

Titre: Classification et quantification de COV gazeux par apprentissage automatique avec un capteur Fabry-Pérot sur puce fonctionnalisée par polymères
Title:

Auteur: Nicolas Malo
Author:

Date: 2026

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Malo, N. (2026). Classification et quantification de COV gazeux par apprentissage automatique avec un capteur Fabry-Pérot sur puce fonctionnalisée par polymères
Citation: [Mémoire de maîtrise, Polytechnique Montréal]. PolyPublie.
<https://publications.polymtl.ca/76620/>

 **Document en libre accès dans PolyPublie**
Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/76620/>
PolyPublie URL:

Directeurs de recherche: Yves-Alain Peter
Advisors:

Programme: Génie physique
Program:

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

**Classification et quantification de COV gazeux par apprentissage automatique
avec un capteur Fabry-Pérot sur puce fonctionnalisé par polymères**

NICOLAS MALO

Département de génie physique

Mémoire présenté en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
Génie physique

Mai 2026

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Ce mémoire intitulé :

**Classification et quantification de COV gazeux par apprentissage automatique
avec un capteur Fabry-Pérot sur puce fonctionnalisé par polymères**

présenté par **Nicolas MALO**

en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
a été dûment accepté par le jury d'examen constitué de :

Frédéric LEBLOND, président

Yves-Alain PETER, membre et directeur de recherche

Bruno AGARD, membre

DÉDICACE

À ma famille et mes amis pour leur soutien tout au long de ce projet.

REMERCIEMENTS

Je tiens tout d’abord à remercier mon directeur de recherche, Yves-Alain Peter, de m’avoir accepté dans son groupe de recherche. Son expertise, ses conseils et son soutien ont grandement contribué à la réalisation de ce projet de maîtrise.

Je remercie également les membres du laboratoire, Régis Guertin, Marc-Antoine Bianki, Thomas Grenon, Cédric Lemieux-Leduc, Richard Bienvenu Bomba, Nicole Tebchrany, Salvador Poveda Hospital et Louis-Frédéric Racicot, pour leur aide et leur soutien tout au long de ce projet.

Je souhaite également remercier Yves Drolet pour son aide précieuse dans la mise en place et le bon fonctionnement des montages expérimentaux de gaz.

Je tiens également à exprimer ma gratitude envers Rosin Claude Ngueveu pour son aide et ses conseils dans la partie apprentissage automatique de ce projet.

Je remercie aussi le personnel du LMF, Christophe Clément, Marie-Hélène Bernier, Alexandre Perron et André Hoang Gaona, qui m’ont aidé et formé aux procédés de microfabrication en salle blanche.

J’aimerais également remercier ma famille, mes amis ainsi que ma copine Lucie pour tout le soutien apporté au cours de ce projet.

Enfin, je tiens à remercier les membres du jury pour avoir accepté d’évaluer ce travail.

RÉSUMÉ

La détection des composés organiques volatils (COV) représente un enjeu important dans de nombreux domaines, notamment la surveillance environnementale ainsi que les secteurs industriel et médical. Parmi les différentes technologies de capteurs de gaz, les capteurs optiques de type Fabry-Pérot fonctionnalisés par des polymères constituent une approche prometteuse en raison de leur forte sensibilité, de leur opération à température ambiante et de leur capacité de miniaturisation. Cependant, les signaux générés par ces capteurs sont extrêmement complexes et présentent souvent des dynamiques temporelles non linéaires en raison des phénomènes de diffusion et d'absorption des analytes gazeux dans les polymères. L'interprétation de ces signaux constitue donc un défi important pour l'estimation des gaz présents ainsi que de leurs concentrations.

Dans ce mémoire, différentes méthodes d'apprentissage automatique ont été étudiées afin de réaliser la classification et la quantification de COV à partir des signaux d'un capteur Fabry-Pérot sur puce fonctionnalisé par des polymères. Les performances de plusieurs modèles, tels que les modèles linéaires, les machines à vecteurs de support (SVM), les réseaux de neurones convolutionnels et récurrents, ont été évaluées. Les modèles ont été entraînés et testés à l'aide d'une validation croisée imbriquée afin d'assurer une estimation robuste de leurs performances. Les deux régimes de fonctionnement du capteur, soit les régimes quasi statique et transitoire, ont également été caractérisés afin d'étudier les modèles les mieux adaptés à chacun d'eux.

Des mesures ont été réalisées pour quatre COV (acétone, isopropanol, méthanol et toluène) à l'aide de quatre polymères sensibles (PDMS, PMMA, PU et PVP), dont certains ont été étudiés pour la première fois dans des cavités Fabry-Pérot. Parmi les différents modèles évalués, les réseaux de neurones convolutionnels bidimensionnels (CNN2D) ont obtenu les meilleures performances globales, avec un coefficient de détermination moyen de $R^2 = 0.78$ pour la régression des concentrations sur l'ensemble des gaz et jusqu'à $R^2 = 0.92$ pour l'acétone. De plus, les résultats de classification obtenus avec ce modèle ont atteint une précision de 94 %.

Ce travail démontre le potentiel de la combinaison entre les cavités Fabry-Pérot fonctionnalisées par polymères et les techniques d'apprentissage automatique pour la détection et la quantification de COV. Les résultats obtenus contribuent à améliorer l'interprétation des signaux et à favoriser le développement de capteurs optiques intelligents pour la détection de gaz.

ABSTRACT

The detection of volatile organic compounds (VOCs) is an important challenge in many fields, including environmental monitoring as well as industrial and medical applications. Among the various gas sensing technologies, optical sensors based on Fabry-Pérot cavities and functionalized with polymers constitute a promising approach due to their high sensitivity, low operating temperature, and potential for miniaturization. However, the signals generated by these sensors are extremely complex and often show nonlinear temporal dynamics due to gas diffusion and absorption phenomena within the polymers. Interpreting these signals therefore represents a significant challenge for estimating the gases present and their concentrations.

In this thesis, different machine learning methods were investigated to perform the classification and quantification of VOCs from the signals of a polymer-functionalized on-chip Fabry-Pérot sensor. The performance of several models such as linear models, support vector machines (SVM), convolutional neural networks, and recurrent neural networks, was evaluated. The models were trained and tested using nested cross-validation in order to ensure a robust estimation of their performance. The two operating regimes of the sensor, quasi-static and transient, were also characterized to study the models best suited for each of them.

Measurements were performed for four VOCs (acetone, isopropanol, methanol, and toluene) using four sensitive polymers (PDMS, PMMA, PU, and PVP), some of which were investigated for the first time in Fabry-Pérot cavities. Among the evaluated models, two-dimensional convolutional neural networks (CNN2D) achieved the best overall performance, with an average coefficient of determination of $R^2 = 0.78$ for concentration regression across all gases and up to $R^2 = 0.92$ for acetone. Also, the classification results obtained with this model reached an accuracy of 94 %.

This work demonstrates the potential of combining polymer-functionalized Fabry-Pérot cavities with machine learning techniques for the detection and quantification of VOCs. The obtained results contribute to improving signal interpretation and support the development of intelligent optical sensors for gas detection.

TABLE DES MATIÈRES

DÉDICACE	iii
REMERCIEMENTS	iv
RÉSUMÉ	v
ABSTRACT	vi
LISTE DES TABLEAUX	x
LISTE DES FIGURES	xii
LISTE DES SIGLES ET ABRÉVIATIONS	xiv
LISTE DES ANNEXES	xv
CHAPITRE 1 INTRODUCTION	1
1.1 Problématique	2
1.2 Objectifs de recherche	3
1.3 Plan du mémoire	3
CHAPITRE 2 REVUE DE LITTÉRATURE	5
2.1 Technologies de capteur de gaz	5
2.1.1 Capteurs électriques	5
2.1.2 Capteurs mécaniques	6
2.1.3 Capteurs optiques	6
2.2 Principes des cavités Fabry-Pérot pour la détection de gaz	7
2.2.1 Interféromètre Fabry-Pérot	7
2.2.2 Les polymères	10
2.2.3 Physique des polymères appliquée aux cavités Fabry-Pérot pour la détection de gaz	11
2.3 Apprentissage Automatique Supervisé	17
2.3.1 Algorithmes classiques de régression et classification	17
2.3.2 Algorithmes avancés de classification et régression supervisées	21
2.3.3 Travaux récents	26
CHAPITRE 3 MÉTHODOLOGIE	28

3.1	Jeu de données publiques et apprentissage préliminaire	28
3.1.1	Description du jeu de données	28
3.1.2	Sélection préliminaire des modèles	29
3.1.3	Prétraitement des données	29
3.1.4	Optimisation des hyperparamètres	32
3.1.5	Métriques de comparaison des modèles	33
3.2	Acquisition et utilisation des données expérimentales	33
3.2.1	Fabrication des capteurs	34
3.2.2	Montage expérimental	37
3.2.3	Prétraitement et analyse des données expérimentales	41
3.3	Application des modèles et optimisation	41
CHAPITRE 4 RÉSULTATS ET DISCUSSION		44
4.1	Comparaison des réponses du jeu de données en ligne avec les capteurs Fabry-Pérot	44
4.2	Jeu de données de UCSD	45
4.2.1	Analyse de l'uniformité des données	46
4.2.2	Étude temporelle des capteurs	46
4.2.3	Choix des 4 meilleurs capteurs	48
4.2.4	Analyse de la séparabilité des gaz par projection PCA	49
4.2.5	Entraînement des différents modèles d'apprentissage automatique	50
4.3	Cavités Fabry-Pérot	54
4.3.1	Validation expérimentale des capteurs Fabry-Pérot fonctionnalisés par polymères	54
4.3.2	Performances des polymères	55
4.3.3	Modèle de régression	56
4.3.4	Modèles de classification	59
4.3.5	Comparaison des performances avec la littérature	61
4.3.6	Limitations et problématique des capteurs	62
CHAPITRE 5 CONCLUSION		66
5.1	Synthèse des travaux	66
5.2	Limitations de la solution proposée	67
5.3	Améliorations futures	68
RÉFÉRENCES		69

ANNEXES	77
-------------------	----

LISTE DES TABLEAUX

Tableau 2.1	Sensibilités des polymères et limites de détection pour CH ₄ et CO ₂ , d'après Guertin et al., 2024.	16
Tableau 2.2	Résumé de travaux récents utilisant l'apprentissage automatique pour la détection de COV	27
Tableau 3.1	Plages de concentration étudiées pour chaque gaz.	29
Tableau 3.2	Familles de modèles évaluées dans cette étude.	29
Tableau 3.3	Modèles évalués et hyperparamètres optimisés	32
Tableau 3.4	Polymères utilisés pour la fonctionnalisation des cavités Fabry-Pérot .	36
Tableau 4.1	Sensibilités, coefficients de détermination (R^2) et écarts-types des quatre meilleurs capteurs pour le CO, le méthane et l'éthylène.	49
Tableau 4.2	Comparaison des performances de régression (moyenne $\pm$ écart-type sur 3 plis externes). Le NRMSE est normalisé sur les données d'entraînement. Le MAE est exprimé en ppm. G1 représente l'éthylène dans la série de mesure et G0 l'autre gaz (Méthane/CO).	51
Tableau 4.3	Comparaison des performances en classification 3 classes (moyenne $\pm$ écart-type, 3 plis externes). Les classes sont : G1 seul (éthylène), G0 seul (méthane ou CO) et mélange G0+G1. Le Recall _{mix} correspond au rappel de la classe de mélange (G0+G1).	53
Tableau 4.4	Performances des polymères sélectionnés pour la détection des quatre COV.	55
Tableau 4.5	Performance des modèles (moyenne $\pm$ écart-type) selon deux régimes d'entrée : Quasi statique et Transitoire. Validation croisée imbriquée par bloc (3 plis externes). Les modèles sont triés (par gaz) du plus faible NRMSE au plus élevé.	57
Tableau 4.6	Comparaison des performances pour les modèles : LR, SVR à noyau RBF et le réseau CNN2D pour la prédiction de la concentration des gaz. Les résultats correspondent à la moyenne et à l'écart-type obtenus sur les plis externes de la validation croisée imbriquée. Les modèles sont triés selon le NRMSE équilibré. Le NRMSE _{eq} correspond au NRMSE équilibré, défini comme la moyenne du NRMSE en régime quasi statique (NRMSE _{qs}) et du NRMSE en régime transitoire (NRMSE _{tr}). .	59

Tableau 4.7	Comparaison des modèles pour la classification des gaz (Acétone, IPA, Méthanol et Toluène) à l'aide d'une validation croisée imbriquée bloquée (3 plis externes). Les valeurs sont présentées sous la forme moyenne $\pm$ écart-type.	60
Tableau 4.8	Résumé de travaux récents utilisant l'apprentissage automatique pour la détection de COV	61
Tableau A.1	Domaines de recherche des hyperparamètres utilisés lors de l'optimisation par Optuna pour les différents modèles d'apprentissage machine.	77
Tableau B.1	Sensibilités des capteurs, résultats de régression et répétabilité pour le CO, le Méthane et l'Éthylène, regroupés par type de capteur. Les meilleurs capteurs sont surlignés	79
Tableau C.1	Principaux hyperparamètres optimaux obtenus pour les modèles de régression lors de l'optimisation dans les plis internes de la validation croisée imbriquée.	80
Tableau C.2	Principaux hyperparamètres optimaux obtenus pour les modèles de classification lors de l'optimisation dans les plis internes de la validation croisée imbriquée.	80
Tableau D.1	Caractérisation des capteurs pour IPA. Les capteurs sélectionnés (chip1 : 2,8 ; chip3 : 1,3) sont surlignés.	81
Tableau D.2	Caractérisation des capteurs pour Acetone. Les capteurs sélectionnés (chip1 : 2,8 ; chip3 : 1,3) sont surlignés.	82
Tableau D.3	Caractérisation des capteurs pour Toluene. Les capteurs sélectionnés (chip1 : 2,8 ; chip3 : 1,3) sont surlignés.	82
Tableau D.4	Caractérisation des capteurs pour Méthanole. Les capteurs sélectionnés (chip1 : 2,8 ; chip3 : 1,3) sont surlignés.	83
Tableau E.1	Principaux hyperparamètres optimaux obtenus pour les modèles de régression lors de l'optimisation dans les plis internes de la validation croisée imbriquée. Ces valeurs ont été sélectionnées à partir d'une métrique pondérée combinant les régimes quasi-statiques et transitoires.	84
Tableau E.2	Principaux hyperparamètres optimaux obtenus pour les modèles de classification lors de l'optimisation dans les plis internes de la validation croisée imbriquée.	84

LISTE DES FIGURES

Figure 2.1	Interféromètre de Fabry-Pérot	8
Figure 2.2	Exemple de franges d'interférence produites par un étalon de Fabry-Pérot.	8
Figure 2.3	Spectre de transmission des cavités Fabry-Pérot au mode fondamental TEM_{00} avec coefficients de transmission et de réflexion $(t, r) = \{(0.9, 0.1), (0.1, 0.9)\}$ et $I_0 = 1$	9
Figure 2.4	Interféromètre Fabry-Pérot sur puce avec parties principales identifiées.	11
Figure 2.5	Décalage spectral $\Delta\lambda_{res}$ du capteur recouvert de polymère PU lors de l'exposition à 96 sccm d'acétone. Le décalage de la réponse spectrale ($\Delta\lambda = 6.32$ nm) est estimé par corrélation croisée du spectre complet.	12
Figure 2.6	Schéma résumant les principaux phénomènes à la source de la détection d'analyte dans les cavités Fabry-Pérot à l'aide de polymères.	13
Figure 3.1	Schéma de validation croisée imbriquée : 3 plis externes (évaluation) et 2 plis internes (optimisation) en respectant l'ordre temporel.	31
Figure 3.2	Schéma du procédé de microfabrication des cavités Fabry-Pérot.	34
Figure 3.3	Tranche de silicium microfabriquée avant l'étape de métallisation.	35
Figure 3.4	Représentation des capteurs Fabry-Pérot suite à la fabrication	37
Figure 3.5	Représentation graphique de la recette de l'analyte (MF2) appliquée aux polymères	39
Figure 3.6	Montage optique et d'injection de gaz pour effectuer les mesures. L'ensemble des systèmes sont contrôlés par un ordinateur.	40
Figure 4.1	Comparaison de la réponse temporelle d'un capteur du jeu de données en ligne, avec ajustement à l'équation de diffusion à deux constantes (<i>two-stage diffusion</i>)	44
Figure 4.2	Exemple de relation linéaire du déplacement en phase stable des capteurs chimiques pour différente concentration d'éthylène pour le capteur de TGS-2600	45
Figure 4.3	Top 5 des paires de concentrations (issues des deux mesures) pour chaque couple de gaz	46
Figure 4.4	Exemple des décalages temporels des capteurs sensibles au CO lorsqu'une concentration de 533 ppm est insérée au système. Le début de réaction des capteurs est identifié en rouge.	47
Figure 4.5	Comparaison du capteur TGS2602 défectueux (TGS2602_2) avec un capteur fonctionnel (TGS2602_1) de même type, lorsqu'exposé à un gaz.	48

Figure 4.6	Projection PCA des échantillons purs d'éthylène, de méthane et de CO pour deux campagnes de mesure indépendantes.	50
Figure 4.7	Matrice de confusion du meilleur modèle CNN2D pour la classification des gaz (Acétone, IPA, Méthanol et Toluène). La matrice correspond à la somme des prédictions des trois plis externes de la validation croisée imbriquée, puis les valeurs sont normalisées par ligne.	60
Figure 4.8	Dérive du signal pour deux types de polymères. La dérive lente du signal est estimée à l'aide de l'algorithme <i>Asymmetric Least Squares</i> (ALS). La recette de cette mesure est celle de la Figure 3.5	63
Figure 4.9	Comparaison du signal normalisé d'un capteur en polyuréthane (PU) lors d'expositions à l'acétone et au toluène. La dérive lente du niveau de base est estimée à l'aide de l'algorithme <i>Asymmetric Least Squares</i> (ALS).	64
Figure 4.10	Exemple d'évolution du spectre d'un capteur Fabry-Pérot utilisant le PU comme polymère après 167 jours	64
Figure B.1	Comparaison de la réponse temporelle du capteur Fabry-Pérot avec polymère PDMS et du jeu de données, avec ajustement à l'équation de diffusion à deux constantes (<i>two-stage diffusion</i>).	78

LISTE DES SIGLES ET ABRÉVIATIONS

IA	Intelligence Artificielle
ML	Machine Learning
PCA	Principal Component Analysis
KPCA	Kernel Principal Component Analysis
SVM	Support Vector Machine
SVR	Support Vector Regression
CNN	Convolutional Neural Network
CNN2D	Two-Dimensional Convolutional Neural Network
LR	Linear Regression
FP	Fabry-Pérot
FPI	Fabry-Pérot Interferometer
VOC	Volatile Organic Compound
COV	Composé Organique Volatil
IPA	Isopropanol
PDMS	Polydimethylsiloxane
PEVA	Polyethylene Vinyl Acetate
PU	Polyurethane
PVP	Polyvinylpyrrolidone
SNR	Signal-to-Noise Ratio
NRMSE	Normalized Root Mean Square Error
MAE	Mean Absolute Error
RMSE	Root Mean Square Error
MSE	Mean Squared Error
CV	Cross Validation
NCV	Nested Cross Validation
ALS	Asymmetric Least Squares
ppm	Parts per million
sccm	Standard cubic centimeters per minute
SOI	Silicon-On-Insulator

LISTE DES ANNEXES

Annexe A	Domaine des hyperparamètres pour les algorithmes d'apprentissage automatique	77
Annexe B	Ajustement de la courbe du modèle <i>two stage diffusion</i> sur les capteurs Fabry-Pérot et performance des capteurs chimiques	78
Annexe C	Résultats d'optimisation des hyperparamètres pour les modèles de régression et de classification pour le jeu de données	80
Annexe D	Performances Polymères Fabry-Pérot	81
Annexe E	Résultats des hyperparamètres pour les cavités Fabry-Pérot	84

CHAPITRE 1 INTRODUCTION

Le sens de l'odorat chez les humains est basé sur un système biologique complexe permettant la détection et l'identification d'un grand nombre de composés chimiques présents dans l'air. C'est plus précisément le système olfactif, composé notamment de récepteurs olfactifs, qui est responsable de ce phénomène par sa capacité à interagir avec différentes molécules odorantes [1]. Lorsqu'une molécule odorante se lie à ces récepteurs, un ensemble de signaux neuronaux est généré et transmis vers le bulbe olfactif puis vers différentes régions du cerveau impliquées dans leur traitement.

L'identification d'une odeur ne repose pas uniquement sur la détection d'une seule molécule, mais bien sur la reconnaissance d'une signature unique de récepteurs activés. Cette signature est alors traitée par le cortex olfactif, notamment le cortex piriforme, qui joue un rôle important dans la détection des relations entre différentes odeurs [2].

Ce principe a inspiré le développement de plusieurs systèmes de détection chimique, soit les nez électroniques, qui reposent sur l'utilisation d'un ensemble de capteurs sensibles et de méthodes de traitement de données pour interpréter leurs signaux. Dans ce cas, les capteurs simulent les récepteurs olfactifs alors que les algorithmes remplissent une fonction semblable à celle des neurones du cerveau humain.

Selon les différentes technologies de capteurs de gaz développées au cours des dernières décennies, les capteurs optiques présentent plusieurs avantages, notamment une bonne sensibilité, stabilité et sélectivité [3]. Parmi les différentes approches optiques, on retrouve les capteurs basés sur l'interférométrie. En particulier, les cavités de Fabry-Pérot reposent sur l'interférence multiple de la lumière entre deux surfaces parallèles semi-réfléchissantes pour former une cavité optique [4]. Toute modification de l'indice de réfraction ou de la distance entre les surfaces parallèles entraîne un décalage du spectre permettant ainsi la détection de molécules gazeuses.

Dans ces dispositifs, la cavité est fonctionnalisée avec un matériau sensible, soit un polymère, capable d'interagir avec des molécules gazeuses. Le processus d'absorption et de diffusion des analytes entraîne des changements au niveau de ses propriétés, telles que son épaisseur et son indice de réfraction.

Cependant, les signaux générés par ces capteurs peuvent présenter des comportements très complexes en raison de leurs réponses transitoires non linéaires associées aux phénomènes de diffusion et d'adsorption dans les polymères. Dès lors, les méthodes d'apprentissage automa-

tique constituent un outil puissant pour interpréter ces données et améliorer les performances de classification et de quantification des gaz.

1.1 Problématique

La détection de composés organiques volatils (COV) est un enjeu important dans de nombreux domaines, notamment environnemental, médical et le secteur industriel. Bien que plusieurs technologies de capteurs de gaz aient été développées, chacune présente certaines limitations qui restreignent leur application et leurs performances dans des conditions spécifiques.

Les capteurs optiques de type Fabry-Pérot fonctionnalisés par des polymères représentent une approche prometteuse pour la détection des COV, en raison de leur sensibilité élevée, de leur fonctionnement à température ambiante, de leur potentiel de miniaturisation et de leur faible coût. De plus, contrairement à plusieurs autres technologies, ces capteurs peuvent être configurés en matrices, permettant ainsi la détection simultanée de plusieurs gaz.

Toutefois, les signaux générés par ces capteurs présentent des réponses temporelles complexes et non linéaires en régime transitoire, en raison des phénomènes d'absorption et de diffusion des analytes dans les polymères. Ceci rend difficile l'extraction d'informations pertinentes, telles que l'identification et la quantification des gaz.

Dès lors, les approches traditionnelles basées sur des modèles analytiques ou des méthodes de traitement simple du signal sont insuffisantes pour exploiter pleinement le potentiel de ces capteurs, notamment en temps réel. Il devient alors nécessaire de développer des méthodes capables de modéliser ces comportements complexes et de discriminer efficacement les différents gaz.

Ainsi, l'apprentissage automatique semble être une solution particulièrement adaptée à cette problématique, en permettant l'analyse de signaux multidimensionnels et non linéaires, et en améliorant les performances en classification et en quantification des COV.

1.2 Objectifs de recherche

L'objectif principal de ce mémoire est de développer et d'évaluer différentes méthodes d'apprentissage automatique pour la classification et la quantification de composés organiques volatils (COV) à partir des signaux générés par un capteur optique de type Fabry-Pérot sur puce fonctionnalisé par des polymères.

Plus spécifiquement, ce travail vise à :

1. Identifier et trouver de nouveaux polymères sensibles aux COV pouvant être utilisés comme matériaux fonctionnels pour le capteur.
2. Analyser les réponses temporelles du capteur lors de l'exposition à différents gaz afin de caractériser leur comportement dynamique.
3. Développer et comparer plusieurs modèles d'apprentissage automatique pour la classification et la régression à partir des signaux mesurés.
4. Évaluer les performances des modèles à l'aide de métriques adaptées et comparer les résultats obtenus avec ceux de la littérature.

1.3 Plan du mémoire

La structure de ce mémoire est organisée en quatre chapitres principaux.

Le deuxième chapitre présente les bases théoriques nécessaires à la compréhension de ce travail. Il introduit notamment les principes de détection des capteurs de gaz, le fonctionnement des interféromètres de Fabry-Pérot ainsi que les interactions entre les polymères et les molécules gazeuses. Une introduction aux concepts fondamentaux de l'apprentissage automatique est également présentée afin d'établir les bases nécessaires à l'analyse des signaux issus des capteurs. Par ailleurs, une revue des travaux récents de la littérature portant sur l'utilisation d'algorithmes d'apprentissage automatique appliqués aux capteurs de gaz est discutée, notamment en ce qui concerne les performances obtenues pour la classification et la quantification de différents analytes à l'aide de diverses technologies de capteurs.

Le troisième chapitre décrit la méthodologie expérimentale utilisée dans ce travail. Il présente la conception des capteurs, le système expérimental employé pour l'exposition aux gaz ainsi que les procédures d'acquisition et de traitement des données. Les différents modèles d'apprentissage automatique utilisés sont également introduits, de même que les méthodes employées pour l'évaluation de leurs performances.

Le quatrième chapitre présente et analyse les résultats obtenus pour la classification et la quantification des gaz à partir des signaux des capteurs. Les performances des différents

modèles sont comparées et discutées en fonction des caractéristiques des données et des régimes de fonctionnement du capteur.

Finalement, le cinquième chapitre présente les conclusions générales de ce travail, les principales contributions de l'étude ainsi que les limitations associées à cette approche et aux capteurs étudiés. Des perspectives de recherche pour des travaux futurs sont également proposées.

CHAPITRE 2 REVUE DE LITTÉRATURE

Dans ce chapitre, il sera question de présenter les différents concepts théoriques et les travaux scientifiques nécessaires à la compréhension du capteur étudié ainsi que les méthodes d'analyse utilisées au cours de ce mémoire. Dans un premier temps, une courte introduction aux différents types de capteurs de gaz sera présentée. Suite à ceci, les principes physiques des cavités Fabry-Pérot seront introduits, notamment au niveau du fonctionnement de la détection de gaz par l'entremise de polymères.

Par la suite, les principales approches d'apprentissage automatique appliquées à l'analyse des signaux de capteurs de gaz seront présentées. Ces méthodes permettent de modéliser efficacement les réponses des capteurs, notamment en présence de comportements non linéaires ou de dynamiques transitoires complexes, afin de réaliser des prédictions pour des tâches de classification et de régression des gaz.

Finalement, un aperçu des travaux récents, combinant des capteurs de gaz pour COV et des techniques d'apprentissage automatique, sera présenté afin d'identifier les approches les plus pertinentes pour l'analyse de données expérimentales et de situer les résultats de ce projet par rapport aux performances de l'état de l'art.

2.1 Technologies de capteur de gaz

Les capteurs de gaz reposent sur différents principes de détection permettant d'identifier et de caractériser la présence de composés dans l'air. Ces technologies peuvent être classées en plusieurs catégories selon leur principe de fonctionnement, notamment les capteurs électriques, mécaniques et optiques. Chacune de ces approches présente des avantages et des limitations qui seront brièvement abordés dans les sous-sections suivantes.

2.1.1 Capteurs électriques

Les capteurs électriques constituent l'une des technologies les plus répandues pour la détection de gaz. Parmi ceux-ci, les capteurs à oxydes métalliques (MOS) fonctionnent en mesurant les variations de conductivité électrique du matériau sensible lors de l'adsorption de molécules gazeuses à la surface [5, 6]. Ces capteurs offrent généralement une bonne sensibilité à faible coût. Toutefois, ils nécessitent souvent des températures de fonctionnement élevées afin d'activer les réactions de surface, ce qui augmente leur consommation d'énergie et peut poser des risques en présence de gaz inflammables ou explosifs [7].

Les capteurs de gaz basés sur des transistors à effet de champ de type MOSFET (*Metal Oxide Semiconductor Field Effect Transistor*) constituent une autre catégorie de capteurs électriques. Dans ces dispositifs, l'adsorption de molécules gazeuses sur un matériau catalytique situé au niveau de la grille modifie le champ électrique du transistor, ce qui entraîne donc une variation du courant drain-source proportionnelle à la concentration du gaz. Malgré leur potentiel d'intégration dans des systèmes compacts, ceux-ci ont aussi besoin d'opérer à des températures relativement élevées pour obtenir des performances adéquates et sont sensibles aux conditions environnementales [8, 9].

2.1.2 Capteurs mécaniques

Les capteurs mécaniques reposent sur la détection de variations de masse associées à l'adsorption de molécules gazeuses dans un matériau sensible. En effet, les microbalances à quartz (QCM) et les capteurs à ondes acoustiques de surface (SAW) mesurent les changements de fréquence de résonance d'une structure vibrante selon la variation de masse du matériau sensible par absorption d'un analyte. Ceux-ci offrent une grande sensibilité aux faibles variations de quantité de gaz et un temps de réponse très rapide, mais peuvent être influencés par les vibrations, leur stabilité à long terme et les conditions environnementales, notamment la température et l'humidité [10, 11].

2.1.3 Capteurs optiques

Les capteurs optiques reposent plutôt sur la mesure de variations des propriétés optiques d'un système en présence d'un analyte. Ces variations sont souvent attribuées au changement d'indice de réfraction ou de longueur d'onde d'interférence. Plusieurs techniques existent, notamment la spectroscopie d'absorption optique, où la présence de gaz est déterminée selon l'absorption sélective de certaines longueurs d'onde lumineuses caractéristiques de la molécule étudiée. Des méthodes, telles que la spectroscopie infrarouge non dispersive (NDIR), sont souvent utilisées en raison de leur excellente sélectivité, car chaque molécule possède un spectre d'absorption spécifique. Cependant, les coûts des systèmes optiques, tels que les sources infrarouges et les spectromètres, les rendent très dispendieux. De plus, ceux-ci sont limités aux gaz possédant des bandes d'absorption infrarouges fortes, limitant leur application [12, 13].

De plus, une autre approche courante est d'utiliser des capteurs à fibre optique, où la lumière se propage à l'intérieur de la fibre et seulement une section de la fibre est fonctionnalisée par un matériau sensible. L'interaction entre le gaz et le matériau entraîne alors un changement de l'intensité, de la phase ou de la longueur d'onde transmise. Ceux-ci ont l'avantage d'être

compacts, résistants aux environnements difficiles et adaptés à des mesures à longue distance. Toutefois, ceux-ci peuvent être sensibles aux perturbations mécaniques et aux variations de température [14, 15].

Les capteurs interférométriques, tels que les capteurs d'interférence Fabry-Pérot, constituent aussi un type de capteur optique. Ceux-ci sont formés de cavités optiques entre deux surfaces réfléchissantes dans lesquelles la lumière est réfléchie pour produire un signal d'interférence dépendamment de la longueur optique de celle-ci. Lorsque le matériau sensible interagit avec un gaz, des phénomènes tels que l'absorption et la diffusion peuvent modifier son indice de réfraction ou son volume, entraînant ainsi un déplacement du spectre d'interférence mesuré. Ces variations peuvent être utilisées pour déterminer la présence et la concentration du gaz [16, 17].

Parmi ces différentes approches, les capteurs Fabry-Pérot présentent une solution particulièrement intéressante pour la détection de gaz. Ceux-ci peuvent être miniaturisés, fonctionner à température ambiante, ont une grande sensibilité, sont peu dispendieux et fonctionnent passivement. Lorsqu'ils sont fonctionnalisés avec des polymères, les réponses de ces derniers sont réversibles et l'utilisation de plusieurs polymères en série leur permet d'obtenir une excellente sélectivité. Néanmoins, les signaux générés par ces capteurs présentent souvent des réponses temporelles complexes et non linéaires liées aux phénomènes de diffusion et d'adsorption dans les polymères. Leur principe de fonctionnement sera détaillé dans la section suivante.

2.2 Principes des cavités Fabry-Pérot pour la détection de gaz

2.2.1 Interféromètre Fabry-Pérot

Les interféromètres de Fabry-Pérot sont des dispositifs optiques constitués de deux miroirs semi-réfléchissants parallèles formant une cavité optique [18]. Ils ont été conçus et développés par Charles Fabry et Alfred Perot en 1899 [19]. Lorsqu'un rayon lumineux traverse ces miroirs, il subit de multiples réflexions internes avant d'être transmis à l'extérieur de la cavité.

Les différentes ondes transmises sont parallèles entre elles. Le déphasage entre ces ondes résulte des différences de chemin optique associées aux multiples réflexions de la lumière entre les miroirs [20]. Ces ondes réfléchies interfèrent donc entre elles en raison de leur déphasage dans le plan focal illustré à la Figure 2.1, c'est-à-dire au niveau du détecteur. Le résultat du déphasage résulte donc du signal d'interférence représenté à l'aide de la Figure 2.2. Les maxima d'intensité apparaissent lorsque la condition d'interférence constructive $2nd \cos(\theta) = m\lambda$ est satisfaite, où n représente l'indice du milieu, d la distance entre les miroirs, θ l'angle

d'incidence par rapport au plan normal et m un entier. Puisque θ varie radialement, on observe l'apparition de franges circulaires concentriques comme montré à la Figure 2.2.

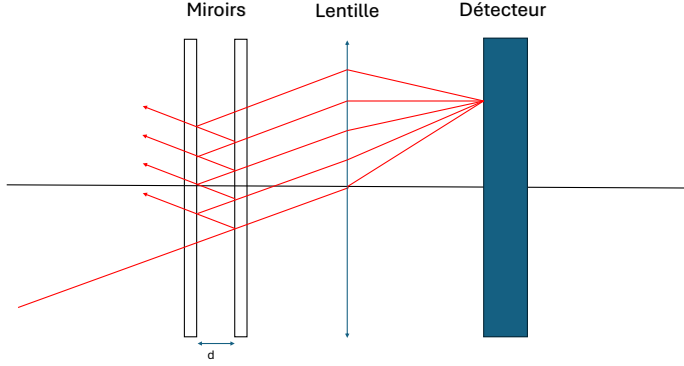


FIGURE 2.1 Interféromètre de Fabry-Pérot

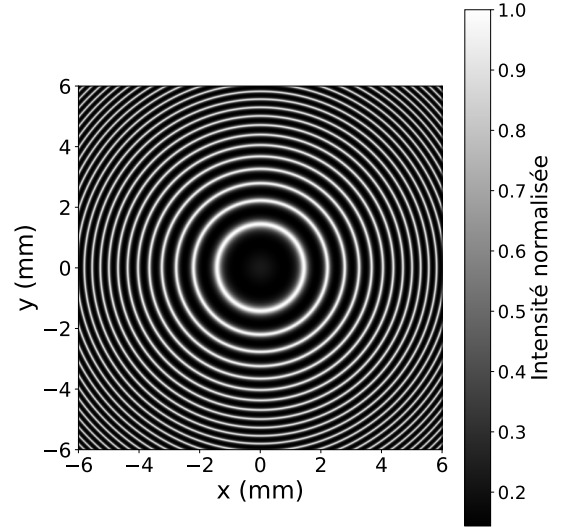


FIGURE 2.2 Exemple de franges d'interférence produites par un étalon de Fabry-Pérot.

Le déphasage entre deux rayons passant par l'interféromètre est noté de la manière suivante en optique géométrique :

$$\Delta\phi = \frac{4\pi nd}{\lambda_0} \cos(\theta) \quad (2.1)$$

Ici, λ_0 représente la longueur d'onde dans le vide. On remarque qu'un changement de la longueur optique, tel que d ou n , vient créer un déphasage. Les cavités Fabry-Pérot sont donc particulièrement sensibles aux variations de longueur optique, ce qui les rend intéressantes pour des applications de capteurs.

Dans le cas où la lumière est parfaitement adaptée spatialement au mode fondamental transverse électromagnétique TEM_{00} , correspondant à la distribution spatiale gaussienne fondamentale du faisceau dans la cavité, aucun mode transverse supérieur n'est excité. Dans ce cas, seuls les modes longitudinaux sont présents. Dès lors, l'intensité de transmission des cavités Fabry-Pérot à miroirs droits en fonction de la fréquence du faisceau est donnée par la fonction d'Airy [21, 22].

$$I_t(\Delta v_q) = \frac{t_1 t_2}{\left(1 - \sqrt{r_1 r_2}\right)^2} \frac{I_0}{1 + \frac{4\sqrt{r_1 r_2}}{(1 - \sqrt{r_1 r_2})^2} \sin^2\left(\frac{\pi \Delta v_q}{v_{FSR}}\right)} \quad (2.2)$$

Où (t_1, t_2) et (r_1, r_2) sont les transmissions et réflexions des miroirs 1 et 2 de l'interféromètre et ν_{FSR} représente la distance en fréquence entre les pics d'intensité résonnants. Voici une représentation graphique des spectres de transmission en considérant deux miroirs avec les mêmes facteurs de transmission, de réflexion et aucune perte (voir Figure 2.3).

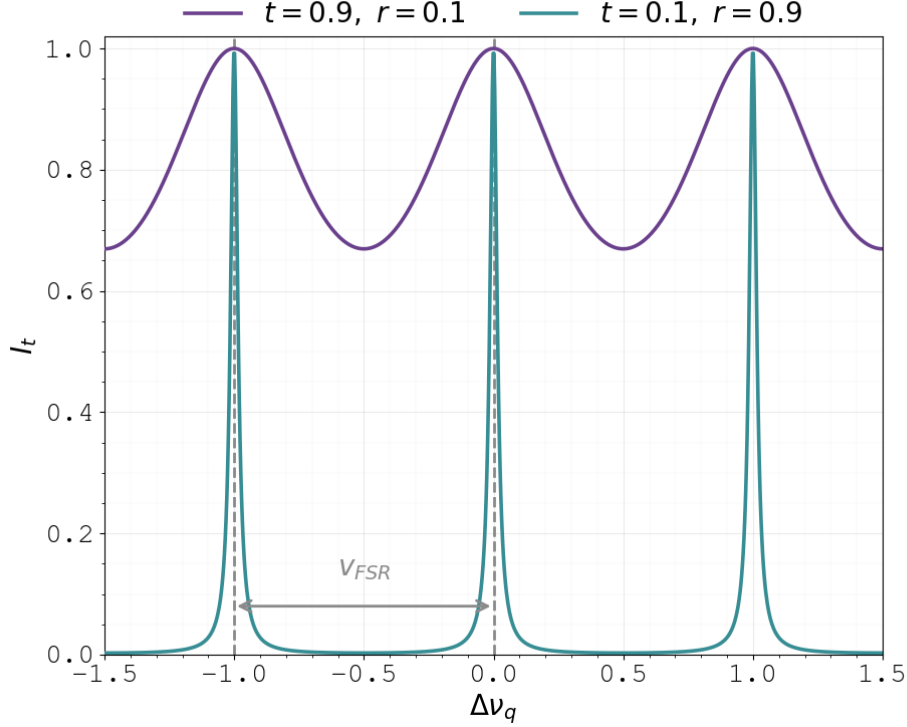


FIGURE 2.3 Spectre de transmission des cavités Fabry-Pérot au mode fondamental TEM_{00} avec coefficients de transmission et de réflexion $(t, r) = \{(0.9, 0.1), (0.1, 0.9)\}$ et $I_0 = 1$.

De ce graphique, on observe que la finesse ainsi que le facteur de qualité (Q), qui mesure la finesse de la résonance comme le rapport entre l'énergie stockée et dissipée, sont grandement affectés par ces coefficients. La finesse du capteur est directement reliée au coefficient de réflexion dans le cas où les deux miroirs semi-réfléchissants sont les mêmes.

$$F = \frac{\pi\sqrt{r}}{(1-r)} \quad (2.3)$$

Un miroir à haute réflectivité produira des pics de transmission beaucoup plus étroits. La réflectivité joue un rôle important dans la capacité de l'interféromètre à distinguer les caractéristiques du spectre de transmission.

Les cavités Fabry-Pérot utilisent donc ces principes à leur avantage. En effet, le signal issu de

cet interféromètre est notamment lié aux propriétés du milieu ainsi qu'à la distance séparant les deux miroirs. Ainsi, l'insertion de matériaux sensibles à différents gaz entre ceux-ci permet de moduler le signal et donc de créer un capteur. Dans ce contexte, des polymères sont utilisés comme matériaux sensibles [4].

2.2.2 Les polymères

Les polymères sont une classe de matériaux caractérisés par des chaînes de macromolécules formées à partir d'unités de faible masse molaire que l'on appelle monomères [23, 24]. La capacité de ces derniers à former des chaînes de monomères correspond à leur degré de polymérisation [25].

La configuration et les types de monomères qui constituent ces derniers permettent à chacun d'avoir des propriétés chimiques, mécaniques et optiques intéressantes avec diverses applications. Les plus connus sont les fibres naturelles, telles que les protéines et les acides nucléiques (ADN et ARN). On y retrouve aussi les plastiques, les caoutchoucs, les colles et bien d'autres [26].

Grâce à leur diversité chimique, ceux-ci sont utilisés dans des domaines variés tels que les biomatériaux, les dispositifs électroniques et le textile. Récemment, leur utilisation s'est étendue aux matériaux dits «intelligents» capables de réagir à une contrainte externe telle que la température, l'humidité, le pH ou bien la présence de certaines molécules [27].

Parmi ces stimuli, les gaz constituent une cible d'intérêt croissante. Certains d'entre eux sont capables d'absorber certains gaz et de subir une expansion volumique, des changements d'indice de réfraction et de conductivité électrique [28–30]. Ces mécanismes sont donc exploités pour la détection optique ou électrique. Plusieurs polymères sont couramment utilisés à cet effet tel que le PDMS, le PEVA, le PHMB, le PEG2000 et bien d'autres [31–34].

Les interactions entre polymères et analytes de gaz dépendent fortement de la structure chimique du polymère ainsi que des propriétés physicochimiques du gaz, telle que sa polarité, sa structure moléculaire ou sa solubilité dans le matériau. Lorsqu'un gaz est absorbé ou diffusé dans un polymère, il modifie les propriétés du matériau. La mesure de ces variations constitue alors le mécanisme exploitable dans les capteurs de gaz à base de polymères.

2.2.3 Physique des polymères appliquée aux cavités Fabry-Pérot pour la détection de gaz

Les cavités Fabry-Pérot intégrées avec des polymères sont une solution prometteuse pour la détection de gaz, en raison de leur sensibilité, de leur réversibilité et de leur faible coût [17]. Ces capteurs comprennent des réservoirs latéraux qui, par capillarité, permettent au polymère déposé de s'infiltrer entre les miroirs semi-réfléchissants de la cavité. Une fibre optique est collée de chaque côté de la cavité afin d'y collecter le signal optique généré par une émission laser à une longueur d'onde de 1450-1650 nm (voir Figure 2.4).

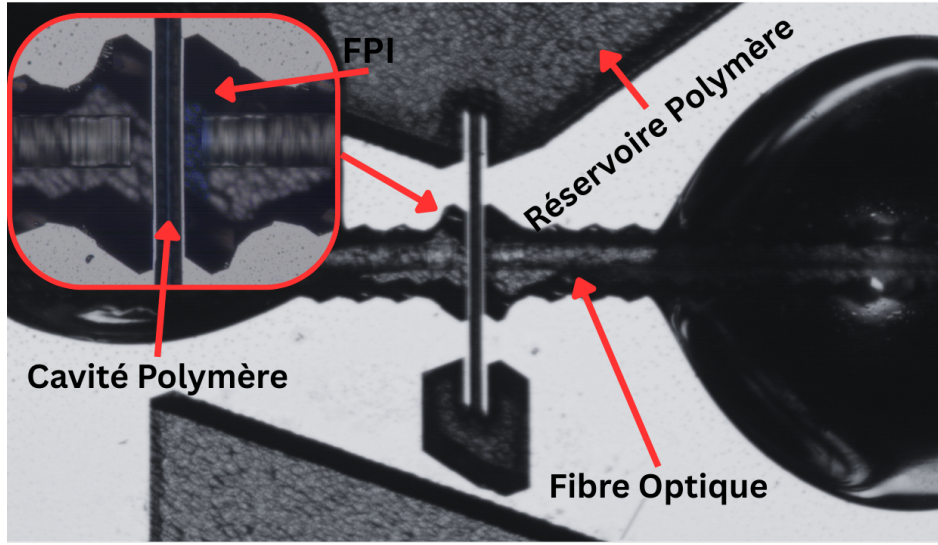


FIGURE 2.4 Interféromètre Fabry-Pérot sur puce avec parties principales identifiées.

Le mécanisme de détection de gaz utilisé dans ces microcavités repose sur le suivi des variations de longueur d'onde de résonance de la cavité optique. Celles-ci sont directement corrélées à la quantité d'analytes absorbée par le polymère. Ceci produit deux effets principaux, soit la modification de l'indice de réfraction effectif du matériau et un gonflement volumique du polymère sous des contraintes mécaniques internes (les miroirs). L'équation générale décrivant ce rapport est la suivante :

$$\frac{\Delta\lambda_{\text{res}}}{\lambda_{\text{res}}} = \alpha \frac{\Delta L}{L} + \beta \frac{\Delta n_{\text{mix}} + \Delta n_{\sigma}}{n} \quad (2.4)$$

La Figure 2.5 illustre un exemple de décalage spectral observé pour une cavité Fabry-Pérot fonctionnalisée avec un polymère lors d'une exposition à un analyte.

L'équation 2.4 relie le décalage relatif de la longueur d'onde de résonance $\Delta\lambda_{\text{res}}/\lambda_{\text{res}}$ aux

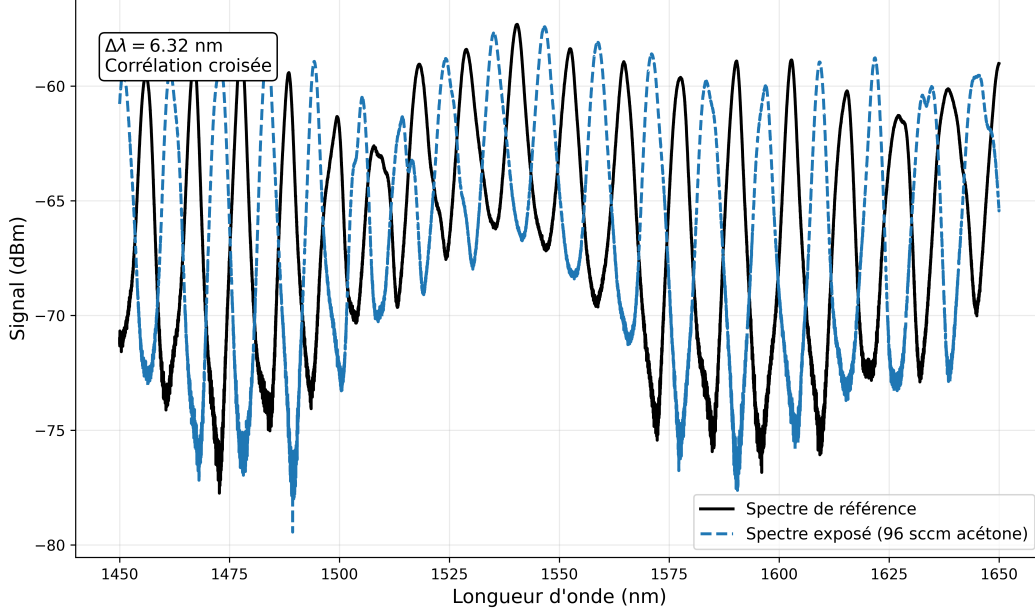


FIGURE 2.5 Décalage spectral $\Delta\lambda_{res}$ du capteur recouvert de polymère PU lors de l'exposition à 96 sccm d'acétone. Le décalage de la réponse spectrale ($\Delta\lambda = 6.32$ nm) est estimé par corrélation croisée du spectre complet.

effets de l'absorption d'un gaz par le polymère. Le premier terme, $\alpha \Delta L/L$, représente la contribution mécanique due au gonflement du polymère, en fonction du facteur α associé à la partie mobile de la cavité. Le second terme, $\beta(\Delta n_{mix} + \Delta n_{\sigma})/n$, décrit la variation de l'indice de réfraction, provenant du mélange polymère/analyte (Δn_{mix}) et des contraintes internes (Δn_{σ}), avec β tenant compte du confinement partiel du mode optique. Des travaux antérieurs ont montré que la contribution mécanique domine le signal mesuré et la variation d'indice est négligeable [35]. Les paramètres α et β ont été calculés à l'aide de l'algorithme de simulation optique développé par St-Gelais et al. pour les interféromètres de Fabry-Pérot multicouches gravés en profondeur [36].

Le gonflement du polymère et la variation de son indice de réfraction sont directement liés à la fraction volumique d'analyte absorbée ϕ_a dans le polymère, comme illustré à la Figure 2.6. Cette fraction peut être reliée à la concentration volumique de l'analyte dans l'air C_{air} à l'aide de la loi des gaz parfaits et du coefficient de partage polymère/analyte K_{p-a} :

$$\phi_a = K_{p-a} \frac{PV_m}{RT} C_{air}, \quad (2.5)$$

où P est la pression, T la température, R la constante des gaz parfaits et V_m le volume molaire de l'analyte à l'état liquide. Sous l'hypothèse d'additivité des volumes, le gonflement

relatif du polymère est proportionnel à la fraction volumique absorbée :

$$\frac{\Delta L}{L} \approx \frac{\Delta V}{V} = \phi_a. \quad (2.6)$$

Quant à la variation d'indice de réfraction du polymère, elle peut être estimée avec la relation de Gladstone-Dale [37] tel que :

$$\frac{\Delta n_{\text{mix}}}{n} \approx \phi_a \frac{n_a - n_p}{n_p}, \quad (2.7)$$

où n_a et n_p sont respectivement les indices de réfraction de l'analyte et du polymère. La contribution Δn_σ due aux contraintes mécaniques étant généralement négligeable pour les polymères souples, elle peut souvent être ignorée.

Ainsi, la concentration de l'analyte dans l'air influence directement le décalage spectral via ϕ_a , qui module à la fois le gonflement du polymère et la variation d'indice de réfraction, selon :

$$\frac{\Delta \lambda_{\text{res}}}{\lambda_{\text{res}}} = \alpha \phi_a + \beta \frac{\phi_a (n_a - n_p)}{n_p}. \quad (2.8)$$

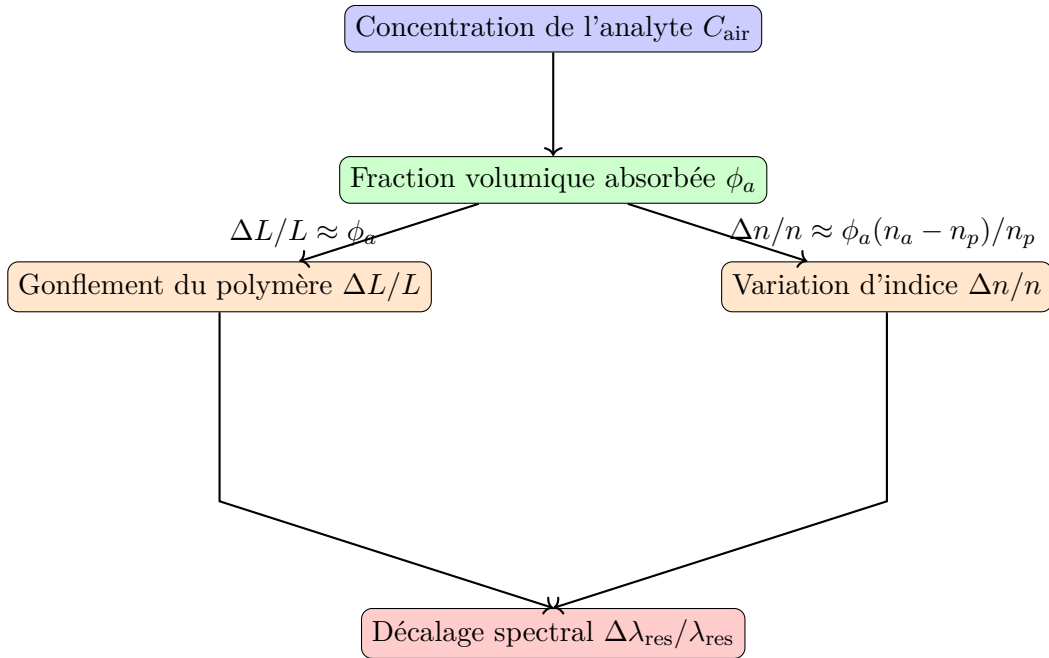


FIGURE 2.6 Schéma résumant les principaux phénomènes à la source de la détection d'analyte dans les cavités Fabry-Pérot à l'aide de polymères.

En substituant l'expression de l'équation 2.5 dans l'équation 2.8, la relation peut être réécrite en regroupant les constantes, ce qui permet de montrer plus clairement que la variation relative de la longueur d'onde de résonance dépend linéairement de la concentration de l'analyte dans l'air, comme présenté à l'équation 2.9. Les termes A et B , définis respectivement aux équations 2.10 et 2.11, représentent les contributions associées au gonflement du polymère et à la variation de l'indice de réfraction.

$$\frac{\Delta\lambda_{\text{res}}}{\lambda_{\text{res}}} = (A + B) C_{\text{air}} \quad (2.9)$$

$$A = \alpha K_{p-a} \frac{PV_m}{RT} \quad (2.10)$$

$$B = \beta \frac{(n_a - n_p)}{n_p} K_{p-a} \frac{PV_m}{RT} \quad (2.11)$$

Physique des dynamiques transitoires des polymères

Lorsqu'un polymère est exposé à un composé organique volatil (COV), l'évolution temporelle de la réponse du capteur est généralement caractérisée par une phase transitoire suivie d'un régime quasi statique. Cette dynamique est principalement générée par les phénomènes de sorption et de diffusion du gaz à l'intérieur du matériau polymérique [38].

Dans un premier temps, les molécules de gaz sont adsorbées à la surface du matériau puis diffusent progressivement dans la matrice polymérique. Ce phénomène est décrit par les lois de diffusion de Fick, selon lesquelles la concentration du gaz dans le polymère évolue en fonction du temps et de la distance à la surface. Selon la deuxième loi de Fick, pour une diffusion à une dimension [39], définie à l'équation 2.12 :

$$\frac{\delta\phi}{\delta t} = D \frac{\delta^2\phi}{\delta y^2} \quad (2.12)$$

où ϕ représente la concentration du gaz dans le polymère, D est le coefficient de diffusion du gaz dans le polymère, y correspond à la position dans le polymère et t représente le temps. Cette équation décrit l'évolution de la distribution spatiale de la concentration dans le polymère sous l'effet de la diffusion [40].

Dans les cavités Fabry-Pérot étudiées dans ce travail, les polymères sensibles sont déposés dans des réservoirs puis insérés dans la cavité par capillarité. La diffusion du gaz dans le polymère peut alors être approximée par un modèle unidimensionnel dans une couche de

largeur finie d reposant sur un substrat imperméable. La solution analytique de la seconde loi de Fick peut alors être obtenue par séparation des variables. La concentration moyenne dans le polymère peut alors être exprimée sous la forme d’une série infinie [40], comme présenté à l’équation 2.13 :

$$\frac{\phi(t)}{\phi_\infty} = 1 - \frac{8}{\pi^2} \sum_{n=0}^{\infty} \frac{1}{(2n+1)^2} \exp\left(-\frac{(2n+1)^2 \pi^2 D t}{4d^2}\right) \quad (2.13)$$

où ϕ_∞ représente la concentration à l’équilibre dans le polymère. Cette expression montre que la cinétique d’absorption est constituée d’une somme de termes exponentiels.

Cependant, dans de nombreux polymères utilisés pour la détection de gaz, les cinétiques expérimentales ne sont pas toujours parfaitement décrites par un seul coefficient de diffusion.

Plusieurs travaux ont montré que les courbes d’absorption présentent plusieurs constantes de temps. Dans ces cas, des modèles comportant plusieurs coefficients de diffusion sont utilisés pour décrire de manière empirique les dynamiques transitoires observées. C’est ce qui a été observé par J. Saunders, H. Chen, C. Brauer, M. Clayton et H.-P. Loock dans leur étude intitulée “Two distinct mechanisms upon absorption of volatile organic compounds into siloxane polymers” [41]. Ceux-ci ont démontré que l’exposition de COV au polymère PDMS pouvait être décrite par la présence de deux termes de diffusion (*two-rate diffusion*). Ce modèle s’applique à la fois au phénomène d’absorption ($X(t)_{abs}$) et au phénomène de désorption ($X(t)_{des}$). Ces phénomènes sont décrits par les équations suivantes [41] :

$$\begin{aligned} X(t)_{abs} = & X_1 \left[1 - \frac{8}{\pi^2} \sum_{n=0}^{\infty} \frac{1}{(2n+1)^2} \exp\left(-\frac{(2n+1)^2 D_1 \pi^2 (t-t_0)}{4d^2}\right) \right] \\ & + X_2 \left[1 - \frac{8}{\pi^2} \sum_{n=0}^{\infty} \frac{1}{(2n+1)^2} \exp\left(-\frac{(2n+1)^2 D_2 \pi^2 (t-t_0)}{4d^2}\right) \right] \end{aligned} \quad (2.14)$$

et

$$\begin{aligned} X(t)_{des} = & X_1 \left[\frac{8}{\pi^2} \sum_{n=0}^{\infty} \frac{1}{(2n+1)^2} \exp\left(-\frac{(2n+1)^2 D_1 \pi^2 (t-t_0)}{4d^2}\right) \right] \\ & + X_2 \left[\frac{8}{\pi^2} \sum_{n=0}^{\infty} \frac{1}{(2n+1)^2} \exp\left(-\frac{(2n+1)^2 D_2 \pi^2 (t-t_0)}{4d^2}\right) \right] \end{aligned} \quad (2.15)$$

où $X(t)$ représente la fraction molaire de l’analyte absorbé dans le polymère en fonction du

temps. Les paramètres X_1 et X_2 correspondent aux contributions respectives des deux mécanismes de diffusion. Les coefficients D_1 et D_2 représentent les coefficients de diffusion associés à chacun de ces mécanismes. Ces deux constantes de diffusion sont généralement associées à des mécanismes physiques différents, tels qu'une absorption rapide dans le polymère suivie d'une diffusion plus lente.

À partir de ces relations, on observe que la variation temporelle de la réponse du polymère en régime transitoire ne suit pas une simple dépendance linéaire, mais plutôt une somme de plusieurs termes exponentiels associés aux différents mécanismes de diffusion.

Choix et propriétés des polymères sensibles aux gaz

Afin de choisir un polymère adapté à la détection de gaz, il y a plusieurs facteurs à prendre en considération. L'un des plus importants est la sélectivité du polymère, c'est-à-dire sa capacité à réagir préférentiellement à un gaz donné. Une sélectivité élevée permet donc de réduire considérablement le nombre de polymères requis pour faire la détection d'un gaz en particulier.

En effet, puisque les polymères présentent des sensibilités croisées ainsi qu'une sensibilité à d'autres facteurs environnementaux, il est nécessaire d'avoir plusieurs polymères avec des sensibilités différenciées de sorte à mieux discrétiser les gaz présents et à tirer avantage de leur sensibilité variable. Par exemple, pour les cavités Fabry-Pérot, il a été démontré que le Polydimethylsiloxane (PDMS), le Polyéthylène acétate de vinyle (PEVA) ainsi que le Polypropylène (PP) auraient des sensibilités et des limites de détection différentes, ce qui permettrait de bien détecter le méthane (CH_4) ainsi que le dioxyde de carbone (CO_2) [42]. Le Tableau 2.1 présente les sensibilités et les limites de détection rapportées par cette étude pour différents polymères utilisés avec des cavités Fabry-Pérot. La sensibilité est exprimée en décalage spectral (pm) par variation de concentration de gaz (ppm), tandis que la LDD correspond à la limite de détection du capteur exprimée en ppm.

TABLEAU 2.1 Sensibilités des polymères et limites de détection pour CH_4 et CO_2 , d'après Guertin et al., 2024. [42]

Polymère	Sensibilité (pm ppm ⁻¹)		LDD (ppm)	
	CH ₄	CO ₂	CH ₄	CO ₂
PDMS	$-2,04 \times 10^{-3}$	$6,44 \times 10^{-4}$	$1,54 \times 10^3$	$4,98 \times 10^3$
PEVA	$-3,36 \times 10^{-2}$	Faible	$5,65 \times 10^2$	-
PP	Faible	$4,07 \times 10^{-4}$	-	$1,02 \times 10^4$

Cette caractéristique est importante pour le traitement des données des capteurs. Pour un modèle linéaire, il faut absolument au minimum N capteurs présentant chacun une sensibilité dite « orthogonale » pour N gaz, afin de pouvoir résoudre de manière unique les concentrations de chaque analyte. En revanche, pour les modèles non linéaires ou utilisant des algorithmes d'apprentissage automatique, cette contrainte est moins stricte, mais reste importante pour maximiser la précision des prédictions [43].

En plus de la sensibilité, il est important de choisir des polymères hydrophobes [44]. Ceci est crucial pour minimiser l'adsorption de l'humidité ambiante et éviter une expansion ou un changement d'indice non corrélé à la présence d'un analyte. De la même façon, il est important d'avoir un polymère peu sensible aux changements de température afin d'éviter les interférences dues à ce paramètre environnemental non contrôlé.

2.3 Apprentissage Automatique Supervisé

L'apprentissage automatique est une branche de l'intelligence artificielle associée à la conception d'algorithmes capable d'extraire automatiquement des connaissances ainsi que des modèles à partir de données [45]. Contrairement aux méthodes traditionnelles de traitement de données, l'apprentissage automatique permet à l'algorithme de s'adapter et de s'améliorer sans aucune intervention humaine directe. Ce type d'apprentissage facilite grandement la résolution de problèmes complexes et variés, tels que la classification, la régression, la reconnaissance de forme, d'image [46]. Dans le contexte des capteurs de gaz, plusieurs études ont démontré l'efficacité de ces approches pour améliorer la précision [47] et le temps d'analyse [48].

Les sections suivantes présenteront les principes d'apprentissage supervisé utilisés pour ce type d'application.

2.3.1 Algorithmes classiques de régression et classification

Régression linéaire

La régression linéaire est une méthode d'apprentissage supervisé simple cherchant à modéliser une relation linéaire entre une variable dépendante et une ou plusieurs variables indépendantes. Malgré sa simplicité, celle-ci constitue un modèle de référence important puisqu'elle permet d'évaluer si les tendances principales présentes dans les données peuvent être décrites à l'aide de relations approximativement linéaires. Malgré cela, les réponses des capteurs polymériques présentent généralement des comportements non linéaires complexes, limitant les

performances de cette approche [49].

Cette méthode cherche à ajuster une droite permettant de représenter la tendance des données tout en minimisant l'erreur entre les prédictions et les valeurs réelles, comme illustré à l'équation 2.16.

$$\hat{y} = \mathbf{w}^\top \mathbf{x} + b \quad (2.16)$$

où

- $\mathbf{x} \in \mathbb{R}^n$: vecteur des variables d'entrée,
- $\mathbf{w} \in \mathbb{R}^n$: vecteur des poids,
- $b \in \mathbb{R}$: biais,
- $\hat{y}$: valeur prédite.

Dès lors, nous cherchons à déterminer les vecteurs de poids et de biais permettant de minimiser la fonction de coût (MSE) définie à l'équation 2.17 :

$$J(\mathbf{w}, b) = \frac{1}{m} \sum_{i=1}^m \left(\hat{y}^{(i)} - y^{(i)} \right)^2 = \frac{1}{m} \sum_{i=1}^m \left(\mathbf{w}^\top \mathbf{x}^{(i)} + b - y^{(i)} \right)^2 \quad (2.17)$$

Régression logistique

La régression logistique est un algorithme d'apprentissage supervisé permettant de résoudre des problèmes simples de classification binaire. Elle est souvent utilisée comme modèle de référence dans les problèmes de classification. Dans le contexte des capteurs de gaz, ce modèle permet notamment d'évaluer si les différentes signatures générées par les analytes sont facilement séparables. Puisque les réponses des capteurs présentent souvent des dynamiques fortement non linéaires et complexes, les performances de ce modèle demeurent généralement limitées comparativement aux approches plus avancées. Cette méthode permet de déterminer une fonction séparant au mieux les données en deux classes. À partir de cette fonction, il est alors possible d'estimer la probabilité qu'une donnée appartienne à une classe donnée, comme présenté à l'équation 2.18.

$$\hat{y} = \sigma(\mathbf{w}^\top \mathbf{x} + b) \quad \text{avec} \quad \sigma(z) = \frac{1}{1 + e^{-z}} \quad (2.18)$$

où :

- $\hat{y} \in [0, 1]$: probabilité que $y = 1$,
- $\sigma(z)$: fonction sigmoïde.

La fonction $\sigma(z)$ permet de ramener les valeurs de sortie dans l'intervalle $[0, 1]$. Afin d'entraîner le modèle, il suffit alors de minimiser la fonction de coût d'entropie croisée [50], définie à l'équation 2.19.

$$J(\mathbf{w}, b) = -\frac{1}{m} \sum_{i=1}^m \left[y^{(i)} \log(\hat{y}^{(i)}) + (1 - y^{(i)}) \log(1 - \hat{y}^{(i)}) \right] \quad (2.19)$$

Arbres de décision

L'arbre de décision est un modèle d'apprentissage supervisé qui permet la classification ou la régression. Celui-ci fonctionne en divisant les données en sous-ensemble à l'aide de règles logiques sur les données entrantes. Il repose sur les structures arborescentes composées de nœuds de décision permettant de partitionner l'espace des variables d'entrée en sous régions [51]. Chacun de ces nœuds correspond à des règles de décision qui sont appliquées par rapport aux données d'entrée. Cette approche présente l'avantage d'être relativement intuitive et interprétable, tout en étant capable de modéliser certaines relations non linéaires présentes dans les données dans les capteurs de gaz.

Pour ce type de problèmes, un processus de partitionnement récursif binaire des données est effectué de manière à maximiser la réduction de l'erreur des sous-ensembles obtenus. Pour les problèmes de régression, le critère de séparation consiste à minimiser la somme des erreurs quadratiques dans les deux régions résultantes, comme défini à l'équation 2.20 :

$$\min_{j,s} \left[\sum_{x_i \in R_1(j,s)} (y_i - \hat{y}_{R_1})^2 + \sum_{x_i \in R_2(j,s)} (y_i - \hat{y}_{R_2})^2 \right] \quad (2.20)$$

où y_i représente la valeur observée pour l'observation i , tandis que R_1 et R_2 correspondent aux deux régions de l'espace obtenues après la séparation. Les valeurs $\hat{y}_{R_1}$ et $\hat{y}_{R_2}$ représentent les prédictions associées à chacune des régions. Les prédictions dans chaque région sont généralement données par la moyenne des observations qu'elles contiennent.

Pour les problèmes de classification, l'objectif dépend du critère de séparation utilisé, notamment des mesures d'impureté telles que l'indice de Gini, défini à l'équation 2.21 :

$$\sum_{k \neq k'} \hat{p}_{mk} \hat{p}_{mk'} = \sum_{k=1}^K \hat{p}_{mk} (1 - \hat{p}_{mk}) \quad (2.21)$$

où p_k représente la proportion d'observations appartenant à la classe k dans un certain nœud.

Malheureusement, ces modèles sont sensibles au surapprentissage lorsque les profondeurs des arbres deviennent importantes [52]. Afin d'éviter ce phénomène et d'améliorer les perfor-

mances prédictives de ce modèle, plusieurs méthodes ont été proposées. Parmi celles-ci, on y retrouve l'algorithme *XGBoost*, qui constitue une implémentation de *gradient boosting*. Cette implémentation consiste à entraîner séquentiellement plusieurs arbres de décision, dans lesquels chaque nouvel arbre est optimisé pour corriger les erreurs du modèle précédent. Ce modèle repose sur l'optimisation d'une fonction de coût définie à l'équation 2.22 :

$$\mathcal{L} = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (2.22)$$

ou $\mathcal{L}$ représente la fonction de perte face aux prédictions et $\Omega(f_k)$ est un terme de régularisation appliqué aux arbres pour limiter la complexité du modèle [53].

Ceux-ci ont été évalués à plusieurs reprises pour la classification et de régression de capteurs de gaz. Dans le cadre d'une étude, les auteurs Wangze *et al.* [54] ont proposé des modèles à l'aide d'un système de nez électronique basé sur un réseau de capteurs de gaz, capable de prédire à la fois le type ainsi que les concentrations présentes de manière efficace à l'aide d'arbres de décision. Le système est basé principalement d'un classificateur XGBoost et d'un algorithme de quantification Random Forest pour reconnaître et quantifier six composés organiques volatils (VOCs) différents. Dans le cadre de cette étude, ils ont utilisé une recherche d'hyperparamètre par validation croisée à 5 plis afin d'optimiser chaque modèle. Ils ont obtenu des performances dépassant les méthodes classiques comme les réseaux de neurones à rétropropagation ou les arbres de décision. Le classificateur a obtenu un score de 96.0% d'exactitude en classification et un score moyen R^2 de 0.923 en régression des concentrations.

k plus proches voisins (k-NN)

L'algorithme des k plus proches voisins (*k-Nearest Neighbors*, k-NN) est une méthode d'apprentissage supervisé utilisée pour des problèmes de classification et de régression. Cette algorithme ne requière pas d'entraînement. Les prédictions sont plutôt effectuées directement à partir des données d'apprentissage en se basant sur la similarité entre les observations [55]. Cette approche est particulièrement intéressante lorsque les frontières de séparation entre les classes sont complexes et non linéaires.

Le principe consiste à identifier les k points de l'ensemble d'entraînement qui sont les plus proches d'une observation selon une mesure de distance. La distance euclidienne, définie à l'équation 2.23, est souvent utilisée :

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\sum_{m=1}^p (x_{im} - x_{jm})^2} \quad (2.23)$$

où p correspond au nombre de variables. Dans le cas de la classification, les classes prédites correspondent généralement aux classes qui revient le plus souvent parmi les k voisins les plus proches. Pour un problème de régression, la valeur prédite est souvent obtenue en calculant la moyenne des valeurs associées à ces voisins.

2.3.2 Algorithmes avancés de classification et régression supervisées

Machines à vecteurs de support (SVM)

Les machines à vecteurs de support (SVM) sont des algorithmes d'apprentissage supervisé utilisés pour des tâches de classification et de régression. Ceux-ci cherchent à construire une frontière de séparation optimale entre les différentes classes en maximisant la marge entre celles-ci. Cette approche est généralement utilisée pour des jeux de données de taille modérée présentant des frontières de séparation complexes, ce qui correspond aux caractéristiques des capteurs étudiés dans ce travail.

Dans le cas de la régression, on utilise alors *Support Vector Regression* (SVR) qui est dérivé de SVM. L'objectif de cet algorithme est de déterminer une fonction $f(x)$ permettant d'approximer les données, en conservant une complexité minimale du modèle [56].

La fonction de régression est généralement définie à l'équation 2.24 :

$$f(x) = \langle w, x \rangle + b \quad (2.24)$$

où w est le vecteur de poids, b le biais et $\langle w, x \rangle$ le produit scalaire entre les vecteurs. L'objectif de l'apprentissage est de minimiser la norme de w tout en permettant une erreur de prédiction inférieure à une marge ε pour la majorité des observations.

Lorsque certaines observations dépassent cette marge, des variables de relâchement ξ_i et ξ_i^* sont introduites afin de permettre une certaine tolérance aux erreurs. Le problème d'optimisation est alors défini à l'équation 2.25 :

$$\min_{w, b, \xi_i, \xi_i^*} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (2.25)$$

où C est un paramètre de régularisation entre la complexité du modèle et l'erreur de prédiction.

Quant à la classification, on utilise alors *Support Vector Classification* (SVC). Le principe de cette méthode consiste à déterminer un hyperplan qui sépare les différentes classes en maximisant l'espacement entre les frontières de ceux-ci.

Afin de modéliser des relations non linéaires, des fonctions noyaux (*kernels*) sont utilisées pour transformer les données dans un espace de dimension plus élevé. Ces transformations permettent de conserver une formulation linéaire du modèle dans l'espace transformé [57]. Cet algorithme est souvent utilisé pour le problème de classification et quantification des capteurs de gaz.

Zhang *et al.* [58] proposent une méthode de détection de gaz mixtes basée sur la machine à vecteur de support (SVM) optimisé par l'algorithme de *Sparrow Search* (SSA). Dans cette étude, la dimensionnalité des données est réduite par l'entremise d'une analyse en composantes principales (PCA). Ensuite, les hyperparamètres du SVM sont ajustés par SSA. Il a été démontré que cette approche surpasse les approches classiques telles que la forêt aléatoire (RF) ainsi que les réseaux de neurones à rétropropagation (BP) avec leurs données. La précision de classification atteint des valeurs élevées, soit un taux d'ajustement maximal de 99.34% pour reconnaître un gaz unique et 99.55% pour des mélanges de gaz.

Réseaux de neurones convolutionnels (CNN)

Les réseaux de neurones convolutionnels (CNN) sont des algorithmes d'apprentissage supervisé profonds composés de neurones artificiels organisés en couches séquentielles. Par l'entremise de filtres de convolution appris durant l'entraînement, ces réseaux sont capables d'extraire automatiquement des caractéristiques pertinentes à partir des données d'entrée. Les CNN sont particulièrement efficaces pour l'analyse de données avec de grandes dimensions, telles que les images ou certaines représentations matricielles de signaux [59].

Les réseaux de neurones convolutionnels trouvent leur origine dans les travaux de LeCun *et al.* avec l'architecture *LeNet-5*, développée pour la reconnaissance de chiffres manuscrits [60]. Leur popularité a fortement augmenté avec l'introduction de l'architecture *AlexNet*, qui a démontré l'efficacité des réseaux convolutionnels profonds lors de la compétition ImageNet en vision par ordinateur [61]. Depuis, les CNN sont devenues une architecture souvent utilisée dans de nombreuses applications d'apprentissage automatique.

Les réseaux CNN sont généralement composés de plusieurs types de couches hiérarchiques, notamment des couches de convolution, des fonctions d'activation, des opérations de pooling, des couches entièrement connectées ainsi que des couches de sortie. Les couches de convolution permettent d'extraire des caractéristiques locales des données d'entrée en appliquant un ensemble de filtres appris durant l'entraînement. Le résultat de cette opération est une carte de caractéristiques (*feature map*). Pour un tenseur d'entrée tridimensionnel, la convolution peut être exprimée comme suit [62] :

$$G_{x,y,z,n} = \sum_i \sum_j \sum_k \sum_m K_{i,j,k,m,n} \cdot F_{x+i-1, y+j-1, z+k-1, m} \quad (2.26)$$

où

- $G_{x,y,z,n}$ représente la valeur de sortie de la carte de caractéristiques à la position spatiale (x, y, z) pour le filtre n ;
- $F_{x+i-1, y+j-1, z+k-1, m}$ correspond à la valeur d'entrée à la position $(x+i-1, y+j-1, z+k-1)$ dans le canal m ;
- $K_{i,j,k,m,n}$ représente le poids du noyau de convolution reliant le canal d'entrée m au canal de sortie n ;
- les indices i, j, k parcourent les dimensions spatiales du noyau et m les canaux d'entrée.

En plus de la convolution, des opérations de *pooling* sont généralement appliquées afin de réduire la dimension spatiale des cartes de caractéristiques tout en conservant l'information la plus pertinente. Ces opérations simplifient les représentations internes et de réduire la charge computationnelle lors de l'entraînement. Une des méthodes les plus courantes est le *max pooling*, qui sélectionne la valeur maximale dans chaque région.

Après l'extraction des caractéristiques par les couches convolutionnelles et les opérations de pooling, les cartes de caractéristiques sont généralement traitées par une ou plusieurs couches entièrement connectées. Ces couches permettent de combiner les caractéristiques extraites afin de produire une représentation globale des données.

Soit $\mathbf{x} \in \mathbb{R}^n$ le vecteur d'entrée de la couche précédente, $\mathbf{W} \in \mathbb{R}^{m \times n}$ la matrice de poids et $\mathbf{b} \in \mathbb{R}^m$ le vecteur de biais. La sortie de la couche peut être exprimée à l'équation 2.27 :

$$\mathbf{y} = f(\mathbf{W}\mathbf{x} + \mathbf{b}) \quad (2.27)$$

où f représente la fonction d'activation. Une fonction d'activation couramment utilisée est la fonction ReLU (*Rectified Linear Unit*), définie à l'équation 2.28 :

$$f(z) = \max(0, z) \quad (2.28)$$

Finalement, la dernière couche du réseau correspond à la couche de sortie et dépend directement de la tâche considérée. Dans le cas d'un problème de classification multi-classe, une fonction d'activation *softmax* est généralement utilisée afin de produire une distribution de probabilité sur les différentes classes. Pour les problèmes de régression, la couche de sortie

est généralement linéaire.

Grâce à leur capacité à extraire automatiquement des caractéristiques complexes à partir de données structurées, les réseaux CNN ont été appliqués dans plusieurs travaux récents pour l'analyse de signaux de capteurs chimiques et la détection de gaz. Dans le cadre de la classification et de la régression, de récents travaux ont porté sur les capteurs à base d'oxydes métalliques semi-conducteurs (SMO), qui sont prometteurs pour la détection environnementale. Malgré cela, leur faible sélectivité limite leur utilisation. Kang *et al.* [63] ont proposé une approche pour pallier à cette difficulté : la fabrication de capteurs très uniformes par l'entremise de dépôt à angle rasant (GLAD) et l'application d'un réseau de neurones convolutif afin de traiter les caractéristiques des signaux acquises et réaliser une classification en temps réel. Les résultats expérimentaux montrent que cette approche atteint une précision de classification de 98% et une erreur de régression de moins de 10% pour plusieurs gaz d'intérêt (CO, NH₃, NO₂, CH₄ et acétone) avec un temps de réponse de moins de quelques secondes.

L'utilisation d'un CNN permettait à ce capteur d'exploiter les parties transitoires, ce qui rend possible la prédiction en temps réel. De plus, l'utilisation de plusieurs matériaux sensibles, fabriqués par GLAD, améliore la sélectivité et facilite ainsi la classification.

Une autre étude propose un réseau convolutif profond nommé GasNet, spécifiquement conçu pour la classification de gaz à partir de données brutes. Ce modèle exploite la capacité des DCNN à apprendre simultanément les caractéristiques et frontières de décision. L'architecture de GasNet comprend 6 blocs convolutifs composés chacun de couches de convolution, de normalisation, de fonction d'activation ReLU suivies par une couche de *pooling* et d'une couche entièrement connectée pour un total de 38 couches.

Les résultats démontrent que GasNet surpasse les approches classiques comme le SVM, le KNN ou le MLP en raison de sa capacité à mieux extraire les caractéristiques pertinentes et à réaliser l'apprentissage directement à partir des données brutes, ce qui réduit la perte d'information et améliore la précision globale de la classification. Ce travail met en évidence le fort potentiel des réseaux profonds pour l'analyse et le traitement des signaux complexe des capteurs de gaz [64].

Réseaux de neurones récurrents (RNN et LSTM)

Les réseaux de neurones récurrents (RNN) sont des algorithmes conçus pour traiter des données séquentielles ou temporelles. Grâce à leurs connexions récurrentes, ils possèdent une mémoire interne permettant de prendre en compte les états passés lors du traitement d'une nouvelle observation. Ceci les rend particulièrement adaptés à l'analyse de séries temporelles [59]. Puisque les signaux générés par les capteurs étudiés dans ce travail présentent des dynamiques temporelles complexes et non linéaires, ces modèles semblent théoriquement particulièrement bien adaptés à cette problématique.

Dans le cas d'un RNN classique, l'état caché $\mathbf{h}_t$ est mis à jour à chaque instant t en fonction de l'entrée courante $\mathbf{x}_t$ et de l'état précédent $\mathbf{h}_{t-1}$, comme défini à l'équation 2.29 :

$$\mathbf{h}_t = f(W_{xh}\mathbf{x}_t + W_{hh}\mathbf{h}_{t-1} + \mathbf{b}_h) \quad (2.29)$$

Cependant, les RNN simples souffrent de problèmes de gradients évanescents ou explosifs lors de l'entraînement, ce qui limite leur capacité à apprendre des dépendances sur de longues séquences temporelles.

Pour remédier à ce problème, les réseaux *Long Short-Term Memory* (LSTM) ont été proposés par Hochreiter et Schmidhuber [65]. Ces architectures introduisent des cellules de mémoire et des portes de contrôle permettant de réguler le flux d'information et de conserver des dépendances temporelles sur de plus longues périodes.

Dans le domaine des capteurs de gaz et des nez électroniques, les RNN et les LSTM sont utilisés pour analyser la dynamique temporelle des réponses des capteurs, permettant d'exploiter l'information contenue dans les phases transitoires d'adsorption et de désorption afin d'améliorer la classification ou la quantification des gaz.

Une étude plus récente s'est intéressée à l'utilisation d'un modèle d'apprentissage profond multitâche (DF-MT1DCL) combinant un 1D-CNN avec un modèle LSTM pour un capteur sensible aux décharges partielles dans les appareillages à isolation gazeuse via la détection de NO_2 . Ceci permet donc de calibrer dynamiquement les signaux et de prédire en même temps les concentrations de NO_2 ainsi que le taux d'humidité. Avec cette approche, ils ont atteint un coefficient de détermination R^2 de 99.1% pour l'humidité, de 93.5% pour le NO_2 et permettent une classification précise des types de décharges partielles avec 100% de précision [66].

2.3.3 Travaux récents

Plusieurs études récentes ont exploré l'utilisation de différents modèles d'apprentissage automatique pour la classification et la prédiction de la concentration de gaz, notamment pour les composés organiques volatils (COV). Cette section présente certains de ces travaux récents et met en évidence les approches expérimentales ainsi que les performances obtenues. Le tableau 2.2 résume les principales caractéristiques de ces études, notamment le type de capteur utilisé, les modèles d'apprentissage automatique appliqués et les performances rapportées.

Une première étude réalisée par Cho *et al.* [67] propose l'intégration d'une micro-LED (μ LED) directement dans le capteur afin de créer un système photoactivé (μ LP). Ce dispositif utilise une illumination temporellement variable pour générer des réponses transitoires, qui sont ensuite traitées par un réseau de neurones profond (DNN). Cette stratégie permet à la fois la classification et la régression des concentrations de différents gaz cibles (méthanol, éthanol, acétone et NO_2). Les auteurs exploitent également les parties transitoires des signaux, ce qui permet une prédiction rapide et améliore la précision de la détection. Les performances obtenues atteignent 96,99% pour la classification et une erreur absolue moyenne en régression (MAPE) de 31,99 % .

En comparaison avec différentes méthodes d'apprentissage supervisé, le D-CNN proposé démontre une performance de détection des gaz nettement supérieure au perceptron multicouche (MLP), les k-plus proches voisins (KNN) et les machines à vecteurs de support (SVM).

De plus, Mei *et al.* [68] ont proposé un système de reconnaissance de COV basé sur un seul capteur MEMS à oxyde métallique utilisant une modulation de température afin de générer des signatures dynamiques pour les gaz auxquelles il était exposé. Les réponses temporelles du capteur sont analysées à l'aide d'un réseau de neurones convolutionnel (CNN) pour la classification des gaz, alors qu'une méthode de transfert d'apprentissage est utilisée pour la prédiction des concentrations. Les résultats obtenus montrent une précision moyenne de classification de 98.38% pour l'identification de 10 COV différents et une précision de 88.65% pour l'estimation de la concentration.

Li *et al.* [69] ont développé un réseau de capteurs à base d'oxydes métalliques avec des nanostructures afin de détecter plusieurs composés organiques volatils. Les réponses de sept capteurs ont été utilisées pour entraîner un modèle basé sur les machines à vecteurs de support (SVM) pour la classification et la prédiction de la concentration des gaz. Après augmentation du jeu de données, la précision de classification a atteint 86.67%. Pour les gaz correctement

identifiés, la prédiction de concentration réalisée par régression SVM a obtenu un coefficient de détermination de $R^2 = 0.96$.

TABLEAU 2.2 Résumé de travaux récents utilisant l'apprentissage automatique pour la détection de COV

Référence	Type de capteur	Modèle ML	Gaz étudiés	Performance
Cho <i>et al.</i> [67]	Capteur photoactif μ LED	D-CNN	Méthanol, éthanol, acétone, NO ₂	96.99%; MAPE 31.99%
Mei <i>et al.</i> [68]	Capteur MOS MEMS	CNN + TL	10 COV	98.38%; 88.65%
Li <i>et al.</i> [69]	Réseau de capteurs MOS	SVM	5 COV	86.67%; $R^2 = 0.96$
Nayak <i>et al.</i> [70]	Capteur MOS (LiFe ₅ O ₈)	AdaBoost	Acétone, éthanol, méthanol	98%; $R^2 = 0.84$
Wang <i>et al.</i> [71]	Capteur électrochimique à potentiel mixte	Random Forest	COV	99%
Singh <i>et al.</i> [72]	Réseau de capteurs MOS (NiO, CuO, ZnO)	KNN / RF	Acétone, éthanol, toluène, chloroforme	>99%; $R^2 > 0.99$

Ces travaux démontrent que les approches basées sur l'apprentissage automatique, et particulièrement les architectures profondes, permettent d'améliorer significativement les performances de classification et de quantification des gaz. Ils montrent également l'intérêt d'exploiter les réponses dynamiques des capteurs afin d'extraire davantage d'information pour la détection et l'identification des gaz. D'autres travaux issus de la littérature ont également été ajoutés au Tableau 2.2 afin de présenter un aperçu plus global des approches proposées dans le domaine.

CHAPITRE 3 MÉTHODOLOGIE

Dans ce chapitre, il sera question de présenter la méthodologie adoptée pour développer et évaluer des modèles d'apprentissage automatique pour la détection de gaz à l'aide de capteurs Fabry-Pérot. La première approche repose sur l'utilisation de jeux de données publiques afin de réaliser un apprentissage préliminaire et de valider différents algorithmes.

En effet, cette étape s'appuie sur des données provenant de capteurs présentant des performances établies et une dynamique de réponse comparable au Fabry-Pérot. Ce choix s'explique aussi par la volonté de faire progresser le projet dans ses premières phases, alors que les polymères sensibles n'étaient pas encore disponibles et que les performances des premières cavités Fabry-Pérot présentaient une dégradation rapide.

Elle permet aussi de comparer les performances dans un cadre contrôlé à partir de données issues d'expérience réelle. Ceci permet aussi d'évaluer si le capteur développé atteint un niveau de performance cohérent avec l'état de l'art pour une dynamique de réponse semblable. Les résultats permettront ainsi d'orienter le choix de l'approche la plus appropriée pour l'acquisition et le traitement des données expérimental.

Ensuite, les modèles retenus seront évalués à l'aide des mesures expérimentales obtenues en chambre à gaz à partir du capteur Fabry-Pérot. Les critères de performance considérés incluent notamment la précision de classification ainsi que les erreurs de régression.

3.1 Jeu de données publiques et apprentissage préliminaire

3.1.1 Description du jeu de données

Pour ce projet, le jeu de données utilisé provient du *ChemoSignals Laboratory* à l'Université de Californie à San Diego. Celui-ci contient des séries temporelles enregistrées à partir de 16 capteurs chimiques de 4 types différents (TGS-2600, TGS-2602, TGS-2610, TGS-2620) exposés à des mélanges gazeux d'éthylène/méthane ainsi que l'éthylène/ monoxyde de carbone [73]. Les données utilisées dans ce projet ont été rendues disponibles via la plateforme Kaggle [74].

Chaque acquisition a été effectuée sur une plage de 12 heures continues avec une fréquence d'échantillonnage de 100Hz. Les concentrations des gaz varient de manière aléatoire tous les 80 à 120 secondes de sorte à couvrir toutes les transitions possibles. Le Tableau 3.1 présente les plages de concentrations des différents gaz ainsi que les classes auxquelles elles

sont associées. La classe G1 correspond à l'éthylène pour les deux mesures, tandis que la classe G0 correspond au méthane pour l'une des mesures et au monoxyde de carbone pour l'autre.

TABLEAU 3.1 Plages de concentration étudiées pour chaque gaz.

Gaz	Classe	Plage de concentration (ppm)
Éthylène	G1	0–20
Méthane	G0	0–300
Monoxyde de carbone	G0	0–600

Les données sont structurées selon 19 colonnes comprenant le temps (colone 1), les concentrations des deux gaz pour chaque mesure (colone 2-3) ainsi que les réponses des 16 capteurs (colone 4-19). Au total, plus de 4208261 échantillons sont présents pour la première mesure avec l'éthylène et le monoxyde de carbone et 4178504 échantillons sont présents pour la deuxième mesure de l'éthylène et du méthane.

3.1.2 Sélection préliminaire des modèles

Dans le cadre de ce projet, plusieurs modèles issus de la revue de littérature (Section 2) ont été retenus. Le Tableau 3.2 énumère l'ensemble de ces modèles qui seront testés et comparés, de manière à couvrir différentes familles de méthodes d'apprentissage automatique adaptées aux tâches de régression et de classification appliquées à des capteurs de gaz.

TABLEAU 3.2 Familles de modèles évaluées dans cette étude.

Famille de modèles	Algorithmes évalués
Modèles linéaires	Régression linéaire multivariée
Méthodes à base d'arbres	XGBoost
Méthodes à base de distance et de marges	k-NN, SVC, SVR
Réseaux de neurones	CNN, RNN, LSTM

3.1.3 Prétraitement des données

Avant l'entraînement des modèles, les données seront traitées et préparées pour optimiser la qualité de l'apprentissage. En premier lieu, une sélection des quatre capteurs les plus performants et de type différent sera effectué à partir d'une étude de sensibilité et de répétabilité selon différents gaz. La quantité de 4 capteurs permettra de bien représenter l'étude des cavités Fabry-Pérot, celles-ci étant limitées à seulement quatre polymères. Un nombre limité de

capteurs seront utilisés lors des mesures expérimentales des capteurs Fabry-Pérot en raison des limitations du montage expérimental et du nombre de polymères sensibles disponibles.

Initialement, les séries temporelles seront ré-échantillonnées à une fréquence de 1Hz de sorte à réduire le temps de calcul lors de l'entraînement tout en gardant une très grande résolution temporelle au niveau du signal des capteurs. Ceci permet aussi de réduire le bruit à haute fréquence du capteur et d'obtenir une meilleure généralisation.

Suivant cette étape, un décalage temporel est appliqué aux capteurs afin de compenser le temps de transit requis entre la sortie du système d'injection et la surface des capteurs. Elle permet donc un alignement des réponses de capteurs avec les changements de concentrations.

Les signaux sont alors séparés à l'aide d'une validation croisée temporelle imbriquée (*Nested Cross-Validation*) afin de bien respecter la nature séquentielle des données et puisque la distribution des régimes le permet. En effet, puisque ceux-ci sont temporels et que la distribution des variations aléatoires de concentration demeure uniforme au sein des mesures, il est important de préserver l'ordre chronologique des observations lors de l'entraînement et des évaluations. Pour s'y faire, la méthode *TimeSeriesSplit* a été utilisée. Celle-ci divise les données en segments successifs où chaque ensemble d'entraînement contient uniquement des observations antérieures à celles du segment de test. Dès lors, le modèle est toujours testé sur les données futures par rapport à ceux utilisés pour l'apprentissage. Cela permet d'évaluer la capacité des modèles à généraliser correctement, plutôt que de simplement mémoriser les données d'apprentissage.

Dans un premier temps, une validation croisée externe à 3 plis est utilisée pour estimer la performance réelle du modèle. À l'intérieur de chaque pli externe s'y retrouvent 2 plis internes dans lesquels nous avons une optimisation des hyperparamètres du modèle en question à l'aide de l'optimisateur Optuna (voir Figure 3.1).

Les données ici sont donc segmentées selon l'architecture des modèles. Pour les modèles limités au vecteur à 2 dimensions en entrée (SVM, KNN, XGBOOST et linéaire), les données d'entraînement ont alors la forme suivante :

$$y(t_i) = f(S_1(t_i), \dots, S_{n_{\text{capteurs}}}(t_i)),$$

Alors que pour les modèles à dynamique temporelle (LSTM/CNN), chaque entrée est une fenêtre temporelle par capteur et la concentration prédite est associée au dernier instant de la fenêtre tel que :

$$y(t_i) = f(S_j(t)), \quad t \in [t_i - T + 1, t_i], \quad j = 1, \dots, n_{\text{capteurs}}.$$

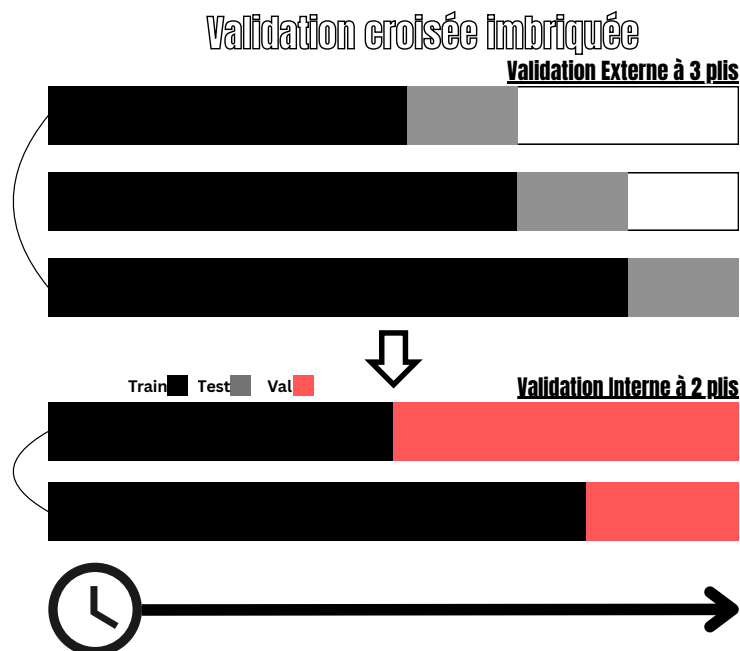


FIGURE 3.1 Schéma de validation croisée imbriquée : 3 plis externes (évaluation) et 2 plis internes (optimisation) en respectant l'ordre temporel.

Les données temporelles sont obtenues à l'aide de fenêtres glissantes et disjointes de 5 points sur tout le domaine. Ceci permet de maximiser le nombre de données tout en minimisant la forte corrélation entre les fenêtres voisines.

Suivant cette étape, les réponses des capteurs seront filtrées et normalisées par pli pour éviter la fuite de données d'un pli à l'autre. Un filtre Savitzky-Golay à polynôme de degré 2 est appliqué séparément sur l'ensemble des plis pour éviter la fuite d'information entre plis. De plus, les modèles SVM, KNN et XGBOOST sont normalisés par standardisation (StandardScaler) alors que les autres modèles sont normalisés à l'aide d'une mise à l'échelle Min-Max.

Les modèles SVR, XGBOOST et K-NN ont entraîné chacun 1 modèle par prédiction de gaz en raison de leur limitation au niveau de leur fondement. Les modèles LSTM, CNN et MLR ont quant à eux été entraînés pour permettre la prédiction des 2 gaz avec un seul modèle. En raison des plages de concentrations différentes pour chacun des gaz, les concentrations de gaz (y_{label}) sont normalisées pour ses modèles à régression multivariée (CNN/LSTM,MLR). Cela empêche qu'un gaz qui atteint de plus hautes concentrations ait une influence disproportionnée sur l'apprentissage par rapport à un autre gaz à plus faible amplitude.

Avant l'apprentissage et la segmentation en plis, les cas où toutes les concentrations de gaz sont nulles ($[0,0]$) ont été retirés du jeu de données. Ces points correspondent à des situations où aucun signal utile n'est présent pour les capteurs, et leur inclusion pourrait biaiser l'apprentissage en favorisant la prédiction de ce cas particulier puisqu'il s'agit de la combinaison de gaz mixte la plus présente dans les 2 séries de mesure. En les retirant, les modèles apprendront donc les variations réelles de concentrations et obtiennent une plus grande précision sur sa capacité à estimer les niveaux de gaz présent.

3.1.4 Optimisation des hyperparamètres

Afin d'optimiser les performances de chaque modèle, une recherche d'hyperparamètres sera effectuée à l'aide d'Optuna, une structure d'optimisation bayésienne moderne adaptée aux espaces de recherche complexes. Elle permet d'explorer les configurations d'hyperparamètres au moyen de stratégies d'échantillonnage intelligentes (TPE) et de mécanismes de pruning arrêtant automatiquement les essais non prometteurs pour réduire le temps de calcul [75].

Contrairement aux méthodes classiques de recherche d'hyperparamètres, telles que les recherches exhaustives par grille (*grid search*) ou les recherches aléatoires (*random search*), cette approche permet d'orienter progressivement l'exploration vers les régions les plus prometteuses de l'espace des hyperparamètres.

L'optimisation des hyperparamètres est réalisée à l'intérieur d'une validation croisée imbriquée. Les hyperparamètres sont sélectionnés à l'aide des plis internes, alors que l'évaluation finale des performances est effectuée sur les plis externes pour d'éviter toute fuite d'information. Pour chacun des modèles ci-dessous, les hyperparamètres pertinents seront ceux recommandés par la littérature, comme présenté au Tableau 3.3.

TABLEAU 3.3 Modèles évalués et hyperparamètres optimisés

Famille de modèles	Hyperparamètres optimisés
Modèles linéaires	α (régularisation), type de normalisation, degré polynomial.
Méthodes à base d'arbres	Profondeur maximale, nombre d'arbres, min samples leaf, critère d'impureté.
Méthodes distance/marges (k-NN, SVM/SVR)	k , métrique de distance, C , γ , ε , choix du noyau.
Réseaux de neurones (CNN, RNN, LSTM)	Nombre de couches, nombre de filtres/unités, taille des noyaux, taux d'apprentissage, dropout, taille des lots, taille de séquence, fonction d'activation, optimiseur.

L'objectif d'optimisation retenu pour les problèmes de régression est la moyenne des NRMSE

calculées sur les cibles et normalisées selon l'intervalle de variation observées dans le pli d'entraînement correspondant. *NRMSE* permet de comparer les performances entre gaz avec une plage de concentration différente et démontre à quel point le modèle pénalise les grandes erreurs. Quant à la classification, l'objectif d'optimisation est le paramètre *F1 macro*.

Finalement, les domaines des hyperparamètres pour ces études sont donnés au Tableau A.1. Ces plages ont été utilisées à la fois pour les problèmes de régression et de classification.

3.1.5 Métriques de comparaison des modèles

Suite à l'optimisation dans le pli interne, Optuna sélectionne la meilleure combinaison d'hyperparamètre offrant les meilleures performances. Celle-ci est ensuite utilisée pour entraîner le modèle sur tout le pli d'entraînement externe correspondant et l'évaluation est donc sur les données de test de ce pli externe. Les performances sont alors stockées et la moyenne des résultats sur les trois plis est comptabilisée. Les performances rapportées correspondent donc à la moyenne et à l'écart-type obtenus sur l'ensemble des plis. Ceci garantit une estimation robuste et non biaisée des performances de prédiction.

Pour la régression, les métriques retenues sont *NRMSE*, *MAE* et R^2 . Ceux-ci sont rapportés séparément pour les deux gaz de la mesure. *NRMSE* permet de comparer les performances entre gaz avec une plage de concentration différente et démontre à quel point le modèle pénalise les grandes erreurs. *MAE* (en ppm) fournit plutôt une mesure intuitive de l'erreur moyenne absolue du modèle. Finalement R^2 quantifie la proportion de variances expliquées par le modèle.

Pour la classification, celle-ci est faite à l'aide d'un problème à 3 classes (gaz 1 seul, gaz 2 seul et mélange). Les métriques retenues sont *Acuracy* (%), *F1-macro* (%) et *Recall_{mix}* (%).

L'*Acuracy* (%) mesure ici la proportion globale des prédictions correctes, alors que *F1-macro* (%) attribue un poids égal à chacune des classes pour la mesure des prédictions puisque le nombre de cas pour chacune des classes est non uniforme. Ensuite, *Recall_{mix}* (%) rapporte plutôt le cas de prédiction correcte de la classe de mélange puisqu'elle représente le cas expérimental le plus complexe et pertinent.

3.2 Acquisition et utilisation des données expérimentales

L'acquisition des données fiables dépend à la fois de la rigueur des procédures de mesure, mais aussi sur la qualité de microfabrication des capteurs Fabry-Pérot. Dans cette section, il sera question d'aborder le procédé de fabrication des capteurs développés ainsi que la

méthodologie mise en œuvre pour l'acquisition et le traitement des mesures expérimentales.

3.2.1 Fabrication des capteurs

La fabrication des capteurs de gaz pour ce projet a impliqué trois étapes principales : la microfabrication, l'intégration des polymères sensibles aux composés organiques volatils ainsi que l'alignement de fibres optiques dans les cavités afin de faire l'acquisition de spectre du Fabry-Pérot.

Microfabrication des cavités Fabry-Pérot

Les différentes étapes de microfabrication, de métallisation et de caractérisation des cavités Fabry-Pérot ont été réalisées dans le cadre de ce mémoire. Les capteurs ont été fabriqués en salle blanche en suivant le procédé illustré à la Figure 3.2 [76] [42]. Des gaufres de silicium de type Czochralski avec orientation cristallin (110) ont été sélectionnées pour leur compatibilité avec les étapes de gravure anisotrope.

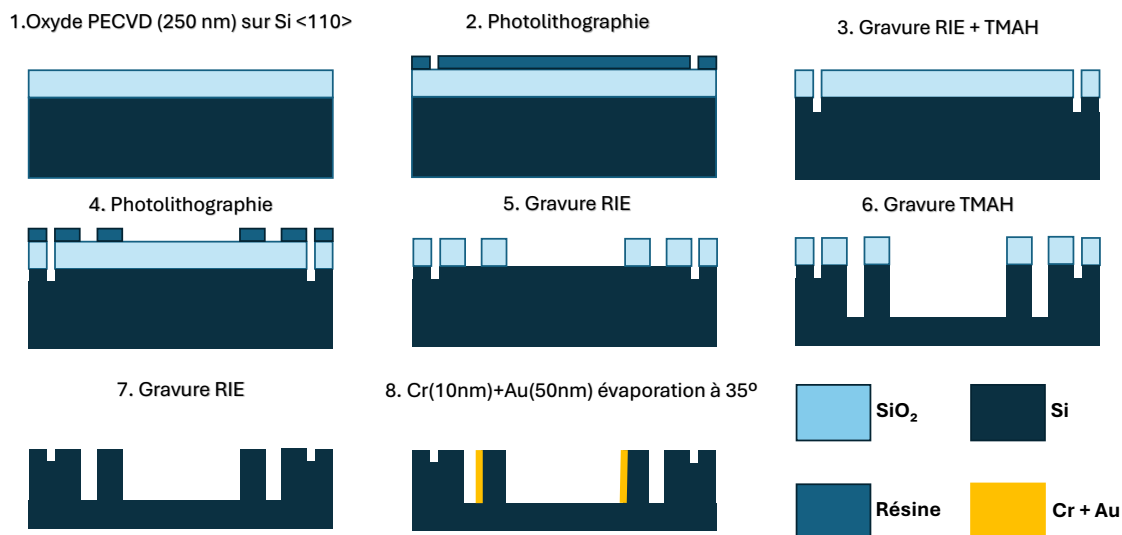


FIGURE 3.2 Schéma du procédé de microfabrication des cavités Fabry-Pérot.

Une couche d'oxyde de silicium de 250 nm a ensuite été déposée par PECVD pour protéger la surface de la gaufre et servir de masque de gravure. Une première photolithographie a été réalisée afin de définir un masque d'alignement, permettant de repérer l'orientation des plans cristallins. La gaufre a ensuite été placée dans un système de gravure réactive plasma (RIE)

pour graver les surfaces exposées par la photolithographie, à l'aide d'un mélange gazeux de 12 sccm de CHF_3 et 2 sccm de O_2 pendant 9 minutes. La résine a ensuite été retirée par gravure O_2 (10 sccm) pendant environ 5 minutes.

Le substrat a été immergé dans une solution de TMAH à 90 °C pendant 15 minutes afin d'exposer les plans cristallins (110). Une deuxième photolithographie a ensuite été réalisée pour définir les motifs des capteurs, en alignant soigneusement le masque selon l'orientation des plans cristallins. Les étapes de gravure RIE ont été répétées afin de transférer le motif sur la couche d'oxyde du wafer.

La gravure TMAH a été poursuivie jusqu'à obtenir l'épaisseur désirée des miroirs semi-réfléchissants, comprise entre 1 et 3 μm . Le résultat obtenu est présenté à la Figure 3.3.

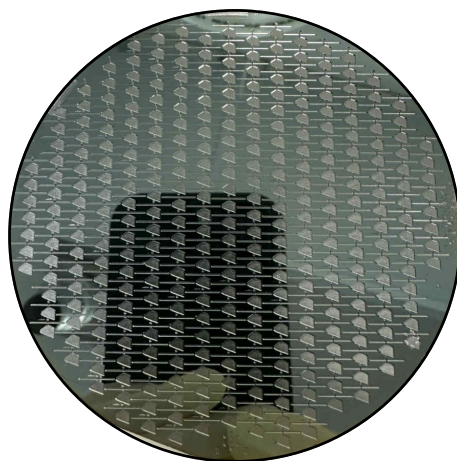


FIGURE 3.3 Tranche de silicium microfabriquée avant l'étape de métallisation.

Enfin, une métallisation a été effectuée par évaporation de 8 μm de chrome suivie de 22 μm d'or, déposés sous un angle de 35° par rapport à la normale des miroirs pour obtenir la réflectivité souhaitée.

Choix et déposition des polymères

Pour ce projet, dix polymères complémentaires ont été sélectionnés afin de fonctionnaliser les cavités Fabry-Pérot pour la détection de différents composés organiques volatils (COV) :

TABLEAU 3.4 Polymères utilisés pour la fonctionnalisation des cavités Fabry-Pérot

Abréviation	Polymère
PEG	poly(ethylene glycol)
PHMB	poly(hexamethylene biguanide)
PEVA	poly(ethylene-co-vinyl acetate)
PDMS	poly(dimethylsiloxane)
PEO	poly(ethylene oxide)
PVA	poly(vinyl alcohol)
PVP	poly(vinylpyrrolidone)
PU	polyurethane
PMMA	poly(methyl methacrylate)
SU-8	SU-8

Le PDMS utilisé est la formulation Sylgard 184. Cette formulation se compose de deux composants : la base, qui est le prépolymère de silicone, et l'agent de durcissement, qui permet la polymérisation par réticulation. Les deux composants ont été mélangés dans un ratio standard de 10 :1 (base :durcisseur) puis injectés dans le réservoir des cavités. Le PDMS a ensuite été distribué entre les miroirs par capillarité et durci complètement après une période d'environ 24h.

Les polymères de type PU et PEVA ont été introduit sous forme de gélules fondues directement dans les réservoirs, à une température de 200 °C, pour s'introduire par capillarité dans les cavités.

Le PMMA a été dissous dans un solvant approprié avec un poids moléculaire de 950 kDa, puis introduit dans le réservoir afin de recouvrir uniformément la cavité. Le solvant a ensuite été évaporé à 180 °C, et quatre dépôts successifs ont été nécessaires pour obtenir une épaisseur et une couverture homogènes du polymère dans la cavité.

Finalement, le PEG, PHMB, PVP, PVA et PEO ont été préparés en solution de 5% massique à partir d'eau DI. Tout comme le PMMA, plusieurs dépôts successifs ont été réalisés afin de remplir les cavités.

Ces étapes ont été répétées de 2-4 fois pour chacun des polymères de sorte à avoir une certaine redondance et répétabilité. Ceci permet donc de pouvoir continuer les séquences de mesures si l'un des capteurs ou polymères devient défectueux.

Suite à ceci, des fibres optiques ont été alignées et collées de chaque côté des miroirs à l'aide de l'adhésif Norland 68 afin d'établir le spectre de transmission des cavités Fabry-Pérot telles que montré sur la Figure 3.4. Par la suite, le tout est installé avec cette même colle sur un support en PLA pour faciliter les déplacements.

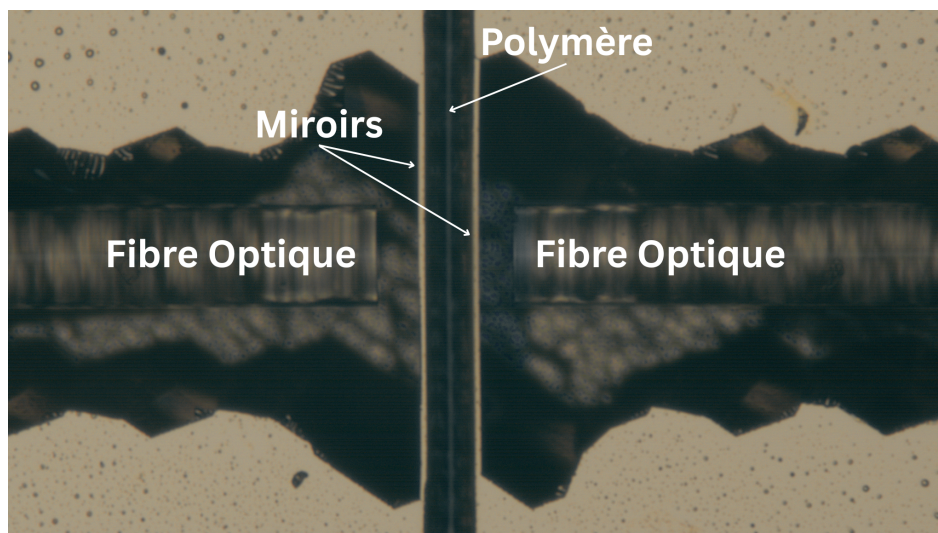


FIGURE 3.4 Représentation des capteurs Fabry-Pérot suite à la fabrication

3.2.2 Montage expérimental

Le montage expérimental se compose d'un système d'injection de gaz et d'un système optique permettant d'exposer différentes cavités à des analytes liquides et gazeux, tout en enregistrant à une certaine fréquence d'échantillonnage les spectres de forme d'onde de chacune des cavités. Le développement du système d'acquisition des données et du contrôle expérimental des injections de gaz a également été réalisé dans le cadre de ce mémoire.

Montage d'injection de gaz

Le montage d'injection de gaz permet à la fois l'injection d'analyte sous forme gazeuse et liquide. Cependant, comme ce projet est concentré sur les COV, seulement le système d'injection d'analytes liquide sera utilisé. Les analytes liquides nécessitent l'utilisation d'un bulleur. Le bulleur est un contenant hermétique dans lequel un gaz porteur traverse le liquide sous forme de bulle afin de favoriser son évaporation et d'acheminer les vapeurs vers la sortie du bulleur. Le gaz porteur choisi à cet effet est le diazote. Le diazote a été sélectionné pour plusieurs raisons. Premièrement il s'agit d'un gaz inerte qui ne favorise pas la dégradation

du polymère, car il est non réactif. De plus, il permet de bien simuler les conditions du gaz atmosphérique, puisque celui-ci est le principal constituant de l'air ambiant.

Pour déterminer les concentrations auxquelles le capteur Fabry-Pérot est exposé à l'aide d'un bulleur, nous devons utiliser l'équation suivante [17] :

$$C_n(P, T) = \frac{\Phi_n}{\sum_{m=1}^M \phi_m + \sum_{n=1}^N \Phi_n} \quad (3.1)$$

où :

- N est le nombre total de bulleurs,
- M est le nombre total de débitmètres,
- Φ_n représente le flux de l'analyte n ,
- ϕ_m désigne le débit du gaz porteur (azote) associé au débitmètre m ,
- $C_n(P, T)$ correspond à la concentration de l'analyte n à la température T et à la pression P à l'entrée de la boîte de mesure.

et

$$\Phi_n = \frac{P_n^{vsat}(T)}{P - P_n^{vsat}(T)} \phi_m \quad (3.2)$$

Dans ces équations, la pression est fixée à 760mmHg et la température à 21.5°C, car ceux-ci ne peuvent pas être mesurés dans le bulleur. Le coefficient $P_n^{vsat}(T)$ peut être déterminé par l'entremise des coefficients A,B,C d'Antoine [77] tel que

$$\log_{10} \left(\frac{P^{vsat}}{1mmHg} \right) = A - \frac{B}{T + C} \quad (3.3)$$

L'ensemble du flux des gaz est contrôlé directement à l'aide de débitmètres de type *Alicats*, les débitmètres pour les analytes liquides permettent des débits allant jusqu'à 100 sccm et celui de dilution jusqu'à 4000 sccm. Ceux-ci ont une erreur de $\pm 0.6\%$. Un cube en cuivre est utilisé pour stabiliser la température du gaz à la sortie, assurant un flux thermique uniforme. La pression et l'humidité du gaz sont également enregistrées pour le suivi des conditions expérimentales.

Pour l'étude des effets de COV sur les différents polymères, les gaz suivants ont été retenus : l'isopropanol (IPA), l'acétone, le méthanol et le toluène. Ceux-ci ont été sélectionnés en raison de leur propriété physicochimique distincte qui permettra d'explorer les différentes interactions entre les polymères et ces molécules.

La recette de l'analyte (MF2) a été appliquée selon des fonctions échelon ascendantes et descendantes à 5 niveaux, avec différents pas de débit : 3 sccm, 5 sccm, 10 sccm, 15 sccm et

20 sccm. Une série complémentaire a ensuite été réalisée avec des débits aléatoires. Le gaz de dilution (MF1) quant à lui a été maintenu à 2000 sccm.

Les faibles variations de concentration à l'aide des fonctions échelon produisent des réponses transitoires de polymère très court et sont rapidement suivies d'un plateau quasi statique. Les variations de débits aléatoires imposent au capteur des changements rapides et imprévisibles des concentrations. La durée entre deux variations de flux est souvent insuffisante pour permettre l'atteinte l'équilibre. La réponse des capteurs est donc maintenue dans un régime transitoire, qui représente la phase la plus complexe du comportement étudié. Par ailleurs, la partie aléatoire permet d'observer la stabilité et la réversibilité du matériau face à des changements rapides ou imprévisibles. Une durée de 15 minutes par flux garantit l'atteinte de la phase stationnaire du polymère, tandis qu'une période expérimentale totale de 55 heures assure la collecte de données robustes et représentatives (Voir Figure 3.5).

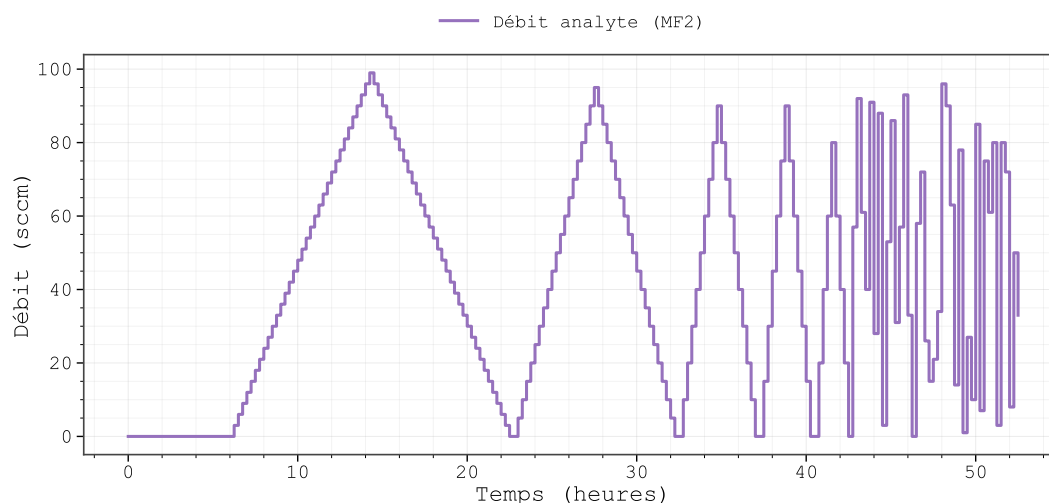


FIGURE 3.5 Représentation graphique de la recette de l'analyte (MF2) appliquée aux polymères

Après le prétraitement et la segmentation des signaux, les acquisitions expérimentales ont permis de générer environ 13000 points de données temporelles de décalage spectral exploitables ($\Delta\lambda$) pour l'apprentissage automatique par analyte, pour un total d'environ 52000 points pour l'ensemble des quatre analytes étudiés.

Montage optique

Le montage optique a été conçu afin de mesurer simultanément, en transmission, le spectre d'interférence de huit cavités en parallèle sur une plage de longueurs d'onde donnée. Pour s'y faire, un laser accordable ainsi que 8 photodiodes ont été utilisés.

Le laser accordable est le *Keysight N7776C* avec une puissance de sortie allant jusqu'à 12 dBm ainsi qu'une plage spectrale de 1450-1650 nm. La résolution spectrale de celui-ci est de 0.1 pm avec une répétabilité de ± 0.5 pm. De plus, la vitesse de balayage maximale de celui-ci est de 200 nm/s.

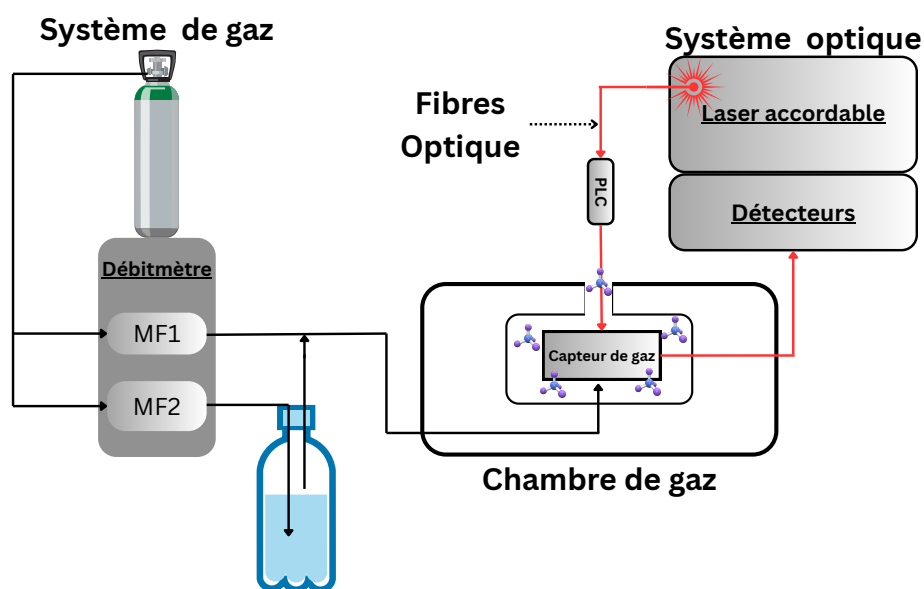


FIGURE 3.6 Montage optique et d'injection de gaz pour effectuer les mesures. L'ensemble des systèmes sont contrôlés par un ordinateur.

La sortie du laser est ensuite divisée en 8 à l'aide d'un coupleur optique *Planar light circuit* (PLC) 1x8. Suite à la transmission du laser dans les cavités Fabry-Pérot, les signaux sont alors mesurés par un puissance mètre *Keysight N7744C* avec 8 canaux, une plage de longueur d'onde de 1450 à 1650 nm, une résolution en puissance de 0.001 dB, une précision de 0.01 dB et un temps d'intégration minimal de 1 μ s.

Les paramètres expérimentaux du balayage laser choisi pour l'ensemble des mesures sont résumés ci-dessous :

- **Plage spectrale** : 1450 à 1650 nm
- **Puissance de sortie** : 13 dBm
- **Vitesse de balayage** : 40 nm/s
- **Pas spectral (résolution)** : 10 pm

Ces paramètres permettent d'acquérir un spectre d'environ 20 000 points.

La combinaison des deux systèmes, contrôlés par un ordinateur sous le logiciel *LabVIEW*, permet donc d'acquérir les données nécessaires à ce projet. Le montage correspondant est affiché à la Figure 3.6. L'alignement des fibres optiques et l'optimisation de l'acquisition des signaux d'interférence ont également constitué une partie importante des travaux expérimentaux réalisés.

3.2.3 Prétraitement et analyse des données expérimentales

Dans les mesures, il sera question d'évaluer le déplacement des résonances des cavités Fabry-Pérot en fonction des concentrations dans le temps. Pour s'y faire, 2 techniques s'offrent : La corrélation croisée ainsi que le suivi de pic. L'utilisation de la méthode de corrélation croisée s'est avérée plus robuste et plus performante que la méthode de suivi de pic (*peak following*). Cette dernière présente une sensibilité accrue aux pics associés à des modes secondaires, lesquels peuvent, dans certaines conditions, dominer les modes fondamentaux, conduisant ainsi à des estimations moins fiables de déplacement mesuré.

La corrélation croisée permet, en revanche, de mettre en évidence la tendance globale des signaux, minimisant l'influence des modes secondaires et fournissant ainsi des résultats plus robustes aux anomalies et représentatifs du comportement du polymère.

3.3 Application des modèles et optimisation

Dans cette section, il est question d'évaluer les modèles retenus à l'issue de l'analyse du jeu de données sur les données des capteurs Fabry-Pérot. Ceux-ci ont été examinés dans différents régimes dynamiques. Plus précisément, les mesures expérimentales ont été réparties en deux régimes distincts, définis en fonction du régime majoritairement observé : un régime quasi statique et un régime transitoire.

La première partie correspond à des expositions graduelles au gaz selon une fonction échelon. Dans ce cas, les capteurs atteignent un plateau stable très rapidement en raison des petits

changements de concentrations. Cette portion de données est donc considérée comme quasi-statique et représente les premiers 80% des mesures. C’est dans cette partie qu’on observe les réponses majoritairement linéaires des capteurs.

La deuxième partie correspond à une séquence de concentrations de gaz choisies de manière aléatoire, plutôt que suivant une progression en échelon. Dans cette situation, les concentrations évoluent beaucoup et rapidement, plaçant ainsi le capteur majoritairement en régime transitoire. Cette section représente les derniers 20% des séries de mesures. C’est dans cette partie qu’on observe les réponses majoritairement non linéaires des capteurs.

Dans un premier temps, les modèles ont été entraînés séparément sur ces deux régimes afin d’évaluer leurs performances dans chacun des cas. Ici, il était question de comparer les performances de chacun de ces modèles s’ils étaient spécialement faits majoritairement pour un de ces deux régimes (quasi statique et transitoire).

Pour ces deux analyses, une validation croisée imbriquée a été utilisée, tout comme pour le jeu de donnée, afin d’obtenir une estimation robuste des performances. Les plis externes ont servi à l’évaluation des performances alors que les plis internes sont utilisés pour l’optimisation des hyperparamètres. Les normalisations des données sont effectuées uniquement à partir des données d’entraînement de chaque pli externe afin d’éviter toute fuite d’information. Les types de normalisation et les domaines d’hyperparamètre analysé par modèles seront les mêmes que ceux des jeux de données.

La stratégie de validation utilisée ici repose sur une approche de type *Group K-fold*. Chaque groupe correspond à une mesure complète, ce qui permet d’éviter d’avoir des segments provenant d’une même acquisition se retrouve simultanément dans les ensembles d’entraînement et de tests. Elle permet d’éviter les corrélations temporelles entre les ensembles pour avoir des estimations plus réalistes au niveau des généralisations.

Suite à ceci, une dernière évaluation a été réalisée en utilisant cette fois-ci l’ensemble des mesures afin d’obtenir des modèles pouvant prédire à la fois les concentrations dans les deux régimes. En raison de la distribution non uniforme de ceux-ci, une pondération 4-1 a été introduite pour éviter que l’entraînement soit dominé par le régime quasi statique en raison de son plus grand nombre de cas.

Pour cette étape, une validation croisée de type *stratified Group K-Fold* a été utilisée. Celle-ci permet de préserver les proportions des données quasi statique et transitoire dans les plis tout en respectant les séparations par groupe. Ceci permet donc d’assurer une distribution uniforme des régimes dynamique dans l’ensemble des entraînements et évite de favoriser.

Afin de faciliter la comparaison des résultats entre les gaz, les valeurs de prédiction des

concentrations pour la tâche de régression ainsi que l'erreur moyenne absolue (MAE) seront maintenues en SCCM plutôt qu'en PPM. Puisque les coefficients d'Antoine sont propres à chaque analyte, la conversion des débits en concentration ferait apparaître des plages de mesure différentes pour chaque gaz, ce qui rendrait la comparaison des résultats moins évidente visuellement.

La tâche de classification des gaz a aussi été réalisée sur l'ensemble des mesures avec une validation *Group K-Fold*. Encore une fois, les mêmes trois modèles pour la régression seront retenus pour cette étude.

Les mêmes métriques de performance que celles utilisées pour l'analyse du jeu de données ont été retenues pour ces évaluations.

CHAPITRE 4 RÉSULTATS ET DISCUSSION

Dans cette section, il sera question de présenter les résultats obtenus pour le jeu de données publiques ainsi que ceux des résultats expérimentaux pour les cavités Fabry-Pérot.

4.1 Comparaison des réponses du jeu de données en ligne avec les capteurs Fabry-Pérot

Malgré les principes de détection différente, ces capteurs chimiques ont une réponse temporelle non linéaire similaire à ceux des Fabry-Pérot.

Tout comme les Fabry-Pérot, ceux-ci ont une phase transitoire caractérisée par une réponse exponentielle ascendante ou descendante suite à une variation de concentration ainsi qu'une phase stable correspondant à un plateau lorsque le capteur atteint un équilibre entre l'adsorption/désorption et la diffusion dans le matériau sensible.

Ce comportement peut être modélisé par les modèles de *two-rate diffusion* défini par les équations de la revue de littérature (section 2). Par l'entremise d'un ajustement du modèle de diffusion à deux constantes (*two-rate diffusion*) sur les données des capteurs chimiques, nous obtenons les résultats présentés à la Figure 4.1 .

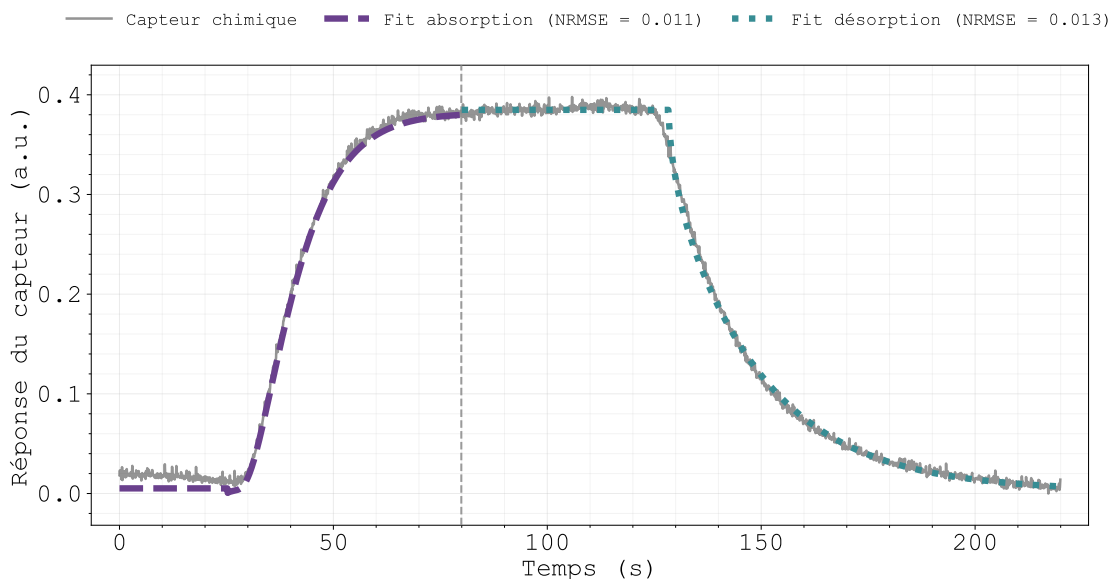


FIGURE 4.1 Comparaison de la réponse temporelle d'un capteur du jeu de données en ligne, avec ajustement à l'équation de diffusion à deux constantes (*two-stage diffusion*) .

Une comparaison équivalente réalisée sur les cavités Fabry-Pérot est présentée en annexe (Figure B.1).

Selon ces graphiques, on conclut que les parties non linéaires de ces capteurs ont les mêmes tendances et peuvent être décrites par le même modèle. Cependant, il est important de noter que la partie transitoire du capteur chimique est plus rapide que celui des capteurs Fabry-Pérot. Il faudra donc prendre ceci en considération pour l'évaluation des modèles.

De même, on remarque que la relation des réponses du capteur dans la phase stable en fonction de la concentration est linéaire tout comme ceux des Fabry-Pérot [16] selon la Figure 4.2.

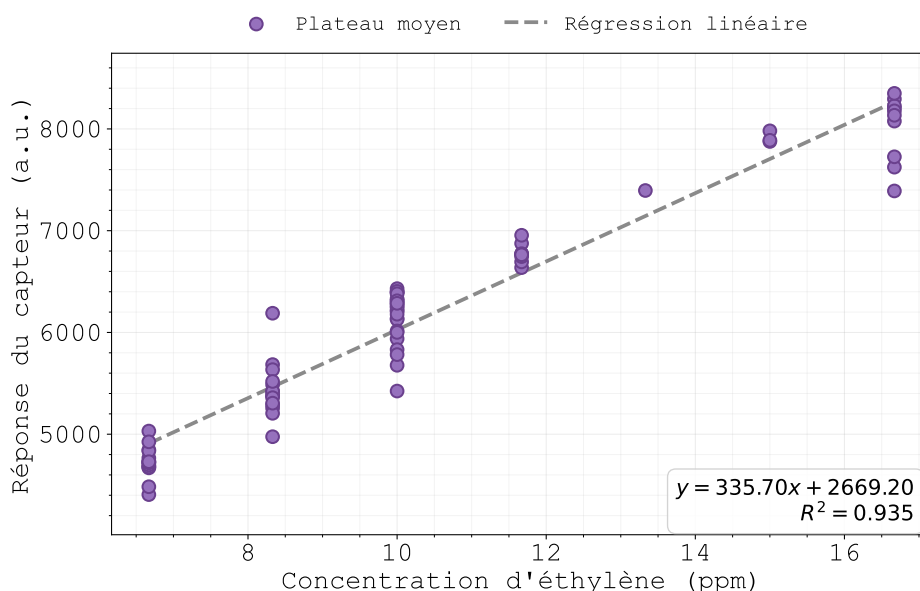


FIGURE 4.2 Exemple de relation linéaire du déplacement en phase stable des capteurs chimiques pour différente concentration d'éthylène pour le capteur de TGS-2600

Nous pouvons donc justifier l'utilisation des données de jeu de l'UCSD afin de tester, trier et optimiser les différents algorithmes.

4.2 Jeu de données de UCSD

Avant d'appliquer le modèle aux données expérimentales issues du montage Fabry-Pérot, nous avons analysé le jeu de données afin de valider notre pipeline d'analyse et de prédiction.

4.2.1 Analyse de l'uniformité des données

En premier lieu, il a été question de regarder la distribution des combinaisons de concentration pour chacune des mesures, soit d'éthylène/CO et d'éthylène/CH₄ afin de caractériser la distribution des données.

Pour chaque couple de gaz, il a été question d'identifier les combinaisons les plus fréquentes, c'est-à-dire les paires de concentrations les plus représentées dans la série de mesures. Voici une représentation des cinq combinaisons les plus courantes sous forme d'histogrammes (Figure 4.3)

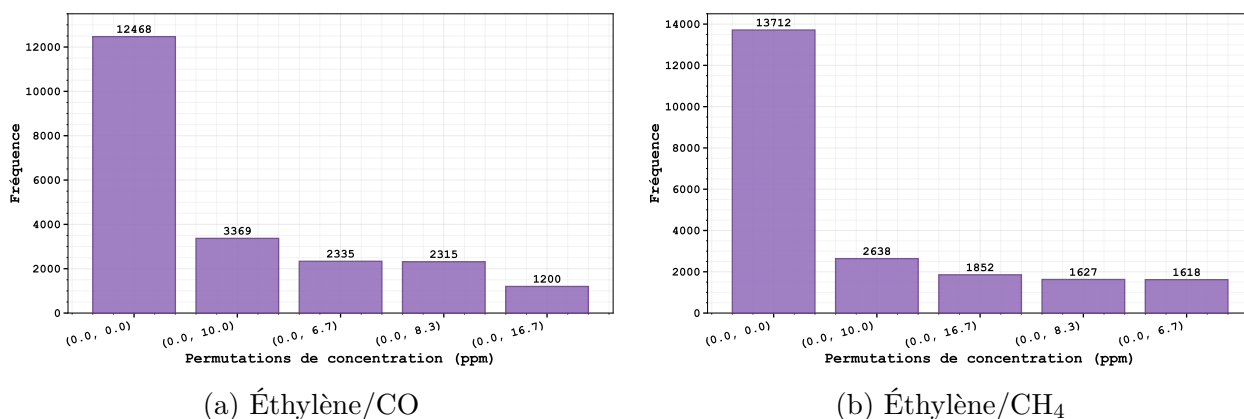


FIGURE 4.3 Top 5 des paires de concentrations (issues des deux mesures) pour chaque couple de gaz

Ces Figures mettent en évidence la non-uniformité des données, notamment au niveau du cas (0,0) ppm, ce qui fournit un aperçu des biais potentiels que pourraient développer les modèles durant l'apprentissage. En effet, selon la Figure 4.3a, on observe environ 12 468 occurrences correspondant à une concentration d'éthylène et de CO à (0,0) ppm, contre 3 369 pour le deuxième cas le plus fréquent, soit (0,10). Cette surreprésentation peut amener les modèles d'apprentissage automatique à surapprendre ce cas spécifique, ce qui peut nuire à leur capacité à bien généraliser. Puisque les cas (0,0) ppm présentent l'absence de gaz, ils ont été retirés de l'étude.

4.2.2 Étude temporelle des capteurs

Lors de l'analyse des données, il a été révélé que les capteurs avaient un décalage d'environ 12.29 s par rapport à la concentration envoyée avec la fréquence d'échantillonnage de 1Hz. Ceci est possiblement dû en raison de la distance que le gaz doit parcourir une fois son

insertion dans le système pour arriver aux capteurs.

Ceci a été déterminé en calculant la dérivée discrète de chacun des capteurs dans un intervalle de temps où la concentration d'un seul gaz passe de 0 à X ppm. Les premiers points de l'intervalle sont utilisés pour déterminer le bruit de fond (baseline) en calculant un écart-type. Un seuil qui est 5 fois supérieur à celui du bruit de fond est alors utilisé pour identifier quand le capteur réagit réellement. On trouve alors le point où la pente (dérivée discrète) dépasse ce seuil (voir Figure 4.4). Les valeurs de chacun des capteurs sont alors moyennées pour déterminer le temps moyen que ceux-ci prennent pour répondre à l'insertion du gaz. Ce temps est donc tronqué aux données des capteurs pour rectifier les signaux.

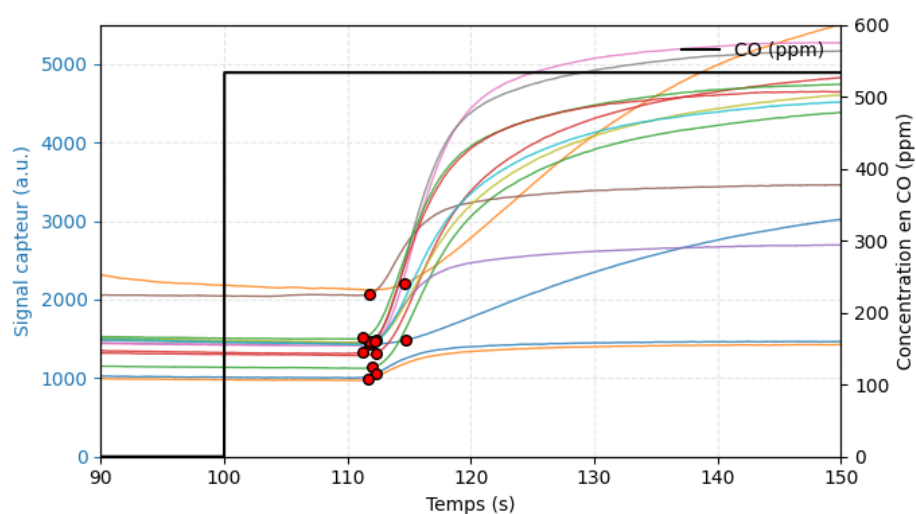


FIGURE 4.4 Exemple des décalages temporels des capteurs sensibles au CO lorsqu'une concentration de 533 ppm est insérée au système. Le début de réaction des capteurs est identifié en rouge.

4.2.3 Choix des 4 meilleurs capteurs

Pour choisir les 4 capteurs les plus performants afin de comparer les différents modèles, une étude de sensibilité et de répétabilité a été réalisée (voir Tableau B.1). En observant les données des capteurs, il a été observé que l'un des capteurs de type TGS2602 était défectueux. Celui-ci a donc été retiré de l'étude.

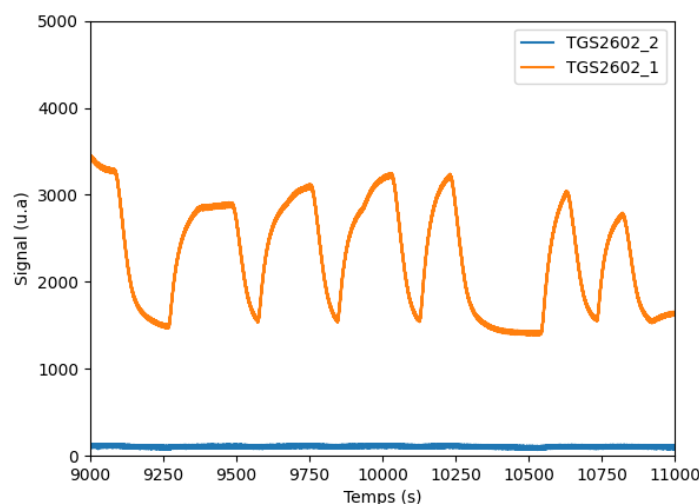


FIGURE 4.5 Comparaison du capteur TGS2602 défectueux (TGS2602_2) avec un capteur fonctionnel (TGS2602_1) de même type, lorsqu'exposé à un gaz.

La sensibilité a été évaluée en identifiant les points où seulement un des deux gaz d'une mesure était présent. Le dernier point de chaque période a donc été retenu pour chacun des capteurs. Une régression linéaire du signal des capteurs à ces points pour les différentes concentrations isolées de chacun des gaz permet donc d'associer au capteur leurs sensibilités avec le facteur de qualité $\mathbf{R}^2$. Le facteur de déviation standard (SD) permet quant à lui de quantifier la répétabilité de chacun des capteurs pour chacune des concentrations de ces gaz. L'ensemble des capteurs ont été analysés au Tableau B.1.

À partir de ces données, les 4 meilleurs capteurs de types différents ont été sélectionnés (voir Tableau 4.1).

Afin de choisir ces capteurs, une emphase sur la répétabilité et la sensibilité a été mise en place. En effet, la sensibilité des capteurs aux gaz permet de quantifier l'amplitude de la réponse du capteur en fonction des variations de concentration, alors que l'écart-type est un indicateur de la répétabilité des mesures pour une même concentration. Ces caractéristiques seront très importantes pour déterminer la précision et les erreurs de prédiction des modèles.

TABLEAU 4.1 Sensibilités, coefficients de détermination (R^2) et écarts-types des quatre meilleurs capteurs pour le CO, le méthane et l'éthylène.

Capteur	CO			CH_4			C_2H_4		
	Sens.	R^2	σ	Sens.	R^2	σ	Sens.	R^2	σ
TGS2602	-0.011	0.514	0.691	0.004	0.085	0.166	-0.342	0.174	2.140
TGS2600	-0.006	0.760	0.202	-0.030	0.230	0.096	-0.305	0.874	0.174
TGS2610	-0.007	0.410	0.555	-0.032	0.177	0.044	-0.369	0.622	0.693
TGS2620	-0.006	0.735	0.206	-0.032	0.301	0.122	-0.276	0.896	0.792

4.2.4 Analyse de la séparabilité des gaz par projection PCA

Suite à la correction du décalage temporel des capteurs et à la sélection des quatre polymères optimaux, il est essentiel d'évaluer si les réponses mesurées contiennent une information suffisamment discriminante pour distinguer les différents gaz présents dans les expériences. À un instant donné t , les réponses simultanées des capteurs peuvent être représentées sous la forme d'un vecteur. L'analyse de ces vecteurs nécessite une méthode de réduction de dimension permettant de mieux visualiser les relations entre les échantillons.

Une analyse en composantes principales (PCA) a donc été réalisée afin de projeter les données dans un espace bidimensionnel tout en maximisant la variance expliquée. Cette méthode transforme linéairement les variables originales en nouvelles variables orthogonales (composantes principales), ordonnées selon la quantité de variances qu'elles expliquent. Les deux premières composantes principales correspondent ainsi aux directions de variation dominante dans les données.

La Figure 4.6 démontre la projection des échantillons purs d'éthylène, de méthane et de CO provenant des deux campagnes de mesure. Les deux premières composantes expliquent respectivement 50.9 % et 26.8 % de la variance totale des signaux bruts, soit plus de 77 % de l'information globale dans les données.

L'analyse de la figure montre que chaque gaz occupe une région spécifique de l'espace des composantes principales. Puisqu'il n'y a pas beaucoup de recouvrement, ceci suggère que les réponses des polymères contiennent une information discriminante suffisante pour permettre à de modèles simples de différencier les gaz étudiés. Donc, l'utilisation de modèle linéaire pour la classification des gaz pourrait être suffisante.

Par ailleurs, les trajectoires observées dans l'espace PCA correspondent à l'évolution temporelle des réponses des capteurs. Les formes des courbes démontrent les différentes phases dynamiques du système (adsorption, régime quasi stationnaire, désorption). Cette géométrie

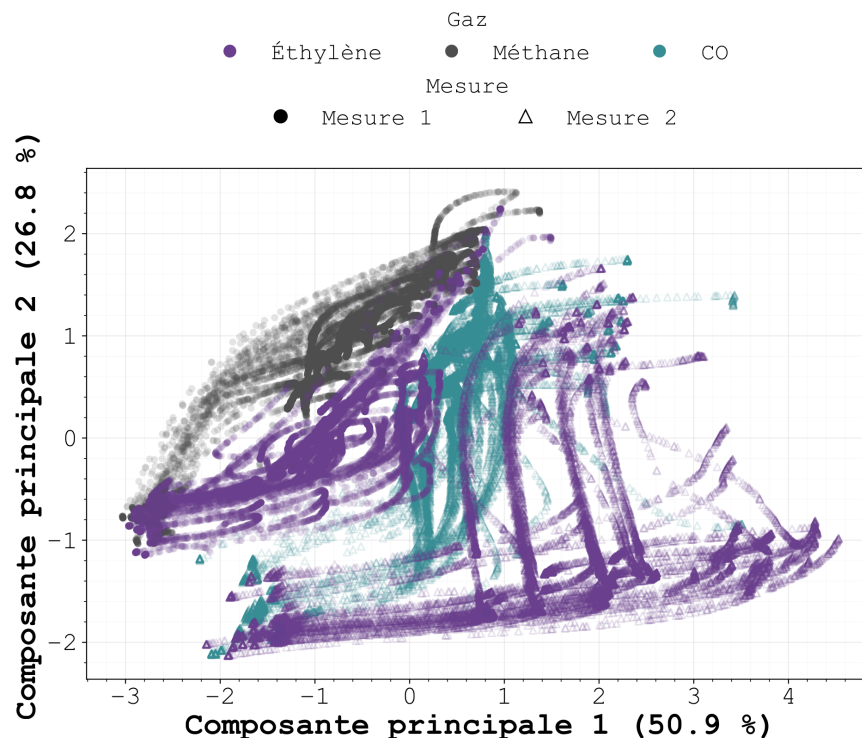


FIGURE 4.6 Projection PCA des échantillons purs d'éthylène, de méthane et de CO pour deux campagnes de mesure indépendantes.

indique que la structure des données n'est pas strictement linéaire.

Bien que la PCA dans ce cas soit une transformation linéaire, la forme non linéaire des trajectoires projetées suggère que la relation entre les réponses des capteurs et la concentration des gaz est complexe. Par conséquent, l'utilisation de modèles strictement linéaires pour les prédictions de concentration serait insuffisante pour capturer ces dynamiques.

Ces observations justifient le choix de modèles d'apprentissage automatique non linéaires comme discuté dans la section 3, capables de modéliser des interactions complexes.

4.2.5 Entraînement des différents modèles d'apprentissage automatique

Suite au traitement des données décrit dans la section 3, il devient possible de procéder à l'entraînement des différents modèles d'apprentissage automatique. L'objectif de cette section est de présenter les résultats de la démarche d'optimisation de ces modèles, notamment par l'entremise de leurs hyperparamètres, puis de comparer de manière rigoureuse les performances obtenues par chacun d'eux.

Modèles de régression

La première partie porte sur les prédictions de concentrations à l'aide des modèles de régression. Afin de comparer ces différents modèles, la moyenne sur les résultats de $NRMSE$, $NRMSE_{CV}$ et fit moyen (R^2) a été calculé sur l'ensemble des gaz de chacune des séries de mesures.

Pour les modèles à série temporelle (CNN et LSTM), l'optimisation des hyperparamètres a été effectuée pour différent nombre de points de mesures consécutives allant de 10 à 100 secondes, avec bond de 10. L'objectif de cette étude est de déterminer le nombre minimal de points nécessaire pour obtenir des résultats de haute qualité. Pour le modèle CNN-1D, le nombre de points optimal était de 20. Quant à LSTM, le nombre de 60 a été retenu, car il correspond au point à partir duquel les améliorations du $NRMSE_{CV}$ commençaient à stagner.

Dès lors, les hyperparamètres des modèles présentés à la section 3 ont été optimisés à l'aide d'une étude conduite avec *Optuna*. Pour chaque configuration d'hyperparamètres proposée, la performance correspondante a été évaluée au moyen d'une validation croisée à trois plis. Une fois les hyperparamètres optimaux identifiés, c'est-à-dire ceux minimisant le $NRMSE$, le modèle final a été réentraîné sur les données prétraitées issues des fichiers de mesures tout en suivant la méthodologie proposée. Les résultats présentés au Tableau 4.2 ont ainsi été obtenus.

TABLEAU 4.2 Comparaison des performances de régression (moyenne $\pm$ écart-type sur 3 plis externes). Le NRMSE est normalisé sur les données d'entraînement. Le MAE est exprimé en ppm. G1 représente l'éthylène dans la série de mesure et G0 l'autre gaz (Méthane/CO).

Mesure	Modèle	NRMSE G0	MAE G0	R^2 G0	NRMSE G1	MAE G1	R^2 G1
Éthylène-Méthane	CNN2D	0.0878 $\pm$ 0.0042	11.74 $\pm$ 1.08	0.909	0.1195 $\pm$ 0.0148	1.15 $\pm$ 0.29	0.859
	SVR	0.0851 $\pm$ 0.0136	12.87 $\pm$ 1.61	0.919	0.1479 $\pm$ 0.0149	1.47 $\pm$ 0.19	0.786
	CNN1D	0.1024 $\pm$ 0.0268	12.54 $\pm$ 1.00	0.889	0.1330 $\pm$ 0.0271	1.15 $\pm$ 0.19	0.828
	XGB	0.1035 $\pm$ 0.0305	17.72 $\pm$ 3.46	0.888	0.1510 $\pm$ 0.0113	1.86 $\pm$ 0.19	0.775
	LSTM	0.1086 $\pm$ 0.0211	16.31 $\pm$ 1.99	0.871	0.1465 $\pm$ 0.0230	1.45 $\pm$ 0.20	0.791
	LR	0.1128 $\pm$ 0.0218	23.53 $\pm$ 1.77	0.862	0.1444 $\pm$ 0.0085	1.85 $\pm$ 0.10	0.794
	KNN	0.1111 $\pm$ 0.0198	16.16 $\pm$ 1.60	0.862	0.1532 $\pm$ 0.0168	1.54 $\pm$ 0.17	0.770
Éthylène-CO	CNN1D	0.1324 $\pm$ 0.0286	42.43 $\pm$ 13.20	0.797	0.1187 $\pm$ 0.0274	1.39 $\pm$ 0.42	0.805
	CNN2D	0.1578 $\pm$ 0.0441	59.30 $\pm$ 23.09	0.706	0.1257 $\pm$ 0.0246	1.63 $\pm$ 0.41	0.783
	LSTM	0.1620 $\pm$ 0.0165	57.35 $\pm$ 12.39	0.699	0.1486 $\pm$ 0.0064	2.02 $\pm$ 0.22	0.690
	SVR	0.1945 $\pm$ 0.0418	71.68 $\pm$ 18.48	0.561	0.1179 $\pm$ 0.0169	1.49 $\pm$ 0.21	0.809
	LR	0.1975 $\pm$ 0.0302	77.15 $\pm$ 15.75	0.554	0.1330 $\pm$ 0.0155	1.88 $\pm$ 0.14	0.756
	XGB	0.1925 $\pm$ 0.0255	75.22 $\pm$ 15.60	0.577	0.1503 $\pm$ 0.0248	2.13 $\pm$ 0.24	0.687
	KNN	0.2129 $\pm$ 0.0160	66.99 $\pm$ 9.39	0.483	0.1525 $\pm$ 0.0263	1.78 $\pm$ 0.18	0.679

Selon ce tableau, nous observons que la régression linéaire multivariée apparaît moins performante que les modèles non linéaires, notamment en termes de MAE et de $NRMSE$. Ceci

confirme que cette dernière ne peut modéliser le comportement non linéaire des capteurs.

Pour la première mesure, les modèles CNN2D et SVR présentent les meilleures performances globales. Le modèle linéaire demeure compétitif pour le paramètre *NRMSE*, mais l’erreur moyenne absolue (*MAE*) très élevée au niveau des prédictions. Les modèles non linéaires permettent de réduire davantage cette dernière, ce qui montre une capacité de prédiction supérieure. Toutefois, nous observons peu de différences entre ces modèles et le modèle linéaire de base en termes de *NRMSE*, ce qui indique qu’il est relativement simple de construire un modèle qui pénalise fortement les erreurs de grande amplitude.

Au niveau de la deuxième mesure, les performances se dégradent globalement, notamment avec des valeurs de R^2 significativement plus faibles. Le modèle CNN1D semble avoir de meilleurs résultats pour cette série de mesure. Les modèles plus complexes tels que CNN et LSTM semblent encore une fois être mieux adaptés à ce type de données.

Au final, nous observons que la complexité semble dépendre de la combinaison gazeuse. Aussi, les modèles qui exploitent la dynamique temporelle des capteurs ont en général de meilleures performances. Cependant, le coût en temps de calcul et le nombre de données requis pour entraîner ceux-ci sont beaucoup plus élevés.

En raison des performances supérieures dans l’ensemble sur les deux séries de mesures, le modèle CNN-2D sera retenu pour la suite des expérimentations en termes de régression. De plus, le modèle SVR sera aussi conservé comme référence de modèle non linéaire et la régression multivariable comme référence simple.

Modèles de classification

La deuxième partie porte plutôt sur l’étude de la classification des gaz par l’entremise des modèles. La tâche de classification repose sur les trois catégories suivantes : G1 seul (Éthylène), G0 seul (CO où Methane) et un mélange (G0+G1). Les métriques afin de comparer les différents modèles sont celles qui ont été mentionnées à la section 3, soit l’exactitude (*Accuracy*), la F1-macro ainsi que le *Recall_{mix}*.

Ici, l’Accuracy (%) mesure la proportion globale de prédictions correctes. La F1-macro (%) attribue un poids égal à chaque classe, ce qui permet de tenir compte du déséquilibre entre celles-ci. Finalement, le *Recall_{mix}* (%) évalue la capacité du modèle à identifier correctement la classe de mélange, qui constitue le cas expérimental le plus complexe et pertinent.

À la suite de l’optimisation des hyperparamètres réalisée par validation croisée interne à 2 plis, les performances moyennes suivantes ont été obtenues sur les 3 plis externes. (Voir Tableau 4.3).

TABLEAU 4.3 Comparaison des performances en classification 3 classes (moyenne $\pm$ écart-type, 3 plis externes). Les classes sont : G1 seul (éthylène), G0 seul (méthane ou CO) et mélange G0+G1. Le $\text{Recall}_{\text{mix}}$ correspond au rappel de la classe de mélange (G0+G1).

Mesure	Modèle	Acc. (%)	F1-macro (%)	$\text{Recall}_{\text{mix}}$ (%)
Éthylène–Méthane	CNN2D	94.98 ± 0.76	94.60 ± 1.15	89.78 ± 4.28
	CNN1D	94.48 ± 0.38	94.23 ± 0.41	92.02 ± 0.59
	LSTM	92.80 ± 0.25	92.38 ± 0.32	89.16 ± 2.10
	KNN	92.09 ± 1.53	91.43 ± 1.01	86.47 ± 2.42
	SVC	89.89 ± 2.28	89.29 ± 2.67	83.63 ± 6.43
	XGB	89.12 ± 1.70	88.56 ± 2.18	84.61 ± 7.39
	LogReg	86.32 ± 2.63	85.99 ± 2.39	86.34 ± 4.06
Éthylène–CO	CNN2D	93.79 ± 2.08	92.73 ± 2.93	92.87 ± 0.80
	LSTM	91.18 ± 2.57	89.62 ± 4.14	90.74 ± 0.42
	KNN	85.25 ± 3.96	81.24 ± 6.26	83.96 ± 5.05
	LogReg	84.75 ± 3.24	83.04 ± 3.90	78.12 ± 9.74
	CNN1D	84.28 ± 8.34	82.49 ± 7.77	72.94 ± 15.11
	SVC	83.14 ± 3.61	80.84 ± 3.02	80.46 ± 8.80
	XGB	82.65 ± 2.36	79.51 ± 3.75	78.29 ± 10.96

Nous observons des performances qui sont élevées dans l'ensemble pour tous les modèles. Cependant, les modèles à structure temporelle sont plus dominants pour les deux mesures.

Dans le cas de la mesure d'éthylène-méthane, les architectures à convolution obtiennent les meilleurs résultats avec $94.98 \pm 0.76\%$ d'exactitude moyenne pour CNN2D et $94.48 \pm 0.38\%$ d'exactitude moyenne pour CNN1D. Ceux-ci ont d'ailleurs les résultats *F1-macro* et de $\text{Recall}_{\text{mix}}$ les plus élevés. Ceci signifie donc que leur capacité supérieure de classification s'étend sur les 3 classes du problème.

Quant à la seconde mesure d'éthylène-CO, le CNN2D conserve encore sa performance supérieure avec une diminution considérable pour le CNN1D, notamment au niveau du paramètre $\text{Recall}_{\text{mix}}$ qui suggère que celui-ci éprouvait beaucoup de difficulté à détecter la présence de mélange.

Nous observons par la suite pour les deux mesures que les modèles plus simples (KNN, SVC, XGBoost) suivent de près les performances des réseaux convolutifs. Leur limitation semble être au niveau de la classification de la classe de mélange.

Le modèle LSTM quant à lui demeure compétitif pour les deux mesures, mais s'avère légèrement inférieur au réseau convolutif.

Finalement, les écarts-types des modèles démontrent que ceux-ci arrivent à bien généraliser

dans l'ensemble et qu'ils possèdent une bonne stabilité à travers les plis externes. Dès lors, pour les mesures expérimentales, le modèle CNN2D sera retenu comme modèle principal en raison de ses performances alors que SVC et le modèle linéaire seront utilisés à titre de comparaison pour quantifier les gains à l'aide de l'apprentissage profond.

4.3 Cavités Fabry-Pérot

Dans cette section, il sera question d'évaluer les modèles sélectionnés dans la phase précédente sur les capteurs Fabry-Pérot. La sélection des modèles à l'aide du jeu de données de référence, plutôt qu'à partir des réponses des cavités Fabry-Pérot encore en développement, permet d'identifier plus clairement les approches les plus pertinentes et d'obtenir une meilleure indication quant aux types de modèle à utiliser.

4.3.1 Validation expérimentale des capteurs Fabry-Pérot fonctionnalisés par polymères

Afin d'évaluer la performance des cavités Fabry-Pérot dans un environnement exposé à différents COV (IPA, Méthanol, Acétone et Toluène) une banque de 10 polymères a été sélectionnée pour évaluation expérimentale à l'aide de la littérature : PEG2000, PHMB, PEVA, PDMS, PEO, PVA, PVP, PU, PMMA et SU-8. Ceux-ci ont des propriétés physico-chimiques très différentes les uns des autres permettant d'explorer les différents mécanismes d'adsorption des COV et d'évaluer ceux qui sont idéals pour ce type de capteurs.

Critères d'évaluations des polymères

Afin de sélectionner les différents polymères à utiliser pour cette étude, une analyse préliminaire de ceux-ci en fonction de plusieurs critères a été mise en place :

- Gaz principalement détecté
- Limite de détection (LOD)
- Sélectivité
- Stabilité du signal
- Sensibilité analytique
- Sensibilité croisée
- Réversibilité

L'ensemble des tests ont été réalisés dans des conditions expérimentales contrôlées et semblables pour tous les polymères tel que discuté dans le chapitre 3

4.3.2 Performances des polymères

À la suite des courtes séries de mesures réalisées avec les polymères au sein de la chambre de gaz en présence des 4 COV de cette étude, des résultats qualitatifs ont été obtenus (voir section D en Annexe). Les polymères PEG2000, PHMB, PEVA, PVA et PEO n'ont pas présenté de signal exploitable pour les quatre COV étudiés. Chaque matériau a été évalué sur un minimum de deux capteurs distincts afin d'assurer la répétabilité expérimentale. L'absence de réponse ne permet toutefois pas de conclure à une insensibilité de ces polymères face aux COVs, mais pourrait être liée à des limitations du procédé de fabrication. Les quatres meilleurs polymères suivant ont été sélectionnées (voir Tableau 4.4).

TABLEAU 4.4 Performances des polymères sélectionnés pour la détection des quatre COV.

Polymère	Sensibilité ($\text{pm} \cdot \text{sccm}^{-1}$)	R^2	LOD (sccm)	$\text{SNR}_{\max}$	Rép. asc. (%)	Rép. desc. (%)
Acetone						
PU	38.762	0.985	2.432	111.73	6.88	7.21
PMMA	0.864	0.919	7.673	36.80	11.43	18.38
PDMS	0.120	0.524	37.262	8.88	33.27	40.55
PVP	-0.891	0.540	7.908	-36.66	79.29	54.04
IPA						
PU	0.876	0.812	1.576	165.64	21.21	9.43
PMMA	0.607	0.931	2.575	99.18	24.36	18.44
PDMS	0.110	0.567	27.364	7.99	35.60	48.88
PVP	-2.915	0.090	1.787	160.97	76.77	26.02
Toluene						
PU	1.629	0.734	1.712	203.88	11.65	11.98
PMMA	0.681	0.818	3.451	56.44	22.69	39.05
PDMS	0.342	0.241	9.886	46.82	20.30	16.91
PVP	0.468	0.049	16.854	2.10	1355.74	1332.50
Methanol						
PU	0.584	0.886	7.649	35.31	9.44	16.58
PMMA	0.125	0.834	11.738	28.53	14.67	25.94
PDMS	0.006	0.001	509.524	-6.63	101.30	109.44
PVP	-0.600	0.034	20.376	-3.58	551.24	363.28

Il est important de noter que l'unité de mesure des gaz ici est en fonction du débit, c'est à dire en SCCM plutôt qu'en PPM.

Les paramètres retenus pour l'étude des polymères sont la sensibilité ($\text{pm} \cdot \text{sccm}^{-1}$) aux gaz, le coefficient de détermination R^2 , qui caractérise la linéarité du décalage spectral du polymère en régime quasi-statique en fonction du débit, ainsi que la limite de détection (LOD, en sccm), définie comme trois fois l'écart-type du bruit à concentration nulle divisé par la sensibilité du capteur. Le rapport signal sur bruit maximal ($\text{SNR}_{\max}$) correspond au rapport entre l'amplitude maximale du signal (différence entre le niveau maximal et le niveau de base) et le bruit de fond. Finalement, les répétabilités ascendante et descendante (%) sont définies comme l'écart-type des réponses mesurées sur les cycles d'absorption et de désorption pour un même niveau de concentration, normalisé par l'amplitude maximale moins l'amplitude minimale du capteur.

4.3.3 Modèle de régression

À l'aide des 4 polymères sélectionnés suite à l'analyse des performances des polymères, une étude des modèles de régression sur les capteurs Fabry-Pérot a été réalisée.

Étant donné qu'un seul gaz a été mesuré à la fois sur une période de 55 h, un modèle distinct est utilisé pour chaque mesure. L'utilisation d'un modèle à multiple sortie en joignant les mesures de gaz distincts entraînerait systématiquement la prédiction de la classe nulle pour les gaz absents, ce qui introduirait un biais dans l'évaluation des performances du modèle.

De plus, en raison de la distribution non uniforme des régimes des séries de mesure, c'est à dire les premiers 80% en régime quasi statique et les derniers 20% en régime dynamique/transitoire, les données ont été segmentées selon ces 2 parties. Cette distribution n'était pas intentionnelle et résulte plutôt d'une hypothèse initiale concernant la structure des mesures. Les modèles ont donc été évalués séparément sur chacun de ces régimes (voir Tableau 4.5).

Pour l'entraînement du modèle CNN2D, une étude des résultats de *NRMSE* en fonction du nombre de points dans la série temporelle a été effectuée. Le nombre optimal de points déterminés à cet effet était de 10, soit 2 minutes 30 d'acquisition. Ceci représente environ 25% du temps d'absorption ou de désorption d'un polymère suite à un changement important de concentrations en gaz.

Selon ce tableau, nous apercevons que pour la partie quasi statique, le modèle de régression linéaire multivarié est très compétitif avec les modèles non linéaires et obtient en général de meilleurs résultats. Ceci est cohérent avec la théorie puisque la réponse des cavités Fabry-Pérot dans ce régime est approximativement linéaire. Il est donc normal qu'un modèle plus adapté à cet effet obtienne de meilleurs résultats.

Cependant, si on analyse le régime transitoire, on observe rapidement la dégradation des performances du modèle linéaire. Les modèles SVR et CNN2D obtiennent alors de meilleurs résultats pour la majorité des gaz. Ce résultat est cohérent avec la théorie de diffusion des polymères (*two-stage diffusion*) qui démontre que la phase transitoire du polymère est non linéaire.

On remarque aussi que les performances varient en fonction des gaz. Pour l'IPA et le toluène, le modèle CNN2D obtient les meilleures performances alors que SVR présente de meilleurs résultats pour l'acétone et le méthanol. Ceci pourrait être liée aux subtilités de différentes interactions des gaz et polymère et leurs effets sur leurs régimes transitoires et la capacité à ces modèles de bien les prédire.

Pour l'ensemble des mesures, les résultats associés au toluène présentent globalement les

TABLEAU 4.5 Performance des modèles (moyenne $\pm$ écart-type) selon deux régimes d'entrée : Quasi statique et Transitoire. Validation croisée imbriquée par bloc (3 plis externes). Les modèles sont triés (par gaz) du plus faible NRMSE au plus élevé.

Régime	Gaz	Modèle	NRMSE	MAE	R^2
Quasi statique	Acetone	LR	0.074 ± 0.002	5.17 ± 0.26	0.933 ± 0.003
		CNN2D	0.097 ± 0.021	7.09 ± 2.04	0.886 ± 0.047
		SVR (RBF)	0.105 ± 0.039	6.31 ± 1.69	0.854 ± 0.083
	IPA	CNN2D	0.133 ± 0.027	8.96 ± 2.09	0.791 ± 0.075
		LR	0.137 ± 0.028	9.78 ± 1.74	0.767 ± 0.083
		SVR (RBF)	0.143 ± 0.027	9.77 ± 2.05	0.718 ± 0.130
	Toluene	LR	0.146 ± 0.013	11.98 ± 0.92	0.740 ± 0.022
		SVR (RBF)	0.165 ± 0.061	11.57 ± 4.97	0.635 ± 0.210
		CNN2D	0.206 ± 0.045	14.71 ± 3.67	0.492 ± 0.214
	Methanol	LR	0.106 ± 0.007	8.01 ± 0.49	0.864 ± 0.008
		SVR (RBF)	0.106 ± 0.009	7.80 ± 0.88	0.851 ± 0.039
		CNN2D	0.154 ± 0.020	10.59 ± 0.11	0.716 ± 0.097
Transitoire	Acetone	SVR (RBF)	0.156 ± 0.015	9.34 ± 1.58	0.687 ± 0.078
		CNN2D	0.181 ± 0.021	12.41 ± 1.57	0.704 ± 0.065
		LR	0.286 ± 0.014	21.44 ± 0.51	0.254 ± 0.105
	IPA	CNN2D	0.148 ± 0.013	10.70 ± 0.84	0.803 ± 0.028
		SVR (RBF)	0.155 ± 0.007	10.88 ± 0.40	0.689 ± 0.063
		LR	0.206 ± 0.009	14.50 ± 0.52	0.615 ± 0.025
	Toluene	CNN2D	0.177 ± 0.019	12.58 ± 1.64	0.713 ± 0.062
		SVR (RBF)	0.213 ± 0.022	15.58 ± 1.88	0.409 ± 0.187
		LR	0.228 ± 0.032	17.41 ± 2.18	0.517 ± 0.141
	Methanol	SVR (RBF)	0.173 ± 0.016	11.95 ± 0.77	0.612 ± 0.113
		CNN2D	0.229 ± 0.018	17.10 ± 2.15	0.531 ± 0.073
		LR	0.303 ± 0.012	23.99 ± 1.50	0.168 ± 0.074

performances les plus faibles. Selon le Tableau 4.4, le toluène est le gaz pour lequel la majorité des polymères ont un haut SNR et le plus faible coefficient de détermination (R^2). Ceci permettrait donc d'expliquer cet écart au niveau des résultats de régression.

Pour l'ensemble des prédictions, on observe un faible écart-type entre les plis de validation. Ceci démontre une bonne capacité de généralisation des modèles et indique que les fuites de données ont été efficacement limitées. De plus, les valeurs de *NRMSE* obtenues demeurent en tout temps supérieures à la répétabilité mesurée des réponses ascendantes et descendantes des polymères, ce qui indique que l'erreur de prédiction est possiblement limitée par la variabilité

intrinsèque du capteur.

En somme, les résultats montrent que les modèles linéaires sont particulièrement adaptés au régime quasi statique, tandis que les modèles non linéaires permettent de mieux représenter les comportements observés dans la phase transitoire. Néanmoins, les modèles non linéaires conservent également de bonnes performances pour la prédiction du régime quasi statique. Des modèles tels que le réseau de neurones CNN2D peuvent donc être utilisés pour modéliser les comportements linéaires. Toutefois, l'estimation de leurs paramètres repose sur des procédures d'optimisation susceptibles de converger vers des minima locaux [78], ce qui peut conduire à des paramètres ne correspondant pas toujours au modèle le plus approprié [43]. Malgré cette limitation, l'utilisation d'un modèle non linéaire pour l'ensemble des régimes apparaît comme une approche plus robuste.

Cette hypothèse a donc été testée à l'aide d'une validation croisée imbriquée combinée à une segmentation temporelle par blocs stratifiés et pondérée. Ceci permet de respecter la structure temporelle des données et de compenser la distribution non uniforme des cas quasi statique et transitoire dans les mesures. La segmentation temporelle par blocs permet donc de retrouver le même pourcentage de cas transitoire dans chacun des plis externe de sorte à bien généraliser. De plus, la pondération 4 pour 1 lors de l'apprentissage des régimes permet de ne pas avantager l'apprentissage du régime quasi statique qui s'avère à représenter environ 80% des données. Cette valeur a été déterminée avec le ratio inverse de la distribution.

Une fois ce traitement terminé, il a été possible de tester les performances des modèles entraînés sur les 2 régimes simultanément (voir Tableau 4.6).

Il est important de noter que la valeur de R^2 correspond au coefficient de détermination calculé sur les données de test des plis externes de la validation croisée. Cette métrique n'est pas pondérée en fonction de la distribution des cas quasi statiques et transitoires et représente donc une mesure non équilibrée de la performance du modèle.

Selon ces résultats, on constate que les modèles SVR et LR éprouvent d'avantages de difficultés à bien généraliser les différents régimes à l'aide d'un seul et unique modèle. On observe un écart important entre les performances des métriques $NRMSE_{qs}$ et $NRMSE_{tr}$, ce qui met en évidence une préférence marquée pour le régime quasi statique dans le cas de ces modèles, et ce malgré la pondération appliquée.

À l'inverse, le modèle CNN2D estime bien les régimes présents selon les données et parvient ainsi à prédire adéquatement les concentrations. Cette meilleure performance est possiblement attribuable à l'exploitation de l'information temporelle supplémentaire, qui lui permet de mieux distinguer les signatures associées aux différents régimes et lui permet alors d'améliorer

TABLEAU 4.6 Comparaison des performances pour les modèles : LR, SVR à noyau RBF et le réseau CNN2D pour la prédiction de la concentration des gaz. Les résultats correspondent à la moyenne et à l'écart-type obtenus sur les plis externes de la validation croisée imbriquée. Les modèles sont triés selon le NRMSE équilibré. Le NRMSE_{eq} correspond au NRMSE équilibré, défini comme la moyenne du NRMSE en régime quasi statique (NRMSE_{qs}) et du NRMSE en régime transitoire (NRMSE_{tr}).

Gaz	Modèle	NRMSE_{eq}	NRMSE_{qs}	NRMSE_{tr}	MAE_{eq}	R^2
Acétone	CNN2D	0.099 ± 0.009	0.057 ± 0.012	0.142 ± 0.018	6.541 ± 0.643	0.924 ± 0.014
	SVR	0.194 ± 0.023	0.081 ± 0.004	0.307 ± 0.051	14.864 ± 1.915	0.720 ± 0.051
	LR	0.196 ± 0.021	0.099 ± 0.013	0.294 ± 0.052	15.520 ± 2.012	0.708 ± 0.035
IPA	CNN2D	0.143 ± 0.005	0.140 ± 0.004	0.146 ± 0.012	10.542 ± 0.378	0.779 ± 0.017
	SVR	0.175 ± 0.015	0.141 ± 0.013	0.209 ± 0.033	12.436 ± 1.119	0.716 ± 0.027
	LR	0.192 ± 0.017	0.153 ± 0.005	0.231 ± 0.037	14.822 ± 1.468	0.665 ± 0.016
Méthanol	CNN2D	0.194 ± 0.016	0.145 ± 0.010	0.243 ± 0.039	14.347 ± 1.251	0.675 ± 0.022
	SVR	0.216 ± 0.029	0.121 ± 0.013	0.311 ± 0.069	17.262 ± 2.827	0.639 ± 0.051
	LR	0.218 ± 0.028	0.131 ± 0.013	0.306 ± 0.065	17.595 ± 2.693	0.624 ± 0.044
Toluène	CNN2D	0.174 ± 0.009	0.136 ± 0.014	0.212 ± 0.022	12.450 ± 0.295	0.733 ± 0.023
	SVR	0.179 ± 0.017	0.136 ± 0.016	0.223 ± 0.042	11.711 ± 0.300	0.714 ± 0.040
	LR	0.236 ± 0.014	0.222 ± 0.015	0.250 ± 0.042	18.345 ± 0.920	0.406 ± 0.043

la précision des prédictions. En utilisant plusieurs points consécutifs du signal, le modèle CNN2D arrive à mieux capturer l'évolution du capteur et de mieux modéliser les phénomènes de diffusion à l'intérieur des polymères.

Au finale, on conclut que l'ajout de l'information temporelle constitue un élément clé pour modéliser à l'aide d'un modèle les régimes quasi statique et transitoire des capteurs Fabry-Pérot. Pour ce faire, il a été démontré que le modèle CNN2D obtenait les meilleures performances.

4.3.4 Modèles de classification

Cette section présente l'évaluation des différents modèles de classification pour l'identification de gaz avec les capteurs Fabry-Pérot. Tout comme pour la régression, les modèles étudiés sont CNN2D, SVC et la régression logistique.

Les modèles obtenus suite au traitement des données et à l'apprentissage machine selon la méthodologie de la section 3 sont inscrits au Tableau 4.7

Selon les résultats, le modèle de CNN2D obtient les meilleures performances globales avec une exactitude moyenne de 94.1% et un F1-score macro de 93.9%. Le modèle SVC avec noyau RBF présente également de bons résultats, bien que légèrement inférieurs au CNN2D. Le

TABLEAU 4.7 Comparaison des modèles pour la classification des gaz (Acétone, IPA, Méthanol et Toluène) à l'aide d'une validation croisée imbriquée bloquée (3 plis externes). Les valeurs sont présentées sous la forme moyenne $\pm$ écart-type.

Modèle	Accuracy	F1-score (macro)
CNN 2D	0.941 $\pm$ 0.015	0.939 $\pm$ 0.017
SVC (RBF)	0.931 $\pm$ 0.015	0.930 $\pm$ 0.014
LogReg	0.884 $\pm$ 0.024	0.883 $\pm$ 0.027

modèle de régression logistique multiclasse obtient des performances un peu plus faible que SVC. Ceci suggère que la séparation des classes de gaz n'est pas strictement linéaire.

Encore une fois, on remarque que les modèles non-linéaires sont mieux adaptés à ce type de capteurs. L'utilisation de séries temporelles offre un avantage non-négligeable au CNN2D afin de distinguer les signatures des capteurs selon chaque gaz.

La matrice de confusion de la Figure 4.7 permet de mieux comprendre l'essence des erreurs de classification.

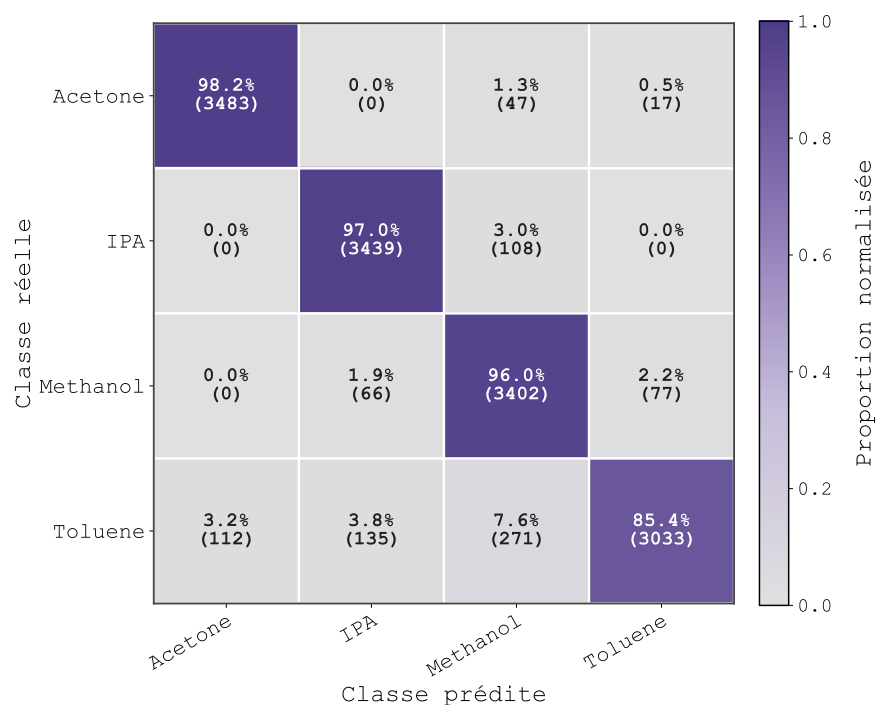


FIGURE 4.7 Matrice de confusion du meilleur modèle CNN2D pour la classification des gaz (Acétone, IPA, Méthanol et Toluène). La matrice correspond à la somme des prédictions des trois plis externes de la validation croisée imbriquée, puis les valeurs sont normalisées par ligne.

Il semblerait que les polymères auraient des signatures dynamiques semblables entre le toluène et le méthanol pour certains cas. D’ailleurs, le toluène représenterait le gaz avec la plus faible précision de classification. Selon le Tableau 4.4, le toluène est le gaz pour lequel la majorité des polymères ont le plus haut SNR et le plus faible coefficient de détermination (R^2). Ceci permettrait donc d’expliquer cet écart au niveau de la précision de classification.

4.3.5 Comparaison des performances avec la littérature

Afin d’évaluer les performances des modèles proposés, les résultats obtenus dans ce travail sont comparés à ceux rapportés dans la littérature pour différents types de système de détection de gaz qui ont été combinés à des méthodes d’apprentissage automatiques.

TABLEAU 4.8 Résumé de travaux récents utilisant l’apprentissage automatique pour la détection de COV

Référence	Type de capteur	Modèle ML	Gaz étudiés	Performance
Cho <i>et al.</i> [67]	Capteur photoactif μ LED	D-CNN	Méthanol, éthanol, acétone, NO ₂	Acc : 96.99%; MAPE : 31.99%
Mei <i>et al.</i> [68]	Capteur MOS MEMS	CNN + TL	10 COV	Acc : 98.38%; 88.65%
Li <i>et al.</i> [69]	Réseau de capteurs MOS	SVM	5 COV	Acc : 86.67%; $R^2 = 0.96$
Nayak <i>et al.</i> [70]	Capteur MOS (LiFe ₅ O ₈)	AdaBoost	Acétone, éthanol, méthanol	Acc : 98%; $R^2 = 0.84$
Wang <i>et al.</i> [71]	Capteur électrochimique à potentiel mixte	Random Forest	COV	Acc : 99%
Singh <i>et al.</i> [72]	Réseau de capteurs MOS (NiO, CuO, ZnO)	KNN / RF	Acétone, éthanol, toluène, chloroforme	Acc : >99%; $R^2 > 0.99$
Ce travail	Cavités Fabry-Pérot sur puce fonctionnalisées par polymères	CNN 2D	Acétone, IPA, méthanol, toluène	Acc : 94.1%; $R^2 = 0.68-0.92$ ($\overline{R^2} = 0.78$)

Comme le montre le Tableau 4.8, plusieurs travaux de la littérature rapportent des performances de classification très élevées, généralement supérieures à 95%. Ces résultats sont obtenus à partir de différents types de capteurs de gaz, notamment des capteurs à oxyde métallique (MOS). Les performances obtenues avec le modèle de classification CNN 2D dans ce travail demeurent compétitives par rapport à celles rapportées dans l’état de l’art. L’analyse des résultats indique toutefois que le principal facteur limitant est la classification du toluène. En l’absence de ces erreurs, la précision de classification atteindrait environ 97%.

Au niveau de la régression avec ce même modèle, les résultats obtenus sont comparables et parfois supérieurs à ceux rapportés dans la littérature uniquement dans le cas de l'acétone, avec un résultats de $R^2 = 0.92$. Pour les autres COV, les performances de prédiction des concentrations demeurent plus faibles et inférieures à celles observées dans les travaux de l'état de l'art. La valeur moyenne du coefficient de détermination obtenue dans cette étude est de $R^2 = 0.78$, alors que la plus faible valeur rapportée dans la littérature est d'environ $R^2 = 0.84$.

Ces différences peuvent s'expliquer par plusieurs facteurs. Elles résultent notamment des limitations et des problématiques associées aux capteurs utilisés, qui seront discutées dans la section suivante.

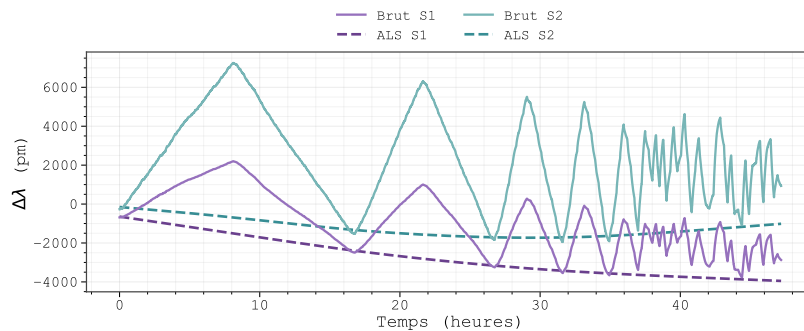
4.3.6 Limitations et problématique des capteurs

Les capteurs Fabry-Pérot ont démontré un grand potentiel pour la détection de gaz grâce à leur sélectivité, sensibilité et fonctionnement passif. Cependant, plusieurs limitations à cette technologie ont affecté la précision et la stabilité des mesures. Parmi ceux-ci, on retrouve notamment la dérive du signal au cours du temps, la dégradation des polymères et les sensibilités aux conditions environnementales tel que la température. Ces phénomènes interfèrent donc avec le traitement des données et compliquent donc l'interprétation des signaux.

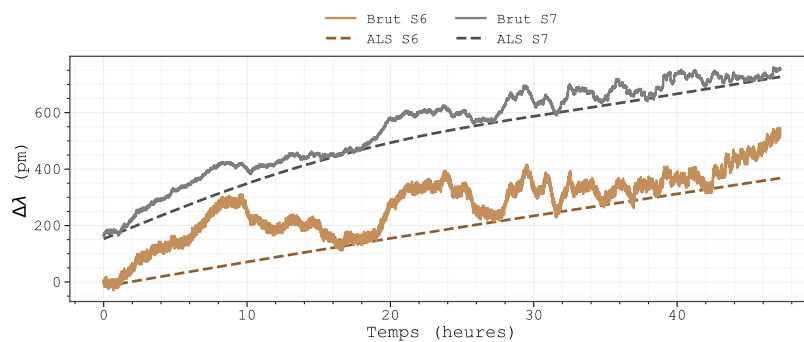
Dans cette section, ces différentes limitations seront présentées et illustrées à l'aide d'exemple issu de mesures expérimentales.

Dérive du signal

Lors de l'acquisition de données expérimentale, on observe une dérive du signal graduel (voir Figure 4.8).



(a) Signaux bruts des capteurs en polyuréthane (PU) lors d'une exposition à l'acétone ainsi que les dérives ajustées.



(b) Signaux bruts des capteurs en PMMA lors d'une exposition à l'acétone ainsi que les dérives ajustées.

FIGURE 4.8 Dérive du signal pour deux types de polymères. La dérive lente du signal est estimée à l'aide de l'algorithme *Asymmetric Least Squares* (ALS). La recette de cette mesure est celle de la Figure 3.5

Ces dérives sont plutôt lentes et influencent le niveau de base du signal au cours de la mesure. Afin de caractériser cette dérive, une estimation par l'algorithme *Asymmetric Least Squares* a été utilisé [79]. Cette variation semble être indépendante de la réponse transitoire du capteur lors de l'exposition au gaz. Nous observons que la tendance de dérive diffère selon le type de polymère, alors qu'elle demeure similaire entre capteurs du même polymère lorsqu'ils sont exposés au même gaz. Nous remarquons aussi que la tendance de la dérive n'est pas la même dépendamment du gaz auquel le polymère est exposé (voir Figure 4.9).

Ceci suggère qu'elle pourrait être liée aux propriétés du matériau, notamment à ses interactions physico-chimiques au gaz exposé et que ceci cause une dérive à long terme.

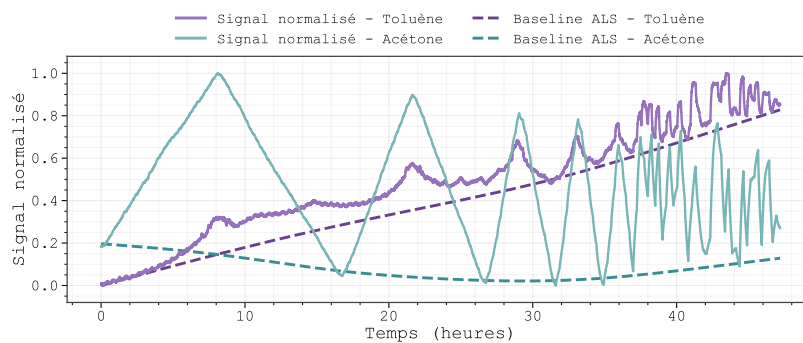


FIGURE 4.9 Comparaison du signal normalisé d'un capteur en polyuréthane (PU) lors d'expositions à l'acétone et au toluène. La dérive lente du niveau de base est estimée à l'aide de l'algorithme *Asymmetric Least Squares* (ALS).

Dégradation des polymères

Les capteurs de gaz ont généralement une évolution de leurs propriétés de détection au cours du temps. Ceci est notamment le cas pour les cavités Fabry-Pérot en raison des polymères utilisés comme matériaux sensibles. Dans les données expérimentales acquises, une dégradation de la sensibilité des polymères ainsi que de leur réponse a été observée.

En analysant les spectres de chacun des polymères après 167 jours de fabrication, la dégradation n'est pas évidente (voir Figure 4.10).

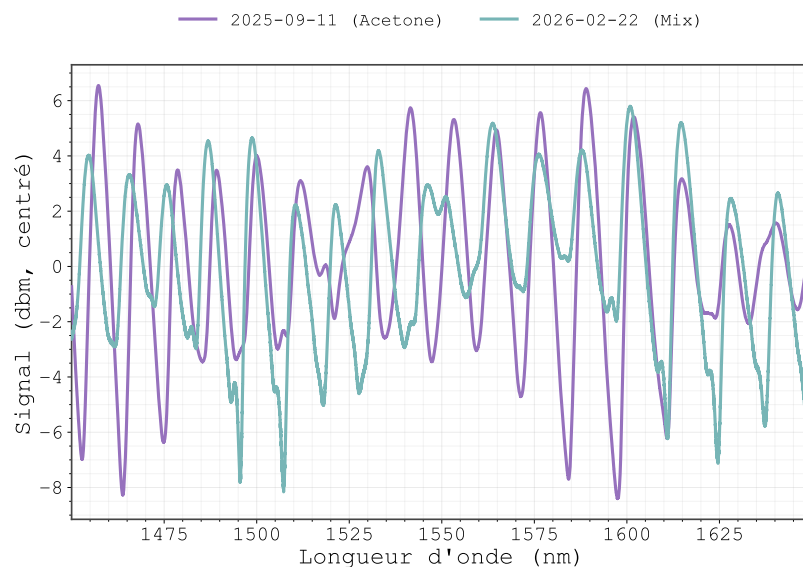


FIGURE 4.10 Exemple d'évolution du spectre d'un capteur Fabry-Pérot utilisant le PU comme polymère après 167 jours

Cependant, cette dégradation apparaît plus nettement lors de l'évaluation du déplacement spectral maximal du capteur ($\Delta\lambda$), pour le même gaz après 167 jours, une réduction de sensibilité à l'acétone pour une exposition à 100 sccm a été mesurée à environ 97.5%.

En raison de cette dégradation des polymères, des mesures en gaz mixte (Acétone/IPA) n'ont pas pu être évaluées dans le cadre de ce projet de recherche. L'objectif aurait été d'évaluer la régression et la classification des gaz avec la présence de 2 gaz simultanément.

Ce changement de sensibilité suggère un changement possible dans les propriétés d'absorption avec le temps. Ceci vient réduire considérablement la durée de vie possible des polymères.

Plusieurs mécanismes physiques et chimiques peuvent expliquer cette dégradation. En effet, les polymères utilisés peuvent subir un vieillissement au cours du temps en raison de phénomènes d'oxydation, de leur exposition à l'humidité ainsi qu'aux variations de température [80].

Dans le cas du polyuréthane (PU), plusieurs travaux rapportent que ce polymère est particulièrement à risque d'une dégradation oxydative en raison de sa structure chimique et de la présence de segments sensibles à l'oxydation [81,82].

Pour le PMMA, la dégradation semble être principalement associée à son exposition à la lumière, à des variations de température ainsi qu'à l'humidité, ces facteurs pouvant changer ses propriétés physiques et chimiques au cours du temps [83].

Concernant le PVP, son exposition aux rayonnements ultraviolets (UV) peut accélérer les processus de photo-oxydation, contribuant ainsi à la dégradation progressive du matériau [84].

Finalement, pour le PDMS, certaines études indiquent que ce polymère peut se dégrader lorsqu'il est exposé à des températures élevées [85]. Toutefois, dans le cadre de cette étude, l'origine exacte de la dégradation observée demeure inconnue, puisque les températures d'exposition ont été maintenues sous les 30 °C durant toute la durée des expériences.

Une calibration des modèles d'apprentissage automatique pourrait être effectuée afin d'essayer de maintenir les performances de prédictions au cours des dégradations.

CHAPITRE 5 CONCLUSION

5.1 Synthèse des travaux

Ce mémoire avait pour objectif d'étudier l'utilisation de méthodes d'apprentissage automatique pour la classification et la quantification de composés organiques volatils (COV) à partir d'un capteur Fabry-Pérot sur puce et fonctionnalisé par des polymères.

Dans un premier temps, les principes physiques basés sur le fonctionnement des capteurs Fabry-Pérot ont été présentés, soit les mécanismes d'interaction entre les polymères et les gaz. Il a été montré que l'absorption d'un gaz dans les polymères peut engendrer un gonflement du matériau ainsi qu'un changement de son indice de réfraction qui se traduit par un déplacement des spectres de la cavité.

Par la suite, plusieurs polymères ont été étudiés afin d'évaluer leur compatibilité avec les cavités Fabry-Pérot pour la détection de composés organiques volatils. En plus de polymères déjà étudiés dans la littérature, tels que le PDMS ou le PEVA, ce travail a permis d'identifier trois polymères supplémentaires, soit le polyvinylpyrrolidone (PVP), le polyméthacrylate de méthyle (PMMA) et le polyuréthane (PU), comme matériaux sensibles prometteurs pour ce type de capteur.

Plusieurs algorithmes d'apprentissage automatique ont été étudiés sur une série de données d'un capteur chimique de l'UCSD et sur les données expérimentales du capteur Fabry-Pérot afin de modéliser les réponses des capteurs et effectuer des tâches de classification et de quantification de gaz. Parmi les modèles évalués, la solution idéale pour ce type de capteur est le CNN2D qui a obtenu des résultats nettement supérieurs aux autres modèles dans l'ensemble.

Une comparaison avec les résultats rapportés dans la littérature a permis de situer les performances obtenues dans ce travail par rapport à l'état de l'art. La moyenne des résultats obtenus pour la régression du modèle CNN2D dans cette étude demeurent légèrement inférieurs à ceux rapportés dans la littérature, ce qui pourrait s'expliquer par les performances limitées des polymères utilisés. Il est toutefois possible d'observer que des résultats comparables, voire supérieurs, à ceux rapportés dans la littérature auraient probablement été obtenus pour la classification si les résultats associés au toluène avaient été exclus de l'analyse. En effet, ce gaz s'est avéré plus difficile à identifier pour les polymères étudiés, ce qui a influencé négativement les performances globales du modèle.

D'ailleurs, les performances obtenues pour le modèle de régression appliqué à l'acétone sont

particulièrement prometteuses. Les résultats obtenus pour ce gaz se révèlent supérieurs à ceux rapportés pour certains autres types de capteurs de gaz dans la littérature.

Dans l'ensemble, les objectifs de ce mémoire ont été atteints. De nouveaux polymères sensibles aux COV ont été identifiés, les réponses dynamiques du capteur ont été analysées, plusieurs modèles d'apprentissage automatique ont été développés et comparés, et leurs performances ont été évaluées et mises en perspective avec l'état de l'art.

Au-delà des aspects liés à l'apprentissage automatique, ce mémoire a également impliqué la microfabrication des cavités Fabry-Pérot, l'étude et la caractérisation de polymères sensibles aux COV, ainsi que le développement de montages expérimentaux optiques et du système d'acquisition des données. Les travaux réalisés comprennent aussi l'alignement des fibres optiques, les acquisitions expérimentales en laboratoire, le traitement des données ainsi que l'analyse des performances des capteurs.

Ces travaux ont également mené à la participation au développement d'un prototype de capteur de gaz destiné à des mesures en conditions réelles au Centre de recherche et de développement de Sherbrooke d'Agriculture et Agroalimentaire Canada, dans le cadre d'une étude sur les émissions de méthane provenant de vaches laitières.

5.2 Limitations de la solution proposée

Malgré les résultats encourageants obtenus dans ce travail, certaines limitations doivent être considérées.

Tout d'abord, les expériences ont été réalisées dans des conditions expérimentales contrôlées, ce qui peut différer des conditions rencontrées dans des environnements réels. Des facteurs tels que l'humidité, la température, la dérive des capteurs ou la présence d'autres gaz peuvent influencer la réponse des polymères et dégrader les performances des modèles.

De plus, le nombre de polymères utilisés dans cette étude demeure limité. L'ajout de polymères supplémentaires présentant des sensibilités différentes permettrait d'augmenter la sélectivité globale du système et d'améliorer la capacité de discrimination entre différents analytes.

Par ailleurs, les polymères étudiés présentent certaines limitations en termes de sélectivité, de sensibilité et de répétabilité ce qui a restreint la qualité des données expérimentales. Cette limitation affecte directement les performances des modèles d'apprentissage automatique, puisque ceux-ci sont directement dépendants de la qualité des données expérimentales disponibles.

Enfin, en raison de la dégradation des polymères et du manque de temps pour refaire des capteurs, les modèles ont été entraînés pour la régression dans des situations expérimentales contrôlées ou seulement un analyte était présent. Ceux-ci n’ont toutefois pas été entraînés ni évalués pour effectuer la régression en présence de plus d’un gaz.

5.3 Améliorations futures

Plusieurs pistes de recherche peuvent être envisagées afin d’améliorer les performances des capteurs et des méthodes d’analyse proposées dans ce projet

La première direction consisterait à poursuivre le développement des capteurs Fabry-Pérot en identifiant de nouveaux polymères présentant une meilleure sélectivité, sensibilité et une plus grande stabilité dans le temps. L’étude de polymères caractérisés par une faible dérive, une excellente répétabilité, une bonne sensibilité et une sélectivité, ainsi qu’une résistance à l’humidité, la température et une dégradation lente pourrait permettre d’améliorer significativement la fiabilité et la durée de vie des capteurs. Il s’agit de l’amélioration future la plus importante et critique au fonctionnement de ce capteur.

De plus, l’acquisition de données expérimentales plus étendue, incluant une plus grande variété de concentrations et de conditions environnementales (température, humidité), permettrait d’améliorer la robustesse des modèles d’apprentissage automatique et leur capacité de généralisation.

Une autre amélioration possible serait l’exploration d’architectures d’apprentissage automatique plus avancées, telles que la combinaison de modèles spécialisés.

Finalement, dans ce travail, la dérive lente du signal des capteurs est estimée à l’aide de l’algorithme *Asymmetric Least Squares* (ALS). Cette approche permet de corriger efficacement les variations lentes du niveau de base dans les signaux expérimentaux. Toutefois, cette méthode demeure une approximation et peut ne pas être suffisante dans certaines situations, notamment lorsque la dérive du capteur présente des comportements plus complexes. Des méthodes de correction plus avancées ou adaptatives pourraient donc être explorées dans des travaux futurs afin d’améliorer les performances.

RÉFÉRENCES

- [1] I. Manzini, D. Schild et C. Di Natale, “Principles of odor coding in vertebrates and artificial chemosensory systems,” *Physiological Reviews*, vol. 102, n°. 1, p. 61–154, 2022, pMID : 34254835. [En ligne]. Disponible : <https://doi.org/10.1152/physrev.00036.2020>
- [2] S. L. Pashkovski, G. Iurilli, D. Brann, D. Chicharro, K. Drummey, K. M. Franks, S. Panzeri et S. R. Datta, “Structure and flexibility in cortical representations of odour space,” *Nature*, vol. 583, n°. 7815, p. 253–258, 2020. [En ligne]. Disponible : <https://doi.org/10.1038/s41586-020-2451-1>
- [3] V. Mishra, Rashmi et Sukriti, “Optical gas sensors,” dans *Metal-Oxide Gas Sensors*, S. Dhara et G. Dutta, édit. London : IntechOpen, 2022, ch. 2. [En ligne]. Disponible : <https://doi.org/10.5772/intechopen.108971>
- [4] P. Jubinville, R. Guertin, L. Erbilgin, W. Skene et Y.-A. Peter, “Selective in-plane fabry-pérot gas sensor functionalized with polymer,” dans *2017 International Conference on Optical MEMS and Nanophotonics (OMN)*, 2017, p. 1–2.
- [5] S. Semancik, R. Cavicchi, M. Wheeler, J. Tiffany, G. Poirier, R. Walton, J. Suehle, B. Panchapakesan et D. DeVoe, “Microhotplate platforms for chemical sensor research,” n°. V. 77, 2001-06-01 2001.
- [6] Y. He et M. Jiao, “A mini-review on metal oxide semiconductor gas sensors for carbon monoxide detection at room temperature,” *Chemosensors*, vol. 12, n°. 4, 2024. [En ligne]. Disponible : <https://doi.org/10.3390/chemosensors12040055>
- [7] A. Staerz, U. Weimar et N. Barsan, “Current state of knowledge on the metal oxide based gas sensing mechanism,” *Sensors and Actuators B : Chemical*, vol. 358, p. 131531, 2022. [En ligne]. Disponible : <https://www.sciencedirect.com/science/article/pii/S0925400522001733>
- [8] L. Khajavizadeh et M. Andersson, “Mosfet-based gas sensors for process industry iot applications,” dans *2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, 2022, p. 1–5.
- [9] H. Shin, J. Park, D. Kim, K. Choi, M.-K. Park, J. Hwang, S. Lee, Y.-C. Joo, G. Jung et J.-H. Lee, “Optimization of co-sputtered zinc indium tin oxide-based mosfet-type sensor for effective no2 gas detection,” *Sensors and Actuators B : Chemical*, vol. 428, p. 137262, 2025. [En ligne]. Disponible : <https://www.sciencedirect.com/science/article/pii/S0925400525000371>

- [10] F. Fauzi, A. Rianjanu, I. Santoso et K. Triyana, “Gas and humidity sensing with quartz crystal microbalance (qcm) coated with graphene-based materials – a mini review,” *Sensors and Actuators A : Physical*, vol. 330, p. 112837, 2021. [En ligne]. Disponible : <https://www.sciencedirect.com/science/article/pii/S0924424721003009>
- [11] S. G. Ramaraj, A. Alrebh, D. Elamaram, H. Zhou, K. Huang, M. Almansoori, H. Yamahara et H. Tabata, “Surface acoustic wave gas sensors : Recent developments and their role in sensing technology,” *Materials Science and Engineering : B*, vol. 317, p. 118157, 2025. [En ligne]. Disponible : <https://www.sciencedirect.com/science/article/pii/S0921510725001801>
- [12] H. Liu, Y. Shi et T. Wang, “Design of a six-gas ndir gas sensor using an integrated optical gas chamber,” *Opt. Express*, vol. 28, n^o. 8, p. 11 451–11 462, Apr 2020. [En ligne]. Disponible : <https://opg.optica.org/oe/abstract.cfm?URI=oe-28-8-11451>
- [13] M. Xu, Y. Xu, J. Tao, L. Wen, C. Zheng, Z. Yu et S. He, “Development of a compact ndir co2 gas sensor for harsh environments,” *Infrared Physics Technology*, vol. 136, p. 105035, 2024. [En ligne]. Disponible : <https://www.sciencedirect.com/science/article/pii/S1350449523004930>
- [14] Y. Liu, Y. Qi, Y. Cai, X. Bao et S. Gao, “Recent advances in optical fiber-based gas sensors utilizing light-induced acoustic/elastic techniques,” *Photoacoustics*, vol. 43, p. 100715, 2025. [En ligne]. Disponible : <https://www.sciencedirect.com/science/article/pii/S2213597925000382>
- [15] L. Guo, H. Bao, F. Chen, P. Zhao, S. Jiang, H. L. Ho et W. Jin, “Ultra-compact optical fiber gas sensor with high sensitivity, fast response and large dynamic range,” *Journal of Lightwave Technology*, vol. 42, n^o. 7, p. 2617–2624, 2024.
- [16] R. Guertin, M.-A. Bianki, C. Lemieux-Leduc et Y.-A. Peter, “Multi-gas detection using fabry-perot interferometers on silicon chip,” *Sensors and Actuators B : Chemical*, vol. 335, p. 129655, 2021. [En ligne]. Disponible : <https://www.sciencedirect.com/science/article/pii/S0925400521002239>
- [17] R. Guertin, “Détection multi-gas avec interféromètre de fabry-perot sur puce de silicium,” Mémoire de maîtrise, Polytechnique Montréal, août 2019. [En ligne]. Disponible : <https://publications.polymtl.ca/3993/>
- [18] G. Hernandez, *Fabry-Perot Interferometers*, ser. Cambridge Studies in Modern Optics. Cambridge University Press, 1988. [En ligne]. Disponible : <https://books.google.ca/books?id=og7uKhM01K8C>
- [19] A. Perot et C. Fabry, “On the application of interference phenomena to the solution of various problems of spectroscopy and metrology,” *The Astrophysical Journal*,

- vol. 9, p. 87, 1899. [En ligne]. Disponible : <https://api.semanticscholar.org/CorpusID:121123138>
- [20] Y. Rao, Z. Ran et Y. Gong, *Fiber-Optic Fabry-Perot Sensors : An Introduction*, ser. Series in Fiber Optic Sensors. CRC Press, 2017. [En ligne]. Disponible : https://books.google.ca/books?id=vta_DgAAQBAJ
- [21] D. Romanini *et al.*, “Cavity-enhanced spectroscopy and sensing,” dans *Cavity-Enhanced Spectroscopy and Sensing*, G. Gagliardi et H. P. Loock, édit. Springer, 2014, vol. 179, ch. 1, p. 1–60.
- [22] Thorlabs, “Fabry-perot interferometer tutorial,” <https://www.thorlabs.com/fabry-perot-interferometer-tutorial?tabName=Fabry-Perot%2520Tutorial>, 2025, tutorial on principles and operation of Fabry-Pérot interferometers.
- [23] H. Namazi, “Polymers in our daily life,” *BioImpacts*, vol. 7, n°. 2, p. 73–74, 2017, epub 2017 Jun 16.
- [24] “Polymer,” <https://en.wikipedia.org/wiki/Polymer>, Wikipedia, n.d., accessed : 2026-01-26.
- [25] A. P. C. Ribeiro, M. O. Martins, A. O. Figueiras et L. M. D. R. S. Martins, *Introduction to Polymerization and Depolymerization*, ch. 1, p. 1–23. [En ligne]. Disponible : <https://pubs.acs.org/doi/abs/10.1021/bk-2025-1498.ch001>
- [26] R. J. Young et P. A. Lovell, *Introduction to Polymers*, 3^e éd. Boca Raton, FL, USA : CRC Press, 2011.
- [27] L. Jingcheng, V. S. Reddy, W. A. D. M. Jayathilaka, A. Chinnappan, S. Ramakrishna et R. Ghosh, “Intelligent polymers, fibers and applications,” *Polymers*, vol. 13, n°. 9, 2021. [En ligne]. Disponible : <https://doi.org/10.3390/polym13091427>
- [28] U. Altenberend, A. Oprea, N. Bârsan et U. Weimar, “Contribution of polymeric swelling to the overall response of capacitive gas sensors,” *Analytical and Bioanalytical Chemistry*, vol. 405, p. 6445–6452, 2013.
- [29] X. Liu, S. Cheng, H. Liu, S. Hu, D. Zhang et H. Ning, “A survey on gas sensing technology,” *Sensors*, vol. 12, n°. 7, p. 9635–9665, 2012. [En ligne]. Disponible : <https://doi.org/10.3390/s120709635>
- [30] Y. C. Wong, B. C. Ang, A. S. M. A. Haseeb, A. A. Baharuddin et Y. H. Wong, “Review—conducting polymers as chemiresistive gas sensing materials : A review,” *Journal of The Electrochemical Society*, vol. 167, n°. 3, p. 037503, sep 2019. [En ligne]. Disponible : <https://doi.org/10.1149/2.0032003JES>
- [31] Z. Gao, G. Song, X. Zhang, Q. Li, S. Yang, T. Wang, Y. Li, L. Zhang, L. Guo et Y. Fu, “A facile pdms coating approach to room-temperature gas

- sensors with high humidity resistance and long-term stability,” *Sensors and Actuators B : Chemical*, vol. 325, p. 128810, 2020. [En ligne]. Disponible : <https://www.sciencedirect.com/science/article/pii/S0925400520311576>
- [32] S. Hoffmann, M. Henfling et S. Trupp, “Investigation of polymers as matrix materials for application in colorimetric gas sensors for the detection of ammonia,” *Sensors*, vol. 25, n°. 9, 2025. [En ligne]. Disponible : <https://doi.org/10.3390/s25092829>
- [33] N. L. Kazanskiy, M. A. Butt et S. N. Khonina, “Carbon dioxide gas sensor based on polyhexamethylene biguanide polymer deposited on silicon nano-cylinders metasurface,” *Sensors*, vol. 21, n°. 2, 2021. [En ligne]. Disponible : <https://doi.org/10.3390/s21020378>
- [34] A. Shuba, T. Kuchmenko et R. Umarkhanov, “Piezoelectric gas sensors with polycomposite coatings in biomedical application,” *Sensors*, vol. 22, n°. 21, 2022. [En ligne]. Disponible : <https://doi.org/10.3390/s22218529>
- [35] R. St-Gelais, G. Mackey, J. Saunders, J. Zhou, A. Leblanc-Hotte, A. Poulin, J. A. Barnes, H.-P. Looock, R. S. Brown et Y.-A. Peter, “Gas sensing using polymer-functionalized deformable fabry-perot interferometers,” *Sensors and Actuators B : Chemical*, vol. 182, p. 45–52, 2013. [En ligne]. Disponible : <https://www.sciencedirect.com/science/article/pii/S0925400513001287>
- [36] R. St-Gelais, A. Poulin et Y.-A. Peter, “Advances in modeling, design, and fabrication of deep-etched multilayer resonators,” *Journal of Lightwave Technology*, vol. 30, n°. 12, p. 1900–1908, 2012.
- [37] T. M. Aminabhavi, “Use of mixing rules in the analysis of data for binary liquid mixtures,” *Journal of Chemical and Engineering Data*, vol. 29, n°. 1, p. 54–55, 1984.
- [38] D. R. Paul et Y. P. Yampol’skii, édit., *Polymeric Gas Separation Membranes*. Boca Raton, FL : CRC Press, 1994.
- [39] J. Crank, *The Mathematics of Diffusion*, ser. Oxford science publications. Clarendon Press, 1979. [En ligne]. Disponible : <https://books.google.ca/books?id=eHANhZwVouYC>
- [40] J. E. Saunders, H. Chen, C. Brauer, M. Clayton, W. Chen, J. A. Barnes et H.-P. Looock, “Quantitative diffusion and swelling kinetic measurements using large-angle interferometric refractometry,” *Soft Matter*, vol. 11, p. 8746–8757, 2015. [En ligne]. Disponible : <http://dx.doi.org/10.1039/C5SM02170G>
- [41] J. Saunders, H. Chen, C. Brauer, M. Clayton et H.-P. Looock, “Two distinct mechanisms upon absorption of volatile organic compounds into siloxane polymers,” *Soft Matter*, vol. 14, 02 2018.

- [42] R. Guertin, M.-A. Bianki, O. Figueroa, C. Amoah, W. Skene et Y.-A. Peter, “Polymer-functionalized on-chip fabry-perot interferometer for co₂ and ch₄ sensing,” dans *2024 International Conference on Optical MEMS and Nanophotonics (OMN)*, 2024, p. 1–2.
- [43] R. Lakhmi, M. Fischer, Q. Darves-Blanc, R. Alammouz, M. Rieu et J.-P. Viricelle, “Linear and non-linear modelling methods for a gas sensor array developed for process control applications,” *Sensors*, vol. 24, n^o. 11, 2024. [En ligne]. Disponible : <https://doi.org/10.3390/s24113499>
- [44] X. Zhang, D. Ma, Y.-Q. Zhao, H. Shi, L. Tao et F. Niu, “Hydrophobic cofs promote nh₃ sensing response of pyridine-based polymers under high humidity,” *ACS Applied Polymer Materials*, vol. 7, n^o. 9, p. 5556–5564, 2025. [En ligne]. Disponible : <https://doi.org/10.1021/acsapm.5c00440>
- [45] K. Razzaq et M. Shah, “Machine learning and deep learning paradigms : From techniques to practical applications and research frontiers,” *Computers*, vol. 14, n^o. 3, 2025. [En ligne]. Disponible : <https://doi.org/10.3390/computers14030093>
- [46] I. H. Sarker, “Machine learning : Algorithms, real-world applications and research directions,” *SN Computer Science*, vol. 2, n^o. 3, p. 160, 2021.
- [47] Y. Yu, X. Cao, C. Li, M. Zhou, T. Liu, J. Liu et L. Zhang, “A review of machine learning-assisted gas sensor arrays in medical diagnosis,” *Biosensors*, vol. 15, n^o. 8, 2025. [En ligne]. Disponible : <https://doi.org/10.3390/bios15080548>
- [48] X. Li, J. Guo, W. Xu et J. Cao, “Optimization of the mixed gas detection method based on neural network algorithm,” *ACS Sensors*, vol. 8, n^o. 2, p. 822–828, 2023, pMID : 36701636. [En ligne]. Disponible : <https://doi.org/10.1021/acssensors.2c02450>
- [49] T. M. Hope, “Chapter 4 - linear regression,” dans *Machine Learning*, A. Mechelli et S. Vieira, édit. Academic Press, 2020, p. 67–81. [En ligne]. Disponible : <https://www.sciencedirect.com/science/article/pii/B9780128157398000043>
- [50] R. M.G, S. Rajanalli, S. Endoli, M. Magi et R. Nagarajan, “Machine learning algorithms for classification of gas sensor array dataset,” *Journal of University of Shanghai for Science and Technology*, vol. 23, p. 721–728, 06 2021.
- [51] L. Breiman, J. H. Friedman, R. A. Olshen et C. J. Stone, *Classification and Regression Trees*. New York : Chapman and Hall/CRC, 1984.
- [52] T. Hastie, R. Tibshirani et J. Friedman, *The Elements of Statistical Learning : Data Mining, Inference, and Prediction*, 2^e éd. New York : Springer, 2009.
- [53] T. Chen et C. Guestrin, “Xgboost : A scalable tree boosting system,” *CoRR*, vol. abs/1603.02754, 2016. [En ligne]. Disponible : <http://arxiv.org/abs/1603.02754>

- [54] W. Ni, T. Wang, Y. Wu, X. Chen, W. Cai, M. Zeng, J. Yang, N. Hu et Z. Yang, "Classification and concentration predictions of volatile organic compounds using an electronic nose based on xgboost-random forest algorithms," *IEEE Sensors Journal*, vol. 24, n° 1, p. 671–678, 2024.
- [55] T. Cover et P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, n° 1, p. 21–27, 1967.
- [56] C. Cortes et V. Vapnik, "Support-vector networks," *Chem. Biol. Drug Des.*, vol. 297, p. 273–297, 01 2009.
- [57] A. J. Smola et B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, n° 3, p. 199–222, 2004.
- [58] H. Zhang et Y. Han, "A new mixed-gas-detection method based on a support vector machine optimized by a sparrow search algorithm," *Sensors*, vol. 22, n° 22, 2022. [En ligne]. Disponible : <https://doi.org/10.3390/s22228977>
- [59] I. Goodfellow, Y. Bengio et A. Courville, *Deep Learning*. MIT Press, 2016, <http://www.deeplearningbook.org>.
- [60] Y. Lecun, L. Bottou, Y. Bengio et P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, n° 11, p. 2278–2324, 1998.
- [61] A. Krizhevsky, I. Sutskever et G. E. Hinton, "Imagenet classification with deep convolutional neural networks," dans *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 1*, ser. NIPS'12. Red Hook, NY, USA : Curran Associates Inc., 2012, p. 1097–1105.
- [62] R. Ye, F. Liu et L. Zhang, "3d depthwise convolution : Reducing model parameters in 3d vision tasks," 2018. [En ligne]. Disponible : <https://arxiv.org/abs/1808.01556>
- [63] M. Kang, I. Cho, J. Park, J. Jeong, K. Lee, B. Lee, D. Del Orbe Henriquez, K. Yoon et I. Park, "High accuracy real-time multi-gas identification by a batch-uniform gas sensor array and deep learning algorithm," *ACS Sensors*, vol. 7, n° 2, p. 430–440, 2022, pMID : 35041384. [En ligne]. Disponible : <https://doi.org/10.1021/acssensors.1c01204>
- [64] P. Peng, X. Zhao, X. Pan et W. Ye, "Gas classification using deep convolutional neural networks," *Sensors*, vol. 18, n° 1, 2018. [En ligne]. Disponible : <https://doi.org/10.3390/s18010157>
- [65] S. Hochreiter et J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, n° 8, p. 1735–1780, 1997.
- [66] Y. Han, H. Zhuang, Z. Yin, Z. Long, Y. Li, Y. Yao, Q. Zheng et Z. Zhu, "A machine-learning-algorithm enhanced multi-functional gas sensor for self-humidity compensation

- and partial discharge detection,” *ACS Sensors*, vol. 10, n°. 8, p. 5882–5891, 2025, pMID : 40802499. [En ligne]. Disponible : <https://doi.org/10.1021/acssensors.5c01214>
- [67] I. Cho, K. Lee, Y. C. Sim, J.-S. Jeong, M. Cho, H. Jung, M. Kang, Y.-H. Cho, S. C. Ha, K.-J. Yoon et I. Park, “Deep-learning-based gas identification by time-variant illumination of a single micro-led-embedded gas sensor,” *Light : Science & Applications*, vol. 12, n°. 1, p. 95, 2023, accessed : 2025-09-08. [En ligne]. Disponible : <https://doi.org/10.1038/s41377-023-01120-7>
- [68] H. Mei, J. Peng, T. Wang, B. Zhang, F. Xuan, L. Wang, M. Zeng et Z. Yang, “Smart vocs recognition system based on single gas sensor and multi-task deep learning model,” *Sensors and Actuators B : Chemical*, vol. 439, p. 137853, 2025. [En ligne]. Disponible : <https://www.sciencedirect.com/science/article/pii/S0925400525006288>
- [69] X. Li, M. Li, R. Li et W. Li, “Copper and iron oxide sensor arrays combined with svm neural networks to achieve classification and concentration prediction of vocs gases,” *IEEE Sensors Journal*, vol. 25, n°. 5, p. 8759–8765, 2025.
- [70] S. Nayak, S. Misra et S. Basu Majumder, “Machine learning-based edge intelligence for reliable gas monitoring networks,” *IEEE Sensors Journal*, vol. 25, n°. 18, p. 35 268–35 277, 2025.
- [71] B. Wang, J. Zhang, T. Wang, W. Li, Q. Lu, H. Sun, L. Huang, X. Liang, F. Liu, F. Liu, P. Sun et G. Lu, “Machine learning-assisted volatile organic compound gas classification based on polarized mixed-potential gas sensors,” *ACS Applied Materials & Interfaces*, vol. 15, n°. 4, p. 6047–6057, 2023.
- [72] S. Singh, S. S. P. Varma, G. Sreelekha, C. Adak, R. P. Shukla et V. B. Kamble, “Metal oxide-based gas sensor array for vocs determination in complex mixtures using machine learning,” *Microchimica Acta*, vol. 191, n°. 4, p. 196, 2024.
- [73] J. Fonollosa, S. Sheik, R. Huerta et S. Marco, “Reservoir computing compensates slow response of chemosensor arrays exposed to fast varying gas concentrations in continuous monitoring,” *Sensors and Actuators B : Chemical*, vol. 215, p. 618–629, 2015. [En ligne]. Disponible : <https://www.sciencedirect.com/science/article/pii/S0925400515003524>
- [74] J. Fonollosa, “Gas sensor array under dynamic gas mixtures [dataset],” UCI Machine Learning Repository / Kaggle, 2015, downloaded from Kaggle : <https://www.kaggle.com/datasets/uciml/gas-sensor-array-under-dynamic-gas-mixtures>. [En ligne]. Disponible : <https://archive.ics.uci.edu/ml/datasets/gas+sensor+array+under+dynamic+gas+mixtures>
- [75] T. Akiba, S. Sano, T. Yanase, T. Ohta et M. Koyama, “Optuna : A next-generation hyperparameter optimization framework,” 2019. [En ligne]. Disponible : <https://arxiv.org/abs/1907.10902>

- [76] L. Sène, R. Guertin et Y.-A. Peter, “Optimization of fabry-perot interferometer performance through low mirror roughness,” dans *2024 International Conference on Optical MEMS and Nanophotonics (OMN)*, 2024, p. 1–2.
- [77] N. P. K. G. C. Yaws, Carl L, *Yaws’ Handbook of Antoine Coefficients for Vapor Pressure (2nd Electronic Edition)*. Knovel, 2009. [En ligne]. Disponible : <https://app.knovel.com/hotlink/toc/id:kpYHACVPEH/yaws-handbook-antoine/yaws-handbook-antoine>
- [78] G. Swirszcz, W. M. Czarnecki et R. Pascanu, “Local minima in training of neural networks,” 2017. [En ligne]. Disponible : <https://arxiv.org/abs/1611.06310>
- [79] S. Oller-Moreno, A. Pardo, J. M. Jiménez-Soto, J. Samitier et S. Marco, “Adaptive asymmetric least squares baseline estimation for analytical instruments,” dans *2014 IEEE 11th International Multi-Conference on Systems, Signals Devices (SSD14)*, 2014, p. 1–5.
- [80] G. Shao, Y. Wang, Z. Lan, J. Wang, J. He, X. Chou, K. Zhu et Y. Zhou, “Polymer-based gas sensors for detection of disease biomarkers in exhaled breath,” *Biosensors*, vol. 16, n° 1, p. 7, 2025.
- [81] C. Wilhelm et J.-L. Gardette, “Infrared analysis of the photochemical behaviour of segmented polyurethanes : aliphatic poly(ether-urethane)s,” *Polymer*, vol. 39, n° 24, p. 5973–5980, 1998. [En ligne]. Disponible : <https://www.sciencedirect.com/science/article/pii/S0032386197100659>
- [82] M. Malíková, J. Rychlý, L. Matisová-Rychlá, K. Csomorová, I. Janigová et H.-W. Wilde, “Assessing the progress of degradation in polyurethanes by chemiluminescence. i. unstabilised polyurethane films,” *Polymer Degradation and Stability*, vol. 95, n° 12, p. 2367–2375, 2010. [En ligne]. Disponible : <https://www.sciencedirect.com/science/article/pii/S0141391010003496>
- [83] E. Borucinska, P. Zamojski, W. Grodzki, U. Blaszcak, I. Zglobicka, M. Zielinski et K. J. Kurzydowski, “Degradation of polymethylmethacrylate (pmma) bioreactors used for algal cultivation,” *Materials*, vol. 16, n° 13, p. 4873, 2023.
- [84] S. M. Louie, J. M. Gorham, J. Tan et V. A. Hackley, “Ultraviolet photo-oxidation of polyvinylpyrrolidone (pvp) coatings on gold nanoparticles,” *Environmental Science : Nano*, vol. 4, n° 10, p. 2030–2040, 2017.
- [85] A. Morawska-Chochół, M. Szumera, A. Młyniec et K. Pielichowska, “The effect of aging process conditions on the thermal properties of poly(dimethylsiloxane)-based silicone rubber,” *Materials*, vol. 17, n° 22, p. 5608, 2024.

ANNEXE A DOMAINE DES HYPERPARAMÈTRES POUR LES ALGORITHMES D'APPRENTISSAGE AUTOMATIQUE

Modèle	Hyperparamètre	Domaine de recherche
SVR	C	$2^{-5} - 2^{10}$
	γ	$2^{-15} - 2^3$
	ϵ	$10^{-4} - 10^{-1}$
XGBoost	n_estimators	100 – 1200
	max_depth	2 – 10
	learning_rate	$10^{-2} - 3 \times 10^{-1}$ (log)
	subsample	0.6 – 1.0
	colsample_bytree	0.6 – 1.0
	min_child_weight	1.0 – 20.0
	reg_alpha	0.0 – 1.0
	reg_lambda	0.0 – 5.0
KNN	n_neighbors	1 – 25
	weights	{uniform, distance}
	p	{1,2}
LSTM	n_layers	1 – 2
	units	{16, 32, 64}
	dropout	0.0 – 0.4
	learning_rate	$10^{-4} - 3 \times 10^{-3}$ (log)
CNN1D	n_points	{20, 40, 60}
	n_conv	1 – 6
	filters	{32, 48, 64}
	kernel_size	3 – 11 (pas de 2)
	dropout	0.1 – 0.3
	learning_rate	$10^{-4} - 3 \times 10^{-3}$ (log)
CNN2D	n_conv	1 – 6
	filters	{16, 32, 48}
	kernel_time	3 – 11
	kernel_feat	1 – 3
	dropout	0.1 – 0.3
	learning_rate	$10^{-4} - 3 \times 10^{-3}$ (log)

TABLEAU A.1 Domaines de recherche des hyperparamètres utilisés lors de l'optimisation par Optuna pour les différents modèles d'apprentissage machine.

ANNEXE B AJUSTEMENT DE LA COURBE DU MODÈLE *TWO STAGE DIFFUSION* SUR LES CAPTEURS FABRY-PÉROT ET PERFORMANCE DES CAPTEURS CHIMIQUES

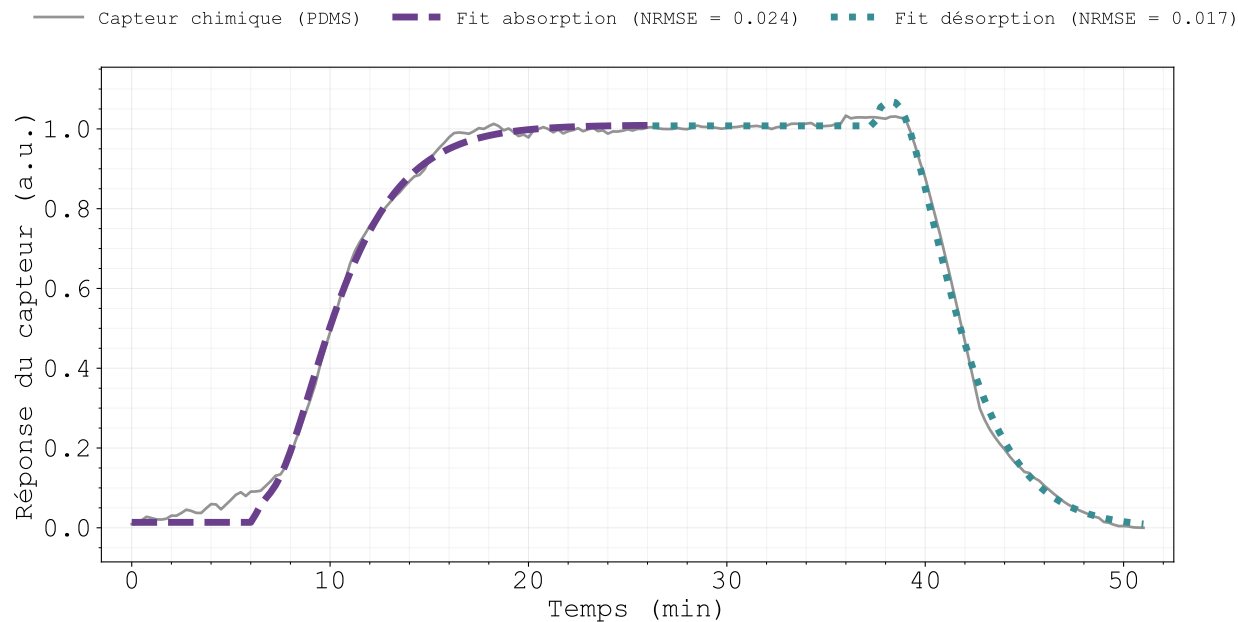


FIGURE B.1 Comparaison de la réponse temporelle du capteur Fabry-Pérot avec polymère PDMS et du jeu de données, avec ajustement à l'équation de diffusion à deux constantes (*two-stage diffusion*).

TABLEAU B.1 Sensibilités des capteurs, résultats de régression et répétabilité pour le CO, le Méthane et l'Éthylène, regroupés par type de capteur. Les meilleurs capteurs sont surlignés .

Type de capteur	Sens CO	R^2 CO	Écart-type CO	Sens CH ₄	R^2 CH ₄	Écart-type CH ₄	Sens C ₂ H ₄	R^2 C ₂ H ₄	Écart-type C ₂ H ₄
TGS2600	-0.007	0.693	0.275	-0.034	0.232	0.114	-0.335	0.871	0.199
TGS2600	-0.006	0.760	0.202	-0.030	0.230	0.096	-0.305	0.874	0.174
TGS2600	-0.006	0.468	0.442	-0.041	0.372	0.233	-0.284	0.744	0.317
TGS2600	-0.008	0.620	0.413	-0.044	0.290	0.155	-0.362	0.797	0.341
TGS2602	-0.011	0.514	0.691	0.004	0.085	0.166	-0.342	0.174	2.140
TGS2602	-0.021	0.015	9.579	0.002	0.078	0.083	-0.722	0.006	18.418
TGS2602	-0.019	0.028	6.479	0.003	0.125	0.100	-0.688	0.011	12.342
TGS2602	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
TGS2610	-0.010	0.418	0.709	-0.042	0.126	0.060	-0.534	0.746	0.756
TGS2610	-0.007	0.410	0.555	-0.032	0.177	0.044	-0.369	0.622	0.693
TGS2610	-0.014	0.146	2.172	-0.045	0.130	0.066	-0.481	0.329	2.505
TGS2610	-0.012	0.239	1.365	-0.045	0.116	0.070	-0.529	0.434	1.575
TGS2620	-0.006	0.735	0.206	-0.032	0.301	0.122	-0.276	0.896	0.792
TGS2620	-0.005	0.746	0.200	-0.029	0.259	0.099	-0.261	0.887	0.756
TGS2620	-0.006	0.568	0.361	-0.035	0.419	0.223	-0.251	0.744	0.781
TGS2620	-0.007	0.608	0.377	-0.045	0.383	0.211	-0.332	0.798	0.290

ANNEXE C RÉSULTATS D'OPTIMISATION DES HYPERPARAMÈTRES POUR LES MODÈLES DE RÉGRESSION ET DE CLASSIFICATION POUR LE JEU DE DONNÉES

TABLEAU C.1 Principaux hyperparamètres optimaux obtenus pour les modèles de régression lors de l'optimisation dans les plis internes de la validation croisée imbriquée.

Modèle	Hyperparamètres principaux
CNN1D	$n_{conv} = 3$, filtres=64, noyau=11, dropout=0.1, $lr = 9.1 \times 10^{-4}$
CNN2D	$n_{conv} = 4$, filtres=32, noyau _t =9, noyau _f =1, dropout=0.1, $lr = 2.6 \times 10^{-3}$
LSTM	$n_{layers} = 1$, unités=64, dropout=0.1, $lr = 2.7 \times 10^{-3}$
SVR	$C = 928$, $\gamma = 1.3 \times 10^{-3}$, $\epsilon = 5.9 \times 10^{-4}$
KNN	$k = 25$, poids=distance, $p = 1$
XGBoost	arbres=330, profondeur=2, $lr = 0.13$, subsample=0.63

TABLEAU C.2 Principaux hyperparamètres optimaux obtenus pour les modèles de classification lors de l'optimisation dans les plis internes de la validation croisée imbriquée.

Modèle	Hyperparamètres principaux
CNN1D	$n_{conv} = 4$, filtres=32, noyau=3, dropout=0.23, $lr = 1.33 \times 10^{-3}$
CNN2D	$n_{conv} = 3$, filtres=64, noyau=6, dropout=0.15, $lr = 1.17 \times 10^{-3}$
LSTM	$n_{layers} = 2$, unités=48, dropout=0.13, $lr = 5.95 \times 10^{-3}$
SVC	$C = 0.175$, $\gamma = 0.033$
KNN	$k = 15$, poids=distance
XGBoost	arbres=416, profondeur=3, $lr = 0.017$, subsample=0.81

ANNEXE D PERFORMANCES POLYMÈRES FABRY-PÉROT

TABLEAU D.1 Caractérisation des capteurs pour IPA. Les capteurs sélectionnés (chip1 : 2,8 ; chip3 : 1,3) sont surlignés.

Chip	Canal	Polymère	Sens.	R^2	LOD	SNR_{max}	Rép. asc. (%)	Rép. desc. (%)
Chip1	1	PU	0.684	0.712	4.480	59.76	28.86	13.49
Chip1	2	PU	0.876	0.812	1.576	165.64	21.21	9.43
Chip1	3	PU	1.925	0.803	3.391	93.80	17.90	6.40
Chip1	4	PU	0.963	0.792	2.737	92.49	24.82	9.95
Chip1	5	PMMA	-0.653	0.850	8.290	-28.57	33.20	23.69
Chip1	6	PMMA	-0.215	0.864	6.796	-50.04	23.56	22.81
Chip1	7	PMMA	0.200	0.230	17.981	1.55	1016.15	965.61
Chip1	8	PMMA	0.607	0.931	2.575	99.18	24.36	18.44
chip3	1	PDMS	0.110	0.567	27.364	7.99	35.60	48.88
Chip3	2	PDMS	0.099	0.308	46.039	-1.43	385.42	511.03
Chip3	3	PVP	-2.915	0.090	1.787	160.97	76.77	26.02
Chip3	4	PVP	-2.536	0.088	3.214	90.74	75.78	26.10
Chip3	5	PVA	0.197	0.277	41.318	9.45	36.72	35.01
Chip3	6	PVA	-4.787	0.019	59.600	-31.07	90.84	88.52
Chip3	7	PEO	-159.045	0.403	9.857	-5.31	182.33	245.07
Chip3	8	PEO	-0.152	0.005	1888.945	-0.72	142.37	83.39

TABLEAU D.2 Caractérisation des capteurs pour Acetone. Les capteurs sélectionnés (chip1 : 2,8 ; chip3 : 1,3) sont surlignés.

Chip	Canal	Polymère	Sens.	R^2	LOD	SNR_{max}	Rép. asc. (%)	Rép. desc. (%)
Chip1	1	PU	47.756	0.996	2.015	139.02	8.02	2.95
Chip1	2	PU	38.762	0.985	2.432	111.73	6.88	7.21
Chip1	3	PU	81.824	0.993	2.070	133.72	5.72	4.46
Chip1	4	PU	50.697	0.995	2.262	121.58	7.40	5.43
Chip1	5	PMMA	0.071	0.067	83.645	11.01	38.28	52.45
Chip1	6	PMMA	0.127	0.197	25.312	17.59	22.15	36.02
Chip1	7	PMMA	1.116	0.752	11.484	28.45	15.15	29.93
Chip1	8	PMMA	0.864	0.919	7.673	36.80	11.43	18.38
Chip3	1	PDMS	0.120	0.524	37.262	8.88	33.27	40.55
Chip3	2	PDMS	0.107	0.110	44.997	12.09	55.25	29.36
Chip3	3	PVP	-0.891	0.540	7.908	-36.66	79.29	54.04
Chip3	4	PVP	-0.674	0.507	7.088	-40.23	86.71	60.42
Chip3	5	PVA	-0.230	0.179	35.651	3.96	108.66	85.75
Chip3	6	PVA	0.013	0.000	14026.768	1.70	63.00	45.61
Chip3	7	PEO	-120.262	0.399	119.749	-2.77	49.05	56.20
Chip3	8	PEO	-1.640	0.050	440.258	-2.51	65.09	43.75

TABLEAU D.3 Caractérisation des capteurs pour Toluene. Les capteurs sélectionnés (chip1 : 2,8 ; chip3 : 1,3) sont surlignés.

Chip	Canal	Polymère	Sens.	R^2	LOD	SNR_{max}	Rép. asc. (%)	Rép. desc. (%)
Chip1	1	PU	0.944	0.432	5.447	66.33	37.24	30.62
Chip1	2	PU	1.629	0.734	1.712	203.88	11.65	11.98
Chip1	3	PU	1.680	0.553	7.094	47.07	33.30	20.65
Chip1	4	PU	1.161	0.542	2.434	129.14	31.20	21.25
Chip1	5	PMMA	-0.693	0.710	5.176	-22.82	46.72	69.28
Chip1	6	PMMA	-0.182	0.612	14.564	-11.90	59.85	79.91
Chip1	7	PMMA	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Chip1	8	PMMA	0.681	0.818	3.451	56.44	22.69	39.05
Chip3	1	PDMS	0.342	0.241	9.886	46.82	20.30	16.91
Chip3	2	PDMS	0.008	0.001	294.321	14.76	25.93	79.66
Chip3	3	PVP	0.468	0.049	16.854	2.10	1355.74	1332.50
Chip3	4	PVP	-86.289	0.141	14.051	17.02	136.97	95.34
Chip3	5	PVA	0.267	0.327	60.752	7.96	29.40	20.58
Chip3	6	PVA	1.180	0.243	131.151	-0.22	640.21	602.97
Chip3	7	PEO	-41.463	0.198	2.228	16.39	242.23	48.41
Chip3	8	PEO	-21.679	0.069	289.288	-3.23	79.19	53.57

TABLEAU D.4 Caractérisation des capteurs pour Méthanole. Les capteurs sélectionnés (chip1 : 2,8 ; chip3 : 1,3) sont surlignés.

Chip	Canal	Polymère	Sens.	R^2	LOD	SNR_{max}	Rép. asc. (%)	Rép. desc. (%)
Chip1	1	PU	0.490	0.771	12.715	20.46	17.40	26.16
Chip1	2	PU	0.584	0.886	7.649	35.31	9.44	16.58
Chip1	3	PU	1.146	0.844	9.547	27.16	13.92	20.09
Chip1	4	PU	0.750	0.903	4.197	65.21	9.62	15.32
Chip1	5	PMMA	0.057	0.724	49.296	8.03	21.43	22.19
Chip1	6	PMMA	-0.010	0.024	370.323	-0.34	668.13	651.95
Chip1	7	PMMA	0.089	0.736	24.849	11.46	25.28	30.29
Chip1	8	PMMA	0.125	0.834	11.738	28.53	14.67	25.94
Chip3	1	PDMS	0.006	0.001	509.524	-6.63	101.30	109.44
Chip3	2	PDMS	0.116	0.202	25.504	19.52	45.79	49.99
Chip3	3	PVP	-0.600	0.034	20.376	-3.58	551.24	363.28
Chip3	4	PVP	-0.448	0.027	22.687	-0.50	3913.10	2641.44
Chip3	5	PVA	-0.121	0.202	65.190	-5.14	52.09	69.17
Chip3	6	PVA	-0.156	0.008	1198.317	-1.33	76.56	67.50
Chip3	7	PEO	-1.152	0.005	381.982	0.60	952.78	968.17
Chip3	8	PEO	-0.347	0.011	720.725	-3.48	44.19	43.45

ANNEXE E RÉSULTATS DES HYPERPARAMÈTRES POUR LES CAVITÉS FABRY-PÉROT

TABLEAU E.1 Principaux hyperparamètres optimaux obtenus pour les modèles de régression lors de l'optimisation dans les plis internes de la validation croisée imbriquée. Ces valeurs ont été sélectionnées à partir d'une métrique pondérée combinant les régimes quasi-statiques et transitoires.

Modèle	Hyperparamètres principaux
CNN2D	$n_{conv} = 3$, $\text{filtres}=32$, $\text{noyau}_t=7$, $\text{noyau}_f=1$, $\text{dropout}_{conv} = 0.4$, $\text{neurones}=32$, $\text{dropout}_{dense} =$ 0.1 , $lr = 1.27 \times 10^{-4}$, $\text{batch}=64$
SVR (RBF)	$C = 933.90$, $\epsilon = 0.351$, $\gamma = 1.03 \times 10^{-4}$

TABLEAU E.2 Principaux hyperparamètres optimaux obtenus pour les modèles de classification lors de l'optimisation dans les plis internes de la validation croisée imbriquée.

Modèle	Hyperparamètres principaux
CNN2D	$n_{conv} = 4$, $\text{filtres}=64$, $\text{noyau}_t=9$, $\text{noyau}_f=1$, $\text{dropout}_{conv} = 0.1$, $\text{neurones}=256$, $\text{dropout}_{dense} =$ 0.3 , $lr = 1.04 \times 10^{-4}$
SVC (RBF)	$C = 215.88$, $\gamma = 0.108$