous-ensemble de valeurs de paramètres spécifiées au préalable afin de sélectionner les valeurs optimales. Etant donné que l'espace des hyper-paramètres n'est pas limité et que les hyper-paramètres peuvent prendre des valeurs réelles dans de nombreux cas, le choix d'un sous-espace optimal n'est pas évident. En effet, même avec seulement trois hyper-paramètres à optimiser et une petite grille d'essais, par exemple de  $10 \times 10 \times 10$ , le processus d'optimisation peut être extrêmement lent compte tenu du nombre de validations croisées à réaliser pour tester les performances et de la vitesse de l'algorithme de régression lui-même.

Dans d'autres domaines de recherche scientifique, de nombreux travaux ont proposé des algorithmes pour optimiser les hyper-paramètres de la régression SVM, notamment : l'algorithme génétique (GA) [Fei et Zhang 2009], l'optimisation par essaims particulaires (PSO) [Lin et al. 2008], le recuit simulé (SA) [Sartakhti et al. 2012], etc... Cependant, ces algorithmes nécessitent eux-mêmes un ajustement de leurs propres paramètres. Par exemple, dans le cas de l'algorithme de recuit simulé, le choix de la température initiale et du programme de refroidissement est très important, car les résultats finaux dépendront directement de ce choix.

Dans cette étude, nous avons cherché un algorithme qui soit facile à mettre en œuvre, qui ne nécessite pas de paramètres supplémentaires à choisir et qui garantit une bonne solution. De ce point de vue, la famille des « algorithmes de recherche directe » s'avère être un bon choix en raison de leur simplicité, de leur efficacité et de leur robustesse.

Nous allons tout d'abord présenter l'impact des hyper-paramètres sur les performances de la SVM. Ensuite, pour l'optimisation des hyper-paramètres de la SVM, nous proposons l'algorithme de recherche par motifs généralisée (*Generalized Pattern Search* GPS) de la famille des algorithmes de recherche directe.

#### III.3.1 Influence de chaque hyper-paramètre sur la performance de la régression SVM

Trois hyper-paramètres impactent la performance de la régression SVM.

$\varepsilon$  : la largeur de tube autour de la fonction modèle ;

$C$  : l'hyper-paramètre de régulation entre la complexité du modèle et sa précision ;  
 $\sigma$  : l'hyper-paramètre du noyau gaussien.

L'hyper-paramètre  $\varepsilon$  contrôle la largeur du tube insensible (figure III-2) autour du modèle de prédiction. Le nombre de données considérées, comme des vecteurs supports, est directement lié à la valeur de  $\varepsilon$ . Le choix d'une valeur élevée de  $\varepsilon$  conduit à sélectionner peu de données comme vecteurs supports, ce qui rend le modèle plus simple. Par contre, une faible valeur de  $\varepsilon$  permet de sélectionner un grand nombre de données en tant que vecteurs supports, ce qui augmente la complexité du modèle.

Concernant l'hyper-paramètre  $C$ , le choix d'une valeur faible rend le modèle simple. En revanche, pour une valeur élevée de  $C$ , le modèle sélectionne plus de données comme vecteurs supports afin d'estimer correctement toutes les données d'apprentissage.

L'hyper-paramètre  $\sigma$  de la fonction noyau gaussien a aussi une influence importante dans la performance du modèle. Si la valeur du paramètre  $\sigma$  est trop faible, l'influence des vecteurs supports est trop forte et aucune marge de la valeur de  $C$  ne pourra empêcher un sur-apprentissage. Lorsque  $\sigma$  est très grand, le modèle est trop contraint et ne peut pas appréhender la variabilité des données.

En conclusion, chacun de ces trois hyper-paramètres affecte la complexité du modèle SVM de manière différente, d'où l'intérêt majeur de leur optimisation.

### III.3.2 Relation entre $C$ et $\sigma$

L'ensemble de données utilisées dans cette section et dans le reste de ce chapitre est collecté sur le terrain à l'aide du dispositif placé à l'intérieur de la station d'ATMO Grand Est. Cet ensemble contient 1200 mesures. L'évaluation des performances est basée sur la moyenne des RMSE issue d'une validation croisée de 10 partitions. Le choix de 10 partitions est basé sur la littérature, en effet le choix du nombre de partitions dépend de la taille de l'ensemble de données car pour chaque validation, les ensembles d'entraînement et de test doivent être suffisamment représentatifs pour l'ensemble de données. Dans notre cas l'ensemble global contient 1200 données, ce qui nous donne pour chaque validation, 1080 données pour l'apprentissage et 120 données pour le test. Ces nombres constituent un bon rapport test/apprentissage pour faire la validation croisée avec 10 partitions.

Pour montrer la relation entre  $C$  et  $\sigma$  avec un  $\varepsilon$  fixe ( $\varepsilon = 0.001$ ), nous avons utilisé une grille de recherche de  $500 * 200$  points où  $C$  varie de 1 à 100 000 avec un pas de 200, et  $\sigma$  de 1 à 100 avec un pas de 0.5. Nous avons fait une étude préliminaire par tâtonnement sur les

grandeurs de ces hyper-paramètres pour déterminer leur plage et leur pas de variation significatifs. Pour chaque point de cette grille, nous avons calculé la moyenne des RMSE obtenus par la validation croisée des 10 partitions. La figure III-3 montre que nous pouvons trouver un ensemble de combinaisons qui assurent un bon modèle. Par exemple, si nous prenons une faible valeur pour l'hyper-paramètre  $C$  (modèle simple), le modèle peut devenir complexe en choisissant une faible valeur de  $\sigma$  (la zone entourée en magenta dans la figure III-3). Nous observons également sur cette figure que, pour une très grande valeur de  $C$ , nous pouvons choisir un grand intervalle de valeurs possibles de  $\sigma$  conduisant à un bon modèle (dans la figure III-3, la zone bleue s'élargit pour des grandes valeurs de  $C$ ). Cependant, un nombre important de vecteurs supports conduit à un sur-apprentissage et l'algorithme occupe alors plus d'espace mémoire.

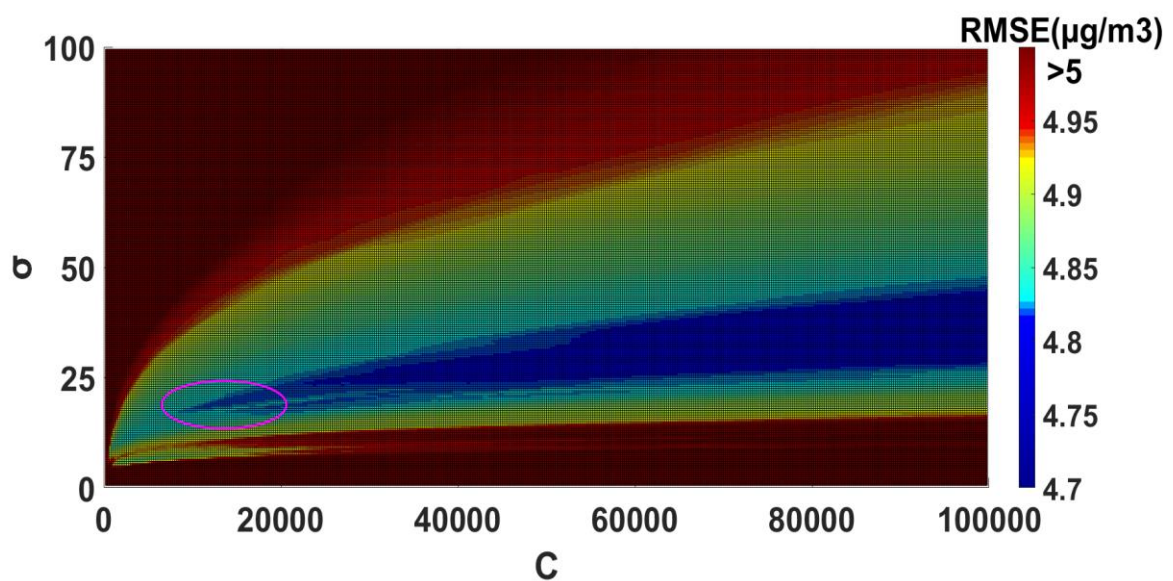


Figure III- 3 Erreur quadratique moyenne de la validation croisée en fonction de  $C$  et  $\sigma$ .

### III.4 Présentation des algorithmes d'optimisation

Nous avons étudié différents algorithmes d'optimisation pour sélectionner les hyper-paramètres de la régression SVM. La figure III-4 illustre la procédure suivie dans le processus d'optimisation des hyper-paramètres de la SVM. Pour chaque itération, l'algorithme SVM reçoit les valeurs des hyper-paramètres fournies par l'algorithme d'optimisation, et puis il calcule la moyenne de l'erreur quadratique moyenne des 10 partitions, que nous appelons moyenne RMSE. Tant que le critère d'arrêt n'est pas atteint, l'algorithme SVM renvoie la valeur de la moyenne RMSE à l'algorithme d'optimisation pour adapter les valeurs des hyper-paramètres.

Dans la partie qui suit, nous allons tout d'abord décrire le principe de chacun des quatre algorithmes d'optimisation étudiés pour ensuite présenter les résultats de la comparaison de leur performance. Notons également que le terme fonction objectif utilisé généralement dans l'optimisation signifie la fonction ou la grandeur que l'on cherche à minimiser. Dans notre cas la fonction objectif est la moyenne des RMSE des 10 validations croisées.

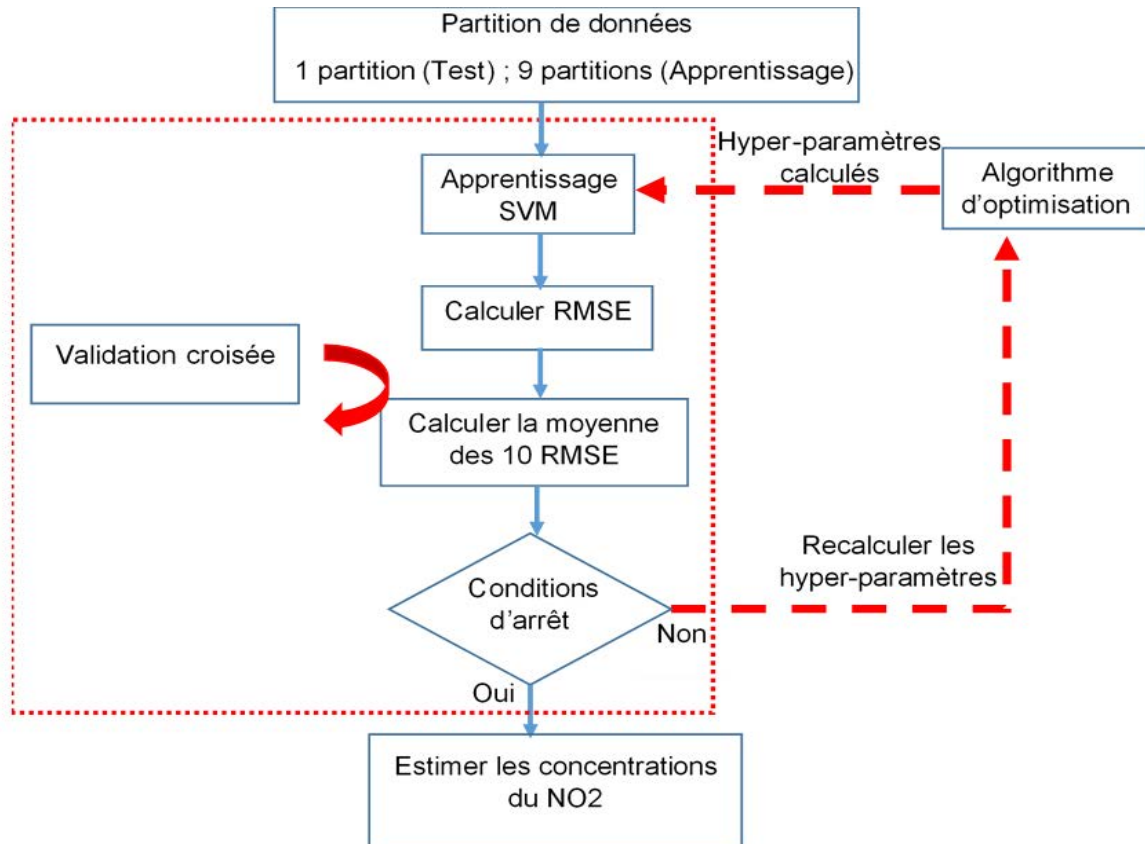


Figure III- 4 Schéma global de la procédure d'optimisation de la SVM.

### III.4.1 Algorithme génétique

Les algorithmes génétiques (AG) imitent l'évolution et la sélection naturelle afin de résoudre les problèmes d'optimisation [Weile et Michielsse 1997]. Les AG sélectionnent les individus les plus aptes à survivre parmi une population au cours d'une génération. Chaque individu représente un point dans un espace de recherche et correspond à une solution possible. Les individus de la population sont ensuite amenés à subir un processus d'évolution. Les AG sont basés sur une analogie avec la structure génétique et le comportement des chromosomes au sein d'une population d'individus en utilisant les fondements suivants.

Les individus qui réussissent le mieux dans chaque compétition produiront plus de progéniture que ceux qui ont de mauvais résultats. Les gènes des bons individus se propagent dans toute la population, de sorte que deux bons parents produiront parfois des descendants meilleurs que

## Présentation des algorithmes d'optimisation

l'un ou l'autre des parents. Ainsi les générations successives deviendront plus adaptées à leur environnement. Après la génération d'une population initiale de manière aléatoire, l'algorithme évolue à travers trois opérations :

1) **Sélection** : l'AG sélectionne les meilleurs individus (combinaison des hyper-paramètres dans notre cas) qui donnent les minimums de la fonction objectif, et transmet leurs gènes à la génération suivante.

2) **Croisement** : un croisement est effectué entre les chromosomes des individus sélectionnés au hasard avec une probabilité définie, pour créer deux nouveaux descendants susceptibles d'être meilleurs que leurs parents. Ces nouveaux individus sont introduits dans la prochaine génération de la population.

3) **Mutation** : une mutation avec une faible probabilité va permettre de maintenir la diversité au sein de la population et d'éviter une convergence prématurée. La mutation consiste alors à changer un bit au hasard dans le chromosome d'un individu.

Le schéma de la figure III-5 illustre le principe de fonction de cet algorithme.

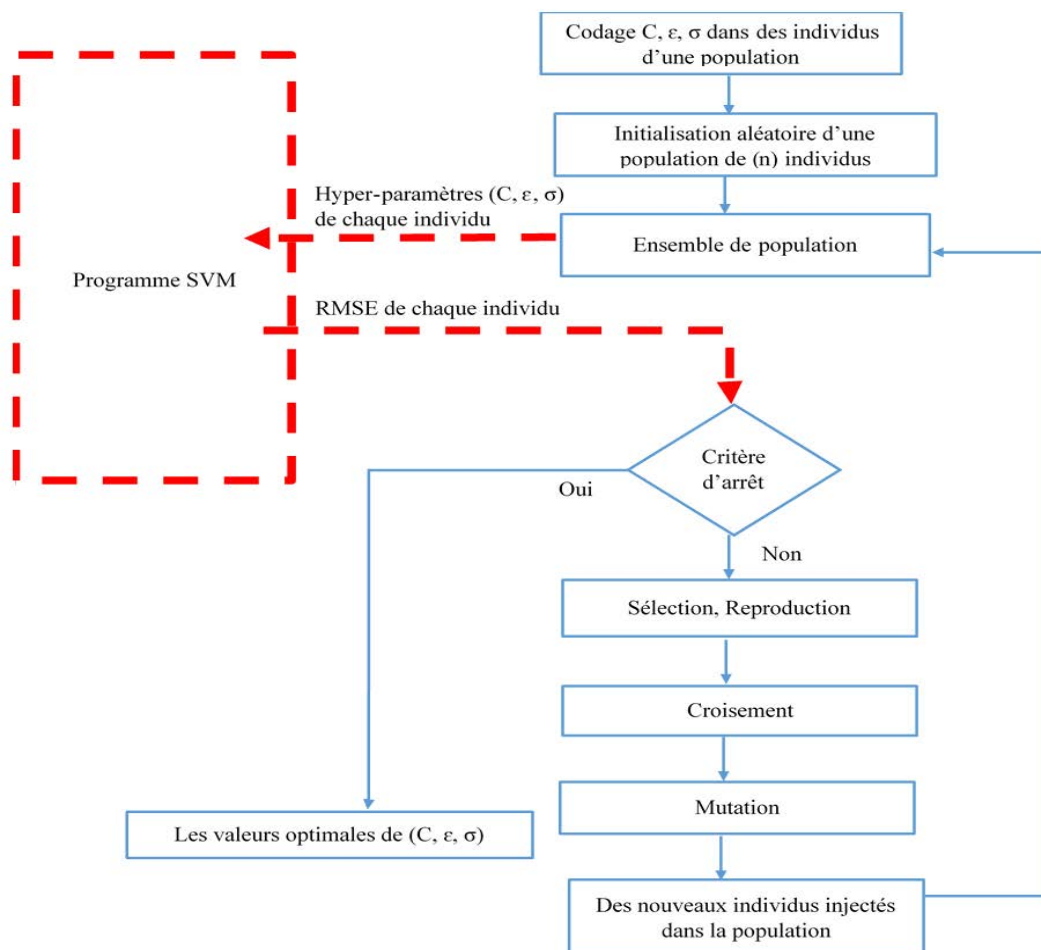


Figure III- 5 Schéma global de la procédure d'optimisation de la SVM par l'algorithme génétique.

### **III.4.2 Recuit simulé**

Le recuit simulé est un algorithme d'optimisation qui simule le comportement des métaux dans le processus du recuit utilisé dans la métallurgie [Kirkpatrick et al. 1983]. Le but du processus du recuit est d'atteindre un état d'équilibre thermodynamique où l'énergie est minimale. Dans l'algorithme de recuit simulé la fonction objectif est représentée par l'énergie du système. L'algorithme de recuit simulé utilise un paramètre fictif de température qui subit une modification selon le programme de refroidissement afin de diminuer la fonction objectif. Le critère « Metropolis » permet de prendre la décision d'accepter ou de refuser la nouvelle solution de la manière suivante :

Si cette modification fait diminuer la fonction objectif (ou l'énergie du système), elle est directement acceptée ; sinon, elle n'est acceptée qu'avec une probabilité égale à  $\exp(\Delta E/T)$  (où  $\Delta E$  est la différence de l'énergie de la solution actuelle et celle de la nouvelle solution, et  $T$  représente la température).

L'algorithme de recuit simulé contient les 5 étapes essentielles suivantes :

- 1) fixer la température de départ  $T = T_0$  et la solution initiale  $x_0$  ;
- 2) comparer la solution actuelle et la solution générée dans le voisinage de la solution actuelle selon le critère de Metropolis ;
- 3) répéter l'étape 2 jusqu'à ce que la condition d'arrêt soit satisfaite ;
- 4) diminuer la température selon le programme de refroidissement comme par exemple :

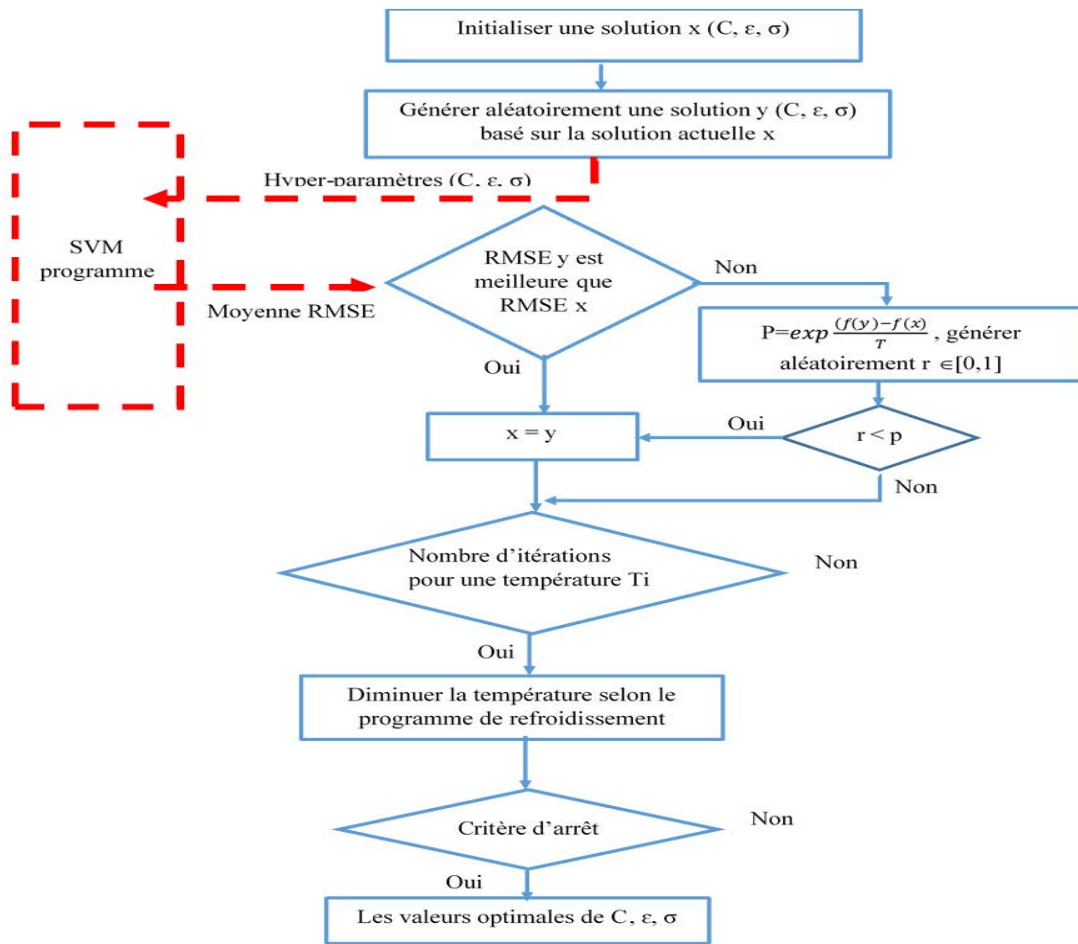
$$T_{i+1} = \alpha T_i \quad \text{avec } (\alpha < 1) \quad \text{(III-24)}$$

- 5) répéter les étapes 2 à 4 jusqu'à ce que le système soit gelé.

On peut constater trois phases dans l'algorithme de recuit simulé.

- La température choisie au début est très grande, ce qui permet d'accepter plusieurs solutions, même celles qui ont une énergie supérieure à la solution principale.
- Quand la température baisse, l'algorithme accepte de temps en temps des solutions qui augmentent l'énergie, ce qui peut permettre d'échapper à un minima local.
- Dans la troisième phase la température baisse nettement, ce qui entraîne à refuser la quasi-totalité des solutions qui augmentent l'énergie, et la méthode de recuit simulé devient un algorithme d'optimisation avec une amélioration itérative classique.

Nous avons illustré le principe de l'algorithme de recuit simulé dans la figure III-6.



**Figure III- 6 Schéma global de la procédure d'optimisation de la SVM par l'algorithme de recuit simulé.**

### III.4.3 Optimisation par essais particuliers

L'algorithme d'optimisation par essais particuliers (PSO) est introduit par James Kennedy et Russell Eberhart. Cet algorithme, inspiré du vol groupé d'oiseaux montre que ses individus harmonisent leurs mouvements en gardant une distance optimale entre eux, et suivent un mouvement global convergeant vers la destination finale [Bai 2010].

Chacune des particules a une position et une vitesse qui lui permet de se déplacer. Chaque particule tente de changer sa position selon sa meilleure position, sa position précédente ainsi que son meilleur voisin grâce à sa capacité de communiquer avec les autres particules afin de converger vers une particule optimale.

Avant de présenter l'algorithme il faut introduire les notions suivantes :

$x_i(t)$  : la position d'une particule dans l'espace de recherche à l'instant  $t$  ;

$v_i(t)$  : la vitesse d'une particule à l'instant  $t$  ;



### Chapitre III : Méthodes multi-variables pour l'estimation de la concentration de gaz à l'aide d'un réseau de capteurs

$x_{pbest_i}(t)$  : la position de la meilleure solution par laquelle la particule est passée ;

$x_{vbest_i}(t)$  : la position de la meilleure solution connue de son voisinage ;

$pbest_i$  : la meilleure solution de la fonction objectif pour la particule ;

$vbest_i$  : la meilleure solution globale connue.

A chaque instant l'algorithme évalue la meilleure position pour chaque particule.

Si  $f(x_i) > pbest_i$  alors  $pbest_i = f(x_i)$  et  $x_{pbest_i} = x_i$

Si  $f(x_i) > vbest_i$  alors  $vbest_i = f(x_i)$  et  $x_{vbest_i} = x_i$

A la fin de chaque itération la position et la vitesse de chaque particule sont modifiées comme suit :

$$v_i = \Psi(v_i + \rho_1(x_{pbest_i} - x_i) + \rho_2(x_{vbest_i} - x_i)) \quad (III-25)$$

$$x_i = x_i + v_i \quad (III-26)$$

où  $\Psi, \rho_1, \rho_2$  sont des variables aléatoires.

La figure III-7 résume l'algorithme d'optimisation par essais particuliers.

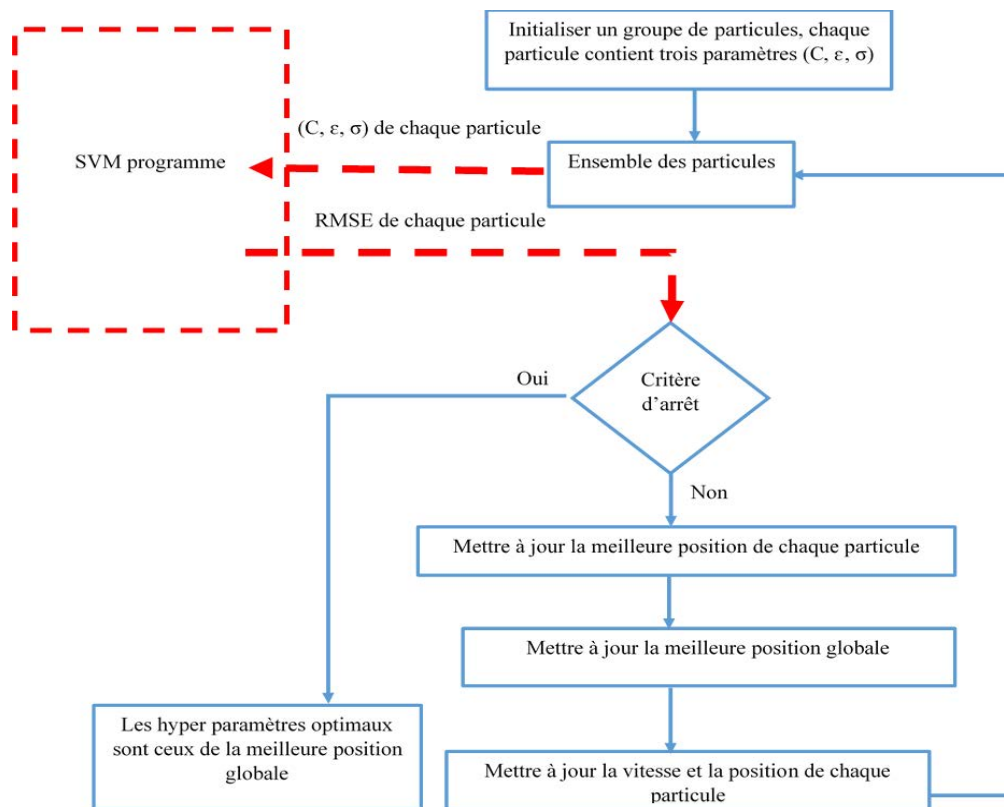


Figure III- 7 Schéma global de la procédure d'optimisation de la SVM par l'algorithme d'optimisation par essais particuliers.

### III.4.4 Recherche directe GPS

Les méthodes de recherche directe sont des méthodes de résolution des problèmes d'optimisation qui ne nécessitent aucune information sur les dérivées de la fonction objectif, contrairement aux méthodes d'optimisation traditionnelles qui utilisent des informations sur la dérivée pour rechercher un point optimal. Un algorithme de recherche directe cherche un ensemble de points autour du point actuel, pour lesquels la valeur de la fonction objectif est inférieure à la valeur au point actuel. Ces méthodes sont très pratiques pour résoudre des problèmes pour lesquels la fonction objectif n'est pas différentiable. L'algorithme de recherche par motifs (GPS) est l'une de ces méthodes de recherche directe introduite et analysée par Torczon [Torczon 1997]. L'idée de base est d'effectuer une exploration dans les différentes directions pour une séquence d'itérations afin de diminuer la valeur de la fonction objectif. Par exemple, si le problème d'optimisation est écrit sous la forme :

$$\min f(x) \text{ avec } x \in \mathfrak{R}^d$$

pour chaque itération  $k$ , l'algorithme évalue la fonction objectif  $f(x)$  à des points autour de la solution courante  $x_k$  dont le but est de trouver un point pour lequel la fonction objectif est plus faible que celle du point courant. Ces points se trouvent à un pas de longueur de  $\Delta_k$  suivant  $2d$  directions ( $d$  étant le nombre de paramètres à optimiser). Nous avons utilisé ce type d'opérations de recherche, appelé « recherche par sonde locale », où son centre est le point courant. L'ensemble des points d'une sonde est alors défini par :

$$P_k = \{x_k \pm \Delta_k D_i, \quad i = 1, \dots, d\}$$

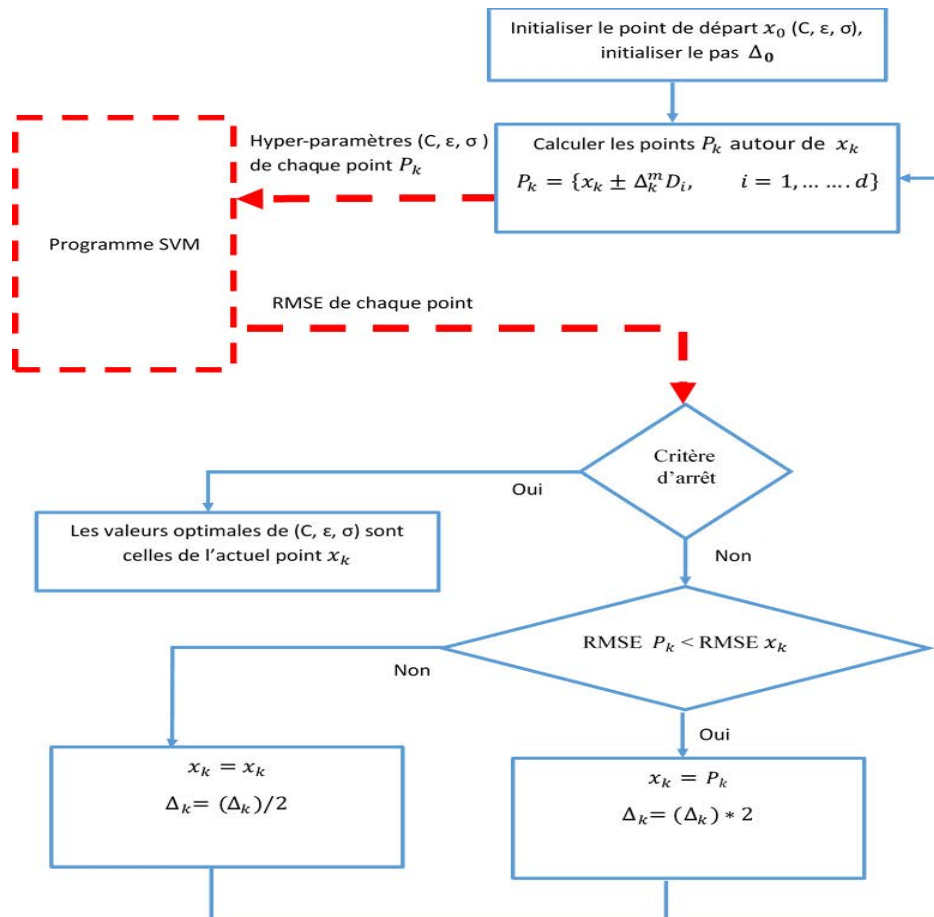
$D_i$  étant la direction.

Une itération est considérée comme réussie si nous trouvons un point qui diminue la fonction objectif. Dans le cas contraire, c.-à-d. si aucun des points ne permet une diminution de la fonction objectif, l'itération est considérée comme itération échouée. Après chaque itération réussie, l'algorithme passe à la nouvelle solution et la taille du pas augmente pour aller rapidement vers une nouvelle solution. Dans le cas d'une itération échouée l'algorithme conserve la solution actuelle et diminue le pas pour évaluer des points plus proches de la solution actuelle. L'adaptation de la taille du pas se fait comme suit :

$$\Delta_{k+1} = \tau \Delta_k \quad \text{(III-27)}$$

tel que  $\tau > 1$  après une itération réussie et  $\tau < 1$  après une itération échouée.

L'algorithme de recherche par motifs est présenté dans la figure III-8.



**Figure III- 8 Schéma global de la procédure d'optimisation de la SVM selon l'algorithme de recherche par motifs.**

### III.5 Application et comparaison des algorithmes d'optimisation

Les données utilisées sont collectées sur la station de surveillance de la pollution atmosphérique durant la période de juillet à septembre 2018. Nous avons utilisé 1500 données réparties uniformément sur cette période (environ une donnée par heure, soit 1\*24 heures \*62.5 jours). Pour évaluer les performances des algorithmes d'optimisation, nous avons calculé la moyenne de l'erreur quadratique (moyenne RMSE des validations croisées en partitionnant les données en 10 partitions).

#### III.5.1 Application de l'algorithme GPS

Nous présentons en détail l'application de la méthode GPS, ensuite nous comparons les performances de cette méthode en termes de RMSE et de temps de calcul avec celles obtenues par l'application des trois autres méthodes citées.

La figure III-9 illustre l'évolution de la minimisation de la fonction objectif (moyenne RMSE) à chaque itération. Nous pouvons constater des améliorations importantes et rapides au début, suivies par des améliorations lentes et marginales au fur et à mesure que nous approchons de la solution optimale. La figure III-10 illustre l'évolution de la taille du pas à chaque itération. En superposant ces deux figures, on remarque qu'après une itération réussie, le pas de la grille augmente, alors qu'il diminue après un échec.

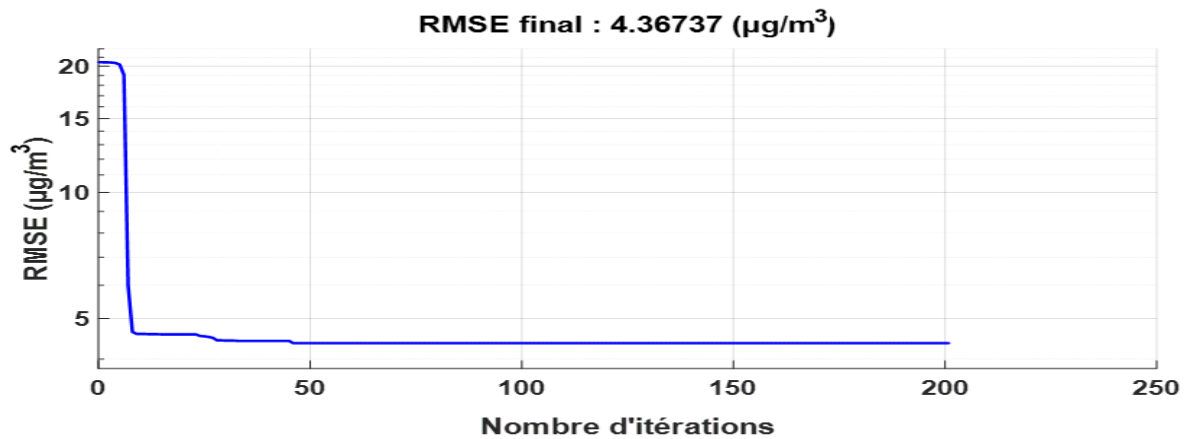


Figure III- 9 Evolution de l'erreur quadratique moyenne à chaque itération.

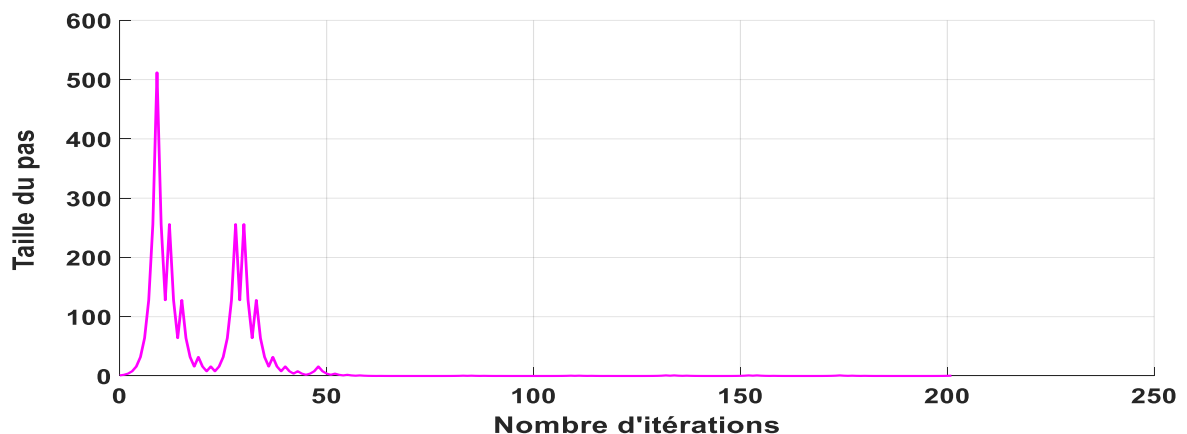


Figure III- 10 Evolution de la taille du pas à chaque itération.

Comme tous les algorithmes d'optimisation, la recherche par motifs (GPS) présente certaines contraintes liées au cas où :1) la fonction objectif est non continue, 2) le nombre des paramètres à optimiser est grand et 3) des minima locaux sont à éviter. Les deux premiers problèmes ne concernent pas notre étude car la fonction objectif est continue pour toutes valeurs des variables dans l'intervalle  $]0, \infty[$ . Le nombre d'hyper-paramètres à optimiser est faible car uniquement les valeurs de deux variables ( $C, \sigma$ ) doivent être optimisées, le troisième hyper-paramètre ( $\varepsilon$ ) est fixe. Par conséquent, dans notre cas, le nombre d'hyperparamètres n'a pas

d'influence significative sur la vitesse de l'algorithme GPS. Le seul problème restant est celui lié aux minimas locaux car la surface de la fonction objectif n'est pas lisse. Ceci peut conduire à piéger l'algorithme dans un minimum local. Par exemple, si le pas de la sonde est trop petit et que toutes les directions ne fournissent pas une amélioration, l'algorithme diminuera encore le pas de la sonde, en rendant impossible d'échapper à cette mauvaise solution.

Pour étudier ce problème, nous avons choisi des cas où la surface n'est pas lisse. La surface représente la fonction objectif où  $\varepsilon$  est fixé à 0.001,  $\sigma$  varie de 1 à 25 et  $C$  de 1 à 20000. Sur la figure III-11, on peut remarquer que la surface n'est pas lisse et contient plusieurs minimas locaux. Dans notre test, nous avons choisi une solution initiale avec  $\varepsilon = 0.001$ ,  $\sigma = 1$  et  $C = 12000$ . Dans ce cas l'algorithme doit éviter des minimas locaux (visibles dans la figure III-11) avant d'arriver à la solution optimale. Les résultats numériques montrent que le GPS peut éviter avec succès plusieurs minimas locaux et converger vers le meilleur minimum local dans cette zone ( $C = 18000$ ,  $\sigma = 13.2$ ,  $\varepsilon = 0.001$ ).

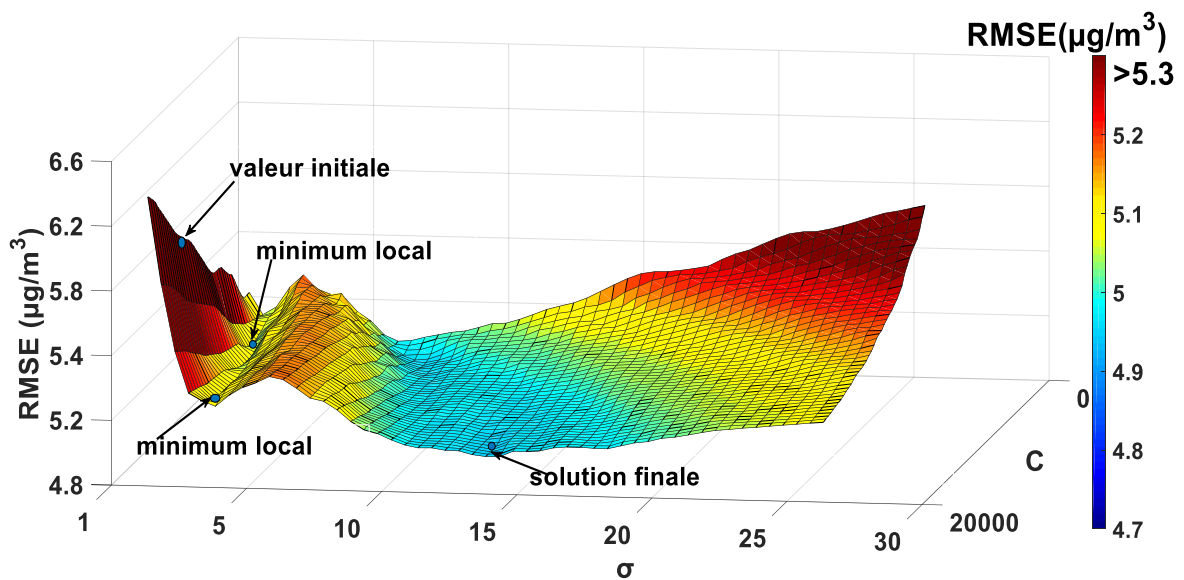


Figure III- 11 Performance de l'algorithme GPS dans le cas des minimas locaux.

Nous avons également testé l'efficacité du GPS. Pour cela différentes valeurs initiales sont sélectionnées pour les deux hyper-paramètres ( $C, \sigma$ ). Ces sélections de valeurs de ( $C, \sigma$ ) permettent de vérifier si l'algorithme peut atteindre la solution optimale. Pour chaque valeur initiale, nous comparons les erreurs quadratiques moyennes de prédiction RMSE. Cette erreur est déterminée par le calcul de la moyenne des RMSE des 10 partitions de la validation croisée.

Les valeurs initiales ont été choisies dans tous les coins d'un vaste espace de recherche  $\{C \in [1: 300000], \sigma \in [1: 200], \varepsilon = 0.001\}$ . Les résultats présentés dans le tableau III-1 montrent qu'à partir des points initiaux différents, le GPS sélectionne différentes valeurs d'hyper-

## Application et comparaison des algorithmes d'optimisation

paramètres optimaux, mais avec la même prédiction d'erreur. Les différentes valeurs optimales d'hyper-paramètres pourraient être justifiées par le fait que plusieurs combinaisons peuvent assurer une bonne solution, comme mentionné précédemment (figure III-3). En effet l'algorithme GPS choisit la solution la plus proche ou la plus rapide à partir du point initial. Nous pouvons conclure que le GPS peut atteindre des valeurs optimales d'hyper-paramètres avec différentes combinaisons, quelle que soit la solution initiale.

Valeurs initiales		Solutions finales		Moyenne RMSE ( $\mu\text{g}/\text{m}^3$ )
$C$	$\sigma$	$C$	$\sigma$	
1	200	273470	15	<b>4.3709</b>
1	1	77163	12	<b>4.3709</b>
100000	200	273470	15	<b>4.3706</b>
100000	1	131693	13.07	<b>4.3734</b>

**Tableau III- 1 Performance de l'algorithme GPS avec différentes valeurs initiales.**

### III.5.2 Applications d'autres algorithmes d'optimisation et comparaison

Afin de confirmer la pertinence de l'algorithme de recherche du modèle, nous avons testé puis comparé cet algorithme avec d'autres algorithmes d'optimisation, soit l'algorithme génétique (GA), l'optimisation PSO et le recuit simulé (SA). Nous avons choisi ces algorithmes dans leur forme de base. Cependant, de nombreuses versions modifiées de ces algorithmes améliorent leurs performances. Mais, l'objectif de cette étude est de proposer un algorithme simple pour le calcul des hyper-paramètres, c'est pourquoi, pour comparer ces algorithmes entre eux, nous avons cherché à limiter les efforts en termes de paramétrage. Dans le cadre de ces comparaisons, nous avons utilisé Matlab et les outils du Toolbox d'optimisation. Les paramètres pris par défaut pour chaque algorithme sont les suivants :

- Algorithme génétique :

Population = 50 individus, taux de croisement = 0.8 ; taux de mutation = 0.05.

- PSO :

Nombre des particules = 30 ; paramètre  $\rho_1 = 1.49$  ; paramètre  $\rho_2 = 1.49$ .

- SA :

Température initiale  $T_0 = 100$  ; protocole de refroidissement =  $T_0 * 0.95 * \text{nombre d'itérations déjà effectuées}$  ; probabilité d'acceptation =

$$\frac{1}{1 + \exp\left(\frac{f(x)\text{actuelle} - f(x)\text{précédente}}{T}\right)}$$

- GPS :

### Chapitre III : Méthodes multi-variables pour l'estimation de la concentration de gaz à l'aide d'un réseau de capteurs

Taille du pas initiale  $\Delta_0 = 1$  ; taille du pas après une itération réussie  $\Delta = 2 * \Delta$  ; le pas après une itération échouée  $\Delta = \frac{\Delta}{2}$ .

Tous les algorithmes ont utilisé le même espace de recherche  $\{C \in [1 : 300000], \sigma \in [1 : 100], \varepsilon = 0.001\}$ . Sur la base des résultats présentés dans le tableau III-2, nous pouvons dire que l'algorithme GPS offre les mêmes performances que celles des autres algorithmes mais avec une meilleure rapidité par rapport aux algorithmes PSO et GA. Nous remarquons que chaque algorithme donne différents hyper-paramètres de la SVM pour sa meilleure performance.

	GPS	GA	PSO	SA
$C$	131489	235150	300000	378
$\sigma$	13.07	14.41	23.6	4.7
$\varepsilon$	0.001	0.001	0.001	0.001
Moyenne RMSE( $\mu\text{g}/\text{m}^3$ )	4.364	4.369	4.363	4.64
Temps de calcul (s)	<b>2550</b>	14470	20160	3900
Nombre d'itération	<b>200</b>	200	200	200

**Tableau III- 2 Comparaison entre les différents algorithmes d'optimisation en termes de précision et de rapidité.**

#### III.6 Application des modèles d'étalonnage et comparaison

Dans cette partie, nous comparons l'efficacité des méthodes PLS, ELM et SVM pour l'estimation de la concentration de dioxyde d'azote en utilisant des données collectées sur le terrain à l'intérieur de la station d'ATMO Grand Est. Les performances de chaque méthode sont évaluées en termes de précision en utilisant comme indicateur l'erreur quadratique moyenne de la validation croisée (RMSE moyenne). Nous comparons également l'influence du nombre d'échantillons nécessaires pour l'apprentissage.

Avant de passer à la comparaison nous commençons d'abord par la présentation du principe de fonctionnement des méthodes PLS et ELM. Nous rappelons que le principe de la méthode SVM a été décrit au début de ce chapitre.

##### III.6.1 Extreme Learning Machine (ELM)

*Extreme learning machine* (ELM) est un algorithme basé sur les réseaux de neurones proposé par Huang [Huang et al.2006]. Cet algorithme est très simple, rapide et offre de bonnes performances en termes de précision et généralisation. ELM a été largement utilisé dans plusieurs domaines comme la vision par ordinateur, la robotique, et l'ingénierie biomédicale.

## Application des modèles d'étalonnage et comparaison

---

C'est un type de réseau de neurones à propagation avant (feed-forwarded), composé d'une couche d'entrée, d'une seule couche cachée et d'une couche de sortie. Comme dans chaque méthode de réseau de neurones, les neurones sont liés à l'aide d'une connexion pondérée par un poids ( $w$ ) et ils sont caractérisés par une fonction d'activation  $g(x)$  et un seuil d'activation  $b$ .

Contrairement aux autres réseaux de neurones qui utilisent la rétro-propagation avec un processus itératif pour déterminer les poids ( $w$ ), ELM sélectionne aléatoirement les poids reliant la couche d'entrée et la couche cachée ainsi que les seuils de chaque neurone de la couche cachée. Ensuite, la méthode des moindres carrés ordinaire est utilisée pour déterminer les poids reliant la couche cachée et la sortie [Huang et al. 2006].

Ce principe se traduit par le résumé suivant :

Pour  $N$  échantillons distincts  $(x_i, y_i)$  d'apprentissage où  $x_i = [x_{i1}, x_{i2}, \dots, x_{id}]^T \in \mathbb{R}^d$  représente le vecteur d'entrée du  $i^{\text{ème}}$  échantillon de dimension  $d$  ;

la sortie  $\hat{y}_i$  estimée par le modèle ELM est donnée par :

$$\hat{y}_i = \sum_{j=1}^k \beta_j g(w_j \cdot x_i + b_j) \quad \text{(III-28)}$$

$y_i$  : la sortie du  $i^{\text{ème}}$  échantillon ;

$k$  : le nombre de neurones dans la couche ;

$g(x)$  : la fonction d'activation ;

$w_j = [w_{j1}, w_{j2}, \dots, w_{jd}]^T$  : le vecteur des poids connectant les couches d'entrée au  $j^{\text{ème}}$  neurone de la couche cachée ;

$b_j$  : le seuil associé au  $j^{\text{ème}}$  neurone de la couche cachée ;

$\beta_j$  : le poids connectant le  $j^{\text{ème}}$  neurone de la couche à la sortie  $y_i$  ;

$w_j \cdot x_i$  : le produit scalaire de  $w_j$  et  $x_i$ .

L'équation III-28 peut être écrite sous la forme matricielle suivante :

$$H\beta = \hat{y} \quad \text{(III-29)}$$

où  $H$  est la matrice de sortie de la couche cachée du réseau ELM soit :



$$H = \begin{bmatrix} g(w_1x_1 + b_{j1}) & \dots & g(w_kx_1 + b_k) \\ \vdots & \ddots & \vdots \\ g(w_1x_N + b_1) & \dots & g(w_kx_N + b_k) \end{bmatrix} \quad \text{(III-30)}$$

$$\text{avec : } \beta = \begin{pmatrix} \beta_1^T \\ \vdots \\ \beta_k^T \end{pmatrix} \text{ et } y = \begin{pmatrix} y_1^T \\ \vdots \\ y_N^T \end{pmatrix}$$

L'apprentissage de la méthode ELM est équivalent à trouver la solution de la norme minimale des moindres carrés comme suit:

$$\beta = H^* y \quad \text{(III-31)}$$

où  $H^*$  est donnée par l'équation suivante :

$$H^* = (H^T H)^{-1} H^T \quad \text{(III-32)}$$

En fonction de  $\beta$ , l'erreur de l'apprentissage peut être améliorée tout en optimisant les paramètres du réseau (nombre de neurones de la couche cachée, type et paramètres de la fonction d'activation).

### III.6.2 Régression des moindres carrés partiels (PLS)

Les méthodes de régression linéaire sont les plus utilisées dans notre type d'application en raison de leur simplicité et fiabilité. Elles correspondent à la création d'un modèle mathématique qui relie les données d'entrées appelées les variables explicatives et les données de sortie appelées les variables à expliquer à travers des coefficients de régression :

$$y_i = a_1x_{i1} + a_2x_{i2} + \dots + a_dx_{id} \quad \text{(III-33)}$$

On peut écrire l'équation III 33 sous la forme matricielle :

$$Y = A * X + \varepsilon \quad \text{(III-34)}$$

$$X^T Y = X^T X A \quad \text{(III-35)}$$

$$A = [X^T X]^{-1} X^T Y \quad \text{(III-36)}$$

La solution est clairement basée sur la matrice  $X^T Y$  et la matrice  $X^T X$ .

Le problème qui se pose pour trouver cette solution est que l'inverse de la matrice  $X^T X$  est instable car les variables sont généralement fortement corrélées.

Pour remédier à ce problème, une solution très intelligente consiste à remplacer la matrice des variables d'entrée  $X$  par d'autres variables latentes totalement non corrélées avant d'appliquer la régression linéaire multiple.

La projection des variables initiales en variables latentes orthogonales linéairement indépendantes peut se faire en utilisant la décomposition en composantes principales PCA. Or l'PCA est basée sur le calcul des directions de plus grande dispersion des échantillons et cela ne correspond pas forcément aux directions qui concernent la variable à expliquer :  $Y$ .

Par contre, la méthode PLS peut résoudre ce problème en tenant compte de la variabilité de la matrice d'entrée  $X$  et la sortie  $Y$  simultanément [Jurs et al. 2000; Geladi et Kowalski 1986]. En d'autres termes, l'ACP est basée sur la décomposition de la matrice  $([X^t][X])$  en cherchant les directions qui maximisent la variance de  $[X]$  alors que PLS est basée sur la décomposition de la matrice  $([X^t Y] [Y^t X])$  en cherchant les axes qui maximisent la covariance  $(X, Y)$ .

### III.6.3 Résultats de la comparaison

Nous avons sélectionné pour chaque méthode les paramètres qui permettent d'obtenir les meilleures performances. Le tableau III-3 résume les hyper-paramètres optimaux de chaque méthode. Ceux de la méthode SVM sont obtenus en utilisant l'algorithme GPS. Ceux de la méthode ELM sont obtenus par une recherche exhaustive en testant les différentes fonctions d'activation les plus utilisées avec une variation du nombre de neurones de la couche cachée entre 20 et 200 neurones. Pour la méthode PLS nous avons trouvé la meilleure performance en prenant toutes les composantes PLS.

Méthode	Paramètres optimaux		
SVM	$C = 220700$	$\varepsilon = 0.001$	$\sigma = 9.4$
ELM	Nombre de neurones de la couche cachée =90		La fonction d'activation sigmoïde : $(1/(1+\exp(-x)))$
PLS	Nombre de composantes PLS =7		

**Tableau III- 3 Paramètres optimaux utilisés pour chaque méthode de regression.**

Pour comparer les trois méthodes, nous avons calculé la moyenne des RMSE obtenus par validation croisée des 10 partitions réalisées sur les données collectées entre février et avril 2019, en utilisant les mesures des capteurs NO2-B43F, OX-B431, MiCS-2714 ainsi que les valeurs de la température et de l'humidité fournies par ATMO Grand Est. La figure III-12 représente les erreurs de prédiction des concentrations de NO<sub>2</sub> obtenues par les méthodes PLS,

### Chapitre III : Méthodes multi-variables pour l'estimation de la concentration de gaz à l'aide d'un réseau de capteurs

SVM et ELM. Nous remarquons que les méthodes non linéaires comme la SVM et l'ELM sont plus performantes que la méthode PLS. La valeur de la moyenne RMSE obtenue est de  $4.43\mu\text{g}/\text{m}^3$  pour la SVM, de  $6.58\mu\text{g}/\text{m}^3$  pour l'ELM et de  $9.77\mu\text{g}/\text{m}^3$  pour la méthode PLS.

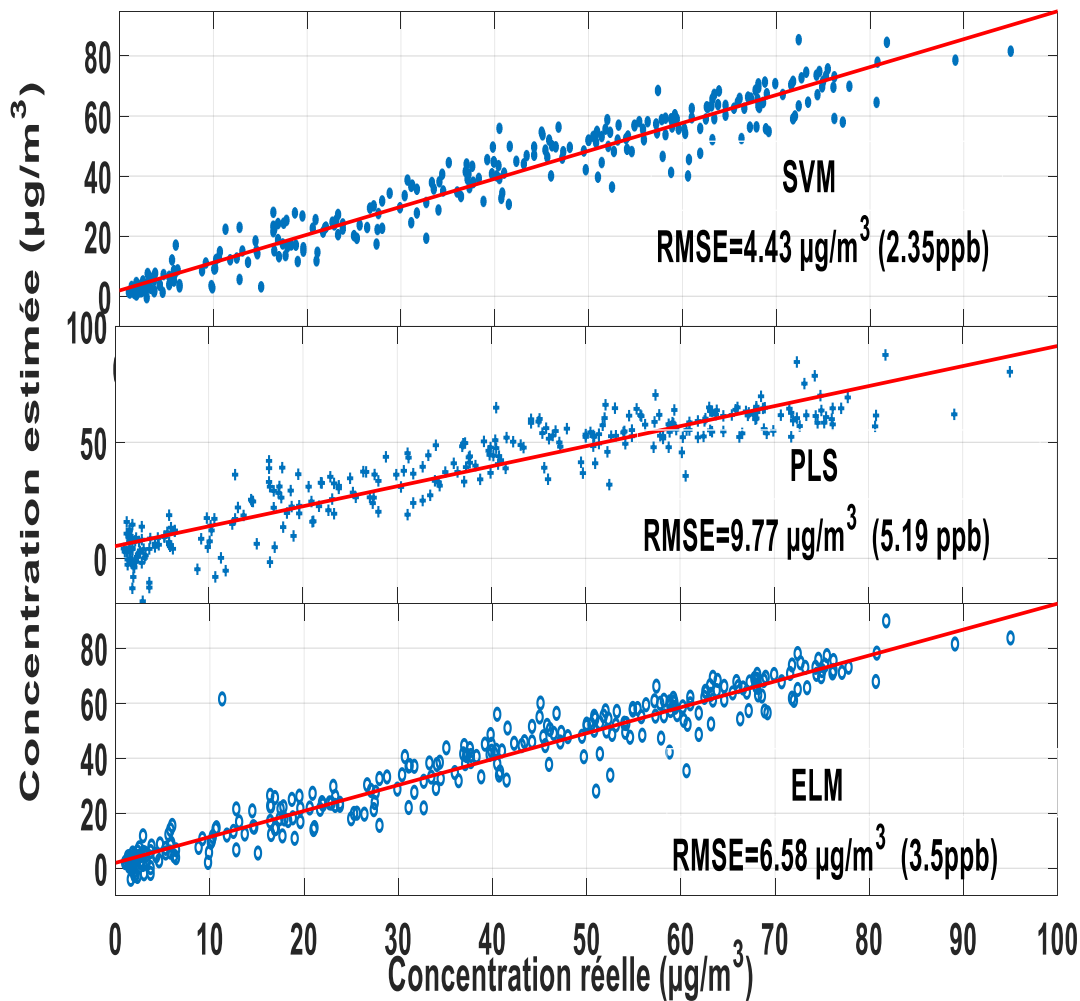


Figure III- 12 Comparaison entre les méthodes SVM, PLS et ELM pour l'estimation des taux de dioxyde d'azote sur le terrain.

Un point important dans le choix de la méthode d'étalonnage des systèmes est le nombre d'échantillons nécessaires, afin d'obtenir un modèle précis. Afin de tester laquelle de ces trois méthodes nécessite moins d'échantillons d'apprentissage pour créer un modèle d'étalonnage stable avec une bonne précision, nous avons construit un sous-ensemble d'échantillons les plus représentatifs en utilisant un algorithme de sélection d'échantillons qui est le Kernel SPXY. Tout d'abord, nous avons construit un modèle à partir de 150 échantillons afin d'avoir une précision acceptable de l'ordre de  $15\mu\text{g}/\text{m}^3$ . Ensuite, à chaque étape, nous avons rajouté un autre échantillon pour l'apprentissage. La figure III-13 illustre l'évolution de la moyenne des RMSE, obtenue par validation croisée des 10 partitions, en fonction du nombre d'échantillons.

## Etude de l'influence de la température

Nous remarquons que les méthodes SVM et PLS donnent des modèles plus stables par rapport au modèle généré par la méthode ELM.

Après un certain nombre d'échantillons (ici environ 700 échantillons) la méthode SVM surpasse PLS et donne un modèle plus stable et plus précis. La méthode ELM génère aléatoirement des poids reliant la couche d'entrée et la couche cachée ce qui explique l'instabilité du modèle. Cependant le modèle obtenu par la méthode ELM s'améliore en fonction du nombre d'échantillons utilisé dans l'apprentissage.

Après cette comparaison nous pouvons conclure que le modèle SVM est le plus adapté à notre application en termes de stabilité et de précision.

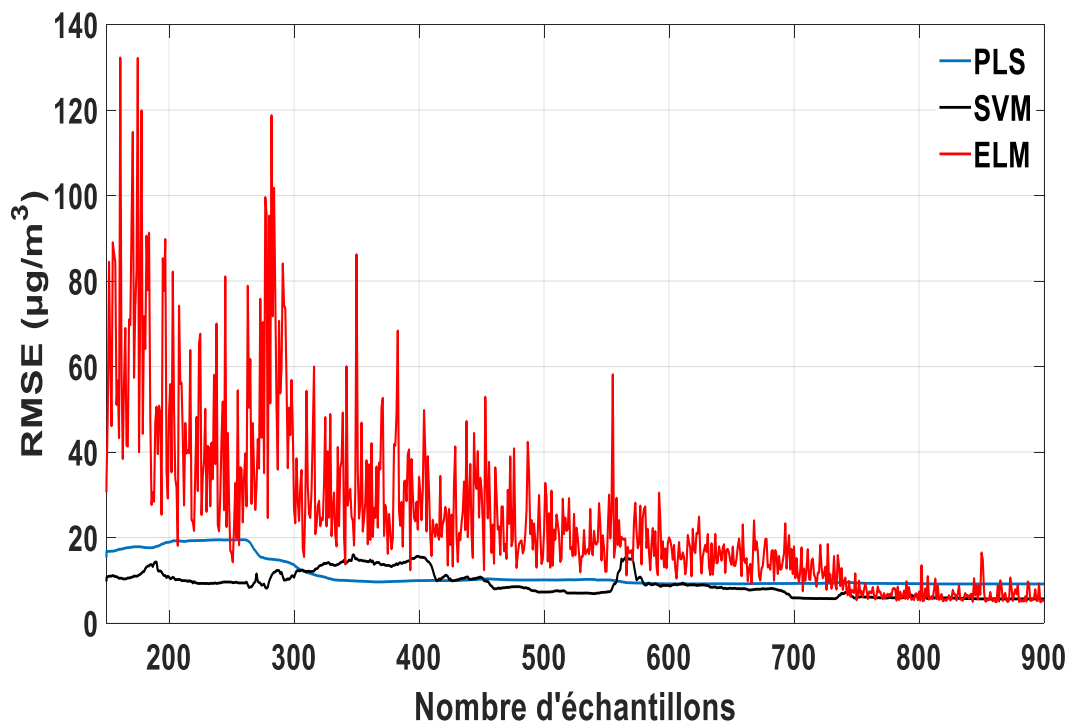


Figure III- 13 Evolution de la moyenne de RMSE en fonction du nombre d'échantillons d'apprentissage.

### III.7 Etude de l'influence de la température

Cette étude est réalisée à partir des données recueillies sur le dispositif placé à l'extérieur de la station sur trois mois, de février à avril 2019. L'algorithme SVM est appliqué sur les données des micro-capteurs avec et sans prise en compte de la variation de température dans le modèle. La projection des concentrations estimées dans chacun des deux cas, en fonction des concentrations réelles (fournies par l'analyseur d'ATMO Grand Est) sont illustrées sur la figure III-14. Le calcul de la moyenne des erreurs montre l'amélioration de l'estimation avec une RMSE moyenne de  $5.36\mu\text{g}/\text{m}^3$  lorsque la température est prise en compte dans le modèle SVM,

alors que l'erreur de l'estimation est de  $9.01\mu\text{g}/\text{m}^3$  lorsque les valeurs de la température ne sont pas incluses dans le modèle.

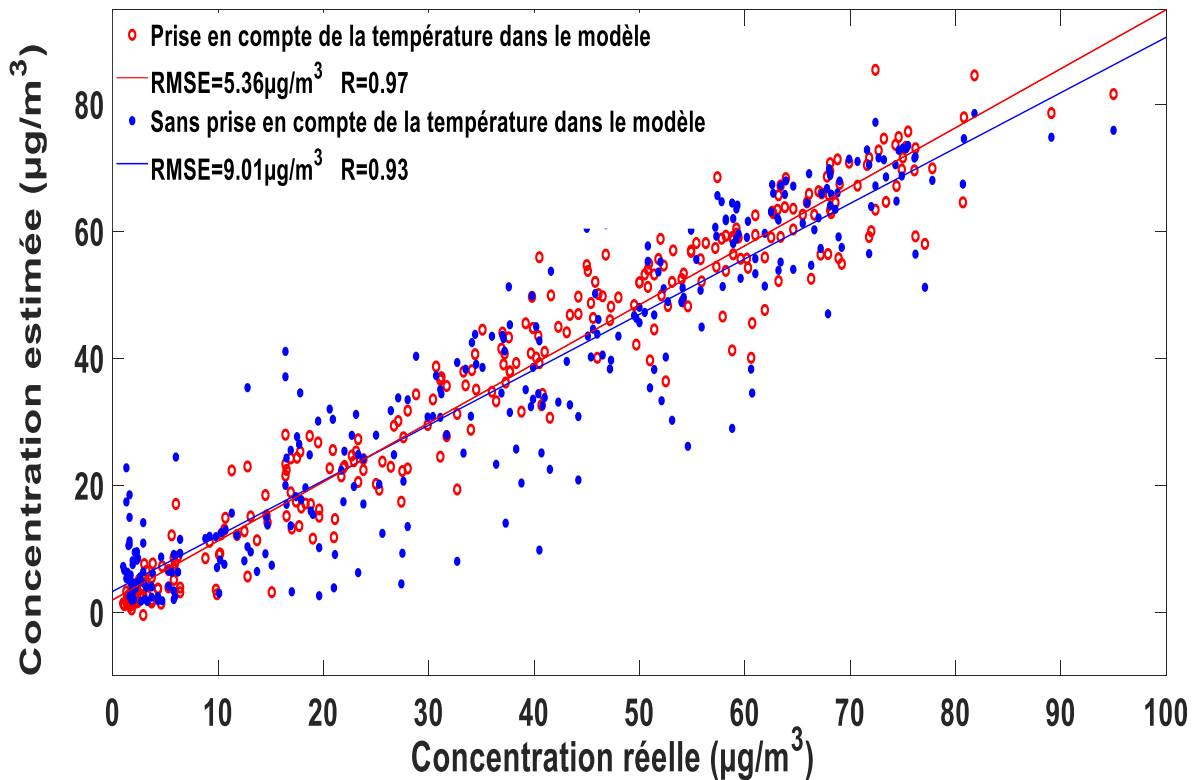


Figure III- 14 Comparaison entre les modèles d'étalonnage qui incluent ou pas les valeurs de la température.

#### III.8 Comparaison entre l'utilisation individuelle des capteurs et la fusion des données des capteurs par les méthodes multi-variables

Nous comparons dans le tableau III-4 les résultats obtenus pour les mesures du dioxyde d'azote et de l'ozone sur le terrain afin de montrer le gain apporté en termes de précision d'estimation lorsqu'on fusionne toutes les données des capteurs à l'aide de méthodes multi-variables. Nous constatons en particulier que pour la mesure du  $\text{NO}_2$ , le capteur électrochimique  $\text{NO}_2\text{-B43F}$  est meilleur que le capteur à oxyde métallique  $\text{MiCS-2714}$ . Bien que le capteur à oxyde métallique ne donne pas de bonne estimation du taux de  $\text{NO}_2$ , il fournit des informations utiles qui permettent d'améliorer les résultats obtenus par le capteur électrochimique. En ce qui concerne l'estimation de l'ozone on peut aussi remarquer que l'utilisation d'un modèle multi-variables améliore considérablement les résultats obtenus lorsqu'on utilise chaque capteur individuellement.

## Conclusion

RMSE de l'estimation du NO <sub>2</sub> à l'intérieur de la station d'ATMO Grand Est		
NO2-B43F	MiCS-2714	NO2-B43F + MiCS-2714 + OX-B431 en utilisant la méthode SVM
<b>9.16µg/m<sup>3</sup></b>	<b>21µg/m<sup>3</sup></b>	<b>4.43µg/m<sup>3</sup></b>
RMSE de l'estimation du O <sub>3</sub> à l'intérieur de la station d'ATMO Grand Est		
NO2-B43F et OX-B431 individuellement		NO2-B43F + MiCS-2714 + OX-B431 en utilisant la méthode SVM
<b>12.12µg/m<sup>3</sup></b>		<b>6.5µg/m<sup>3</sup></b>

**Tableau III- 4 Amélioration obtenue en utilisant la méthode SVM et en combinant les réponses des trois capteurs.**

### III.9 Conclusion

Dans ce chapitre, nous avons étudié l'efficacité de la régression SVM, appelée SVR pour l'estimation de la concentration des gaz polluants à l'aide d'un nez électronique. Les données utilisées sont collectées sur le terrain avec notre dispositif.

D'abord nous avons étudié l'optimisation des hyper-paramètres de la régression SVM et nous avons démontré comment chacun des trois hyper-paramètres affecte les performances de la régression SVM d'une manière différente. L'hyper-paramètre de régulation  $C$  influe sur le compromis entre la complexité du modèle et le risque empirique. L'hyper-paramètre de la fonction noyau  $\sigma$  peut augmenter ou diminuer l'influence des vecteurs supports et enfin l'hyper-paramètre  $\varepsilon$  contrôle directement le nombre de vecteurs supports. Nous avons vu que plusieurs combinaisons d'hyper-paramètres peuvent donner une bonne solution car un hyper-paramètre en excès peut être compensé par un autre hyper-paramètre. Nous avons proposé la méthode de recherche par motifs généralisés (GPS) comme une alternative appropriée à la recherche par grille principalement référencée dans la littérature pour le paramétrage des hyper-paramètres. Pour notre application, l'algorithme GPS est une solution appropriée car il offre une précision, une rapidité, une efficacité et une robustesse optimales. Nous avons mené plusieurs expériences numériques pour illustrer sa robustesse, et les résultats ont montré la capacité de l'algorithme GPS à trouver une combinaison d'hyper-paramètres qui offre de bonnes performances, lorsque

### **Chapitre III : Méthodes multi-variables pour l'estimation de la concentration de gaz à l'aide d'un réseau de capteurs**

---

les valeurs initiales des hyper-paramètres se situent dans différentes régions de l'espace des hyper-paramètres. Cependant l'algorithme GPS a été comparé à trois autres algorithmes d'optimisation en termes de robustesse, d'efficacité et de rapidité de calcul. La comparaison des résultats obtenus montre que l'algorithme GPS est le plus rapide et que sa performance en termes de précision est similaire aux deux algorithmes GA et PSO. Ce travail a été publié dans la revue internationale JSR «Chemometrics and Intelligent Laboratory Systems » [Laref et al. 2019].

Après avoir optimisé le choix des hyper-paramètres de la méthode SVM, nous avons alors confronté les résultats de cette méthode à ceux obtenus par les méthodes PLS et ELM. A l'issue de cette comparaison, nous pouvons confirmer que la méthode SVM est la mieux adaptée à notre application car elle donne une meilleure précision dans l'estimation de la concentration de nos deux gaz polluants, mais aussi un modèle stable avec un minimum d'échantillons nécessaires pour l'apprentissage. Par exemple, pour l'estimation des concentrations de dioxyde d'azote, seulement 200 échantillons ont suffi pour atteindre un RMSE à environ  $5\mu\text{g}/\text{m}^3$ .

Il est important de noter que, lors de la validation de chacun des capteurs sur le terrain (fin du chapitre II), la meilleure précision d'estimation (RMSE moyenne) obtenue était de  $9.19\mu\text{g}/\text{m}^3$  avec le capteur NO<sub>2</sub>-B43F ; ce qui confirme l'intérêt d'utiliser un modèle d'apprentissage multi-variables combinant les informations de tous les capteurs.

La notion de la minimisation du nombre d'échantillons pour l'apprentissage sera utile pour aborder le chapitre suivant, qui porte sur le transfert du modèle d'étalonnage car nous aurons besoin d'une méthode qui peut faire le lien entre deux ensembles de données différents avec un nombre d'échantillons très réduit.

**Chapitre IV : Transfert d'étalonnage pour  
une utilisation à long terme et sur des  
instruments identiques**



### **IV.1 Introduction**

Dans le chapitre précédent nous avons vu que les méthodes d'apprentissage statistique permettent de combiner et d'exploiter les différentes informations pour établir un modèle d'étalonnage capable d'estimer avec précision la concentration des gaz polluants comme le dioxyde d'azote et l'ozone.

Bien que les systèmes à base de micro-capteurs puissent fournir de bons résultats initiaux, ces résultats se détériorent assez rapidement dans le temps [Mueller et al. 2017], ce qui implique des réétalonnages fréquents au cours de leur utilisation. Étant donné que l'étalonnage des nez électroniques est coûteux et prend généralement beaucoup de temps, les réétalonnages complets deviennent un obstacle majeur pour l'exploitation de ces systèmes, et en particulier dans le cadre de la surveillance de la qualité de l'air. Ajoutons à cela que des systèmes identiques doivent être étalonnés individuellement en raison d'une mauvaise reproductibilité des capteurs neufs et de la diversité des conditions d'exploitation qui induisent des dérives différentes souvent non modélisables [Tomic et al. 2004].

C'est pourquoi nous avons étudié des techniques telles que le transfert d'étalonnage qui constitue une solution intéressante pour prolonger la durée de vie du modèle. Dans notre cas, la technique de transfert d'étalonnage permettra également d'adapter le modèle d'étalonnage établi pour un système à d'autres systèmes identiques utilisés dans des conditions d'exploitation identiques [Rudnitskaya et al. 2017]. Ainsi il serait possible de réduire le coût, l'effort et le temps par rapport à un réétalonnage complet [Zhang et al. 2017; Liu et al. 2013].

Dans la première partie de ce chapitre nous abordons les différents facteurs qui entraînent une réduction de la robustesse et de la fiabilité d'un modèle d'étalonnage. Par la suite, dans la deuxième partie, nous décrivons les différentes méthodes et stratégies qui permettent d'adapter un modèle. La troisième partie présente l'approche que nous proposons pour améliorer la méthode de « standardisation directe », basée sur la régression SVM. Pour finir, nous comparons notre approche avec la standardisation directe classique.

### **IV.2 Les causes d'invalidité d'un modèle d'étalonnage**

Lorsqu'on utilise un modèle d'étalonnage qui a été établi sur un instrument et qui fonctionne bien, ces performances sont en général beaucoup moins bonnes à long terme ou sur un autre instrument identique [Tomic et al. 2002]. Dans le cas d'un nez électronique basé sur des capteurs chimiques, ce problème est encore très présent. La raison principale est que les

réponses de ces capteurs varient légèrement en fonction du temps ou d'un capteur à l'autre de même type [Fernandez et al. 2016].

Des paramètres qui n'ont pas été pris en compte dans l'étape d'étalonnage, peuvent également introduire des variations dans les réponses des capteurs. La présence de ces variations non modélisées dans les nouvelles données peut alors conduire à des prédictions biaisées [Maag et al. 2018].

Il existe essentiellement trois situations pouvant rendre un modèle d'étalonnage invalide [Wang et al. 1991].

La première comprend les modifications de la constitution physique et / ou chimique des échantillons gazeux [Zhang et al. 2011; Esposito et al. 2017]. Par exemple, si le système est étalonné en laboratoire sous une exposition au seul dioxyde d'azote, il ne pourra alors pas fournir de bonnes estimations des concentrations de ce gaz polluant lorsqu'il est exposé à l'air ambiant qui contient de nombreux autres gaz interférents. La deuxième situation susceptible de rendre un modèle invalide se produit lorsque l'environnement du système change au fil du temps [Macías et al. 2014]. Par exemple, les variations de la température et de l'humidité peuvent avoir une forte influence sur les concentrations estimées [Popoola et al. 2016; Korotcenkov et Cho 2011].

La troisième situation est liée aux changements dans les capteurs à cause de leurs dérives, ou suite au remplacement d'un ou plusieurs capteurs [Hui et al. 2003; Ahmadou et al. 2017; Sun et al. 2017].

- **Dérives des capteurs**

Les raisons possibles de la dégradation plus au moins rapide des capteurs de gaz sont nombreuses : 1) les capteurs vieillissent en raison de la fatigue thermomécanique, de la dérive en température de l'élément chauffant ou de la modification de la surface active due à son fonctionnement à haute température (cas des capteurs à oxyde métallique) ; 2) les capteurs peuvent être empoisonnés du fait de leur exposition à des produits chimiques agressifs ; 3) l'absorption et la désorption sur les tubes et les chambres du capteur peuvent provoquer des effets de mémoire.

- **Remplacement d'un capteur**

À la suite des processus de dégradation, un capteur commence à fournir des réponses moins précises, puis complètement erronées, ce qui nécessite son remplacement. En raison des tolérances technologiques inhérentes à la fabrication de capteurs chimiques, et ceci quelle que soit la technologie utilisée, le nouveau capteur aura des caractéristiques différentes du capteur

initial. Ainsi, des variations peuvent apparaître dans la ligne de base (la réponse à l'air pur) et dans la sensibilité des capteurs identiques. Ce manque de reproductibilité dans la fabrication des capteurs peut entraîner une forte baisse de la précision du modèle d'étalonnage entre deux systèmes identiques.

### **IV.3 Les stratégies d'adaptation d'un modèle d'étalonnage**

Le transfert d'étalonnage peut être réalisé en suivant deux stratégies différentes [Rudnitskaya 2018; Marco et Gutierrez-Galvez 2012] :

1. En supprimant la différence entre la réponse de l'instrument maître (système initialement étalonné) et celle de l'instrument esclave (système à étalonner) pour qu'ils deviennent similaires, avant de créer le modèle d'étalonnage [Ziyatdinov et al. 2010; Di Carlo et Falasconi 2012]. Les méthodes les plus fréquemment utilisées sont la Correction des Composantes (CC) [Artursson et al. 2000; Ziyatdinov et al. 2009], l'Analyse en Composantes Indépendantes (ICA) [Di Natale et al. 2002], la Correction Orthogonale du Signal (OSC) [Laref et al. 2017; Padilla et al. 2010; Sjöblom et al. 1998] et la Pondération Généralisée des Moindres Carrés (GLSW) [Fu et al. 2012]. Toutes ces méthodes montrent une grande capacité à supprimer la variabilité d'un instrument à l'autre.
2. En transformant les données de l'instrument esclave pour que le modèle d'étalonnage de l'instrument maître reste valide sur l'instrument esclave [Feudale et al. 2002]. Cette stratégie est généralement appelée la standardisation. La standardisation directe (DS) [Bouveresse et Massart 1996; Pereira et al. 2008] et la standardisation directe par morceaux (PDS) [Yan et Zhang 2015] sont des méthodes les plus répandues pour standardiser la réponse des instruments esclaves.

La méthode de transfert d'étalonnage utilisée dans ce travail est basée sur la standardisation.

### **IV.4 Méthodes de standardisation**

Les méthodes de standardisation font le lien entre l'instrument maître sur lequel le modèle d'étalonnage a été établi et l'instrument esclave pour lequel on veut adapter ce même modèle. La standardisation peut se faire sur les coefficients du modèle de régression, sur les

valeurs prédites ou encore sur les réponses des capteurs en utilisant un sous-ensemble de données pour faire le lien entre les instruments [Feudale et al. 2002].

### IV.4.1 Standardisation des coefficients du modèle

Dans cette catégorie de méthodes de standardisation, le modèle original de l'instrument maître est transformé pour qu'il devienne adéquat pour l'instrument esclave.

On considère que :

- $X_1, X_2$  sont les matrices des données des capteurs de l'instrument maître et de l'instrument esclave respectivement ;
- $Y$  est le vecteur des valeurs prédites calculées avec un modèle de régression linéaire défini par le vecteur de coefficients de régression  $B_1$  et  $B_2$  respectivement ;
- $s_1$  et  $s_2$  sont deux sous-ensembles de données de  $X_1, X_2$  respectivement.

La réponse de l'instrument maître est :

$$Y = X_1 * B_1 \quad (\text{IV-1})$$

De même, la réponse de l'instrument esclave est :

$$Y = X_2 * B_2 \quad (\text{IV-2})$$

On peut alors écrire

$$Y = X_2 * B_2 = X_2 * (B_1 + \Delta b) \quad (\text{IV-3})$$

où  $\Delta b$  est le vecteur de la différence entre  $B_1$  et  $B_2$

$\Delta b$  peut être exprimé par l'expression suivante :

$$\Delta b = B_2 - B_1 = X_2^+ * Y - X_1^+ * Y = (X_2^+ - X_1^+) * Y \quad (\text{IV-4})$$

où  $X_1^+$  et  $X_2^+$  représentent respectivement les matrices inverses de la matrice  $X_1$  et  $X_2$ .

$\Delta b$  peut également être obtenu à partir des sous-ensembles  $s_2$  et  $s_1$  soit:  $\Delta b = (s_2^+ - s_1^+) * Y$

On a alors :

$$B_2 = B_1 + \Delta b = X_1^+ * Y + (s_2^+ - s_1^+) * Y \quad (\text{IV-5})$$

On peut écrire l'équation VI-5 sous la forme :

$$B_2 = X_1^+ * Y + s_2^+ * Y - s_1^+ * Y = B_1 + \overline{B_2} - \overline{B_1} \quad (\text{IV-6})$$

## Chapitre IV : Transfert d'étalonnage pour une utilisation à long terme et sur des instruments identiques

---

où  $\overline{B}_2$  et  $\overline{B}_1$  sont les vecteurs de régression estimés à partir des sous-ensembles  $s_2$  et  $s_1$ .

Cependant la différence entre  $B_1$  et  $\overline{B}_1$  est négligeable puisque  $s_1$  est un sous ensemble de l'instrument maître  $X_1$ . Ce qui nous permet de dire que le vecteur des coefficients de régression  $B_2$  est tout simplement calculé à partir du sous ensemble  $s_2$ , autrement dit que nous avons étalonné l'instrument esclave séparément et qu'aucun transfert réel n'a été réalisé.

### IV.4.2 Standardisation des valeurs prédites

Cette méthode consiste à corriger les valeurs prédites en se basant sur une relation linéaire entre les valeurs prédites par l'unité maître et par l'unité esclave, et en utilisant le biais et la pente de la droite de régression. Pour cela on calcule tout d'abord les valeurs prédites d'un sous ensemble de chacun des deux instruments, soient  $s_2$  et  $s_1$ .

$$y_{s1} = s_1 * b \quad (\text{IV-7})$$

$$y_{s2} = s_2 * b \quad (\text{IV-8})$$

Ensuite, on calcule la droite de régression entre  $y_{s1}$  et  $y_{s2}$  en utilisant la régression linéaire. La pente et le biais de cette droite vont permettre par la suite de corriger les valeurs prédites sur l'instrument esclave selon l'équation suivante :

$$y_2 \text{ corrigée} = \text{biais} + \text{pente} * y_2 \quad (\text{IV-9})$$

Généralement la standardisation des valeurs prédites est utilisée dans le cas où la différence entre les instruments est linéaire et simple.

### IV.4.3 Standardisation des réponses des capteurs

La standardisation des réponses des capteurs a pour but de corriger les réponses des capteurs de l'instrument esclave (instrument 2) pour qu'elles ressemblent à celles de l'instrument maître (instrument 1). On peut distinguer deux manières pour faire la standardisation des réponses des capteurs.

- **Standardisation uni-variable**

La standardisation uni-variable s'applique à chaque capteur individuellement [HAUGEN et al. 2000]. Elle consiste à corriger la réponse de chaque capteur esclave par une simple régression linéaire.

$$r_{i,1} = b_i + a_i r_{i,2} \quad (\text{IV-10})$$

où  $r_{i,1}$  et  $r_{i,2}$  sont les réponses du  $i^{\text{ème}}$  capteur des instruments 1 et 2 respectivement ;  
et  $b_i$  et  $a_i$  sont respectivement le biais et la pente du modèle de correction.

Cette méthode est utilisée dans le cas où les réponses des capteurs ne sont pas corrélées. Dans le cas contraire la standardisation multi-variables notamment la standardisation directe est plus pertinente.

- **Standardisation directe**

La standardisation directe est une méthode de transfert multivariée. Elle consiste à relier directement la réponse d'un échantillon mesurée avec un instrument, à la réponse obtenue avec un autre instrument. C'est une relation linéaire qui est décrite par la matrice de transformation  $F$  comme suit :

$$s_1 = s_2 * F \quad (\text{IV-11})$$

La matrice  $F$  est calculée par :

$$F = s_2^+ * s_1 \quad (\text{IV-12})$$

où  $s_2^+$  est l'inverse généralisée de  $s_2$ .

Lorsque la matrice de transformation  $F$  est calculée, il serait possible de projeter les données de l'instrument esclave (instrument 2) dans l'espace de l'instrument maître (instrument 1) et obtenir une nouvelle matrice  $X_{2\_standardisé}$  :

$$X_{2\_standardisé} = X_2 * F \quad (\text{IV-13})$$

La standardisation directe suppose que la relation entre les deux systèmes est linéaire et ne prend pas en considération certaines non-linéarités entre les systèmes. Un autre inconvénient de cette approche est qu'on utilise un nombre très petit d'échantillons pour le calcul de la matrice de transformation  $F$ . Ceci conduit à un sur-apprentissage et la matrice  $F$  ne représente plus réellement les différences entre les deux systèmes. De plus, le calcul de l'inverse généralisé  $s_2^+$  qui rentre dans le calcul de  $F$  est instable dans le cas où les variables sont corrélées.

### IV.4.4 Conclusion

Dans le cadre de ce travail, nous avons choisi d'utiliser la méthode de standardisation directe pour réaliser le transfert d'étalonnage. A partir de l'analyse présentée en IV.4.3 on peut conclure qu'une méthode de standardisation directe efficace nécessite deux éléments essentiels :

1. Un algorithme qui peut sélectionner un sous-ensemble des échantillons de sorte qu'il soit le plus représentatif pour l'ensemble de données.
2. Une technique de projection qui prend en considération les non-linéarités et qui permet de relier au mieux l'ensemble de données de l'unité esclave à celles de l'unité maître.

Pour notre approche, nous proposons l'utilisation de l'algorithme Kernel SPXY pour sélectionner un sous-ensemble d'échantillons et la régression SVM pour faire la projection des données de l'instrument esclave dans l'espace de l'ensemble de données de l'instrument maître.

## IV.5 Algorithmes de sélection d'échantillons

Tous les algorithmes de sélection de données utilisent le même principe basé sur le calcul de la distance entre les échantillons : plus les échantillons sont distants mieux ils peuvent couvrir l'espace de l'ensemble de données. Nous commençons par définir l'algorithme Kennard-Stone qui sert de base aux autres algorithmes.

### IV.5.1 L'algorithme Kennard-Stone (K-S)

L'algorithme K-S classique vise à sélectionner un sous-ensemble représentatif dans un ensemble de données de  $N$  échantillons. Pour assurer une distribution uniforme de ce sous-ensemble le long de l'espace des données  $X$ , l'algorithme suit une procédure pas à pas dans laquelle de nouvelles sélections sont effectuées dans les régions de l'espace situées loin des échantillons déjà sélectionnés [Kennard et Stone 1969]. L'algorithme de Kennard-Stone repose sur le calcul de la distance euclidienne définie par :

$$ED_x(p, q) = \sqrt{\sum_{i=1}^k (x_p(i) - x_q(i))^2} \quad (\text{IV-14})$$

où  $x_p, x_q$  sont les échantillons à  $k$  variables.

La sélection commence par prendre une paire d'échantillons pour lesquels la distance est la plus grande. À chaque itération suivante, l'algorithme sélectionne l'échantillon qui est le

plus distant par rapport à tous les échantillons déjà sélectionnés. Une telle procédure est répétée jusqu'à ce que le nombre d'échantillons requis soit atteint.

L'algorithme Kennard-Stone ne prend pas en considération la variation de la variable à expliquer  $Y$ , mais tient compte uniquement de la variation des variables explicatives  $X$ . Pour introduire la variation de la variable à expliquer dans la sélection des échantillons il existe une version évoluée de l'algorithme K-S appelée l'algorithme SPXY (Sample set Partitioning based on joint  $X - Y$  distances).

### IV.5.2 L'algorithme SPXY

L'algorithme SPXY combine la distance définie dans l'équation IV-14 avec une autre distance qui dépend des variables à expliquer  $y$ . Cette distance est définie par :

$$ED_y(p, q) = \sqrt{\sum_{i=1}^k (y_p(i) - y_q(i))^2} \quad (\text{IV-15})$$

où  $k$  est le nombre de variables à expliquer tandis que  $y_p(i)$  et  $y_q(i)$  sont les  $i^{\text{èmes}}$  variables des échantillons  $p$  et  $q$  respectivement.

Afin d'attribuer la même importance aux distributions d'échantillons dans les espaces des données  $x$  et  $y$ , les distances  $ED_x(p, q)$  et  $ED_y(p, q)$  sont divisées par leurs valeurs maximales dans l'ensemble de données. De cette manière, une distance  $xy$  normalisée est calculée comme suit :

$$ED_{xy}(p, q) = \frac{ED_x(p, q)}{\max(ED_x(p, q))} + \frac{ED_y(p, q)}{\max(ED_y(p, q))} \quad (\text{IV-16})$$

Une procédure de sélection par étapes, similaire à celle de l'algorithme K-S est ensuite appliquée à l'aide des critères de sélection  $ED_{xy}(p, q)$ , au lieu de  $ED_x(p, q)$ .

### IV.5.3 L'algorithme kernel SPXY

Une version de l'algorithme SPXY plus performante est proposée par Gani et Liman [Gani et Liman 2016]. Dans cet algorithme, la distance euclidienne est remplacée par une distance de noyau. La motivation de ce remplacement est que dans un espace de haute dimension (ou de dimension infinie) on obtient une meilleure représentation des données. Cependant, le calcul de la projection des données dans un espace de haute dimension est très coûteux en termes de calcul, d'où le recours à la fonction noyau.



## Chapitre IV : Transfert d'étalonnage pour une utilisation à long terme et sur des instruments identiques

---

La distance noyau entre deux échantillons  $x_p, x_q$  est définie par :

$$KD(x_p, x_q) = \sqrt{\sum_{i=1}^I K(x_p(i), x_p(i)) + \sum_{i=1}^I K(x_q(i), x_q(i)) - 2 \sum_{i=1}^I K(x_p(i), x_q(i))} \quad (\text{IV-17})$$

De même pour la variable à expliquer  $y$ , nous avons :

$$KD(y_p, y_q) = \sqrt{K(y_p, y_p) + K(y_q, y_q) - 2K(y_p, y_q)} \quad (\text{IV-18})$$

En remplaçant les équations VI-17 et VI-18 dans VI-16, on trouve :

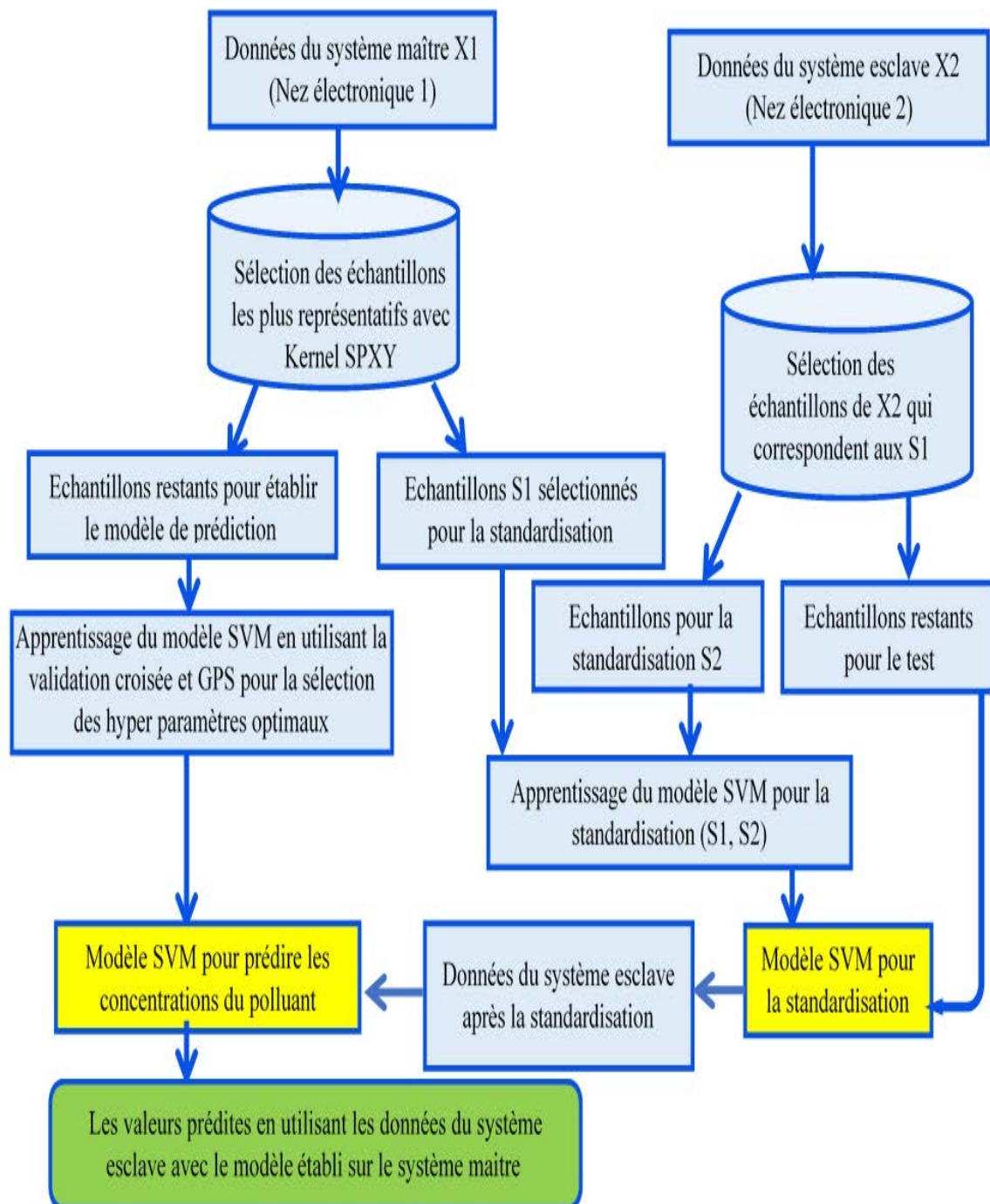
$$KD_{xy}(p, q) = \frac{KD(x_p, x_q)}{\max(KD(x_p, x_q))} + \frac{KD(y_p, y_q)}{\max(KD(y_p, y_q))} \quad (\text{IV-19})$$

$KD_{xy}(p, q)$  est la distance de noyau utilisée pour la sélection d'échantillons dans l'algorithme kernel SPXY.

### IV.6 Algorithme proposé pour la standardisation

L'organigramme présenté dans la figure IV-1 résume la procédure que nous proposons pour le transfert d'étalonnage entre deux systèmes « nez électroniques » utilisés pour estimer la concentration du NO<sub>2</sub>. Notre démarche est la suivante : nous utilisons l'algorithme noyau SPXY pour sélectionner le nombre souhaité d'échantillons destinés à la standardisation à partir du jeu de données de l'instrument maître. Les données restantes sont utilisées pour construire un modèle d'étalonnage en utilisant la régression SVM. Les hyper-paramètres de la régression SVM sont optimisés à l'aide de l'algorithme GPS. Ensuite, à partir de l'instrument esclave, nous sélectionnons les échantillons équivalents à ceux qui sont sélectionnés à partir de l'instrument maître. Nous notons que deux échantillons sont dits équivalents lorsqu'ils sont issus de la même la concentration de gaz.

La régression SVM est également utilisée pour créer un autre modèle afin de standardiser les données de l'instrument esclave. Enfin, nous pouvons prédire la concentration de polluants en utilisant le modèle d'étalonnage construit sur l'unité maître avec les données standardisées de l'unité esclave.



**Figure IV-1 Illustration de la procédure de transfert d'étalonnage entre deux nez électroniques identiques à l'aide de la régression SVM.**

### IV.7 Transfert d'étalonnage entre deux instruments identiques

Cette étude est réalisée sur deux réseaux de capteurs « nez électroniques » identiques, tout d'abord en mettant en évidence leur différence pour ensuite montrer les effets du transfert d'étalonnage à rapprocher les réponses des deux instruments.

## Chapitre IV : Transfert d'étalonnage pour une utilisation à long terme et sur des instruments identiques

---

Les données utilisées dans cette partie sont collectées lors de la caractérisation en laboratoire à partir de deux réseaux de capteurs identiques placés dans la même chambre de mesure et qui subissent les mêmes conditions expérimentales. Nous avons collecté 508 mesures réalisées à différentes concentrations de dioxyde d'azote allant de 25ppb à 250ppb avec un pas de 25ppb.

Pour procéder au transfert d'étalonnage nous avons regroupé les réponses des capteurs du premier réseau (instrument 1), considéré comme instrument maître, dans la matrice notée  $X1_N^D$  et les données du deuxième réseau (instrument 2), considéré comme instrument esclave, dans la matrice notée  $X2_N^D$ . Les concentrations correspondant à ces réponses sont regroupées dans un vecteur  $Y_N^1$ .  $N = 508$  représente le nombre de mesures, et  $D = 900$  la dimension des variables explicatives, soit 300 points de mesure pour chacun des trois capteurs utilisés : NO2-B43F, MiCS-2714, GGS 7530.

### IV.7.1 Nécessité de la standardisation

Les réponses des capteurs obtenues à partir de ces deux réseaux de capteurs identiques et exposés aux mêmes concentrations de gaz sous les mêmes conditions d'exploitation, sont différentes. Ceci est illustré sur la figure IV-2 qui présente les signaux de sortie du capteur NO2-B43F et du capteur MiCS-2714 de chacun des deux instruments maître et esclave. Même si les deux réseaux de capteurs ont subi le même historique d'utilisation, nous constatons que les signaux recueillis à partir du même type de capteurs sont différents aussi bien en termes de ligne de base (réponse sous air synthétique) qu'en termes de sensibilité. Par exemple, concernant la ligne de base, nous observons un décalage d'environ 12mV entre les deux capteurs électrochimiques NO2-B43F d'Alphasense, ce qui correspondrait à une concentration de 75ppb de NO<sub>2</sub>. Pour les deux capteurs MOS MiCS-2714, la différence de la ligne de base est plus importante. Pour les capteurs GGS 7530 elle est encore plus grande. Ce décalage notable de la ligne de base des capteurs identiques MOS peut être dû à la grande variabilité de leur valeur de résistance initiale fournie par le constructeur (par exemple, la résistance initiale du GGS 7530 est de  $50k\Omega \pm 35K\Omega$  d'après son constructeur). En ce qui concerne la sensibilité, nous avons utilisé un indicateur numérique sur la figure IV-2 pour mentionner sa variation en raison de la très faible échelle du signal. Par exemple, dans le cas des capteurs MiCS-2714, les sensibilités sont respectivement de 33mV et 51mV pour 25ppb de NO<sub>2</sub>. Cette différence en termes de sensibilité peut dévier considérablement l'estimation des concentrations de NO<sub>2</sub> d'autant plus que la plage de concentrations correspondant à la pollution atmosphérique est étroite et très

faible. En conséquence, tout modèle établi sur un premier nez électronique ne peut être utilisé pour un autre nez électronique même avec des capteurs identiques à ceux du premier.

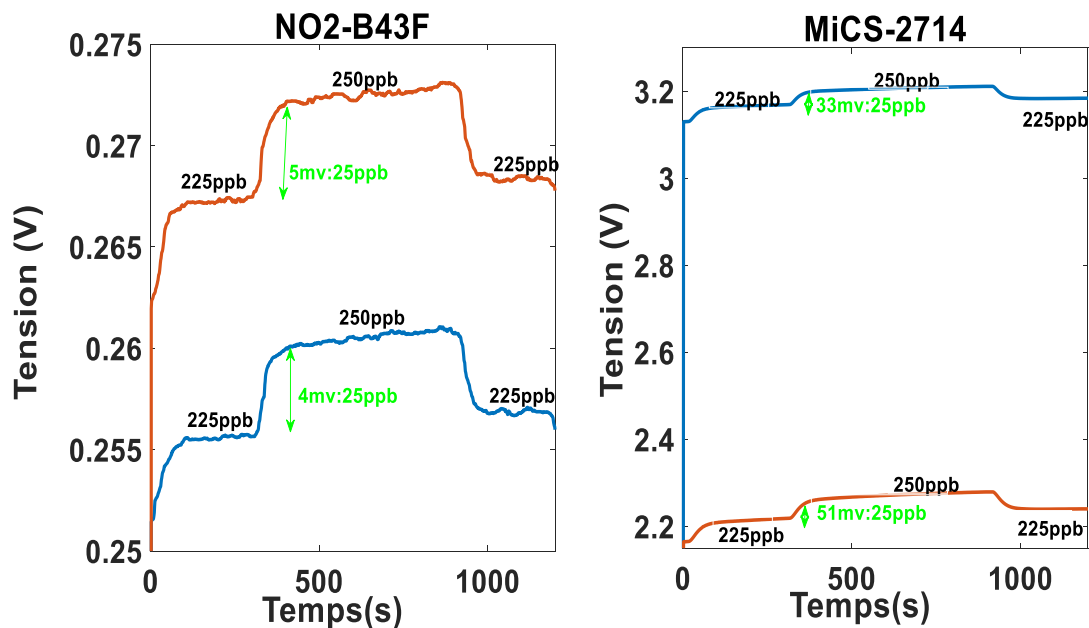


Figure IV-2 Exemple de réponses des capteurs identiques acquises sous mêmes conditions de mesure.

### IV.7.2 Application de la standardisation SVM

Pour montrer la capacité de la méthode proposée à créer un lien entre les réponses des capteurs de l'instrument maître et ceux de l'instrument esclave, nous avons utilisé la méthode PCA (Principal Component Analysis). Cette méthode permet une réduction de dimension et par conséquent la visualisation des données par projection sur les premiers axes principaux. Nous avons tracé sur les deux premières composantes PCA, l'ensemble des données obtenues à l'aide de l'instrument maître et des données de l'instrument esclave avant et après la standardisation. La figure IV-3 montre clairement que sans standardisation les données de chacun des instruments sont rassemblées dans deux régions différentes. Nous avons ensuite utilisé l'algorithme noyau SPXY pour sélectionner 10 échantillons pour effectuer la standardisation. Et nous avons à nouveau tracé sur les deux premières composantes les données de l'instrument maître et celles de l'instrument esclave avant et après la standardisation (figure IV-4). Cette figure illustre clairement d'une part l'intérêt et l'efficacité du transfert d'étalonnage, et d'autre part, le fait que l'utilisation de seulement 10 échantillons pour la standardisation a été suffisante pour décaler les données de l'instrument esclave vers la région des données de l'instrument maître.

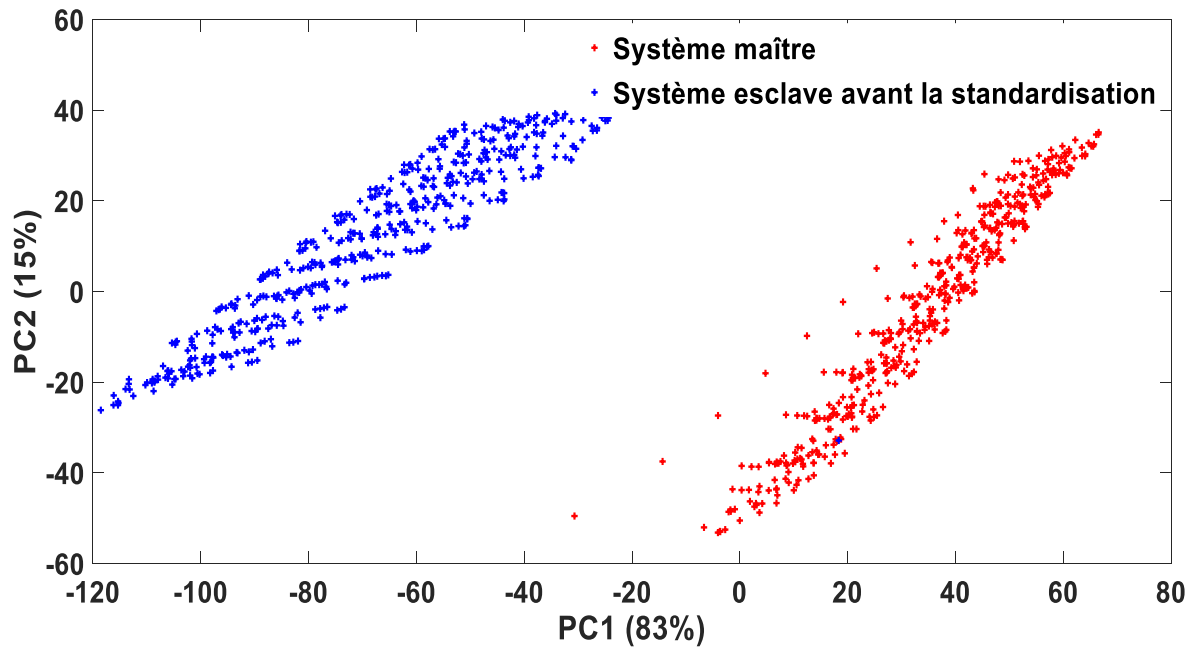


Figure IV-3 Projection des données de l'instrument maître et de l'instrument esclave avant la standardisation, sur les deux premières composantes principales de l'ACP.

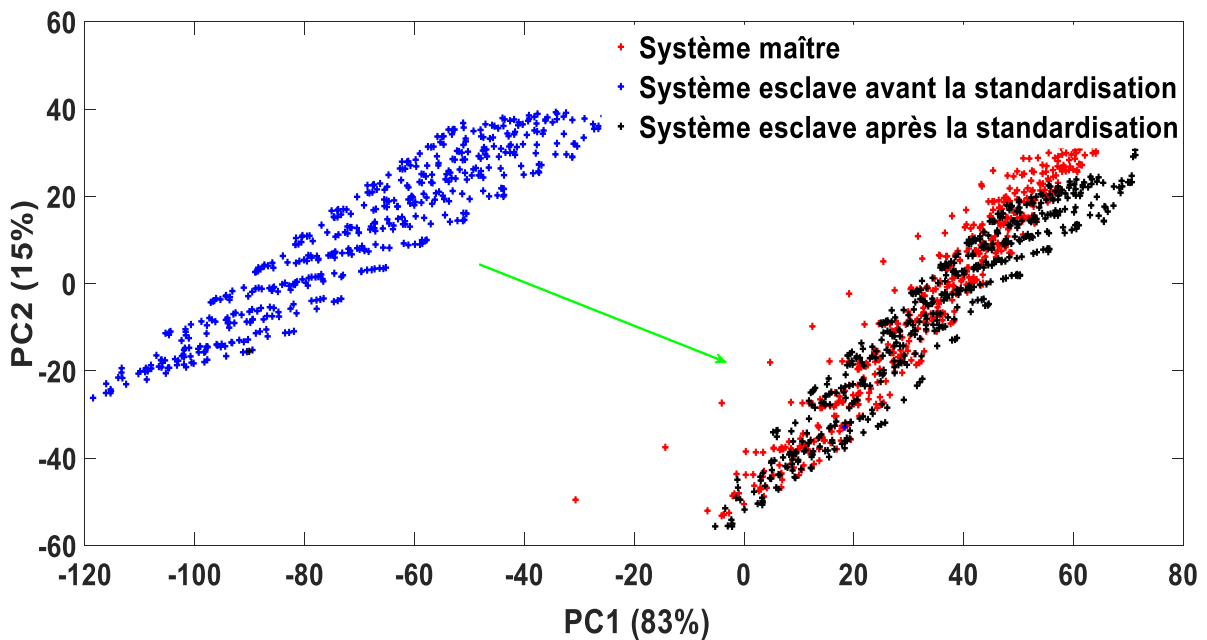


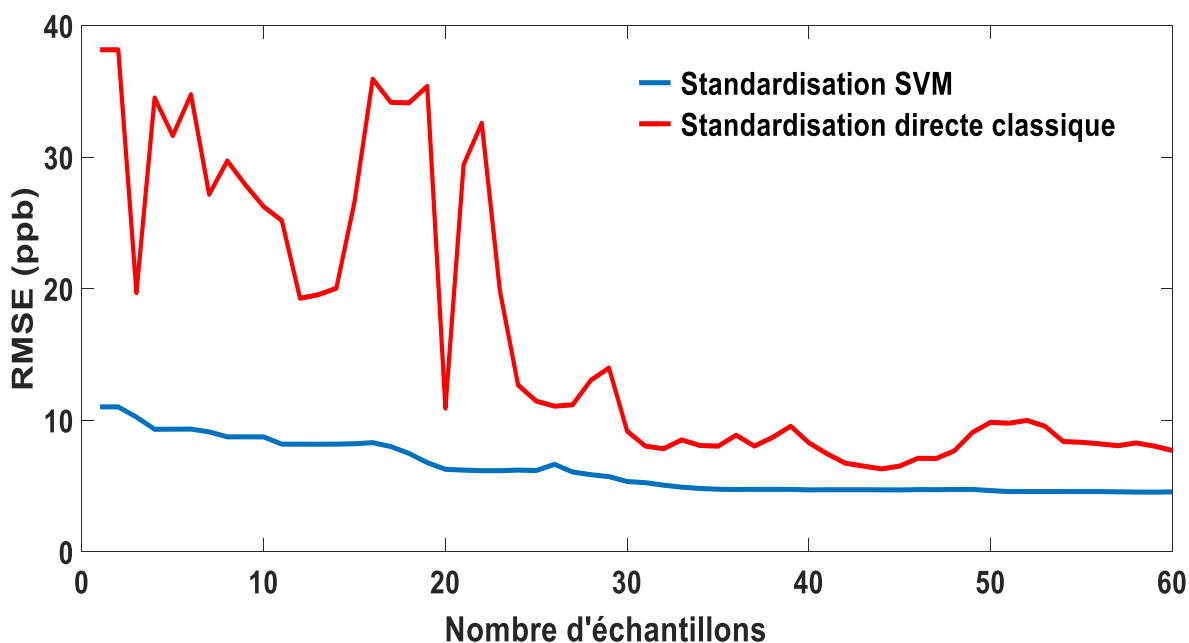
Figure IV-4 Projection des données de l'instrument maître et de l'instrument esclave après la standardisation, sur les deux premières composantes principales de l'ACP.

#### IV.7.3 Comparaison avec la standardisation directe classique

L'approche que nous avons proposée pour de la standardisation a été comparée à la standardisation directe classique. Tout d'abord, nous avons construit un modèle d'étalonnage sur l'unité maître en utilisant la régression SVM. Nous avons ensuite utilisé l'algorithme noyau SPXY pour sélectionner un certain nombre d'échantillons pour la standardisation. Après la

## Standardisation pour éviter un réétalonnage complet

standardisation des données de l'unité esclave, nous avons utilisé le modèle d'étalonnage établi sur l'unité maître pour prévoir les concentrations de NO<sub>2</sub>. Nous avons répété cette procédure 60 fois en ajoutant à chaque fois un échantillon pour la standardisation. Nous avons calculé l'erreur quadratique moyenne RMSE en utilisant des validations croisées sur l'ensemble des données segmentées en 10 partitions. Dans la figure IV-5, nous pouvons visualiser le gain de performance de notre méthode par rapport à la standardisation directe classique. Notre méthode nécessite 10 échantillons pour atteindre une précision de prédiction d'environ 8.7ppb, tandis que la standardisation directe classique nécessite au moins 40 échantillons pour atteindre une précision stable voisine de 10ppb.



**Figure IV-5 Evolution de la moyenne RMSE en fonction du nombre d'échantillons utilisé pour la standardisation dans les deux cas : standardisation directe classique et standardisation SVM.**

### IV.8 Standardisation pour éviter un réétalonnage complet

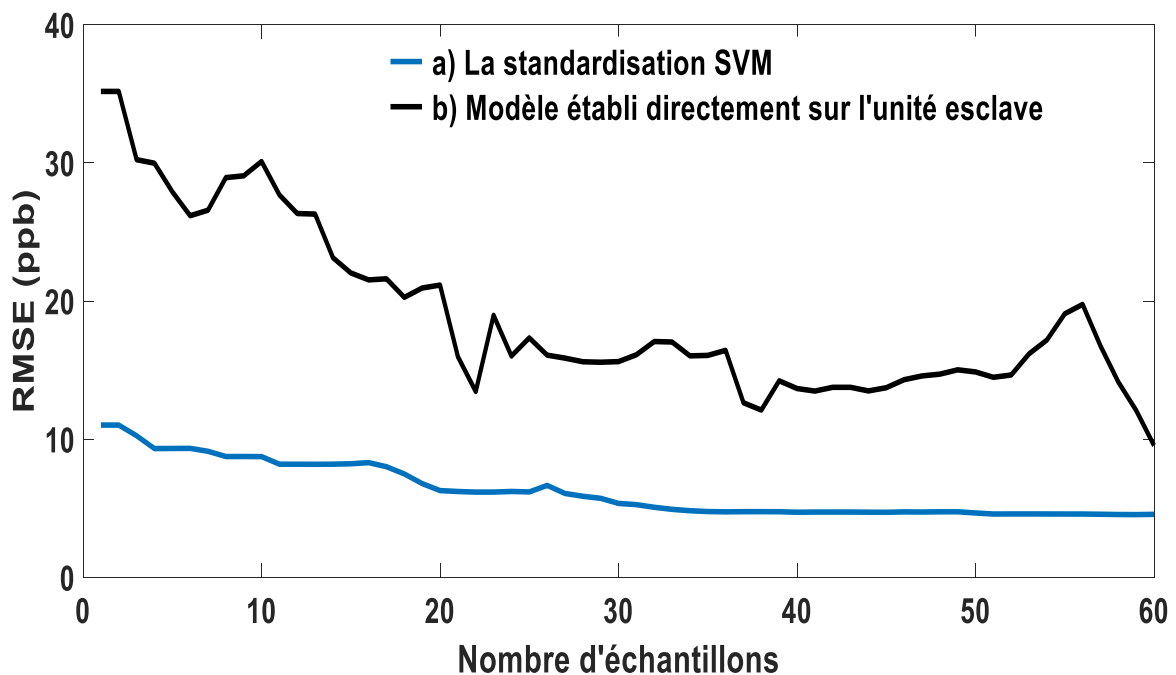
L'objectif de toute méthode de transfert d'étalonnage est de préserver le temps et d'éviter l'effort et le coût d'un réétalonnage complet. En fait, si la procédure de la standardisation directe nécessite autant d'échantillons pour construire un nouveau modèle pour l'instrument esclave, il n'est pas utile de procéder à un transfert d'étalonnage. Dans ce cas, il serait préférable d'étalonner chaque instrument individuellement.

Nous avons donc vérifié si le nombre d'échantillons utilisé précédemment pour la standardisation SVM est suffisant pour créer un nouveau modèle d'étalonnage établi directement sur l'instrument esclave. C'est pourquoi, nous avons utilisé le même sous ensemble

## Chapitre IV : Transfert d'étalonnage pour une utilisation à long terme et sur des instruments identiques

destiné à la standardisation pour construire directement un nouveau modèle d'étalonnage sur l'instrument esclave. Nous avons démarré l'opération avec un échantillon en rajoutant à chaque fois un échantillon, jusqu'à 60 échantillons. A chaque fois, nous avons calculé l'erreur quadratique moyenne de prédictions croisées de 10 partitions selon deux circonstances : a) dans le cas des échantillons sélectionnés utilisés pour la standardisation et b) dans le cas de l'utilisation de ces échantillons pour construire un nouveau modèle d'étalonnage sur l'instrument esclave.

La figure IV-6 présente l'évolution de l'erreur quadratique moyenne de prédition RMSE en fonction du nombre d'échantillons. Cette figure montre que la construction d'un nouveau modèle d'étalonnage nécessite plus que 60 échantillons pour obtenir un modèle stable avec une précision acceptable correspondant à un RMSE de 10ppb.



**Figure IV-6 Evolution de l'erreur quadratique moyenne en fonction du nombre d'échantillons utilisé pour : a) la standardisation SVM ; b) modèle établi directement sur l'unité esclave.**

Ainsi, afin d'éviter d'utiliser trop d'échantillons (plus de 60), il faut avoir recours à la standardisation qui nécessite moins d'échantillons. Le résultat obtenu est illustré sur la figure IV-7 qui présente les concentrations estimées en fonction des concentrations réelles. Les valeurs prédites sont obtenues en utilisant un modèle d'étalonnage construit sur l'unité maître et les données standardisées de l'unité esclave avec 10 échantillons utilisés pour la standardisation.

## Transfert d'étalonnage pour un seul dispositif entre deux périodes

Les valeurs estimées et les valeurs réelles sont bien corrélées. Ceci est confirmé numériquement en calculant la moyenne des RMSE de 10 validations croisées, qui est autour de 8.7ppb.

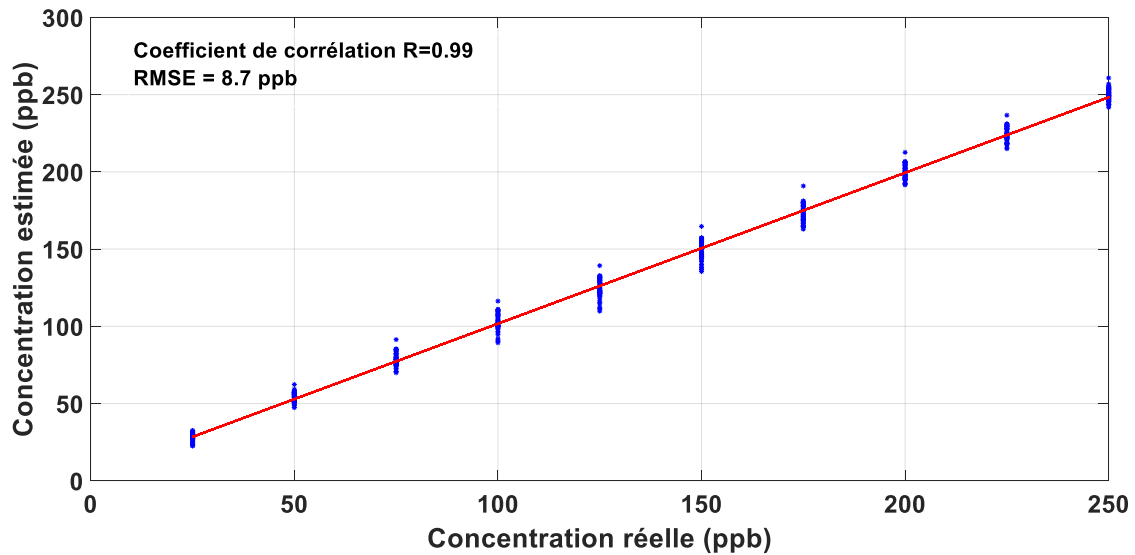


Figure IV-7 Performance de la standardisation SVM avec 10 échantillons illustrée par la corrélation des concentrations réelles et estimées.

### IV.9 Transfert d'étalonnage pour un seul dispositif entre deux périodes

Ce transfert d'étalonnage est très utile pour répondre au vieillissement des micro-capteurs de gaz au cours de leur utilisation. Pour l'application du transfert d'étalonnage sur un seul instrument à des temps différents, il nous a été possible d'utiliser les données collectées dans la station de mesure entre les mois de juillet et décembre 2018. Nous avons considéré que les données collectées durant le mois de juillet correspondent à l'instrument maître, et que celles collectées avec le même instrument au mois de décembre correspondent à l'instrument esclave.

#### IV.9.1 Nécessité d'un étalonnage fréquent ou d'une standardisation pour corriger les dérives

Avant de tester la fiabilité de la méthode de standardisation directe, nous avons mis en évidence la dégradation des performances d'un modèle d'étalonnage au fil du temps. D'abord, nous avons établi un modèle d'étalonnage sur des données collectées durant le mois de juillet. Nous avons ensuite testé ce modèle pour prédire les concentrations du dioxyde d'azote à des périodes équidistantes allant du mois d'aout jusqu'au mois de décembre 2018. La figure IV-8 présente l'erreur quadratique moyenne et les coefficients de corrélation entre les valeurs prédites et réelles pour chacun des ensembles de données collectées sur un mois. On remarque que le modèle établi à partir des données du mois de juillet reste fiable en fournissant des



## Chapitre IV : Transfert d'étalonnage pour une utilisation à long terme et sur des instruments identiques

résultats acceptables pendant le mois d'août et le mois de septembre (le RMSE moyen du mois d'août =  $6.65\mu\text{g}/\text{m}^3$ , et le RMSE moyen du mois de septembre =  $6.88\mu\text{g}/\text{m}^3$ ). Par contre, à partir du mois d'octobre sa performance se dégrade au fil du temps. L'erreur sur la prédiction de la concentration de  $\text{NO}_2$  devient alors égale à  $23.48\mu\text{g}/\text{m}^3$  au mois de décembre, alors qu'elle était de  $6.65\mu\text{g}/\text{m}^3$  au mois de juillet.

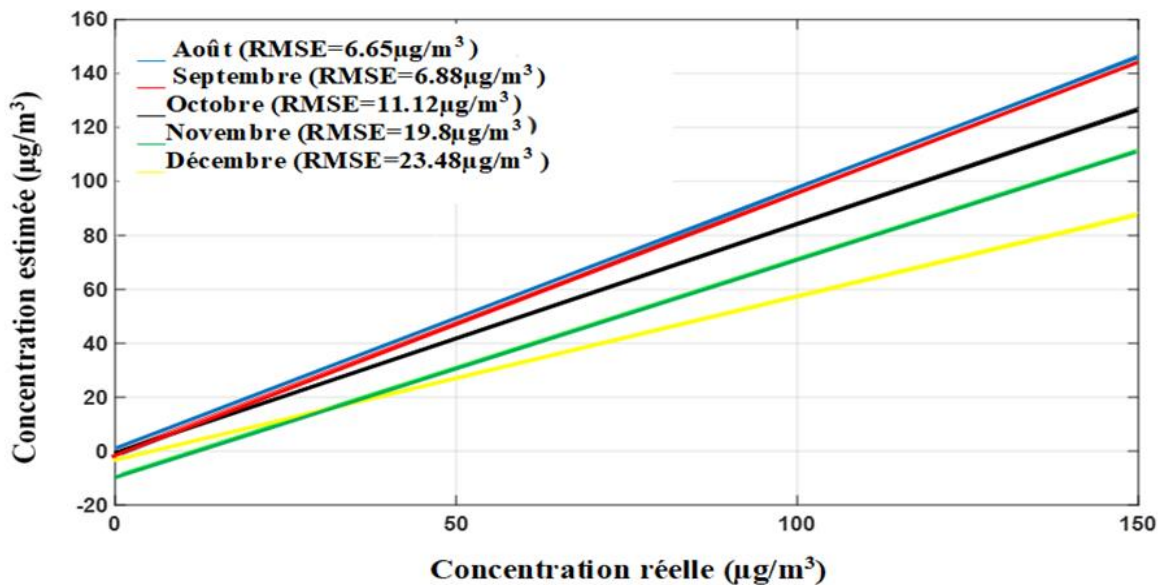


Figure IV-8 Mise en évidence de la dégradation des capteurs au cours du temps.

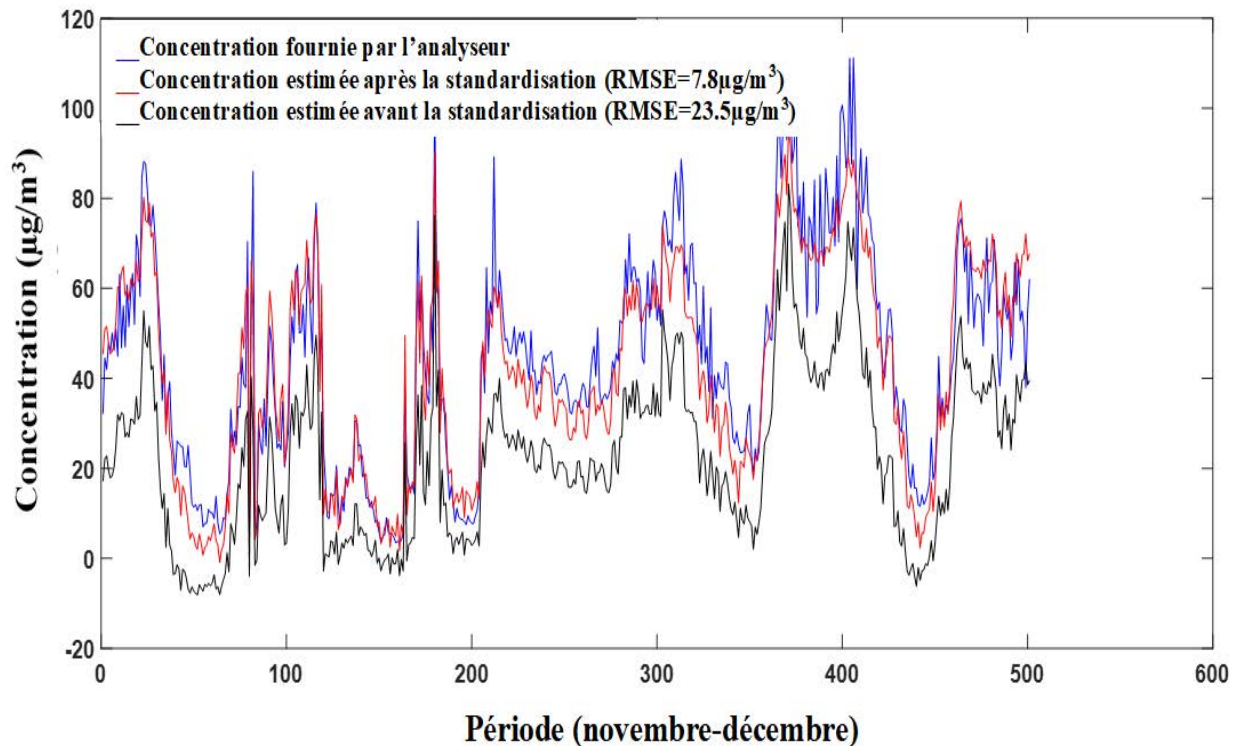
### IV.9.2 Application de la standardisation

Afin d'appliquer la standardisation dans le cas d'un seul instrument à des périodes de temps différentes, nous avons sélectionné et classé 15 échantillons les plus représentatifs en utilisant l'algorithme Kernel SPXY, pour l'instrument maître (instrument du mois de juillet). Ensuite nous avons cherché leurs analogues en concentration dans l'ensemble des données collectées durant le mois de décembre en se basant sur leurs concentrations réelles fournies par l'analyseur d'ATMO Grand Est. Nous avons ainsi sélectionné 15 échantillons destinés à la standardisation, et nous avons repris la même procédure que précédemment utilisée dans le cas de deux instruments identiques.

La figure IV-9 présente les concentrations réelles fournies par l'analyseur durant la période novembre-décembre et les concentrations estimées pour la même période, avant et après la standardisation du modèle établi sur des données du mois de juillet. On voit clairement comment la standardisation a pu améliorer l'estimation des concentrations tout en gardant le même modèle de prédiction, c'est-à-dire le modèle établi cinq à six mois auparavant. L'écart

## Conclusion

quadratique moyen (RMSE moyen) qui était de  $23.5\mu\text{g}/\text{m}^3$  avant l'application de la méthode de standardisation est réduit à  $7.8\mu\text{g}/\text{m}^3$  après l'application de notre méthode de standardisation.



**Figure IV-9 Concentrations réelles et concentrations estimées avant et après la standardisation sur une période de 2 mois.**

### IV.10 Conclusion

Dans ce chapitre nous avons étudié les méthodes de standardisation. Nous avons constaté que deux nez électroniques identiques, fonctionnant dans les mêmes conditions, fournissent des signaux différents en termes de ligne de base (leur réponse à l'air pur) et de sensibilité. Ces différences sont aussi présentes pour un même nez électronique à des périodes de temps différentes durant son utilisation.

Le but de notre étude a été de développer une nouvelle approche de transfert d'étalonnage en utilisant un nombre réduit d'échantillons.

Lors d'un transfert d'étalonnage, deux éléments essentiels sont à prendre en considération :

- La sélection d'un sous-ensemble d'échantillons pour qu'il soit le plus représentatif possible pour l'ensemble des données.
- Une méthode nécessitant le moins d'échantillons possible pour établir la relation entre les données des deux systèmes.

## **Chapitre IV : Transfert d'étalonnage pour une utilisation à long terme et sur des instruments identiques**

---

Dans notre approche nous avons utilisé l'algorithme Kernel SPXY pour la sélection des échantillons et la régression SVM pour la standardisation des données. Nous avons démontré que le modèle d'étalonnage de l'unité maître peut être utilisé avec succès sur l'unité esclave après la standardisation aussi bien dans le cas de deux instruments identiques, que dans le cas d'un même instrument à des périodes différentes tout au long de son utilisation en continu.

Les résultats ont aussi montré que cette méthode offre de meilleures performances par rapport à la standardisation directe classique. Les résultats de ce chapitre nous ont permis de publier un article sur la standardisation dans une revue internationale JCR « Sensors » [Laref et al. 2018].

## **Conclusion générale**

# Conclusion générale

Nos travaux de recherche ont pour objectif d'étudier la possibilité de l'utilisation d'un dispositif à base de micro-capteurs de gaz (nez électronique) pour la surveillance du dioxyde d'azote et de l'ozone en milieu urbain sur un « site trafic ». Il s'agit d'une contribution à la possibilité de compléter les réseaux de surveillance de la qualité de l'air nationaux avec des instruments à coût modéré, simples d'utilisation et portables, afin de permettre une meilleure résolution aussi bien spatiale que temporelle de la mesure des polluants tels que le dioxyde d'azote et l'ozone.

A l'issue d'une étude bibliographique, nous avons tout d'abord sélectionné des capteurs commerciaux de type électrochimique et à base d'oxyde métallique pour leurs performances, leur fiabilité et leur sensibilité aux polluants ciblés. Puis, nous avons adapté un équipement de mesure expérimental pour la caractérisation précise en laboratoire des capteurs choisis.

Les tests en laboratoire ont montré que les capteurs sélectionnés (sauf un capteur à oxyde métallique, le GGS 7530) sont capables de détecter le dioxyde d'azote à des concentrations de l'ordre de la partie par billion (ppb) avec un comportement linéaire en fonction de la concentration de NO<sub>2</sub>. Les mesures en laboratoire ont également montré que les capteurs électrochimiques sont plus stables et plus précis que les capteurs à oxyde métallique. Les études de la répétabilité, du temps de réponse, et des dérives à court terme ont montré que les caractéristiques des capteurs utilisés sont acceptables dans le cadre de notre application où les concentrations des gaz et leurs variations sont très faibles. En effet, les réponses de nos capteurs respectent les normes européennes concernant les mesures indicatives des polluants étudiés puisqu'elles fournissent des valeurs de concentration avec une incertitude inférieure à 25% pour le dioxyde d'azote et inférieure à 30% pour l'ozone.

Pour les tests sur le terrain, nous avons réalisé deux dispositifs identiques de type nez électroniques, équipés chacun de deux capteurs électrochimiques et deux capteurs à base d'oxyde métallique. Concernant l'échantillonnage de l'air à analyser, nous avons choisi de travailler à flux dynamique constant en utilisant une pompe et un contrôleur de débit massique en aspiration. Le débit d'air est fixé à 500ml/min identique à celui utilisé par les analyseurs d'ATMO Grand Est. Ces dispositifs ont été placés dans une station de surveillance d'ATMO Grand Est près de l'autoroute A31 qui traverse la ville de Metz. Le premier dispositif « nez électronique » est placé à l'intérieur de la station, dans une armoire où la température est

## Conclusion générale

---

contrôlée, alors que le second est mis sur le toit de cette armoire subissant les modifications environnementales notamment la température, l'humidité de l'air et le vent.

Les mesures réalisées sur site ont montré que les capteurs à oxyde métallique sont influencés par d'autres gaz interférents présents dans l'air ambiant, alors que les capteurs électrochimiques, plus sélectifs, fournissent toujours des mesures très satisfaisantes même à l'extérieur de la station de surveillance où les variations de la température et de l'humidité sont très importantes.

L'analyse de ces résultats de caractérisation nous a conduit à combiner, grâce aux méthodes d'apprentissage multi-variables, les informations de différents capteurs ainsi que les valeurs de la température de l'air afin d'obtenir un modèle d'étalonnage robuste et précis.

Concernant les méthodes d'apprentissage nous avons réalisé une comparaison entre les modèles statistiques les plus utilisées dans le cas des nez électroniques, notamment la méthode des Moindres Carrés Partiels (PLS), *extreme learning machine* (ELM) et *support vector machine* (SVM). Nous avons démontré que la régression SVM est la plus performante à la fois en termes de précision qu'en termes de nombre de données utilisées pour l'apprentissage.

Les méthodes d'apprentissage statistique nécessitent une optimisation préalable de leurs hyper-paramètres pour obtenir les meilleures performances possibles. Pour cela, nous avons étudié l'influence de chaque hyper-paramètre de la méthode *support vector machine*, en montrant leur influence sur la performance de la méthode. Ensuite, nous avons comparé entre plusieurs algorithmes d'optimisation : l'algorithme génétique, le recuit simulé, l'optimisation par essais particuliers et la recherche par motifs. Cette comparaison a été basée sur la simplicité de la mise en œuvre de la méthode ainsi que la précision de la solution obtenue. Nous avons conclu que l'algorithme de recherche par motifs est le plus simple à mettre en œuvre puisqu'il ne nécessite pas de paramètres à ajuster et donne des solutions comparables voir meilleures que celles des autres algorithmes.

Les données de caractérisation en laboratoire et les données recueillies sur le terrain ont été modélisées à partir de la méthode SVM avec des hyper-paramètres optimisés. Les résultats obtenus sur le terrain sont très prometteurs, la précision obtenue pour le dioxyde d'azote est de  $5.43\mu\text{g}/\text{m}^3$ .

Un autre point important que nous avons abordé dans ce travail, est la problématique du transfert d'étalonnage ou de modèle. En effet, les mesures effectuées en laboratoire ont montré que les capteurs de deux nez électroniques identiques exposés aux mêmes concentrations de gaz et sous les mêmes conditions de mesure ne fournissaient pas les mêmes réponses. C'est pourquoi chaque instrument doit être étalonné individuellement.

## Conclusion générale

---

Pour surmonter cette problématique nous avons utilisé une approche de transfert d'étalonnage pour permettre d'adapter un modèle d'étalonnage à plusieurs nez électroniques identiques, tout en utilisant un nombre d'échantillons faible pour effectuer ce transfert d'étalonnage. Nous avons étudié les différentes approches du transfert d'étalonnage. Ces approches sont regroupées en deux types. Le premier type est basé sur la suppression des différences entre les réponses des capteurs des nez électroniques identiques exposés aux mêmes conditions d'utilisation pour pouvoir utiliser un modèle global. Tandis que le deuxième type s'articule sur la projection des données d'un nez électronique dans l'espace de données d'un nez électronique déjà étalonné.

Nous avons proposé une amélioration de la standardisation directe basée sur la régression SVM, qui à notre connaissance n'a jamais été citée dans la littérature, pour effectuer la projection de données à estimer dans l'espace de données où le modèle d'étalonnage a été établi. Nous avons appliqué notre approche sur les données de deux réseaux de capteurs identiques collectées lors de la caractérisation des capteurs en laboratoire. Les résultats ont montré que l'utilisation d'une dizaine d'échantillons est suffisante pour faire la standardisation du modèle d'étalonnage. Ainsi, le modèle d'étalonnage d'un nez électronique devient valide pour d'autres appareils identiques moyennant une dizaine de mesures d'apprentissage supplémentaires à intégrer dans la base de données initiale.

Cette approche est également utilisée pour corriger les dérives à long terme des micro-capteurs de gaz qui altèrent les résultats de mesures. Nous avons vu que le modèle établi à partir des données collectées sur le terrain durant le mois de juillet n'était plus valable pour les données recueillies en décembre de la même année. Avec l'application de la standardisation directe proposée, il a été possible de corriger les données collectées en décembre et continuer à utiliser le modèle établi en juillet, soit durant six mois d'utilisation consécutive.

Notre recherche a montré la possibilité de réaliser un dispositif fiable et à bas coût destiné à des mesures indicatives du dioxyde d'azote et de l'ozone. Le travail réalisé a illustré que l'utilisation des techniques de traitement de données et des méthodes d'apprentissage adéquates est l'élément crucial pour ce type d'application.

La méthodologie mise au point dans cette thèse peut être généralisée à d'autres systèmes pour pallier aux problèmes de dérives et de reproductibilité dans la fabrication des capteurs.

Nous avons abordé dans cette étude deux aspects essentiels d'un dispositif micro-capteurs de gaz : la caractérisation du réseau de capteurs et l'application de méthodes d'étalonnage pertinentes. D'autres aspects peuvent être étudiés notamment l'influence du flux d'air sur les réponses des capteurs ou encore l'impact de l'emplacement des systèmes micro-

## Conclusion générale

---

capteurs sur leurs réponses. Une étude comparative entre un système statique sans flux d'air et un système dynamique à flux d'air constant serait nécessaire afin de mettre en évidence la nécessité d'équiper ce type de système avec un module d'aspiration d'air. Il serait également nécessaire de collecter des données avec des systèmes identiques placés dans plusieurs sites différents afin d'étudier l'impact de la diversité de la composition d'air sur les performances de notre système.

L'étude du vieillissement des capteurs est notre prochaine étape. La base de données récoltée durant un an et demi est un élément précieux pour étudier le vieillissement des capteurs et proposer d'autres méthodes de correction de dérives non supervisées. Cette analyse du vieillissement permet également de prédire la durée de vie des capteurs afin de mieux planifier le remplacement d'un capteur défaillant.

Aujourd'hui, l'évolution technologique conduit à l'émergence des smart cities. L'internet des objets permet de collecter un énorme volume de données à travers les objets connectés (big data). Et grâce au *machine learning* et plus spécialement les approches de *deep learning* (apprentissage statistique profond), nous pouvons traiter des données afin de mieux les exploiter dans plusieurs domaines d'application. Dans le domaine de la surveillance de la pollution atmosphérique, l'internet des objets va ouvrir la porte vers un système intelligent basé sur plusieurs systèmes micro-capteurs connectés, déployés à l'échelle du quartier. Les algorithmes de *deep learning* vont permettre de traiter la masse des données collectées et de modéliser d'une façon plus précise la dispersion de la pollution pour une intervention rapide et une meilleure sensibilisation des citoyens. Il serait également important d'étudier les stratégies de maintenance pour garantir la performance en exploitation de ces systèmes à des coûts raisonnables.



**Annexe: Normes de surveillance pour la  
pollution de l'air**

## Annexe : Normes de surveillance pour la pollution de l'air

### Seuils d'évaluation et régimes de surveillance

Seuils d'évaluation inférieurs « SEI » (ou minimaux) et supérieurs « SES » (ou maximaux) exprimés en pourcentage de l'objectif environnemental considéré défini à l'article R. 221-1 du code de l'environnement pour déterminer le régime de surveillance

POLLUANT	DIOXYDE DE SOUFRE		DIOXYDE D'AZOTE ET OXYDES D'AZOTE			PARTICULES (PM10/PM2,5)			MONOXYDE DE CARBONE
	Valeur limite sur 24 heures pour la protection de la santé	Niveau critique hivernal pour la protection de la végétation	Valeur limite horaire pour la protection de la santé (NO <sub>2</sub> )	Valeur limite annuelle pour la protection de la santé humaine (NO <sub>2</sub> )	Niveau critique annuel pour la protection de la végétation et des écosystèmes naturels (NO <sub>x</sub> )	Valeur limite en moyenne sur 24 heures (PM10)	Valeur limite en moyenne annuelle (PM10)	Valeur limite en moyenne annuelle (PM <sub>2,5</sub> )	Valeur limite en moyenne sur 8 heures
Seuil d'évaluation supérieur	60 % 75µg/m <sup>3</sup> , à ne pas dépasser plus de 3 fois par année civile	60 % 12µg/m <sup>3</sup>	70 % 140 µg/m <sup>3</sup> , à ne pas dépasser plus de 18 fois par année civile	80% 32µg/m <sup>3</sup>	80% 24µg/m <sup>3</sup>	70 % 35µg/m <sup>3</sup> , à ne pas dépasser plus de 35 fois par année civile	70 % 28µg/m <sup>3</sup>	70 % 17µg/m <sup>3</sup>	70 % 7mg/m <sup>3</sup>
Seuil d'évaluation inférieur	60% 75µg/m <sup>3</sup> à ne pas dépasser plus de 3 fois par année civile	40% 8µg/m <sup>3</sup>	50% 100 µg/m <sup>3</sup> , à ne pas dépasser plus de 18 fois par année civile	65% 26µg/m <sup>3</sup>	65% 19,5 µg/m <sup>3</sup>	50% 25µg/m <sup>3</sup> , à ne pas dépasser plus de 35 fois par année civile	50 % 20µg/m <sup>3</sup>	50 % 12µg/m <sup>3</sup>	50 % 5mg/m <sup>3</sup>

## Annexe : Normes de surveillance pour la pollution de l'air

POLLUANT	PLOMB	BENZÈNE	ARSENIC	CADMIUM	NICKEL	B [A] P
Objectif environnemental	Valeur limite en moyenne annuelle	Valeur cible en moyenne annuelle	Valeur cible en moyenne annuelle	Valeur cible en moyenne annuelle	Valeur cible en moyenne annuelle	Valeur cible en moyenne annuelle
Seuil d'évaluation supérieur	70% 0,35µg/m <sup>3</sup>	70% 3,5µg/m <sup>3</sup>	60% 3,4ng/m <sup>3</sup>	60% 3ng/m <sup>3</sup>	70% 14ng/m <sup>3</sup>	60% 0,6ng/m <sup>3</sup>
Seuil d'évaluation inférieur	50% 0,25µg/m <sup>3</sup>	40% 2µg/m <sup>3</sup>	40% 2ng/m <sup>3</sup>	40% 2ng/m <sup>3</sup>	50% 10ng/m <sup>3</sup>	40% 0,4ng/m <sup>3</sup>

### Objectifs de qualité des données sur la qualité de l'air

Les objectifs minimums de qualité à respecter pour l'évaluation de la qualité de l'air ambiant pour les polluants définis dans les tableaux suivants.

	SO <sub>2</sub> , NO <sub>2</sub> , NO <sub>x</sub> et CO	C <sub>6</sub> H <sub>6</sub>	PM <sub>10</sub> , PM <sub>2,5</sub> et PB	O <sub>3</sub> , NO et NO <sub>2</sub>
1. Mesures fixes (1)				
Incertitude	15%	25%	25%	15%
Saisie minimale de données (2)	90%	90%	90%	90% en été 75% en hiver
Période minimale :				
- sites urbains de fond et sous influence du trafic routier	-	35% (3)	-	-
- sites sous influence industrielle	-	90%	-	-
2. Mesures indicatives				
Incertitude	25%	30%	50%	30%
Saisie minimale de	90%	90%	90%	90%

## Annexe : Normes de surveillance pour la pollution de l'air

données				
Période minimale	14% (5)	14% (4)	14% (5)	> 10% en été
3. Incertitude du modèle				
Par heure	50%	-	-	50%
Moyenne sur 8 heures	50%	-	-	50%
Moyennes journalières	50%	-	Non défini	-
Moyennes annuelles	30%	50%	50%	-
4. Incertitude de l'estimation objective	75%	100%	100%	75%

(1) Des mesures aléatoires peuvent être utilisées pour le benzène, le plomb et les particules au lieu de mesures continues, s'il est démontré que l'incertitude, y compris l'incertitude liée à l'échantillonnage aléatoire, respecte l'objectif de qualité des données de 25% et que la période prise en compte reste supérieure à la période minimale fixée pour les mesures indicatives. L'échantillonnage aléatoire est réparti uniformément sur l'année pour éviter de biaiser les résultats. L'incertitude liée à l'échantillonnage aléatoire peut être quantifiée selon la procédure décrite dans la norme ISO 11222 (2002), « Qualité de l'air - détermination de l'incertitude de mesure de la moyenne temporelle de mesurages de la qualité de l'air ». Si des mesures aléatoires sont utilisées pour évaluer les exigences liées à la valeur limite applicable aux PM10, le 90,4e percentile est inférieur ou égal à 50µg/m<sup>3</sup>.

(2) Les exigences concernant la saisie minimale de données pour les mesures fixes ne comprennent pas les pertes d'information dues à l'étalonnage régulier ou à l'entretien normal des instruments.

(3) Réparti sur l'année pour être représentatif des diverses conditions de climat et de trafic routier.

(4) Une mesure journalière aléatoire par semaine, répartie uniformément sur l'année, ou huit semaines réparties uniformément sur l'année.

(5) Une mesure aléatoire par semaine, répartie uniformément sur l'année, ou huit semaines réparties uniformément sur l'année.

## **Annexe : Normes de surveillance pour la pollution de l'air**

	BENZO [A] PYRENE	ARSENIC, CADMIUM ET NICKEL	HAP AUTRES QUE LE BENZO [A] PYRENE, MERCURE GAZEUX TOTAL	DÉPÔT TOTAL
1. Incertitude				
- Mesures fixes et indicatives	50%	40%	50%	70%
- Modélisation	60%	60%	60%	60%
2. Saisie minimale de données (1)	90%	90%	90%	90%
3. Période minimale de prise en compte				
- Mesures fixes (2)	33%	50%		
- Mesures Indicatives (2) (3)	14%	14%	14%	33%

(1) Les exigences concernant la saisie minimale de données pour les mesures fixes ne comprennent pas les pertes d'information dues à l'étalonnage régulier ou à l'entretien normal des instruments.

(2) Réparties sur l'année pour être représentatives des diverses conditions climatiques et activités anthropiques.

(3) Les mesures indicatives sont des mesures effectuées avec une régularité réduite mais qui correspondent aux autres objectifs en matière de qualité des données.

### **Méthodes de référence et rapports d'essais**

La liste des méthodes de référence pour l'évaluation des concentrations des polluants est la suivante :

NF EN 14211 - Air ambiant - Méthode normalisée pour le mesurage de la concentration en dioxyde d'azote et monoxyde d'azote par chimiluminescence

NF EN 14212 - Air ambiant - Méthode normalisée pour le mesurage de la concentration en dioxyde de soufre par fluorescence UV

NF EN 14625 - Qualité de l'air ambiant - Méthode normalisée de mesurage de la concentration d'ozone par photométrie UV.

NF EN 14626 - Air ambiant - Méthode normalisée de mesurage de la concentration en monoxyde de carbone par spectroscopie à rayonnement infrarouge non dispersif

## **Annexe : Normes de surveillance pour la pollution de l'air**

---

NF EN 14662-1 - Qualité de l'air ambiant - Méthode normalisée pour le mesurage des concentrations en benzène - Partie 1 : échantillonnage par pompage suivi d'une désorption thermique et d'une méthode chromatographie en phase gazeuse

NF EN 14662-2 - Qualité de l'air ambiant - Méthode normalisée pour le mesurage de la concentration en benzène - Partie 2 : prélèvement par pompage suivi d'une désorption au solvant et d'une méthode de chromatographie en phase gazeuse

NF EN 14662-3 - Qualité de l'air ambiant - Méthode normalisée pour le mesurage de la concentration en benzène - Partie 3 : prélèvement par pompage automatique avec analyse chromatographique en phase gazeuse sur site

NF EN 12341 - Air ambiant - Méthode normalisée de mesurage gravimétrique pour la détermination de la concentration massique MP10 ou MP2.5 de matière particulaire en suspension

NF EN 14902 - Qualité de l'air ambiant - Méthode normalisée de mesure du plomb, du cadmium, de l'arsenic et du nickel dans la fraction MP10 de matière particulaire en suspension

NF EN 15841 - Qualité de l'air ambiant - Méthode normalisée pour la détermination des dépôts d'arsenic, de cadmium, de nickel et de plomb

NF EN 15549 - Qualité de l'air - Méthode normalisée de mesure de la concentration du benzo [a] pyrène dans l'air ambiant

NF EN 15980 - Qualité de l'air - Détermination du benzo [a] anthracène, benzo [b] fluoranthène, benzo [j] fluoranthène, benzo [k] fluoranthène, benzo [a] pyrène, dibenzo [a, h] anthracène et indéno [1,2,3-cd] pyrène dans les dépôts atmosphériques

NF EN 15852 - Qualité de l'air ambiant - Méthode normalisée pour la détermination du mercure gazeux total

NF EN 15853 - Qualité de l'air ambiant - Méthode normalisée pour la détermination des dépôts de mercure

## **Références bibliographiques**

## Références bibliographiques

- [Ahmadou et al. 2017] Ahmadou, D., Laref, R., Losson, E., & Siadat, M. (2017, March). Reduction of drift impact in gas sensor response to improve quantitative odor analysis. In 2017 IEEE International Conference on Industrial Technology (ICIT) (pp. 928-933). Toronto, Canada.
- [Ahmadou 2015] Ahmadou, M. D. (2015). Contribution au développement d'un dispositif robuste de détection-diffusion d'huiles essentielles à concentration contrôlée (Doctoral dissertation, Université de Lorraine. Metz, France).
- [Al Barakeh 2012] Al Barakeh, Z. (2012). Suivi de pollution atmosphérique par système multi-capteurs–méthode mixte de classification et de détermination d'un indice de pollution. (Doctoral dissertation, EMSE. Saint-Etienne, France).
- [Aleixandre et Gerboles 2012] Aleixandre, M., & Gerboles, M. (2012). Review of small commercial sensors for indicative monitoring of ambient gas. *Chem. Eng. Trans*, 30, 169-174.
- [Apetrei et al. 2010] Apetrei, C., Apetrei, I. M., Villanueva, S., De Saja, J. A., Gutiérrez-Rosales, F., & Rodriguez-Mendez, M. L. (2010). Combination of an e-nose, an e-tongue and an e-eye for the characterisation of olive oils with different degree of bitterness. *Analytica chimica acta*, 663(1), 91-97.
- [Artursson et al. 2000] Artursson, T., Eklöv, T., Lundström, I., Mårtensson, P., Sjöström, M., & Holmberg, M. (2000). Drift correction for gas sensors using multivariate methods. *Journal of chemometrics*, 14(5-6), 711-723.
- [Awang et al. 2000] Awang, M. B., Jaafar, A. B., Abdullah, A. M., Ismail, M. B., Hassan, M. N., Abdullah, R., ... & Noor, H. (2000). Air quality in Malaysia: impacts, management issues and future challenges. *Respirology*, 5(2), 183-196.
- [Babich et Stotzky 1972] Babich, H., & Stotzky, G. (1972). Ecologie ramifications of air pollution. *SAE Transactions*, Vol. 81, pp. 1955-1971.
- [Bai 2010] Bai, Q. (2010). Analysis of particle swarm optimization algorithm. *Computer and information science*, 3(1), 180.
- [Baron et Saffell 2017] Baron, R., & Saffell, J. (2017). Amperometric gas sensors as a low cost emerging technology platform for air quality monitoring applications: A review. *ACS sensors*, 2(11), 1553-1566.
- [Basak et al. 2007] Basak, D, Srimanta, P., Dipak C., et Ra P. (2007). Support vector regression. *Neural Information Processing Letters and Reviews*, Vol. 11, No. 10, 203–224.
- [Bigi et al. 2018] Bigi, A., Mueller, M., Grange, S. K., Ghermandi, G., & Hueglin, C. (2018, April). Performance of NO/NO<sub>2</sub> low cost sensors and of three calibration algorithms within a long-term real-world application. In EGU General Assembly Conference Abstracts (Vol. 20, p. 12448). Vienne, Autriche.



## Références bibliographiques

---

- [Blanco et al. 1999] Blanco, M., Coello, J., Iturriaga, H., Maspoch, S., & Pages, J. (1999). Calibration in non-linear near infrared reflectance spectroscopy: a comparison of several methods. *Analytica Chimica Acta*, 384(2), 207-214.
- [Blanco-Rodríguez et al. 2018] Blanco-Rodríguez, A., Camara, V. F., Campo, F., Becherán, L., Durán, A., Vieira, V. D., ... & Garcia-Ramirez, A. R. (2018). Development of an electronic nose to characterize odours emitted from different stages in a wastewater treatment plant. *Water research*, 134, 92-100.
- [Bouveresse et Massart 1996] Bouveresse, E., & Massart, D. L. (1996). Improvement of the piecewise direct standardisation procedure for the transfer of NIR spectra for multivariate calibration. *Chemometrics and intelligent laboratory systems*, 32(2), 201-213.
- [Brienza et al. 2015] Brienza, S., Galli, A., Anastasi, G., & Bruschi, P. (2015). A low-cost sensing system for cooperative air quality monitoring in urban areas. *Sensors*, 15(6), 12242-12259.
- [Brook et al. 2002] Brook, R. D., Brook, J. R., Urch, B., Vincent, R., Rajagopalan, S., & Silverman, F. (2002). Inhalation of fine particulate air pollution and ozone causes acute arterial vasoconstriction in healthy adults. *Circulation*, 105(13), 1534-1536.
- [Brunekreef et Forsberg 2005] Brunekreef, B., & Forsberg, B. (2005). Epidemiological evidence of effects of coarse airborne particles on health. *European Respiratory Journal*, 26(2), 309-318.
- [Castell et al. 2017] Castell, N., Dauge, F. R., Schneider, P., Vogt, M., Lerner, U., Fishbain, B., ... & Bartonova, A. (2017). Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates?. *Environment international*, 99, 293-302.
- [Cavaliere et al. 2018] Cavaliere, A., Carotenuto, F., Di Gennaro, F., Gioli, B., Gualtieri, G., Martelli, F., ... & Zaldei, A. (2018). Development of Low-Cost Air Quality Stations for Next Generation Monitoring Networks: Calibration and Validation of PM<sub>2.5</sub> and PM<sub>10</sub> Sensors. *Sensors*, 18(9), 2843.
- [Cevoli et al. 2011] Cevoli, C., Cerretani, L., Gori, A., Caboni, M. F., Toschi, T. G., & Fabbri, A. (2011). Classification of Pecorino cheeses using electronic nose combined with artificial neural network and comparison with GC-MS analysis of volatile compounds. *Food Chemistry*, 129(3), 1315-1319.
- [Charpin et Raheison 2007] Charpin, D., & Raheison, C. (2007). Pollution atmosphérique et pathologies respiratoires. *Revue des maladies respiratoires*, 24(7), 924-926.
- [Clements et al. 2017] Clements, A. L., Griswold, W. G., Rs, A., Johnston, J. E., Herting, M. M., Thorson, J., ... & Hannigan, M. (2017). Low-cost air quality monitoring tools: from research to practice (a workshop summary). *Sensors*, 17(11), 2478.
- [Cordero et al. 2018] Cordero, J. M., Borge, R., & Narros, A. (2018). Using statistical methods to carry out in field calibrations of low cost air quality sensors. *Sensors and Actuators B: Chemical*, 267, 245-254.

## Références bibliographiques

---

- [Curtis et al. 2006] Curtis, L., Rea, W., Smith-Willis, P., Fenyves, E., & Pan, Y. (2006). Adverse health effects of outdoor air pollutants. *Environment international*, 32(6), 815-830.
- [Daudé 2012] Daudé, B. (2012). Développement d'un photomètre UV pour des mesures de référence de gaz d'intérêt atmosphérique: ozone et COVs (Doctoral dissertation, Université Pierre et Marie Curie. Paris, France).
- [De Vito et al. 2018] De Vito, S., Esposito, E., Salvato, M., Popoola, O., Formisano, F., Jones, R., & Di Francia, G. (2018). Calibrating chemical multisensory devices for real world applications: An in-depth comparison of quantitative machine learning approaches. *Sensors and Actuators B: Chemical*, 255, 1191-1210.
- [De Vito et al. 2009] De Vito, S., Piga, M., Martinotto, L., & Di Francia, G. (2009). CO, NO<sub>2</sub> and NO<sub>x</sub> urban pollution monitoring with on-field calibrated electronic nose by automatic bayesian regularization. *Sensors and Actuators B: Chemical*, 143(1), 182-191.
- [Di Carlo et Falasconi 2012] Di Carlo, S., & Falasconi, M. (2012). Drift correction methods for gas chemical sensors in artificial olfaction systems: techniques and challenges. In Tech, Prof. Wen Wang (Ed.), ISBN: 978-953-307-792-5.
- [Di Natale et al. 2002] Di Natale, C., Martinelli, E., & D'Amico, A. (2002). Counteraction of environmental disturbances of electronic nose data by independent component analysis. *Sensors and Actuators B: Chemical*, 82(2-3), 158-165.
- [Elichegaray et al. 2009] Elichegaray, C., Bouallala, S., Maitre, A., & Ba, M. (2009). État et évolution de la pollution atmosphérique. *Revue des maladies respiratoires*, 26(2), 191-206.
- [Esposito et al. 2016] Esposito, E., De Vito, S., Salvato, M., Bright, V., Jones, R. L., & Popoola, O. (2016). Dynamic neural network architectures for on field stochastic calibration of indicative low cost air quality sensing systems. *Sensors and Actuators B: Chemical*, 231, 701-713.
- [Esposito et al. 2017] Esposito, E., Salvato, M., De Vito, S., Fattoruso, G., Castell, N., Karatzas, K., & Di Francia, G. (2017, February). Assessing the Relocation Robustness of on Field Calibrations for Air Quality Monitoring Devices. In *AISEM Annual Conference on Sensors and Microsystems* (pp. 303-312). Lecce, Italie.
- [Estrella et al. 2005] Estrella, B., Estrella, R., Oviedo, J., Narváez, X., Reyes, M. T., Gutiérrez, M., & Naumova, E. N. (2005). Acute respiratory diseases and carboxyhemoglobin status in school children of Quito, Ecuador. *Environmental health perspectives*, 113(5), 607-611.
- [Faleh et al. 2016] Faleh, R., Othman, M., Gomri, S., Aguir, K., & Kachouri, A. (2016). A transient signal extraction method of WO<sub>3</sub> gas sensors array to identify pollutant gases. *IEEE Sensors Journal*, 16(9), 3123-3130.
- [Fei et Zhang 2009] Fei, S. W., & Zhang, X. B. (2009). Fault diagnosis of power transformer based on support vector machine with genetic algorithm. *Expert Systems with Applications*, 36(8), 11352-11357.

## Références bibliographiques

---

- [Fenger 2009] Fenger, J. (2009). Air pollution in the last 50 years—From local to global. *Atmospheric environment*, 43(1), 13-22.
- [Fernandez et al. 2016] Fernandez, L., Guney, S., Gutierrez-Galvez, A., & Marco, S. (2016). Calibration transfer in temperature modulated gas sensor arrays. *Sensors and Actuators B: Chemical*, 231, 276-284.
- [Feudale et al. 2002] Feudale, R. N., Woody, N. A., Tan, H., Myles, A. J., Brown, S. D., & Ferré, J. (2002). Transfer of multivariate calibration models: a review. *Chemometrics and Intelligent Laboratory Systems*, 64(2), 181-192.
- [Fonollosa et al. 2014] Fonollosa, J., Rodríguez-Luján, I., Trincavelli, M., Vergara, A., & Huerta, R. (2014). Chemical discrimination in turbulent gas mixtures with mox sensors validated by gas chromatography-mass spectrometry. *Sensors*, 14(10), 19336-19353.
- [Fu et al. 2012] Fu, Q., Wang, J., Lin, G., Suo, H., & Zhao, C. (2012). Short-wave near-infrared spectrometer for alcohol determination and temperature correction. *Journal of analytical methods in chemistry*, 2012, 728128.
- [Fuchs 2008] Fuchs, S. (2008). Conception et réalisation de méthodes de détection de polluants gazeux atmosphériques à l'aide d'un nez électronique portable (Doctoral dissertation, Université Paul Verlaine. Metz, France).
- [Gani et Limam 2016] Gani, W., & Limam, M. (2016). A kernel distance-based representative subset selection method. *Journal of Statistical Computation and Simulation*, 86(1), 135-148.
- [Gardner et Bartlett 2000] Gardner, J. W., & Bartlett, P. N. (2000). BOOK REVIEW: *Electronic Noses. Principles and Applications*. *Measurement Science and Technology*, 11, 1087.
- [Gardner et Bartlett 1994] Gardner, J. W., & Bartlett, P. N. (1994). A brief history of electronic noses. *Sensors and Actuators B: Chemical*, 18(1-3), 210-211.
- [Geladi et Kowalski 1986] Geladi, P., & Kowalski, B. R. (1986). Partial least-squares regression: a tutorial. *Analytica chimica acta*, 185, 1-17.
- [Gillooly et al. 2019] Gillooly, S. E., Zhou, Y., Vallarino, J., Chu, M. T., Michanowicz, D. R., Levy, J. I., & Adamkiewicz, G. (2019). Development of an in-home, real-time air pollutant sensor platform and implications for community use. *Environmental pollution*, 244, 440-450.
- [Godish et al. 2014] Godish, T., Davis, W. T., & Fu, J. S. (2014). *Air quality*. CRC Press, ISBN:13:978-1-4665-8445-7.
- [Gorji-Chakespari et al. 2016] Gorji-Chakespari, A., Nikbakht, A. M., Sefidkon, F., Ghasemi-Varnamkhasti, M., Brezmes, J., & Llobet, E. (2016). Performance comparison of fuzzy ARTMAP and LDA in qualitative classification of Iranian *Rosa damascena* essential oils by an electronic nose. *Sensors*, 16(5), 636.

## Références bibliographiques

---

- [Hadjiiski et al. 1999] Hadjiiski, L., Geladi, P., & Hopke, P. (1999). A comparison of modeling nonlinear systems with artificial neural networks and partial least squares. *Chemometrics and intelligent laboratory systems*, 49(1), 91-103.
- [Harrou et al. 2018] Harrou, F., Dairi, A., Sun, Y., & Kadri, F. (2018). Detecting abnormal ozone measurements with a deep learning-based strategy. *IEEE Sensors Journal*, 18(17), 7222-7232.
- [Haugen et al. 2000] Haugen, J. E., Tomic, O., & Kvaal, K. (2000). A calibration method for handling the temporal drift of solid state gas-sensors. *Analytica chimica acta*, 407(1-2), 23-39.
- [Hines et al. 1999] Hines, E. L., Llobet, E., & Gardner, J. W. (1999). Electronic noses: a review of signal processing techniques. *IEE Proceedings-Circuits, Devices and Systems*, 146(6), 297-310.
- [Hossain et al. 2016] Hossain, M., Saffell, J., & Baron, R. (2016). Differentiating NO<sub>2</sub> and O<sub>3</sub> at low cost air quality amperometric gas sensors. *ACS Sensors*, 1(11), 1291-1294.
- [Huang et al. 2006] Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: theory and applications. *Neurocomputing*, 70(1-3), 489-501.
- [Jiao et al. 2016] Jiao, W., Hagler, G., Williams, R., Sharpe, R., Brown, R., Garver, D., ... & Weinstock, L. (2016). Community Air Sensor Network (CAIRSENSE) project: evaluation of low-cost sensor performance in a suburban environment in the southeastern United States. *Atmospheric Measurement Techniques*, 9(11), 5281-5292.
- [Jurs et al. 2000] Jurs, P. C., Bakken, G. A., & McClelland, H. E. (2000). Computational methods for the analysis of chemical sensor array data from volatile analytes. *Chemical Reviews*, 100(7), 2649-2678.
- [Kamionka et al. 2006] Kamionka, M., Breuil, P., & Pijolat, C. (2006). Atmospheric pollution measurement with a multi-materials sensing device. *Materials Science and Engineering: C*, 26(2-3), 290-296.
- [Kampa et Castanas 2008] Kampa, M., & Castanas, E. (2008). Human health effects of air pollution. *Environmental pollution*, 151(2), 362-367.
- [Karagulian et al. 2019] Karagulian, F., Barbiere, M., Kotsev, A., Spinelle, L., Gerboles, M., Lagler, F., ... & Borowiak, A. (2019). Review of the Performance of Low-Cost Sensors for Air Quality Monitoring. *Atmosphere*, 10(9), 506.
- [Kennard et Stone 1969] Kennard, R. W., & Stone, L. A. (1969). Computer aided design of experiments. *Technometrics*, 11(1), 137-148.
- [Kirkpatrick et al. 1983] Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220(4598), 671-680.
- [Korotcenkov 2007] Korotcenkov, G. (2007). Metal oxides for solid-state gas sensors: What determines our choice?. *Materials Science and Engineering: B*, 139(1), 1-23.

## Références bibliographiques

---

- [Korotcenkov et Cho 2011] Korotcenkov, G., & Cho, B. K. (2011). Instability of metal oxide-based conductometric gas sensors and approaches to stability improvement (short survey). *Sensors and Actuators B: Chemical*, 156(2), 527-538.
- [Kumar et al. 2015] Kumar, P., Morawska, L., Martani, C., Biskos, G., Neophytou, M., Di Sabatino, S., ... & Britter, R. (2015). The rise of low-cost sensing for managing air pollution in cities. *Environment international*, 75, 199-205.
- [Laref et al. 2018] Laref, R., Losson, E., Sava, A., Adjallah, K., & Siadat, M. (2018, February). A comparison between SVM and PLS for E-nose based gas concentration monitoring. In *2018 IEEE International Conference on Industrial Technology (ICIT)* (pp. 1335-1339). Lyon, France.
- [Laref et al. 2019] Laref, R., Losson, E., Sava, A., & Siadat, M. (2019). On the optimization of the support vector machine regression hyperparameters setting for gas sensors array applications. *Chemometrics and Intelligent Laboratory Systems*, 184, 22-27.
- [Laref et al. 2017] Laref, R., Ahmadou, D., Losson, E., & Siadat, M. (2017). Orthogonal signal correction to improve stability regression model in gas sensor systems. *Journal of Sensors*, 2017, 9851406.
- [Laref et al. 2018] Laref, R., Losson, E., Sava, A., & Siadat, M. (2018). Support vector machine regression for calibration transfer between electronic noses dedicated to air pollution monitoring. *Sensors*, 18(11), 3716.
- [Laref et al. 2019] Laref, R., Losson, E., Sava, A., & Siadat, M. (2019, September). Field Evaluation of Low Cost Sensors Array for Air Pollution Monitoring. In *2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)* (Vol. 2, pp. 849-853). Metz, France.
- [Lauer et Bloch 2006] Lauer, F., & Bloch, G. (2006, November). Méthodes SVM pour l'identification. in *Journées Identification et Modélisation Expérimentale (JIME'2006)*. Poitiers, France.
- [Le Moullec 2018] Le Moullec, A. (2018). Bilan de la qualité de l'air extérieur en France en 2017. Edité par le service de la donnée et des études statistiques (SDES), Laboratoire Central de Surveillance de la Qualité de l'Air (LCSQA).
- [Lelieveld et al. 2015] Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D., & Pozzer, A. (2015). The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature*, 525(7569), 367-371.
- [Lentka et al. 2015] Lentka, Ł., Smulko, J. M., Ionescu, R., Granqvist, C. G., & Kish, L. B. (2015). Determination of gas mixture components using fluctuation enhanced sensing and the LS-SVM regression algorithm. *Metrology and Measurement Systems*, 22(3), 341-350.
- [Lerner et al. 2015] Lerner, U., Yacobi, T., Levy, I., Moltchanov, S. A., Cole-Hunter, T., & Fishbain, B. (2015). The effect of ego-motion on environmental monitoring. *Science of the Total Environment*, 533, 8-16.

## Références bibliographiques

---

- [Li et al. 2007] Li, C., Heinemann, P., & Sherry, R. (2007). Neural network and Bayesian network fusion models to fuse electronic nose and surface acoustic wave sensor data for apple defect detection. *Sensors and Actuators B: Chemical*, 125(1), 301-310.
- [Lin et al. 2008] Lin, S. W., Ying, K. C., Chen, S. C., & Lee, Z. J. (2008). Particle swarm optimization for parameter determination and feature selection of support vector machines. *Expert systems with applications*, 35(4), 1817-1824.
- [Liu et al. 2013] Liu, Q., Li, X., Ye, M., Ge, S. S., & Du, X. (2013). Drift compensation for electronic nose by semi-supervised domain adaption. *IEEE Sensors Journal*, 14(3), 657-665.
- [Liu et al. 2012] Liu, X., Cheng, S., Liu, H., Hu, S., Zhang, D., & Ning, H. (2012). A survey on gas sensing technology. *Sensors*, 12(7), 9635-9665.
- [Lösch et al. 2008] Lösch, M., Baumbach, M., & Schütze, A. (2008). Ozone detection in the ppb-range with improved stability and reduced cross sensitivity. *Sensors and Actuators B: Chemical*, 130(1), 367-373.
- [Lozano et al. 2007] Lozano, J., Santos, J. P., Arroyo, T., Aznar, M., Cabellos, J. M., Gil, M., & del Carmen Horrillo, M. (2007). Correlating e-nose responses to wine sensorial descriptors and gas chromatography–mass spectrometry profiles using partial least squares regression analysis. *Sensors and Actuators B: Chemical*, 127(1), 267-276.
- [Maag et al. 2018] Maag, B., Zhou, Z., & Thiele, L. (2018). A survey on sensor calibration in air pollution monitoring deployments. *IEEE Internet of Things Journal*, 5(6), 4857-4870.
- [Mabahwi et al. 2014] Mabahwi, N. A. B., Leh, O. L. H., & Omar, D. (2014). Human health and wellbeing: Human health effect of air pollution. *Procedia-Social and Behavioral Sciences*, 153, 221-229.
- [Macías et al. 2014] Macías, M. M., Agudo, J. E., Manso, A. G., Orellana, C. J. G., Velasco, H. M. G., & Caballero, R. G. (2014). Improving short term instability for quantitative analyses with portable electronic noses. *Sensors*, 14(6), 10514-10526.
- [Marco et Gutierrez-Galvez 2012] Marco, S., & Gutierrez-Galvez, A. (2012). Signal and data processing for machine olfaction and chemical sensing: A review. *IEEE Sensors Journal*, 12(11), 3189-3214.
- [Martín et al. 2001] Martín, M. A., Santos, J. P., & Agapito, J. A. (2001). Application of artificial neural networks to calculate the partial gas concentrations in a mixture. *Sensors and Actuators B: Chemical*, 77(1-2), 468-471.
- [Mathé et Malherbe 2017] Mathé, F., et Malherbe, L. (2017). Guide méthodologique sur les stations françaises de surveillance de la qualité de l'air. Edité par le Laboratoire Central de Surveillance de la Qualité de l'Air (LCSQA).
- [Mathé et Houdret 2009] Mathé, F. & Houdret, J. L. (2009). Surveillance de la qualité de l'air: le dispositif français. in *Techniques de l'ingénieur : Environnement - Sécurité | Environnement*, Vol. 1, G1632.

## Références bibliographiques

---

- [Mayer 1999] Mayer, H. (1999). Air pollution in cities. *Atmospheric environment*, 33(24-25), 4029-4037.
- [McKercher et al. 2017] McKercher, G. R., Salmond, J. A., & Vanos, J. K. (2017). Characteristics and applications of small, portable gaseous air pollution monitors. *Environmental Pollution*, 223, 102-110.
- [Mead et al. 2013] Mead, M. I., Popoola, O. A. M., Stewart, G. B., Landshoff, P., Calleja, M., Hayes, M., ... & Lewis, A. (2013). The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks. *Atmospheric Environment*, 70, 186-203.
- [Men et al. 2018] Men, H., Fu, S., Yang, J., Cheng, M., Shi, Y., & Liu, J. (2018). Comparison of SVM, RF and ELM on an Electronic Nose for the Intelligent Evaluation of Paraffin Samples. *Sensors*, 18(1), 285.
- [Moltchanov et al. 2015] Moltchanov, S., Levy, I., Etzion, Y., Lerner, U., Broday, D. M., & Fishbain, B. (2015). On the feasibility of measuring urban air pollution by wireless distributed sensor networks. *Science of The Total Environment*, 502, 537-547.
- [Morawska et al. 2018] Morawska, L., Thai, P. K., Liu, X., Asumadu-Sakyi, A., Ayoko, G., Bartonova, A., ... & Gao, J. (2018). Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: How far have they gone?. *Environment international*, 116, 286-299.
- [Mueller et al. 2017] Mueller, M., Meyer, J., & Hueglin, C. (2017). Design of an ozone and nitrogen dioxide sensor unit and its long-term operation within a sensor network in the city of Zurich. *Atmospheric Measurement Techniques*, 10(10), 3783.
- [Mukherjee et al. 2017] Mukherjee, A., Stanton, L. G., Graham, A. R., & Roberts, P. T. (2017). Assessing the utility of low-cost particulate matter sensors over a 12-week period in the Cuyama valley of California. *Sensors*, 17(8), 1805.
- [Munir et al. 2019] Munir, S., Mayfield, M., Coca, D., Jubb, S. A., & Osammor, O. (2019). Analysing the performance of low-cost air quality sensors, their drivers, relative benefits and calibration in cities—a case study in Sheffield. *Environmental monitoring and assessment*, 191(2), 94.
- [Padilla et al. 2010] Padilla, M., Perera, A., Montoliu, I., Chaudry, A., Persaud, K., & Marco, S. (2010). Drift compensation of gas sensor array data by orthogonal signal correction. *Chemometrics and Intelligent Laboratory Systems*, 100(1), 28-35.
- [Pascal 2009] Pascal, L. (2009). Effets à court terme de la pollution atmosphérique sur la mortalité. *Revue Française d'Allergologie*, 49(6), 466-476.
- [Pénard-Morand et Annesi-Maesano 2008] Pénard-Morand, C., & Annesi-Maesano, I. (2008). Maladies allergiques respiratoires et pollution atmosphérique extérieure. *Revue des maladies respiratoires*, 25(8), 1013-1026.
- [Penza et al. 2012] Penza, M., Suriano, D., Cassano, G., Pfister, V., Alvisi, M., & Rossi, R. (2012, May). Portable chemical sensor-system for urban air-pollution monitoring. In

## Références bibliographiques

---

- Proceedings of the 14th International Meeting on Chemical Sensors, (pp. 20-23) Nuremberg, Allemagne.
- [Pereira et al. 2008] Pereira, C. F., Pimentel, M. F., Galvão, R. K. H., Honorato, F. A., Stragevitch, L., & Martins, M. N. (2008). A comparative study of calibration transfer methods for determination of gasoline quality parameters in three different near infrared spectrometers. *analytica chimica acta*, 611(1), 41-47.
- [Peterson et al. 2017] Peterson, P. J., Aujla, A., Grant, K. H., Brundle, A. G., Thompson, M. R., Vande Hey, J., & Leigh, R. J. (2017). Practical use of metal oxide semiconductor gas sensors for measuring nitrogen dioxide and ozone in urban environments. *Sensors*, 17(7), 1653.
- [Piedrahita et al. 2014] Piedrahita, R., Xiang, Y., Masson, N., Ortega, J., Collier, A., Jiang, Y., ... & Shang, L. (2014). The next generation of low-cost personal air quality sensors for quantitative exposure monitoring. *Atmospheric Measurement Techniques*, 7(10), 3325.
- [Popoola et al. 2016] Popoola, O. A., Stewart, G. B., Mead, M. I., & Jones, R. L. (2016). Development of a baseline-temperature correction methodology for electrochemical sensors and its implications for long-term stability. *Atmospheric Environment*, 147, 330-343.
- [Qiu et Wang 2017] Qiu, S., & Wang, J. (2017). The prediction of food additives in the fruit juice based on electronic nose with chemometrics. *Food chemistry*, 230, 208-214.
- [Qiu et al. 2015] Qiu, S., Wang, J., Tang, C., & Du, D. (2015). Comparison of ELM, RF, and SVM on E-nose and E-tongue to trace the quality status of mandarin (Citrus unshiu Marc.). *Journal of food engineering*, 166, 193-203.
- [Rai et al. 2017] Rai, A. C., Kumar, P., Pilla, F., Skouloudis, A. N., Di Sabatino, S., Ratti, C., ... & Rickerby, D. (2017). End-user perspective of low-cost sensors for outdoor air pollution monitoring. *Science of The Total Environment*, 607, 691-705.
- [Redon et al. 2016] Redon, N., Gaudion, V., Crunaire, N., Locoge, N., (2016). Etablissement d'un protocole de détermination en laboratoire des caractéristiques de performance métrologique des micro-capteurs pour la mesure indicative des polluants gazeux inorganiques. Edité par le Laboratoire Central de Surveillance de la Qualité de l'Air (LCSQA).
- [Rudnitskaya 2018] Rudnitskaya, A. (2018). Calibration update and drift correction for electronic noses and tongues. *Frontiers in chemistry*, 6, 433.
- [Rudnitskaya et al. 2017] Rudnitskaya, A., Costa, A. M. S., & Delgadillo, I. (2017). Calibration update strategies for an array of potentiometric chemical sensors. *Sensors and Actuators B: Chemical*, 238, 1181-1189.
- [Sartakhti et al. 2012] Sartakhti, J. S., Zangooei, M. H., & Mozafari, K. (2012). Hepatitis disease diagnosis using a novel hybrid method based on support vector machine and simulated annealing (SVM-SA). *Computer methods and programs in biomedicine*, 108(2), 570-579.



## Références bibliographiques

---

- [Sayago et al. 2019] Sayago, I., Aleixandre, M., & Santos, J. P. (2019). Development of tin oxide-based nanosensors for electronic nose environmental applications. *Biosensors*, 9(1), 21.
- [Simon et Charpin 2010] Simon, I., & Charpin, D. (2010). Fluctuations des taux de polluants atmosphériques et symptômes respiratoires en population générale. *Revue des maladies respiratoires*, 27(6), 625-638.
- [Sjöblom et al. 1998] Sjöblom, J., Svensson, O., Josefson, M., Kullberg, H., & Wold, S. (1998). An evaluation of orthogonal signal correction applied to calibration transfer of near infrared spectra. *Chemometrics and Intelligent Laboratory Systems*, 44(1-2), 229-244.
- [Smola et Schölkopf 2004] Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, 14(3), 199-222.
- [Song et al. 2011] Song, K., Wang, Q., Liu, Q., Zhang, H., & Cheng, Y. (2011). A wireless electronic nose system using a Fe<sub>2</sub>O<sub>3</sub> gas sensing array and least squares support vector regression. *Sensors*, 11(1), 485-505.
- [Song et al. 2010] Song, S., Zhang, X., Hayat, K., Jia, C., Xia, S., Zhong, F., ... & Niu, Y. (2010). Correlating chemical parameters of controlled oxidation tallow to gas chromatography–mass spectrometry profiles and e-nose responses using partial least squares regression analysis. *Sensors and Actuators B: Chemical*, 147(2), 660-668.
- [Spinelle et al. 2013] Spinelle, L., Aleixandre, M., & Gerboles, M. (2013). Protocol of evaluation and calibration of low-cost gas sensors for the monitoring of air pollution. Publication Office of the European Union: Joint Research Centre, Institute for Environment and Sustainability, Luxembourg.
- [Spinelle et al. 2014] Spinelle, L., Gerboles, M., Villani, M. G., Aleixandre, M., & Bonavitacola, F. (2014, November). Calibration of a cluster of low-cost sensors for the measurement of air pollution in ambient air. In *SENSORS, 2014 IEEE* (pp. 21-24). Valence, Espagne.
- [Spinelle et al. 2015] Spinelle, L., Gerboles, M., Villani, M. G., Aleixandre, M., & Bonavitacola, F. (2015). Field calibration of a cluster of low-cost available sensors for air quality monitoring. Part A: Ozone and nitrogen dioxide. *Sensors and Actuators B: Chemical*, 215, 249-257.
- [Spinelle et al. 2017] Spinelle, L., Gerboles, M., Kok, G., Persijn, S., & Sauerwald, T. (2017). Review of portable and low-cost sensors for the ambient air monitoring of benzene and other volatile organic compounds. *Sensors*, 17(7), 1520.
- [Spinelle et al. 2017] Spinelle, L., Gerboles, M., Villani, M. G., Aleixandre, M., & Bonavitacola, F. (2017). Field calibration of a cluster of low-cost commercially available sensors for air quality monitoring. Part B: NO, CO and CO<sub>2</sub>. *Sensors and Actuators B: Chemical*, 238, 706-715.
- [Sun 2005] Sun, H. (2005). Mercer theorem for RKHS on noncompact sets. *Journal of Complexity*, 21(3), 337-349.

## Références bibliographiques

---

- [Sun et al. 2016] Sun, L., Wong, K. C., Wei, P., Ye, S., Huang, H., Yang, F., ... & Ning, Z. (2016). Development and application of a next generation air sensor network for the Hong Kong marathon 2015 air quality monitoring. *Sensors*, 16(2), 211.
- [Sun et Zhu 2019] Sun, Z., & Zhu, D. (2019). Exposure to outdoor air pollution and its human health outcomes: A scoping review. *PloS one*, 14(5), e0216550.
- [Tomic et al. 2004] Tomic, O., Eklöv, T., Kvaal, K., & Haugen, J. E. (2004). Recalibration of a gas-sensor array system related to sensor replacement. *Analytica Chimica Acta*, 512(2), 199-206.
- [Tomic et al. 2002] Tomic, O., Ulmer, H., & Haugen, J. E. (2002). Standardization methods for handling instrument related signal shift in gas-sensor array measurement data. *Analytica Chimica Acta*, 472(1-2), 99-111.
- [Torczon 1997] Torczon, V. (1997). On the convergence of pattern search algorithms. *SIAM Journal on optimization*, 7(1), 1-25.
- [Trincavelli et al. 2008] Trincavelli, M., Reggente, M., Coradeschi, S., Loutfi, A., Ishida, H., & Lilienthal, A. J. (2008, September). Towards environmental monitoring with mobile robots. In *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 2210-2215). Nice, France.
- [Vitousek et al. 1997] Vitousek, P. M., Mooney, H. A., Lubchenco, J., & Melillo, J. M. (1997). Human domination of Earth's ecosystems. *Science*, 277(5325), 494-499.
- [Wang et al. 2010] Wang, C., Yin, L., Zhang, L., Xiang, D., & Gao, R. (2010). Metal oxide gas sensors: sensitivity and influencing factors. *Sensors*, 10(3), 2088-2106.
- [Wang et al. 1991] Wang, Y., Veltkamp, D. J., & Kowalski, B. R. (1991). Multivariate instrument standardization. *Analytical chemistry*, 63(23), 2750-2756.
- [Weile et Michielssen 1997] Weile, D. S., & Michielssen, E. (1997). Genetic algorithm optimization applied to electromagnetics: A review. *IEEE Transactions on Antennas and Propagation*, 45(3), 343-353.
- [White et al. 2012] White, R. M., Paprotny, I., Doering, F., Cascio, W. E., Solomon, P. A., & Gundel, L. A. (2012). Sensors and 'apps' for community-based atmospheric monitoring. *EM: Air and Waste Management Association's Magazine for Environmental Managers*, 5, 36-40.
- [Williams et al. 2013] Williams, D. E., Henshaw, G. S., Bart, M., Laing, G., Wagner, J., Naisbitt, S., & Salmond, J. A. (2013). Validation of low-cost ozone measurement instruments suitable for use in an air-quality monitoring network. *Measurement Science and Technology*, 24(6), 065803.
- [Williams et al. 2019] Williams, R., Duvall, R., Kilaru, V., Hagler, G., Hassinger, L., Benedict, K., ... & Allen, G. (2019). Deliberating performance targets workshop: Potential paths for emerging PM<sub>2.5</sub> and O<sub>3</sub> air sensor progress. *Atmospheric Environment: X*, 2, 100031.

## Références bibliographiques

---

- [Wilson et Baietto 2009] Wilson, A. D., & Baietto, M. (2009). Applications and advances in electronic-nose technologies. *Sensors*, 9(7), 5099-5148.
- [Wilson 2012] Wilson, A. D. (2012). Review of electronic-nose technologies and algorithms to detect hazardous chemicals in the environment. *Procedia Technology*, 1, 453-463.
- [Zellers et al. 1993] Zellers, E. T., Pan, T. S., Patrash, S. J., Han, M., & Batterman, S. A. (1993). Extended disjoint principal-components regression analysis of SAW vapor sensor-array responses. *Sensors and Actuators B: Chemical*, 12(2), 123-133.
- [Zhang et al. 2011] Zhang, L., Tian, F., Kadri, C., Xiao, B., Li, H., Pan, L., & Zhou, H. (2011). On-line sensor calibration transfer among electronic nose instruments for monitoring volatile organic chemicals in indoor air quality. *Sensors and Actuators B: Chemical*, 160(1), 899-909.
- [Zhang et al. 2012] Zhang, L., Tian, F., Nie, H., Dang, L., Li, G., Ye, Q., & Kadri, C. (2012). Classification of multiple indoor air contaminants by an electronic nose and a hybrid support vector machine. *Sensors and Actuators B: Chemical*, 174, 114-125.
- [Zhang et al. 2017] Zhang, D., Guo, D., & Yan, K. (2017). Improving the Transfer Ability of Prediction Models. In *Breath Analysis for Medical Applications* (pp. 91-112). Springer, Singapore.
- [Zimmerman et al. 2018] Zimmerman, N., Presto, A. A., Kumar, S. P., Gu, J., Haurlyiuk, A., Robinson, E. S., ... & Subramanian, R. (2018). A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring. *Atmospheric Measurement Techniques*, 11, 291-313.
- [Ziyatdinov et al. 2009] Ziyatdinov, A., Chaudry, A., Persaud, K., Caminal, P., & Perera, A. (2009, May). Common principal component analysis for drift compensation of gas sensor array data. In *AIP Conference proceedings* (Vol. 1137, No. 1, pp. 566-569). Brescia, Italie.
- [Ziyatdinov et al. 2010] Ziyatdinov, A., Marco, S., Chaudry, A., Persaud, K., Caminal, P., & Perera, A. (2010). Drift compensation of gas sensor array data by common principal component analysis. *Sensors and Actuators B: Chemical*, 146(2), 460-465.
- [Zuidema et al. 2019] Zuidema, C., Afshar-Mohajer, N., Tatum, M., Thomas, G., Peters, T., & Koehler, K. (2019). Efficacy of paired electrochemical sensors for measuring ozone concentrations. *Journal of occupational and environmental hygiene*, 16(2), 179-190.
- [Zuppa et al. 2004] Zuppa, M., Distanto, C., Siciliano, P., & Persaud, K. C. (2004). Drift counteraction with multiple self-organising maps for an electronic nose. *Sensors and Actuators B: Chemical*, 98(2-3), 305-317.