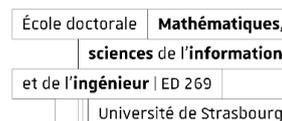




INSA INSTITUT NATIONAL
DES SCIENCES
APPLIQUÉES
STRASBOURG



UNIVERSITÉ DE
STRASBOURG



octapharma
For the safe and optimal use of human proteins

*Ecole doctorale 269 Mathématiques, Sciences de l'Information et de l'Ingénieur
(MSII)*

**Laboratoire des sciences de l'ingénieur, de l'informatique et de l'imagerie
(UMR7357)**

THÈSE présentée par :

Florent WADEL

soutenue le : **27 mai 2025**

pour obtenir le grade de :

Docteur de l'Université de Strasbourg

Discipline (Spécialité) : Génie Industriel

**Contribution à l'élaboration d'une stratégie de
maintenance prévisionnelle basée sur l'IOT :
Application à l'industrie biopharmaceutique**

THÈSE dirigée par :

**HOUSSIN Rémy
COULIBALY Amadou**

MCF-HDR, Université de Strasbourg
MCF-HDR, INSA Strasbourg

RAPPORTEURS :

**DI MASCOLO Maria
HAJEJ Zied**

Directrice de recherche, CNRS- Grenoble Alpes
MCF-HDR, Université de Lorraine-Metz

AUTRES MEMBRES DU JURY :

**SIMEU-ABAZI Zineb
DERIGENT William**

Professeure, Université Grenoble Alpes
Professeur, Université de Lorraine-Nancy

INVITÉS :

**TIGHAZOUI Ayoub
CALMELS Bastien**

MCF, Université de Strasbourg
Directeur Qualité – OCTAPHARMA SAS

***« La vie est la vie, autrement dit un combat,
pour une nation comme pour un homme. »***

Charles de Gaulle, 1965

REMERCIEMENTS

Il me semble important d'ouvrir ces remerciements en revenant sur les circonstances qui ont permis la réalisation de cette thèse. Elle n'est pas le fruit d'un parcours prédéfini, mais celui d'un enchaînement d'opportunités, de soutiens et de décisions partagées. C'est grâce à la confiance et à l'implication de plusieurs personnes clés que ce projet a pu voir le jour.

Mes premiers remerciements s'adressent donc tout naturellement à vous, **Éric Penn**, qui avez réceptionné l'idée de ce projet de thèse pour lui donner ses premières formes et jouer ce rôle de « catalyseur » si je puis dire. Ce projet a par la suite obtenu le soutien de **Fanny Chauvel**, Directrice générale d'Octapharma SAS à cette époque, que je tiens à remercier également. Mes remerciements vont également à **Yann Véronneau**, Directeur technique d'Octapharma SAS, pour avoir accepté d'encadrer cette thèse du côté entreprise. Les circonstances propres aux dispositifs CIFRE étant ce qu'elles sont, lancer cette thèse n'a pas été simple pour toi mais, tu l'as fait. Par la suite, ton implication et ton exigence (constructives, toujours) ont grandement contribué à faire avancer ce travail dans de bonnes conditions. Je n'oublie pas que rien n'aurait été possible sans toi. Enfin, j'ai eu la chance de compter dans mon encadrement interne, des personnes compétentes et issues de différents horizons qui ont enrichi ces travaux. **Sébastien Franz**, **Bastien Calmels** et **Bertrand Gérardin**, je vous remercie de vous être joints à cette aventure pour me permettre de la mener à bien dans les meilleures conditions.

Ma gratitude va également à **Rémy Houssin** et **Amadou Coulibaly**, mes directeurs de thèse. Leur accompagnement, leurs conseils éclairés et leur disponibilité ont été essentiels tout au long du déroulé de cette thèse. Leur expertise et leur exigence scientifique ont offert à cette recherche une orientation claire et solide. Je tiens également à témoigner ma reconnaissance envers **Ayoub Thigazoui**, co-encadrant et alter-égo universitaire, pour son soutien. A mon sens, ton appui tout au long de ces travaux n'est pas reconnu à son entière valeur et j'ai conscience que ces quelques mots ne suffiront pas à t'exprimer toute ma gratitude. Notre collaboration a été bien plus déterminante que je ne l'imaginais au départ (toi aussi, probablement !), et a profondément enrichi ce travail. Mes remerciements vont également aux membres du jury qui ont accepté d'évaluer ce travail : **Maria Di Mascolo** et **Zied Hajej**, qui ont pris le temps d'en être les rapporteurs, ainsi que **Zineb Simeu-Abazi** et **William Derigent**, examinateurs. Je vous remercie encore pour la qualité des échanges lors de ma défense de thèse.

Que ce soient mes collègues d'Octapharma, du laboratoire, les membres des différentes équipes avec lesquelles j'ai collaboré, je souhaite également adresser mes remerciements à toutes celles et ceux qui, de près ou de loin, ont contribué à la réussite de ces travaux. Vous avez été nombreux à apporter votre pierre à l'édifice et je tiens à souligner l'importance de chacune de ces contributions. À toi, **Alexis Ducancel**, qui a joué un rôle de mentor informel ; A vous, **Pierre Linger**, **Adrien Simard**, **Ben Idiri** & **François Jud**, avec qui nous avons constitué pour l'occasion un prototype d'équipe « Data » ; **Laurent Contet**, **Stéphane Maier** & **Christian Nold** pour votre appui technique côté maintenance et à vous, **Thierry Kleiser** et **Mohammed Ouazanni** pour votre aide à la programmation et au déploiement des systèmes d'automatisme. Une mention spéciale et oblige pour toi, Pierre, pour nos débats sur mes publications et sur l'avancée des techniques de traitement de données développées. Vous tous avez rendu possible le bon déroulement de ces travaux et contribué, chacun à votre manière, à leur aboutissement. Que ce soit par votre soutien technique, nos échanges, ou simplement votre présence attentive, votre implication a été un véritable moteur tout au long de ce parcours.

Enfin, je tiens à adresser quelques mots à ma compagne **Lucie**, à ma famille et à mes amis d'enfance. Vous qui m'avez parfois vu m'enfermer malgré moi dans ce que l'on appelle la solitude du chercheur et faire face à des moments de désarroi et de désillusion, que j' imagine propre à tout doctorant, votre soutien indéfectible m'a porté bien au-delà des moments difficiles.

Je conclus ces quelques mots de remerciement avec une pensée pour ma fille **Olivia**, qui a quatre ans aujourd'hui et que je n'ai pas vu grandir ces dernières années. J'espère que tu liras un jour ces quelques mots laissés ici pour toi. Si la vie te donne l'occasion, un jour, de te lancer dans une thèse de recherche, tu feras probablement face toi aussi à des moments de solitude, de désarroi et de désillusion. Puisse ces quelques mots te rappeler que la difficulté forge le caractère et ouvre des horizons insoupçonnés. Que chaque obstacle franchi est une victoire et que l'effort portera toujours ses fruits. Churchill disait que le succès, c'est d'aller d'échec en échec sans perdre son enthousiasme. A posteriori, je pense qu'il avait raison.

Avertissement au lecteur / Warning to the reader

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition des membres de la communauté universitaire. Il est soumis à la propriété intellectuelle de l'auteur. Cela implique une obligation de citation et de référencement lors de l'utilisation de ce document. D'autre part, toute contrefaçon, plagiat, reproduction ou représentation illicite encourt une poursuite pénale.

This document is the result of a long process approved by the jury and made available to members of the university community. It is subject to the intellectual property rights of its author. This implies an obligation to quote and reference when using this document. Furthermore, any infringement, plagiarism, unlawful reproduction or representation will be prosecuted.

[Code de la Propriété Intellectuelle](#)

[Article L122-4](#) :

Toute représentation ou reproduction intégrale ou partielle faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite. Il en est de même pour la traduction, l'adaptation ou la transformation, l'arrangement ou la reproduction par un art ou un procédé quelconque.

Any representation or reproduction in whole or in part without the consent of the author or his successors in title or assigns is unlawful. The same applies to translation, adaptation or transformation, arrangement or reproduction by any art or process whatsoever.

[Articles L335-1 à L335-9](#) : Dispositions pénales / Penal provisions.

Licence attribuée par l'auteur



<https://creativecommons.org/licenses/?lang=fr-FR>

Florent WADEL

Contribution à l'élaboration d'une stratégie de maintenance prévisionnelle basée sur l'IoT : Application à l'industrie biopharmaceutique

Résumé

Cette thèse, réalisée en collaboration entre OCTAPHARMA SAS et l'équipe CSIP du laboratoire iCube, étudie la transition vers la maintenance prédictive dans l'industrie biopharmaceutique, en réponse à l'évolution des besoins du secteur dans le cadre de l'Industrie 4.0. En s'appuyant sur l'Internet des Objets et le cadre générique du Pronostic Health Management, l'étude propose une méthodologie permettant d'anticiper les défaillances des membranes de vannes pneumatiques, des composants essentiels aux systèmes de production de fluides. L'état de l'art met en évidence un retard dans l'adoption des stratégies de maintenance basées sur l'IOT dans le domaine biopharmaceutique, soulignant ainsi une opportunité majeure de contribution scientifique. Au travers d'une analyse de données, de l'extraction de motifs représentatifs et de l'utilisation de techniques d'apprentissage automatique, un modèle de dégradation est développé pour estimer la durée de vie restante des équipements. Enfin, une stratégie de maintenance hybride, combinant approches systématique et prédictive, est proposée afin d'améliorer la fiabilité et de réduire les coûts opérationnels.

Mots clés :

Maintenance prévisionnelle, Internet des Objets, Pronostic Health, Management, Industrie 4.0, Biopharmaceutique, Machine Learning

Abstract

This thesis, conducted in collaboration between OCTAPHARMA SAS and the CSIP team of the iCube laboratory, investigates the transition toward predictive maintenance in the biopharmaceutical industry, addressing the sector's evolving demands in line with Industry 4.0. Leveraging the Internet of Things and Prognostic Health Management, the study proposes a methodology to anticipate failures in pneumatic valve membranes, components essential to fluid production systems. The literature review reveals a lag in the adoption of IoT-based maintenance strategies within the biopharmaceutical sector, underscoring a significant opportunity for scientific advancement. Through data analysis, pattern recognition, and machine learning techniques, a degradation model is developed to estimate the remaining useful life (RUL) of equipment. Finally, a hybrid maintenance strategy, combining both systematic and predictive approaches, is proposed to enhance reliability and reduce operational costs.

Keywords :

Predictive maintenance, Internet of Things, Pronostic Health, Management, Industry 4.0, Biopharmaceutical, Machine Learning

Table des matières

| | |
|---|-----|
| Chapitre 1..... | 15 |
| 1. Introduction..... | 15 |
| 1.1. Contextualisation de la thèse | 15 |
| 1.2. Contexte scientifique de la maintenance industrielle | 22 |
| 1.3. Des objectifs industriels aux questions scientifiques | 24 |
| 1.4. Plan de thèse | 25 |
| Chapitre 2..... | 27 |
| 2. Des pratiques de maintenance dans le milieu du biomédicament | 27 |
| 2.1. Introduction..... | 27 |
| 2.2. Méthodologie de recherche | 31 |
| 2.3. Évolution vers une maintenance IOT / data-driven | 35 |
| 2.4. Modèles pour la maintenance prévisionnelle | 37 |
| 2.5. Discussion : Analyse des modèles de maintenance prévisionnelle | 48 |
| 2.6. Conclusion du chapitre 2..... | 52 |
| Chapitre 3..... | 55 |
| 3. Du modèle de dégradation de l'équipement..... | 55 |
| 3.1. Introduction et contexte industriel..... | 55 |
| 3.2. État de l'art | 57 |
| 3.3. Description du système étudié..... | 60 |
| 3.4. Expérimentation sur banc d'essai..... | 63 |
| 3.5. Analyse des résultats expérimentaux..... | 75 |
| 3.6. Conclusion du chapitre 3..... | 80 |
| Chapitre 4..... | 83 |
| 4. De l'établissement du pronostic..... | 83 |
| 4.1. Approches du pronostic | 83 |
| 4.2. Apprentissage supervisé..... | 85 |
| 4.3. Test des algorithmes | 92 |
| 4.4. Méthode d'entraînement de l'algorithme..... | 98 |
| 4.5. Création du modèle de dégradation..... | 101 |
| 4.6. Conclusion du chapitre 4..... | 103 |
| Chapitre 5..... | 105 |

| | | |
|------|--|-----|
| 5. | De la prise de décision | 105 |
| 5.1. | Introduction..... | 105 |
| 5.2. | Contexte opérationnel et maintenance actuelle | 106 |
| 5.3. | Intégration des prévisions dans la planification | 111 |
| 5.4. | Aide à la décision | 118 |
| 5.5. | Généralisation du modèle..... | 120 |
| 5.6. | Conclusion du chapitre 5..... | 124 |
| 6. | Conclusion générale..... | 125 |
| 7. | Perspectives | 127 |
| 8. | Références | 129 |

Table des figures

| | |
|---|----|
| Figure 1 - Présence d'un plan de maintenance systématique et ratios annuels de maintenance systématiques VS l'ensemble des maintenances préventives | 29 |
| Figure 2 – Présence d'un plan de maintenance conditionnelle et ratios annuels de maintenance conditionnelles VS l'ensemble des maintenances préventives | 30 |
| Figure 3 - Présence d'un plan de maintenance prévisionnelle et ratios annuels de maintenance prévisionnelles VS l'ensemble des maintenances préventives..... | 30 |
| Figure 4 - Taux de présence et type d'équipements de production en ZAC..... | 31 |
| Figure 5 - Diagramme PRISMA | 34 |
| Figure 6 - Distribution des publications pour chaque requête | 35 |
| Figure 7 - Résultats de sondage, Fantana et al (2013) | 36 |
| Figure 8 - Distribution des modèles selon le type d'équipement..... | 51 |
| Figure 9 - Schéma de développement d'un modèle ML pour la maintenance | 52 |
| Figure 10 - Illustration des étapes de la méthode PHM selon Gouriveau et al., (2013) .. | 57 |
| Figure 11 - Vanne à membrane pneumatique GEMÜ Biostar 650 | 61 |
| Figure 12 - Principe générique d'une vanne à membrane dérivé du schéma de Muskid, CC BY-SA 3.0, Wikimedia Commons..... | 62 |
| Figure 13 - Membrane de vanne pneumatique de la marque GEMÜ | 62 |
| Figure 14 - Indicateur de position électrique de la marque GEMÜ..... | 63 |
| Figure 15 - Schéma de la plateforme de test | 64 |
| Figure 16 - Architecture IO-Link pour le pilotage de la plateforme de test + Système de collecte de données Python | 65 |
| Figure 17 - Diagramme de commutation extrait de la documentation technique de fabricant..... | 66 |
| Figure 18 - Etapes de traitement du signal | 67 |
| Figure 19 - Signal de position extrait de l'indicateur de position pour 7 cycles avec mise en évidence d'une dérive | 67 |
| Figure 20 - Signal brut reconstruit avec les valeurs minimales..... | 69 |
| Figure 21 - Algorithme 1, pseudo-code utilisé pour la reconstruction du signal et l'établissement de la moyenne mobile | 69 |
| Figure 22 - Extraction des valeurs caractéristiques des cycles via PolyFit | 71 |
| Figure 23 - Algorithme 2, pseudo-code utilisé pour extraire les valeurs caractéristiques des cycles..... | 71 |
| Figure 24 - Signal normalisé et post-traité représentant les coefficients quadratiques par cycles | 73 |
| Figure 25 - Décomposition t-SNE des valeurs statistiques du signal brut par cycle | 75 |
| Figure 26 – Positions minimales du signal brut reconstruit annoté avec indication de la défaillance..... | 77 |

| | |
|--|-----|
| Figure 27 - Signal normalisé, post-processé et annoté avec indication de la défaillance | 78 |
| Figure 28 - Plateforme d'essai universitaire..... | 79 |
| Figure 29 - Résultats des tests de compression comparatifs de la membrane usée versus une membrane neuve | 79 |
| Figure 30 - Signaux augmentés et bruités | 93 |
| Figure 31 - Visualisation obtenue par fusion des valeurs..... | 93 |
| Figure 32 - Classification des valeurs par l'algorithme RandomForestClassifier | 95 |
| Figure 33 - Mise en évidence des erreurs de classification | 95 |
| Figure 34 - Décompte progressif du RUL..... | 96 |
| Figure 35 - Pseudo-code de création du modèle d'estimation du RUL..... | 102 |
| Figure 36 - Schématisation de la maintenance systématique chez OCTAPHARMA SAS | 106 |
| Figure 37 - Processus de remplacement et de suivi d'une membrane sous "correctif urgent" | 109 |
| Figure 38 - Processus de remplacement et de suivi d'une membrane sous "correctif planifié" | 110 |
| Figure 39 - Schéma du scénario 1 ($TRUL < T_{sys}$) | 114 |
| Figure 40 - Schéma du scénario 2 ($TRUL \approx T_{sys}$) | 114 |
| Figure 41 - Schéma du scénario 3 ($T_{sys} < TRUL$) | 115 |
| Figure 42 - Stratégie de maintenance hybride orientée en fonction de la criticité des équipements | 117 |
| Figure 43 - Proposition de système d'aide à la décision basé sur impacts criticité & business..... | 119 |
| Figure 44 - Processus de validation générique de l'approche prévisionnelle..... | 121 |
| Figure 45 - Processus de système d'aide à la décision générique basé sur impacts criticité & business | 123 |

Liste des tableaux

| | |
|--|----|
| Tableau 1 - Liste des acronymes | 13 |
| Tableau 2 - Critères d'inclusion et d'exclusion | 32 |
| Tableau 3 - Compilation des méthodes vues en section 2.4 | 49 |
| Tableau 4 - Récapitulatif des performances en classification | 94 |
| Tableau 5 - Récapitulatif des performances en régression | 96 |
| Tableau 6 - Comparaison du RUL réel VS RUL prédit..... | 97 |

Liste des acronymes

Tableau 1 - Liste des acronymes

| Acronymes | | | | | |
|-----------|---|---------|---|--------|--|
| AE | Auto Encoding | GBM | Gradient Boosting Machine | PF | Particle Filter |
| ARIMA | AutoRegressive Integrated Moving Average | GMM | Gaussian Mixture Model | PSO | Particle Swarm Optimization |
| BPNN | Back Propagation Neural Network | GOA | Grasshopper Optimization Algorithm | RBM | Restricted Boltzmann Machine |
| CCE | Categorical Cross-entropy | GPR | Gaussian Process Regression | RF | Random Forests |
| CCF | Convolutional Composite Features | GWO | Grey Wolf Optimization | RFE | Recursive Feature Elimination |
| CEEMD | Complementary Ensemble Empirical Mode Decomposition | HHE | Hilbert–Huang Entropy | RFR | Random Forest Regression |
| CNN | Convolutional Neural Network | HSMM | Hidden Semi-Markov Model | RNN | Recurrent Neural Network |
| CVA | Canonical Variate Analysis | IABC | Improved Artificial Bee Colony (algorithm) | Bi-RNN | Bidirectional Recurrent Neural Network |
| DBN | Deep Belief Network | KNN | K-Nearest Neighbors | RPA | Retro-Propagation Algorithm |
| DBSCAN | Density-Based Spatial Clustering of Applications with Noise | LSTM | Long-Short Term Memory | RUL | Remaining Useful Life |
| ELM | Extreme Learning Machine | Bi-LSTM | Bidirectional-LSTM | SFE | Systematic Feature Engineering |
| EM | Expectation & Maximization (algorithm) | LR | Linear Regression | SOM | Self-Organizing Map |
| ESN | Echo State Network | MLE | Multi-Loss Encoder | SR | Sparse Representation |
| FCA | Fuzzy Cognitive Maps | MGMF | Metabolism Grey Forecasting Model | SSA | Sparrow Search Algorithm |
| FE | Feature Engineering | NCNN | Noisy - CNN | SVM | Support Vector Machine |
| FOA | Fly Optimization Algorithm | NFC | Noisy Fusion Center | SWA | Swarm Intelligence Algorithm |
| FFT | Fast Fourier Transform | NPBGRU | Noisy Parallel Bidirectional Gated Recurrent Unit | TCN | Temporal Convolutional Network |
| GA | Genetic Algorithm | PCA | Principal Component Analysis | VAE | Variational Autoencoder |

Chapitre 1

1. Introduction

1.1. Contextualisation de la thèse

Cette thèse est née d'un partenariat entre le laboratoire biopharmaceutique OCTAPHARMA SAS et le laboratoire de recherche iCube de Strasbourg et plus précisément, l'équipe de recherche CSIP (Conception, Système d'Information et Processus inventifs). OCTAPHARMA SAS est une entreprise leader dans la production de biomédicaments issus de la valorisation du plasma sanguin humain. CSIP est une équipe du département de mécanique du laboratoire iCube. Ses thèmes de recherche s'inscrivent en génie industriel et sont axés plus particulièrement vers la conception inventive, l'ingénierie des systèmes d'information et de production, le développement durable et plus largement vers les concepts de l'industrie 4.0.

Ces travaux découlent d'un besoin d'évolution des pratiques de maintenance industrielle exprimé par l'entreprise OCTAPHARMA SAS, en accord avec les principes de l'Industrie 4.0. Ceci implique de démarrer la transition d'un plan de maintenance aujourd'hui largement couvert par l'approche systématique vers des pratiques de maintenance prévisionnelle, impliquant l'emploi de données pertinentes pour estimer l'état des équipements concernés et préconiser des prévisions de maintenance.

1.1.1. Contexte industriel de la thèse

Dans le secteur industriel, les pratiques de maintenance ont évolué au fil du temps. Elles sont tout d'abord essentiellement correctives, puis préventives avec des approches principalement systématiques pour évoluer vers des méthodes plus avancées, telles que la maintenance conditionnelle et prévisionnelle.

Pour ce qui est de la maintenance préventive, l'approche systématique est une forme de maintenance préventive où les interventions sont programmées à intervalles réguliers, indépendamment de l'état réel des équipements. Elle vise à éviter les défaillances, mais elle peut conduire à des interventions inutiles et à des coûts additionnels sans pour autant garantir qu'une panne subite n'apparaisse. La maintenance conditionnelle quant à elle, s'appuie sur l'évaluation de l'état des équipements à l'aide de paramètres mesurables (comme la température, la vibration, ou la pression) pour déterminer si une intervention est nécessaire. Cette méthode permet de réduire les interventions inutiles, tout en s'assurant que les équipements sont maintenus en bon état de fonctionnement. La maintenance prévisionnelle va plus loin en intégrant des outils de surveillance en temps réel et des analyses de données avancées pour prévoir les défaillances avant qu'elles ne surviennent. En surveillant en continu l'état des équipements, elle permet d'anticiper les pannes et de programmer les interventions de manière optimale, réduisant ainsi les coûts et les temps d'arrêt imprévus. Cette évolution vers des méthodes de maintenance prévisionnelles et basées sur l'état réel des équipements présente un réel intérêt dans les secteurs industriels où la fiabilité est essentielle.

Cependant, l'évolution des processus industriels dans des secteurs fortement contraints comme celui du biomédicament, se voit confrontée à un certain nombre d'obstacles parmi lesquels des exigences réglementaires strictes nécessitent un processus rigoureux et documenté pour la mise en œuvre de changement. Contrairement à d'autres secteurs manufacturiers où l'optimisation des procédés est plus agile, toute modification / évolution de

procédés dans ce type d'industrie nécessite des validations rigoureuses, engendrant des coûts élevés et des délais conséquents. La complexité de production des biomédicaments, souvent issus de processus biologiques sensibles, impose une traçabilité exhaustive et une documentation minutieuse à chaque étape de production, ce qui ralentit encore davantage l'adoption de nouvelles technologies. De plus, la gestion des risques y est particulièrement stricte, avec une tolérance au changement limitée en raison des enjeux de sécurité pour les patients. Tandis que d'autres industries adoptent rapidement de nouveaux modes de fonctionnement, l'industrie biopharmaceutique tend à rester prudente face à ces évolutions, privilégiant des approches incrémentales pour assurer la conformité réglementaire et la qualité des produits.

1.1.1.1. Aspects réglementaires du secteur du biomédicament

La production de biomédicaments est régie par un cadre réglementaire strict visant à garantir la qualité, la sécurité et l'efficacité des produits biopharmaceutiques. Ce cadre repose sur des réglementations internationales, européennes et nationales qui imposent des normes de fabrication rigoureuses afin de prévenir tout risque pour la santé publique. Parmi ces réglementations, les Bonnes Pratiques de Fabrication (BPF) et les Good Manufacturing Practices (GMP) occupent une place centrale dans la maîtrise des processus de production et de contrôle de la qualité des biomédicaments. De plus, tout médicament mis sur le marché fait l'objet d'une Autorisation de Mise sur le Marché, processus de contrôle géré par les autorités des pays concernés. Enfin, un certain nombre de réglementations et directives annexes s'appliquent à la fabrication de médicaments.

1.1.1.2. Les Bonnes Pratiques de Fabrication (BPF)

Les Bonnes Pratiques de Fabrication (BPF), définies en France, établissent un ensemble de normes et de recommandations destinées à assurer une production conforme aux exigences de qualité pharmaceutique. Elles sont une traduction littérale des GMP européennes (EU-GMP) et encadrent l'ensemble du cycle de vie d'un produit, depuis la conception jusqu'à la distribution, dont les principaux items sont les suivants :

- Système de gestion de la qualité : Un ensemble de procédures et de contrôles doit être mis en place pour assurer une fabrication conforme aux normes. La documentation, l'audit interne et l'amélioration continue des processus sont des exemples de pratiques employées pour répondre à ce point.
- Personnel qualifié et formé : Tout le personnel impliqué dans la fabrication doit être formé aux bonnes pratiques de fabrication et disposer des compétences nécessaires pour garantir la qualité des produits.
- Locaux et équipements adaptés : Les locaux de production doivent respecter des normes strictes en matière de conception pour éviter toute contamination croisée ou dégradation des produits. Ce point traite également de la maintenance des équipements qui doit être rendue possible « par design ».
- Maîtrise des matières premières et des produits finis : Toutes les matières utilisées doivent être tracées, contrôlées et stockées dans des conditions garantissant leur qualité.
- Documentation et traçabilité : Chaque étape du processus de fabrication doit être enregistrée et documentée de manière claire et précise pour assurer la traçabilité des produits. Cela permet notamment de pouvoir identifier rapidement l'origine d'un problème en cas de non-conformité.

- Contrôle en cours de production : Des tests et des inspections sont réalisés à différentes étapes de la fabrication afin de garantir la conformité aux spécifications et d'éviter toute dérive.
- Gestion des déviations et des risques : Tout écart par rapport aux procédures établies doit être documenté, traité et rectifié. Une approche basée sur l'évaluation des risques permet d'anticiper et de minimiser les incidents.
- Procédures de nettoyage et d'hygiène : Des protocoles stricts de nettoyage et de désinfection doivent être respectés pour limiter les risques de contamination microbiologique, chimique ou croisée.
- Gestion des réclamations et rappels de produits : Un système de gestion des non-conformités doit être mis en place pour traiter les éventuels problèmes posés aux patients (allergies, etc.) et si nécessaire, organiser le rappel des produits non-conformes.
- Audit et amélioration continue : Les entreprises doivent mettre en place des audits internes réguliers pour vérifier le respect des BPF et identifier les axes d'amélioration.

Les EU-GMP sont inscrites dans le Règlement (UE) 1252/2014, qui complète la directive 2001/83/CE du Parlement européen et du Conseil relative aux médicaments à usage humain. En France, le Code de la Santé Publique, l'Agence Européenne du Médicament (EMA) et l'Agence Nationale de Sécurité du Médicament et des produits de santé (ANSM) assurent leur mise en application et leur contrôle à travers des inspections régulières. Il est important de noter que l'ensemble des informations données ici sont des extraits des BPF actuelles et que ces règles sont susceptibles d'évoluer.

Concernant la maintenance, l'exigence 5.10 dispose que les équipements utilisés pour la fabrication des intermédiaires et des substances actives doivent être conçus, dimensionnés et implantés de façon appropriée à leur utilisation, leur nettoyage, leur décontamination (si nécessaire) et à leur entretien. L'exigence 5.20 des EU-GMP quant à elle stipule que des procédures écrites, incluant l'attribution claire des responsabilités, ainsi qu'un planning détaillé, doivent être établis pour la maintenance préventive des équipements.

1.1.1.3. Les current Good Manufacturing Practices (cGMP)

Les current Good Manufacturing Practices (cGMP) sont l'équivalent américain des GMP européennes et sont utilisés aux États-Unis, sous la supervision de la Food and Drug Administration (FDA), ainsi que dans d'autres régions telles que l'Asie et l'Amérique du Sud. Les cGMP sont définies dans la réglementation 21 CFR Part 210-211 aux États-Unis et s'appliquent aux industries pharmaceutiques, biopharmaceutiques et dispositifs médicaux notamment. Malgré quelques différences notables, les cGMP couvrent des domaines similaires aux GMP. Tout comme les GMP, l'objectif des cGMP est d'établir un cadre rigoureux couvrant toutes les étapes de la production, depuis l'approvisionnement en matières premières jusqu'à la distribution des produits finis, afin de réduire les risques liés aux erreurs humaines, aux contaminations et aux variations dans les procédés de fabrication. Globalement, les cGMP couvrent les items suivants :

- Système d'assurance qualité et gestion des risques : Les entreprises doivent établir un Quality Management System (QMS) structuré, intégrant une évaluation continue des risques. L'approche « Quality by Design » (QbD) est encouragée pour anticiper les problèmes et optimiser la production.

- Personnel qualifié et formé : Tous les employés doivent être formés régulièrement aux GMP, avec des mises à jour sur les nouvelles réglementations et technologies.
- Environnement de production et équipements conformes : Les infrastructures doivent être pensées pour limiter les contaminations et faciliter le nettoyage. Chaque équipement utilisé doit être qualifié et maintenu régulièrement pour garantir une fabrication uniforme.
- Maîtrise des matières premières et des produits finis : Les fournisseurs sont contrôlés, les matières premières doivent être testées et certifiées avant leur utilisation. La reproductibilité de chaque étape de fabrication doit être garantie.
- Documentation et traçabilité : Tous les processus de fabrication doivent être consignés dans des Master Batch Records (MBR). Une traçabilité totale est nécessaire pour chaque lot de production de sorte à traiter rapidement les écarts.
- Tests et contrôles qualité : Les entreprises doivent effectuer des tests en cours de production pour s'assurer du respect des spécifications.
- Gestion des déviations et actions correctives/préventives : Tout écart par rapport aux spécifications doit être documenté via un système de gestion des déviations. Des actions correctives et préventives (CAPA – Corrective and Preventive Actions) doivent être mises en place pour éviter que l'erreur ne se reproduise.
- Procédures de rappel et gestion des réclamations : Un plan de rappel doit être établi pour retirer rapidement du marché les lots défectueux. Les retours des patients doivent être analysés pour identifier d'éventuels problèmes de production.

Les sections § 211.63, § 211.65 et § 211.67 des GMP américaines (FDA) définissent les exigences pour la maintenance des équipements. Selon § 211.63, les équipements utilisés dans la fabrication, le traitement, le conditionnement ou le stockage des produits médicamenteux doivent être de conception appropriée, de taille adéquate et correctement implantés pour faciliter leur utilisation, leur nettoyage et leur maintenance. La section § 211.65 exige que les surfaces en contact avec les composants, matériaux en cours de fabrication ou produits finis ne soient ni réactives, ni additives, ni absorbantes, afin de préserver la sécurité, l'identité, la qualité et la pureté du produit. Enfin, § 211.67 stipule que les équipements doivent être nettoyés et maintenus à des intervalles appropriés pour éviter tout dysfonctionnement ou contamination affectant la qualité du produit. Des procédures écrites doivent détailler les responsabilités, les calendriers de maintenance et de nettoyage, les méthodes, matériaux, le démontage/remontage la protection des équipements propres et l'inspection de leur propreté avant usage. Des registres de ces opérations doivent être tenus conformément aux §§ 211.180 et 211.182.

Il est important de noter que l'ensemble des informations données ici sont des extraits des cGMP actuelles et que ces règles sont susceptibles d'évoluer.

1.1.1.4. Différences entre les EU-GMP (ou BPF en France) et les cGMP

Bien que les EU-GMP et les cGMP poursuivent des objectifs similaires, visant à assurer la qualité des médicaments, les entreprises pharmaceutiques opérant à l'échelle internationale doivent se conformer aux référentiels en fonction du marché cible. Par exemple, une entreprise européenne souhaitant commercialiser ses produits aux États-Unis devra se conformer aux GMP américaines, en plus des exigences des EU-GMP.

1.1.1.5. Réglementation Européenne et Autorisation de Mise sur le Marché (AMM)

L'AMM (Autorisation de Mise sur le Marché) est une autorisation délivrée par une agence réglementaire, permettant la mise sur le marché d'un médicament. Elle est accordée après une évaluation des données cliniques, pharmaceutiques et toxicologiques pour garantir l'efficacité, la sécurité et la qualité du produit.

Plus spécifiquement, la délivrance porte en partie sur la composition et la fabrication du médicament pour garantir sa qualité et sa reproductibilité. Cela inclut :

- Substance active : Identification, pureté, stabilité et méthode de synthèse.
- Excipients : Rôle, compatibilité avec la substance active et absence de toxicité.
- Formulation : Justification du choix des formes galéniques (comprimés, injections, etc.).
- Processus de fabrication : Maîtrise des étapes de production et validation industrielle.
- Contrôles en cours de production : Critères de qualité (dosage, impuretés, uniformité, etc.).
- Conditionnement : Compatibilité avec le médicament, protection contre l'humidité/lumière.
- Stabilité : Études de conservation pour déterminer la durée de validité et les conditions de stockage.

En Europe, l'AMM peut être centralisée, délivrée par l'Agence Européenne des Médicaments (EMA) et valable dans tous les États membres, ou nationale, délivrée par une autorité nationale. Il existe également des procédures de reconnaissance mutuelle et décentralisée, permettant l'extension d'une AMM nationale à d'autres pays. Pour les biomédicaments, une AMM centralisée est obligatoire.

1.1.1.6. Normes Internationales

Au niveau international, l'Organisation Mondiale de la Santé (OMS) fournit des directives sur les bonnes pratiques de fabrication pour les produits pharmaceutiques. Ces directives servent de référence pour juger de la conformité aux BPF lors de l'examen des demandes d'autorisation de fabrication et lors des inspections des installations de production.

1.1.1.7. Autres réglementations applicables aux biomédicaments

En plus des EU-GMP et cGMP, plusieurs autres réglementations et directives s'appliquent au secteur biopharmaceutique, notamment :

- Directives ICH (International Council for Harmonization) :
 - ICH Q8 (développement pharmaceutique).
 - ICH Q9 (gestion des risques qualité).
 - ICH Q10 (systèmes qualité pharmaceutique).
 - Normes ISO :
 - ISO 9001 et ISO 13485, utilisées dans la gestion de la qualité des dispositifs médicaux associés aux biomédicaments.
- Pharmacopées (USP, EP, JP, Ch) : établissant les standards de pureté, d'efficacité et de sécurité des produits biopharmaceutiques.

Ces réglementations forment un cadre complexe, évolutif et nécessitant une veille réglementaire constante ainsi qu'une adaptation permanente des processus industriels. Les

entreprises opérant dans ce domaine doivent assurer la qualité et la sécurité de leurs produits, tout en restant compétitives sur le marché international et conformes aux législations.

1.1.1.8. Freins aux évolutions des processus

Comme discuté jusqu'ici, nous pouvons faire le constat que la production de médicaments et particulièrement de biomédicaments fait l'objet d'une régulation très forte de la part des autorités.

De ce fait, l'évolution des stratégies de maintenance dans l'industrie du biomédicament peut se voir ralentie par les exigences strictes des EU-GMP, cGMP et des obligations liées à l'Autorisation de Mise sur le Marché (AMM).

Les EU-GMP (EudraLex, Volume 4, Chapitre 3, 4 & 5) exigent que les équipements soient conçus, installés et entretenus de manière à minimiser les risques de contamination et de déviation. Toute modification des procédures de maintenance doit être validée, documentée et approuvée, ce qui limite l'évolution des pratiques. De plus, les EU-GMP imposent une traçabilité stricte des interventions, rendant la digitalisation complexe en raison des exigences sur l'intégrité des données (Annexe 11).

Les cGMP américaines (21 CFR Part 211.67) quant à elles, exigent un programme de maintenance préventive détaillé, avec des calendriers fixes et une documentation complète des actions réalisées. Cette approche rigoureuse rend difficile l'intégration de stratégies plus flexibles, telles que la maintenance conditionnelle, car nécessitant une revalidation et une démonstration de leur équivalence ou supériorité en matière de fiabilité.

L'AMM ajoute une contrainte supplémentaire car, toute modification significative des équipements ou de leur maintenance peut être considérée comme une variation nécessitant une nouvelle soumission aux autorités (FDA, EMA), retardant ainsi l'évolution des pratiques.

La cybersécurité et les réglementations sur l'intégrité des données (FDA Data Integrity, EudraLex Annexe 11) rendent la transition vers des systèmes connectés plus complexe, en raison des exigences sur la validation des logiciels et la sécurisation des accès.

Enfin, l'impact financier et organisationnel est un facteur clé. Les mises à niveau technologiques et la formation des équipes peuvent nécessiter des investissements lourds, difficiles à justifier face à des réglementations conservatrices privilégiant la stabilité des processus. Ainsi, bien que la modernisation de la maintenance soit un levier d'amélioration de la fiabilité et de la productivité, elle se heurte à des exigences réglementaires qui limitent l'adoption et / ou l'évolution rapide de stratégies / pratiques.

1.1.2. Objectif industriel

L'entreprise OCTAPHARMA SAS a fait connaître un besoin particulier de développement d'une approche prévisionnelle de maintenance. Cette évolution concerne les vannes à membranes pneumatiques. Les vannes sont largement représentées dans les industries aux procédés fluidiques, du fait de leur rôle dans le blocage des fluides. Une variante de ces vannes est équipée de membranes, de sorte à répondre à des critères de non-contamination des produits circulant dans les lignes de transfert des ateliers de production.

La vanne à membrane pneumatique est retrouvée en grande quantité sur les lignes de production de l'entreprise OCTAPHARMA SAS. L'emploi de cet équipement se justifie du fait de :

- La production de l'entreprise qui est dans sa grande majorité une production fluide
- Les exigences réglementaires et techniques strictes décrites plus haut et applicables aux industries du biomédicament

Dans ce contexte, les vannes à membrane jouent un rôle critique, assurant la régulation des flux tout en garantissant l'intégrité des produits. Cependant, ces vannes sont soumises à des sollicitations mécaniques répétées, des variations de température et des interactions chimiques avec les fluides traités, entraînant une usure progressive des membranes. La défaillance de ces composants peut provoquer des arrêts de production coûteux, des risques de contamination de produits et des pertes financières importantes.

Les pratiques actuelles de maintenance chez OCTAPHARMA SAS reposent majoritairement sur des plans préventifs systématiques basés sur des calendriers de maintenance fixes, qui ne tiennent pas compte de l'état réel des équipements. Ces plans sont complétés par une approche corrective venant en complément de la systématique. Ces méthodes présentent plusieurs limites :

- Coût élevé des arrêts de production imprévus, liés à des défaillances non-anticipées.
- Maintenance d'équipements dont les composants sont encore opérationnels.
- Non-optimisation des ressources, avec des interventions programmées sans justification réelle.
- Stock de pièces détachées important.

Ces enjeux mettent en évidence la nécessité de faire évoluer les plans de maintenance vers, lorsque c'est possible, des approches de maintenance prévisionnelle.

Ceci constituera le premier objectif industriel de cette thèse. Il va s'agir de proposer une évolution des pratiques de maintenance permettant d'anticiper les défaillances en exploitant des données pertinentes collectées sur les équipements concernés en temps réel.

Un deuxième objectif industriel consiste à permettre une intégration technique de ce plan de maintenance en proposant une solution compatible avec les systèmes de gestion de production existants (Architecture automatisme, MES, SCADA...)

Un troisième et dernier objectif industriel est de vérifier la compatibilité de ce plan de maintenance prévisionnelle avec l'ensemble des contraintes qualités appliquées à ce secteur. Il sera notamment discuté de la manière dont une telle évolution de pratique doit être validée pour être conforme avec les normes en vigueur.

Pour atteindre ces objectifs, ces travaux s'appuieront sur :

- Une étude technique de l'équipement cible
- Une étude de faisabilité et de compatibilité de l'architecture de pilotage automatisme présente chez OCTAPHARMA SAS pour l'intégration d'une solution basée sur les données
- Des techniques avancées d'analyse de données, incluant des méthodes d'intelligence artificielle et de surveillance conditionnelle.
- Une revue des exigences qualités actuelles auxquelles est soumise l'entreprise.

Ces objectifs industriels nous amènent à faire émerger un certain nombre de questions scientifiques. Pour y répondre, nous allons contextualiser la maintenance industrielle en regard

avec la recherche scientifique puis détailler les leviers scientifiques sur lesquels nous allons travailler.

1.2. Contexte scientifique de la maintenance industrielle

La maintenance industrielle se décline en plusieurs catégories, chacune adaptée aux besoins spécifiques des équipements et aux contraintes opérationnelles. Elle se divise en deux grandes familles : la maintenance corrective et la maintenance préventive, qui comportent chacune plusieurs sous-catégories.

1.2.1. Maintenance Corrective

La maintenance corrective s'applique après l'apparition d'une panne ou d'un dysfonctionnement. Elle vise à remettre un équipement immédiatement en état de fonctionnement. Elle se divise en deux sous-catégories :

Maintenance corrective palliative : Il s'agit d'une intervention temporaire permettant de maintenir un équipement en fonctionnement malgré une anomalie. Cette solution est souvent adoptée en cas d'urgence, lorsque l'arrêt immédiat n'est pas envisageable. L'objectif est de retarder une intervention définitive.

Maintenance corrective curative : Elle consiste à réparer ou remplacer les composants défectueux afin de restaurer les performances initiales de l'équipement. Cette intervention est définitive et vise à garantir un fonctionnement optimal après la panne.

Bien que souvent perçue comme coûteuse en raison des arrêts imprévus de production, la maintenance corrective reste indispensable pour gérer les défaillances imprévues et compléter les stratégies préventives exposées à la suite.

1.2.2. Maintenance Préventive

La maintenance préventive vise à éviter l'apparition des pannes en intervenant avant qu'un problème ne survienne. Elle se décline en trois sous-catégories :

Maintenance préventive systématique : Elle repose sur un calendrier d'interventions planifiées à intervalles réguliers, indépendamment de l'état réel de l'équipement. Cette approche, bien que simple à mettre en œuvre, peut générer des coûts excessifs si les interventions sont réalisées trop fréquemment. Elle ne garantit pas non plus l'apparition de défaillances inopinées.

Maintenance préventive conditionnelle : Elle repose sur la surveillance (temps réel ou non) des équipements à l'aide de capteurs et d'indicateurs de performance. Une intervention est déclenchée dès que certains seuils critiques sont atteints (vibration, température, pression, etc.). Cette approche optimise les coûts de maintenance en permettant une intervention lorsque cela est nécessaire.

Maintenance préventive prévisionnelle : Elle s'appuie sur des modèles de dégradation connus et des algorithmes spécifiques pour anticiper les pannes avant qu'elles ne surviennent, parfois avant même que des symptômes visibles de dégradation n'apparaissent. Cette approche permet de planifier les interventions de manière optimale en s'appuyant sur des pronostics. Elle s'inscrit dans le cadre du Prognostics and Health Management (PHM).

1.2.3. Prognostics Health Management

La maintenance prévisionnelle, intégrée dans le cadre du Prognostics and Health Management (PHM), vise à anticiper les défaillances des systèmes industriels en s'appuyant sur une analyse approfondie des données et des modèles de dégradation. Ce processus est défini et structuré en sept étapes, chacune fournissant des informations nécessaires à l'étapes suivantes (Gouriveau, 2013).

1.2.3.1. Acquisition des Données

La première étape consiste à collecter des données pertinentes sur le système ou l'équipement surveillés. Ces données peuvent être issues de capteurs, d'historiques de maintenance, de mesures externes ou indirectes... L'objectif est de disposer d'un ensemble de données complet et précis pour analyse.

1.2.3.2. Traitement des Données

Une fois les données acquises, elles doivent être traitées pour éliminer le bruit, gérer les valeurs manquantes et extraire des caractéristiques significatives. Des techniques de filtrage, de normalisation et de transformation sont appliquées pour préparer les données à l'analyse.

1.2.3.3. Évaluation de l'État

À cette étape, l'état actuel du système est évalué en comparant les données traitées à des modèles de référence ou à des seuils prédéfinis. Cela permet de déterminer si le système fonctionne normalement ou s'il présente des signes de dégradation.

1.2.3.4. Diagnostic

Le diagnostic vise à identifier les causes profondes des anomalies détectées lors de l'évaluation de l'état. Des techniques telles que l'analyse des modes de défaillance et de leurs effets (AMDE) ou des méthodes basées sur l'intelligence artificielle peuvent être utilisées pour isoler et comprendre les défaillances potentielles.

1.2.3.5. Pronostic

Le pronostic consiste à prédire l'évolution future de l'état de santé du système, notamment en estimant la durée de vie utile restante (Remaining Useful Life - RUL). Trois approches principales sont utilisées pour cette prédiction :

Approche basée sur des modèles physiques : Elle nécessite une compréhension approfondie des mécanismes de dégradation et l'élaboration de modèles mathématiques décrivant le comportement du système. Cette approche est efficace lorsque des modèles précis sont disponibles.

Approche basée sur les données : Elle s'appuie sur des techniques d'apprentissage automatique pour modéliser la dégradation à partir de données historiques. Cette méthode est particulièrement utile lorsque les modèles physiques sont difficiles à obtenir.

Approche hybride : Elle combine les deux approches précédentes pour tirer parti des avantages de chacune, améliorant ainsi la précision des prédictions.

1.2.3.6. Prise de Décision

Les informations issues du pronostic sont utilisées pour planifier des actions de maintenance optimales. L'objectif est de minimiser les coûts associés tout en évitant les défaillances non planifiées. Des stratégies telles que la maintenance conditionnelle ou la maintenance

opportuniste peuvent être mises en œuvre en fonction des estimations de la RUL et des contraintes opérationnelles.

1.2.3.7. Interface Homme-Machine

La dernière étape consiste à présenter les informations pertinentes aux opérateurs et aux décideurs via des interfaces intuitives. Cela inclut des tableaux de bord, des alertes et des recommandations d'actions, facilitant ainsi la prise de décision éclairée.

Ainsi, le PHM devient une trame proposant une application relativement standardisée au développement de la maintenance prévisionnelle dans l'industrie. Cette trame va nous servir de base de travail pour répondre à la majorité des questions scientifiques dont nous allons discuter plus bas.

1.3. Des objectifs industriels aux questions scientifiques

Nous avons discuté au point « 1.2 » de 3 objectifs industriels à cette thèse, avancés par l'entreprise OCTAPHARMA SAS dont le besoin exprimé est une évolution des pratiques de maintenance ciblant les vannes à membranes pneumatiques.

Pour chaque objectif industriel, une question scientifique ainsi qu'une proposition de contribution scientifique est proposée. Les questions scientifiques traitées feront l'objet de chapitres particuliers dans la suite du manuscrit.

1.3.1. Question 1 : *Des pratiques de maintenance dans le milieu du biomédicament*

Le 1^{er} objectif industriel consiste à proposer une évolution des pratiques de maintenance industriel dans le secteur du biomédicament. Nous avons ici à faire à un type d'industrie dont le procédé de fabrication est essentiellement fluide et particulièrement normé.

La 1^{ère} question scientifique que nous souhaitons traiter consiste à nous interroger sur les pratiques de maintenance dans les industries du biomédicament. Il s'agira ici de vérifier si, et de quelle manière, les principes de l'industrie 4.0 et avec eux, l'application de plans de maintenance prévisionnelle ont infusés dans notre secteur d'étude. L'entreprise exprimant le besoin d'évolution, est-elle seule en retard ? D'autres entreprises similaires ont-elles réussi une transition vers une maintenance prévisionnelle tenant compte de contraintes similaires ? A l'inverse, ce retard est-il général dans ce secteur ? Si oui, pourquoi ?

La 1^{ère} contribution proposée consiste donc en la création d'un état de l'art exposant à la fois :

- Les résultats d'une enquête menée auprès d'un certain nombre d'industriels du biomédicament, établissant les principales pratiques de maintenance, les équipements concernés et les volontés d'évolution des plans de maintenance.
- Les résultats d'une recherche dans la littérature scientifique des modèles de maintenance prévisionnelle, dans différents types d'industries et plus particulièrement dans le secteur du biomédicament.

Cet état de l'art aura également pour objectif de nous permettre de positionner nos travaux en mettant en évidence des gaps dans la littérature scientifique.

1.3.2. Question 2 : *Du modèle de dégradation de l'équipement*

En lien avec le 1^{er} objectif industriel, la 2^{ème} question scientifique abordée par cette thèse consiste à établir le modèle de dégradation des membranes de vannes pneumatiques, de sorte à permettre la création d'une maintenance prévisionnelle en regard des principes de la

méthode PHM. Cette dernière dispose que pour être efficace, il nous faut pouvoir traiter des données fiables et représentatives de la dégradation de l'équipement cible. En somme, est-il possible de caractériser la dégradation de l'équipement sur la base de données caractéristiques ? Également, au regard du 2^{ème} objectif industriel dont l'objet est de répondre à la capacité d'intégration technique du plan de maintenance, ces données sont-elles pertinentes et exploitables ? Peuvent-elles être accessibles à partir d'une ligne de production réelle ?

La 2^{ème} contribution proposée dans cette thèse réside dans la fourniture d'un modèle de vie et de comportement de l'équipement étudié basé sur des données concrètes et représentatives. Ce modèle fourni aura été obtenu en respectant les 4 premières briques de la méthode PHM. Les données disponibles et employées pour modéliser le comportement des membranes étant issues d'un dispositif de pilotage des fluides, nous opérerons une mise en lien entre l'état physique du matériau composant les membranes et les signaux recueillis. L'objectif étant de constater physiquement les déformations de l'équipement par-rapport aux dérives des signaux observées de sorte à démontrer la fiabilité du modèle comportemental.

Un complément à cette contribution consistera à couvrir le 2^{ème} objectif industriel par la description d'une architecture fonctionnelle de pilotage et d'acquisition de données en place chez OCTAPHARMA SAS permettant de collecter et traiter les données employées pour constituer le modèle de dégradation.

1.3.3. Question 3 : Du pronostic de la défaillance et de la prise de décision

Enfin, pour compléter le 1^{er} objectif industriel, il s'agira d'employer le modèle de vie de l'équipement pour en établir des pronostics de défaillance. Une fois le comportement de l'équipement et sa dégradation modélisés, comment traiter le modèle de sorte à fournir une approche de maintenance prévisionnelle complète ?

La 3^{ème} contribution proposée par ces travaux consiste à fournir une méthode de traitement du signal de dégradation fourni par le modèle comportemental préalablement proposé dans la question précédente. Il s'agira ici de proposer une méthode de calcul du RUL, de sorte à établir des pronostics fiables d'apparition de défaillances spécifiques aux membranes de vannes pneumatiques.

En complément du traitement de cette question, le 3^{ème} objectif industriel est abordé dans ce chapitre. Le calcul du RUL sera complété par une méthode de prise de décision adaptée aux environnements réglementaires fortement contrôlés. Cette méthode sera basée sur des scénarios concrets issus des pratiques actuelles de maintenance de l'entreprise OCTAPHARMA SAS.

1.4. Plan de thèse

À partir de trois objectifs industriels définis, ce travail propose une démarche scientifique structurée en quatre chapitres.

Un premier chapitre vise à poser les bases scientifiques et industrielles de la thèse en répondant à la première question scientifique : quelles sont les pratiques de maintenance dans le secteur du biomédicament ? Un état de l'art est établi à partir d'une enquête auprès d'industriels du biomédicament pour recenser les pratiques actuelles, les équipements ciblés et les aspirations à évoluer vers une maintenance prévisionnelle. Parallèlement, une revue de

la littérature scientifique explore les modèles de maintenance prévisionnelle existants, en comparant le secteur biopharmaceutique aux autres industries manufacturières. Ces travaux permettent d'identifier les lacunes dans la recherche, notamment sur l'adoption des principes de l'industrie 4.0 dans ce domaine, et de positionner les contributions de la thèse.

Le chapitre suivant traite de la deuxième question scientifique : établir un modèle de dégradation des membranes de vannes pneumatiques pour une maintenance prévisionnelle basée sur la méthode PHM. L'étude débute par une analyse expérimentale du comportement de l'équipement cible dans des conditions définies, représentatives de conditions de procédé réelles. Quatre premières étapes du PHM seront appliquées : acquisition des données via une architecture fonctionnelle existante chez OCTAPHARMA SAS, manipulation des données, monitoring de la condition, et établissement de seuils pour identifier les patterns des phases de vie de l'équipement. Les résultats expérimentaux font l'objet d'une étude comparative physique pour vérifier la fiabilité du modèle. Ce chapitre propose ainsi un modèle de dégradation concret, répondant au deuxième objectif industriel.

Pour faire suite à l'établissement du modèle de dégradation, le chapitre suivant répond à la troisième question scientifique, proposant une méthode d'établissement du pronostic de défaillance basée sur le modèle de dégradation fourni au chapitre précédent. Une méthode de classification des phases de vie ainsi que du calcul du RUL est développée, spécifique aux membranes de vannes pneumatiques, afin de prédire les défaillances. Une approche data-driven est ainsi développée pour traiter ce cas d'usage, basée sur une étude comparative des performances de différentes techniques de machine learning.

Enfin, un dernier chapitre synthétise les contributions de la thèse en proposant un modèle global de maintenance prévisionnelle adapté aux processus fluidiques, avec un focus sur la dernière brique du PHM : l'aide à la décision. En exploitant les prévisions de dégradation et le RUL issus du chapitre 3, il s'agit de transformer ces données en recommandations concrètes, de sorte à planifier les interventions lors des arrêts programmés ou hors production pour maximiser la fiabilité des vannes à membranes, réduire les coûts et éviter les arrêts non planifiés. Une méthode de validation du modèle, basée sur l'évaluation qualitative post-intervention des membranes, est proposée dans ce chapitre. Partant du contexte retrouvé chez OCTAPHARMA SAS, ce modèle est ensuite discuté pour son extensibilité à d'autres industries fluidiques (nucléaire, oil & gas, agroalimentaire), en tenant compte des similitudes d'équipements et des contraintes réglementaires. Des perspectives d'amélioration continue clôturent ce chapitre.

Une conclusion récapitulera les réponses aux trois questions scientifiques et les contributions apportées aux objectifs industriels d'OCTAPHARMA SAS. Elle soulignera l'avancement des pratiques de maintenance dans le biomédicament, la modélisation comportementale des vannes pneumatiques et l'intégration d'une maintenance prévisionnelle via le PHM. Enfin, elle mettra en lumière les perspectives d'évolution et les implications pour le secteur.

Nous entamons ainsi ce travail par une étude approfondie des pratiques de maintenance dans le secteur du biomédicament, de sorte à dresser un état des lieux objectif, identifier les manques et positionner les apports de cette recherche.

Chapitre 2

2. Des pratiques de maintenance dans le milieu du biomédicament

Des suites de l'émergence du concept de l'Industrie 4.0 il y a une douzaine d'années maintenant, la généralisation de la digitalisation, des objets connectés et des bases de données massives révolutionne les processus industriels. Les données générées par les équipements de production deviennent de plus en plus diverses et peuvent être exploitées dans des domaines tels que la maintenance, faisant ainsi évoluer des pratiques auparavant exclusivement correctives vers des approches fondées sur la condition des équipements, voire vers une maintenance prévisionnelle. Cette évolution implique d'enrichir des plans de maintenance centrés sur des plannings établis selon des recommandations constructeurs au profit d'une écoute et d'un monitoring actif des équipements, permettant une maintenance au plus juste, c'est-à-dire exploiter le plus d'informations possibles émises d'un appareil monitoré tout en limitant le risque d'apparition d'une défaillance. Ces nouvelles pratiques, souvent désignées sous le terme de maintenance pilotée par l'IOT (Internet of Things) ou "IOT-driven maintenance", sont abondamment décrites dans la littérature scientifique pour la plupart des secteurs manufacturiers.

Au travers de ce chapitre, nous souhaitons mettre en lumière tout d'abord, le retard des travaux de recherche dans le secteur du biomédicament par-rapport à l'industrie manufacturière générale, et ensuite, l'emploi des méthodes basée sur l'IOT dans le milieu biopharmaceutique. Cette revue de littérature a permis d'identifier une lacune importante dans les travaux consacrés à la maintenance basée sur l'IoT, en particulier dans le secteur biopharmaceutique. Cette constatation suggère qu'il existe une opportunité notable pour apporter une contribution scientifique nouvelle et pertinente à ce domaine spécifique. Le contenu de ce chapitre a fait l'objet d'une publication (Wadel et al., 2024) dans un journal indexé.

2.1. Introduction

L'objectif de cet état de l'art est d'identifier dans la littérature scientifique, les publications traitant des évolutions de l'industrie du biomédicament vers les concepts de la maintenance 4.0 et l'emploi de technologies IOT / data-driven pour la maintenance prévisionnelle. Dans le cadre de notre collaboration avec un acteur majeur du biomédicament, nous souhaitons également, à la suite de ce chapitre, proposer un cas d'usage de maintenance prévisionnelle IOT / data-driven s'appliquant à des équipements génériques largement utilisés sur des lignes de productions biopharmaceutiques, à savoir les vannes à membranes pneumatiques.

2.1.1. Contexte

Depuis l'émergence des principes de l'industrie 4.0 (Pereira and Romero 2017, , Nimawat et al., 2023), l'emploi d'objets connectés (de Oliveira et al, 2022) et l'adaptation des bases de données (Zhang et al, 2019) pour le pilotage des processus industriels, ne cesse d'évoluer.

La maintenance prévisionnelle se définit comme étant un type de maintenance conditionnelle dont la surveillance permet la prévision des défaillances (Kent, 2018). Les travaux actuels dans ce domaine emploient le processus PHM (Prognostics Health Management) process (Baur et al., 2020) dont les étapes essentielles sont décrites au point 1.2.3.

De nombreux exemples dans la littérature scientifique décrivent les différents types d'industries manufacturières ayant franchi le cap de l'utilisation de la donnée émise pour établir les plans de maintenance (Méchin, 2005, Ayo-Imoru et al, 2018, Wongchai, 2022). Généralement, il ne s'agit que d'une fraction du parc machine qui est concernée par la maintenance 4.0. Pour le reste du parc, les responsables de maintenance utilisent d'autres types de maintenance traditionnelle tels que la maintenance corrective ou la maintenance préventive systématique par exemple. Ceci s'explique en partie par la difficulté d'acquérir les données nécessaires auprès des équipements qui disposent ou non de capteurs embarqués, ce qui empêche ou tout du moins ralentit la mise en place de la maintenance prévisionnelle (Compare et al, 2020). Et d'autre part, par la nécessité d'avoir des ressources humaines et matérielles pour maîtriser les outils de monitoring et de l'analyse de données en temps réel. Ce point est abordé par Silvestri et al, (2020) qui décrivent l'impact technique et humain de l'instauration des concepts de l'industrie 4.0 sur les opérateurs de maintenance, induisant une mutation des activités de maintenance.

L'industrie biopharmaceutique, du fait de ses spécificités liées notamment à la nature de sa production, son contexte réglementaire et son système qualité contraignant, rencontre plus de difficultés à évoluer que d'autres industries manufacturières. En effet, un médicament commercialisé doit faire l'objet au préalable d'une autorisation de mise sur le marché (Caulin, 2008). Cette autorisation conditionne notamment la façon de produire un médicament de sorte à garantir la qualité, la pureté, l'identité et la sécurité pour le patient. Ce faisant, toute modification de l'outil de production, par exemple dans le cadre d'une modernisation des équipements, s'avère compliquée puisque cela nécessite une nouvelle approbation des autorités sanitaires. Ces dernières sont tenues de contrôler la conformité de la production et de la maintenance des équipements avant la mise sur le marché.

La maintenance dans le secteur biopharmaceutique s'opère dans un contexte réglementaire et nécessite une assurance qualité rigoureuse (Dollard, 2016, Ahmed, 2013, Christiansen, 2020) propre aux environnements de production biopharmaceutiques. Ceci implique de maintenir des équipements de production dans un environnement contraignant et normé. Outre le fait d'impacter les activités de production, une maintenance doit se dérouler de façon procédurale, de sorte à garantir l'absence de contaminations et plus largement, la qualité du médicament produit. Un industriel de la biopharmaceutique s'orientera naturellement vers des méthodes qui, même n'étant pas parfaites, lui permettront de justifier lors d'une inspection, d'une maîtrise et d'un suivi rigoureux des opérations de maintenance. Aujourd'hui, nous constatons une organisation des activités de maintenance tenant compte des contraintes citées plus haut. Ces activités se déroulent largement selon un plan de maintenance systématique. Traditionnellement, cela veut dire compiler l'ensemble des pièces d'usure d'un équipement et, en se basant sur les recommandations des constructeurs, établir un calendrier de remplacement de ces pièces à échéances fixes. Ceci permet de démontrer une maîtrise de l'équipement contrôlé lors d'une inspection. Ce type de plan permet, sur le papier, un maintien en état des équipements de production de sorte à garantir une limitation de leur usure, moyennant un remplacement ou une remise à neuf fréquents des éléments d'usure. Cependant, ce plan de maintenance ne met pas à l'abri des défaillances subites, liées à un montage non-conforme ou à une pièce défectueuse.

2.1.2. Enquête auprès d'industriels du biomédicament

Pour déterminer si le plan systématique n'est retrouvé que localement ou à l'inverse et si l'ensemble des entreprises du biomédicament est concerné par cette pratique, nous avons adressé un sondage à l'ensemble des grands groupes industriels du biomédicament. La méthodologie employée pour interroger ces entreprises s'inspire des méthodes de Saleh et Bistra (2017) et de Koundinya et al. (2016). Cela implique que l'enquête doit suivre certaines recommandations pour augmenter les chances de réponse du destinataire. Par exemple, envoyer une enquête à des fins académiques, avec des questions claires et concises nécessitant peu de temps de réponse, utiliser un format en ligne, garantir l'anonymat du destinataire et soumettre des questions à choix multiples.

Les résultats ont montré que 100 % des sites de production disposaient d'un plan de maintenance systématique (figure 1). Cependant, les plans de maintenance conditionnelle (figure 2) et prévisionnelle (figure 3) ne sont que complémentaires et dans des proportions marginales ne concernant principalement que les équipements situés hors des Zones à Atmosphère Contrôlée (ZAC). De plus, la majorité des fabricants ont déclaré que leurs équipements de production étaient situés dans une ZAC (figure 4). L'enquête a également montré que les équipements de production étaient globalement similaires d'un site à un autre. Ces mêmes fabricants sont à 100 % favorables à un changement des pratiques de maintenance au cours des cinq prochaines années. Les résultats de l'enquête ont indiqué que, malgré l'émergence des concepts de l'Industrie 4.0 au cours de la dernière décennie, les industriels du biomédicament continuent de maintenir des pratiques de maintenance essentiellement systématiques et la transition vers des concepts de maintenance conditionnelle et prévisionnelle reste marginale. Des constats tirés de cette enquête, ce chapitre porte d'une part sur l'identification des évolutions des pratiques de maintenance IOT / data-driven dans l'industrie du biomédicament et dans un second temps sur les modèles employés issus notamment du PHM qui sont aujourd'hui applicables dans l'industrie.

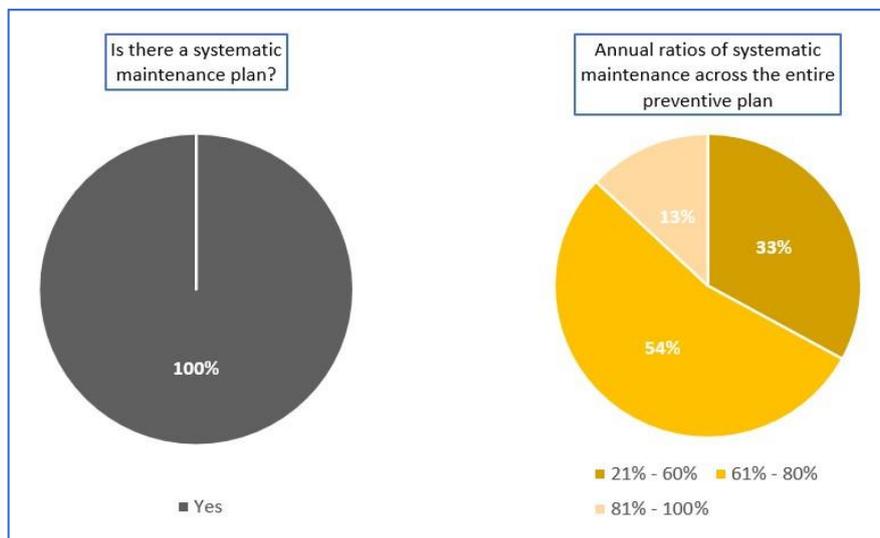


Figure 1 - Présence d'un plan de maintenance systématique et ratios annuels de maintenance systématiques VS l'ensemble des maintenances préventives

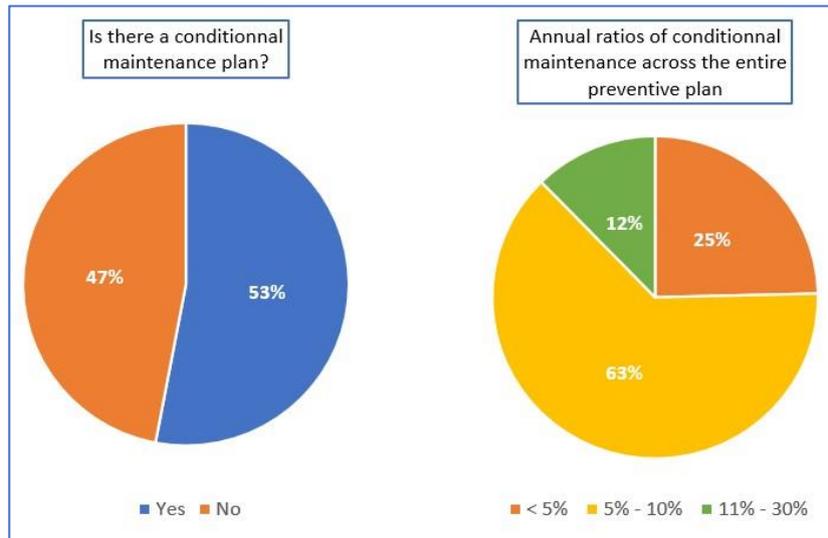


Figure 2 – Présence d'un plan de maintenance conditionnelle et ratios annuels de maintenance conditionnelles VS l'ensemble des maintenances préventives

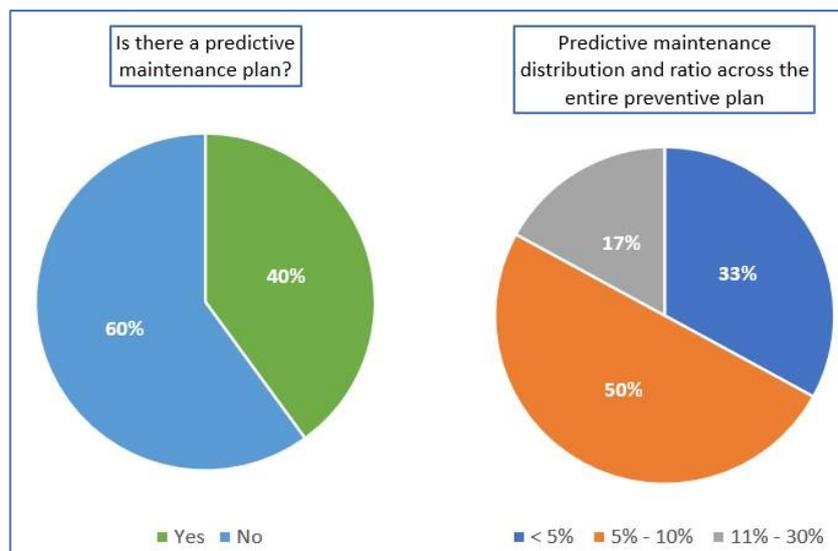


Figure 3 - Présence d'un plan de maintenance prévisionnelle et ratios annuels de maintenance prévisionnelles VS l'ensemble des maintenances préventives

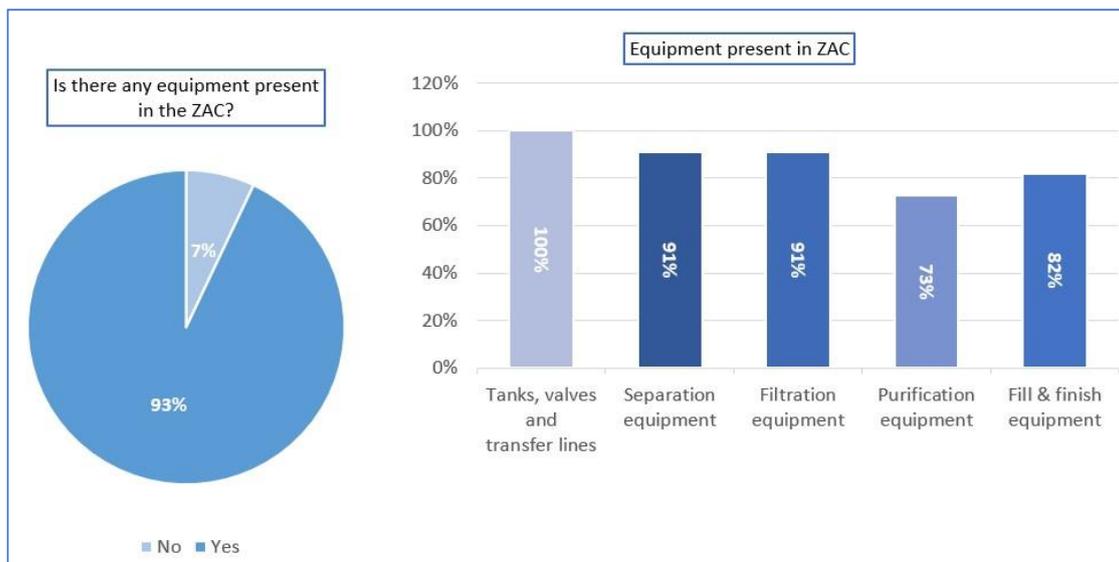


Figure 4 - Taux de présence et type d'équipements de production en ZAC

Pour ce qui suit, la section 2.2. est consacrée à la présentation de notre méthodologie de recherche bibliographique. La section 2.3. est consacrée à la contextualisation de l'utilisation de l'IOT et des méthodes pilotées par les données pour les procédés industriels, en particulier la planification de la maintenance prévisionnelle, de 2011 à nos jours. La section 2.4. présente ensuite les différentes méthodes d'analyse et de modélisation du comportement des équipements de production industrielle. La section 2.5. est consacrée à une discussion des résultats de l'étude actuelle. Enfin, nous concluons cette étude dans la section 2.6.

2.2. Méthodologie de recherche

Il existe une abondante littérature scientifique concernant l'utilisation de l'IOT pour la maintenance prévisionnelle. Afin d'évaluer le nombre de publications spécifiques à l'industrie biopharmaceutique par rapport à l'industrie manufacturière en général, nous avons effectué une recherche bibliographique couvrant la période de 2011 à aujourd'hui.

Cependant, les chiffres pour 2023 sont susceptibles d'être modifiés, car cette étude est menée en cours d'année. La majeure partie de notre littérature a été consultée à l'aide de la base de données Google Scholar, qui croise des articles de plusieurs revues, dont IEEE, Springer, Elsevier & Emerald, Taylor & Francis, Science Direct, etc.

2.2.1. Détail des requêtes

Nous avons commencé avec un ensemble de mots-clés indiquant les thèmes recherchés, ainsi qu'un champ de recherche générique. Les requêtes portent sur l'ensemble du texte et non uniquement sur le titre des publications. Les critères d'inclusion et d'exclusion sont listés dans le tableau suivant :

| | |
|---|---|
| 1 | Identification des domaines industriels d'intérêt pour la recherche Période comprise entre 2011 et 2023 |
| 2 | Articles en anglais conservés. Exclusion des termes relatifs à la maintenance conditionnelle Analyse qualitative des articles concernés par la dernière exclusion |
| 3 | Exclusion des chapitres d'ouvrages, des articles non pertinents et des doublons |
| 4 | Inclusion des articles après vérification des références des articles sélectionnés |

Tableau 2 - Critères d'inclusion et d'exclusion

Le choix de l'ordre des requêtes est le suivant :

Première requête : «"maintenance" AND "PHM" AND "RUL" AND "industry".

Nous avons ensuite sélectionné des domaines industriels spécifiques, similaires en termes de processus, ainsi que le domaine de notre étude de cas.

Deuxième requête : "maintenance" AND "PHM" AND "RUL" AND "food OR pharmaceutical OR refinery OR biopharmaceutical".

Les termes « -CBM » et « -condition-based » font référence à des articles qui ne traitent pas spécifiquement de la maintenance prévisionnelle. Ces termes sont supprimés lors de la troisième requête afin de cibler uniquement les articles liés à la maintenance prévisionnelle. Une analyse qualitative est effectuée afin d'identifier les publications potentiellement éliminées par cette requête, mais toujours pertinentes.

Troisième requête : "maintenance" AND "PHM" AND "RUL" AND "food OR pharmaceutical OR refinery OR biopharmaceutical" -"condition-based" -"CBM".

2.2.2. Diagramme PRISMA

En nous appuyant sur la méthode présentée par Peters et al. (2017), nous avons sélectionné un ensemble de publications que nous jugeons pertinentes à l'aide du diagramme PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*). Cette sélection s'effectue en plusieurs étapes, comme illustré à la figure 5 :

a. Identification : Les bases de données, les revues et les revues de la littérature sont identifiées.

b. Sélection : Examen des premiers résultats de recherche. Une sélection préliminaire est effectuée sur la base de critères d'exclusion prédéfinis. Les exclusions sont justifiées.

c. Éligibilité : Les documents issus de cette première sélection sont soigneusement évalués afin de déterminer leur qualité et leur pertinence par rapport au sujet. Une seconde sélection est effectuée et justifiée.

d. Inclusion : Sélection des documents répondant à tous les critères d'inclusion. Les sources des articles sélectionnés sont examinées afin d'identifier des articles susceptibles d'être intéressants et dignes d'être inclus dans l'article.

La première requête nous a permis d'obtenir un résultat significatif de 4780 publications, car nous avons ciblé le secteur au sens large. En introduisant le terme « PHM » au début de la requête, nous avons centré notre recherche sur les pratiques de « maintenance 4.0 ». Nous avons ensuite restreint le champ de recherche au secteur industriel d'intérêt, ainsi qu'aux secteurs proches en termes de processus, afin de pouvoir comparer les pratiques. Ces secteurs comprennent les produits pharmaceutiques, biopharmaceutiques, les produits alimentaires et le raffinage.

Le nombre de publications est donc réduit à 435. Nous avons appliqué deux critères d'exclusion : conserver uniquement les publications en anglais et supprimer les articles mentionnant la maintenance conditionnelle, et conserver uniquement les publications en anglais traitant exclusivement de la maintenance prévisionnelle. Cependant, la maintenance prévisionnelle étant une forme de maintenance conditionnelle, et afin d'éviter d'exclure des articles susceptibles d'entrer dans le champ de notre étude, nous avons réalisé une analyse qualitative de tous les articles exclus par les mots-clés « CBM » et « conditionnelle ».

Cela a réduit le champ à 150 publications, dont 10 ont été extraites de l'analyse qualitative mentionnée ci-dessus. Une sélection supplémentaire a été effectuée, au cours de laquelle nous avons supprimé les études non-accessibles, les doublons et les thèses, ainsi que les articles non pertinents pour notre étude. Les thèses sont exclues car leurs contributions sont publiées dans des articles. Dans le cas contraire, nous considérons que leur contribution n'est pas suffisamment significative. Nous avons également inclus les articles auxquels nous n'avons pas pu accéder.

Nous avons retenu 43 articles à ce stade de l'analyse. Une étude a été réalisée à partir des références bibliographiques contenues dans ces 43 articles. Conformément à la méthodologie, nous avons identifié huit articles que nous jugeons pertinents, traitant généralement de l'évolution des pratiques de maintenance conventionnelles vers des méthodes IOT-driven. Nous tenons à souligner que, malgré les mots-clés sélectionnés, certains des résultats présentés concernent effectivement des cas d'application de maintenance prévisionnelle, mais dans d'autres domaines comme l'aéronautique. Nous avons décidé d'inclure ces articles dans notre étude car ils traitent de modèles applicables aux équipements mécaniques et transposables au domaine biopharmaceutique. En revanche, nous n'avons pas inclus les publications traitant de modélisation physique, par exemple les études sur la durée de vie utile des batteries au lithium (Catelani et al., 2022).

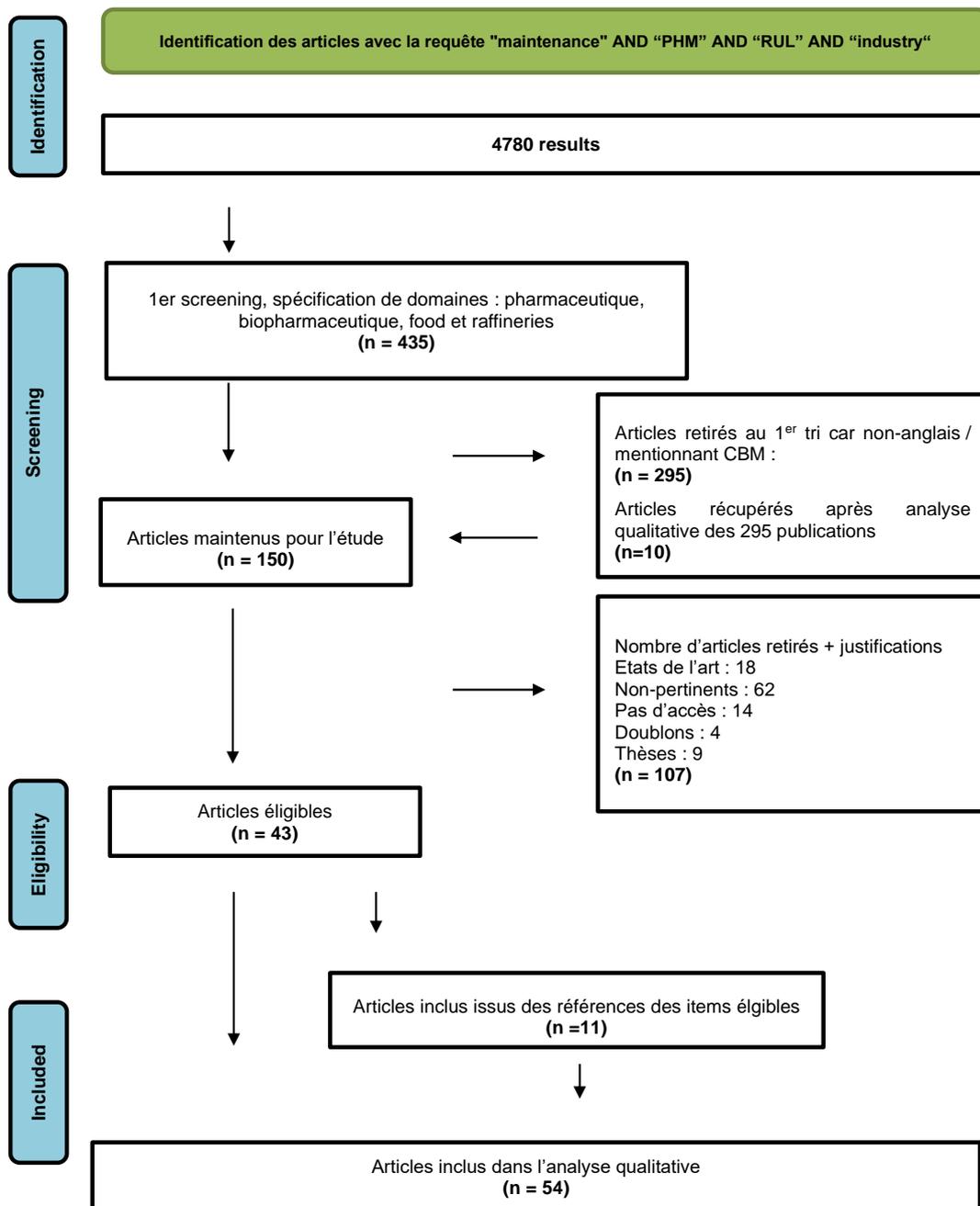


Figure 5 - Diagramme PRISMA

Les résultats de cette recherche bibliographique sont les suivants : 54 articles ont été sélectionnés après traitement. La répartition des publications pour chaque requête est présentée dans la figure 6. Ces publications sont réparties comme suit :

- 8 articles sont analysés dans la section 3 et traitent de l'évolution des pratiques de maintenance vers des méthodes basées sur l'IoT.
- 46 articles sont analysés dans la section 4 et traitent des modèles basés sur l'IoT pour la maintenance prédictive dans l'industrie.

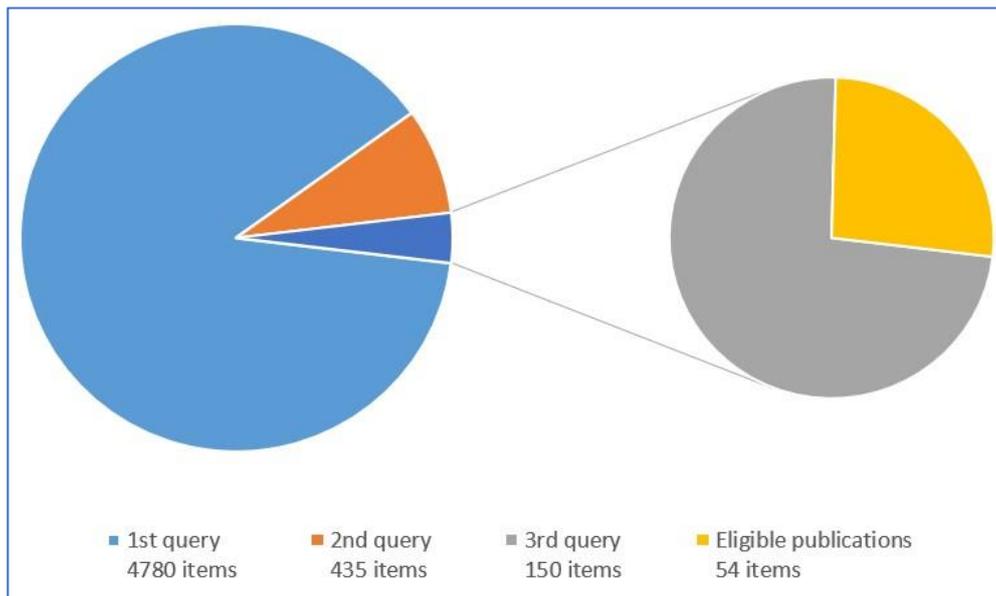


Figure 6 - Distribution des publications pour chaque requête

2.3. Évolution vers une maintenance IOT / data-driven

Dans cette section, nous souhaitons fournir un contexte chronologique pour l'adoption de l'IOT par le monde industriel, et plus particulièrement pour la maintenance des équipements. Fantana et al. (2013) ont décrit les applications de l'IOT dans un contexte industriel général, soulignant les défis techniques liés à son utilisation, notamment dans le domaine de la maintenance. Les auteurs ont souligné la nécessité de standardiser ce type d'équipements connectés. Il est important de considérer que les IOT ont une durée de vie qui doit être définie dès la conception d'une ligne de production et qu'ils doivent également faire l'objet d'un plan de maintenance. Les auteurs ont également souligné l'absence de norme concrète pour l'IoT, qui peut communiquer dans différents langages et à différentes fréquences. L'article de Fantana et al. (2013) présente deux cas d'usage. Le premier concerne l'industrie pharmaceutique et l'utilisation de l'IOT dans la chaîne d'approvisionnement. Il permet d'améliorer les processus de traçabilité afin de minimiser les erreurs d'adressage, la qualité dès la conception et l'approche réglementaire basée sur le risque utilisée pour encadrer l'évolution de ce processus. Le deuxième cas concerne l'industrie du raffinage (pétrole et gaz) et l'utilisation de l'IOT pour la surveillance à distance des processus de production. L'étude s'appuie sur les résultats d'une enquête menée lors d'un atelier sur le thème « IOT » afin de déterminer les attentes des industriels vis-à-vis des technologies IOT par rapport aux applications actuelles. Selon l'enquête, l'évolution de la maintenance arrive en deuxième position des attentes des industriels, juste derrière l'automatisation et le contrôle-commande (voir figure 7). Il convient de noter que la conférence a eu lieu en 2012, juste après l'avènement des concepts de l'industrie 4.0. L'enquête a révélé que la maintenance pilotée par l'IOT a été un domaine d'intérêt précoce pour les industriels. Les auteurs concluent en alertant les entreprises sur la nécessité de prendre en compte les critères de robustesse, de standardisation et de facilité d'accès et de configuration lors de la création d'un processus industriel utilisant l'IOT. Shah et al. (2019) ont décrit l'évolution de l'utilisation de l'IOT dans la chaîne d'approvisionnement. Ils se sont concentrés sur la transformation des pratiques de gestion logistique impactées par l'intégration de l'IOT dans les processus afin de réduire les coûts et de maximiser les profits. Les auteurs ont utilisé une recherche bibliographique par

mots-clés pour évaluer l'intérêt de l'IOT pour les chaînes d'approvisionnement dans les publications scientifiques. Ils ont conclu que, malgré l'usage répandu du terme « IOT », très peu d'articles décrivent des mises en œuvre concrètes de son utilisation dans les chaînes d'approvisionnement.

Selon Reinhardt et al. (2020), la maintenance préventive des fabricants de produits pharmaceutiques en Irlande arrive en première position parmi les processus faisant l'objet de projets planifiés pour une évolution vers les concepts de l'industrie 4.0. L'acquisition et le stockage des données arrivent en deuxième position.

Un article de Filz et al. (2021) a montré que les méthodes basées sur l'IOT peuvent intégrer les informations issues de l'analyse des risques des équipements pour la mise à jour automatisée des calendriers de maintenance et ainsi orienter la prise de décision en fonction des impacts potentiels d'une défaillance d'un équipement plus ou moins critique.

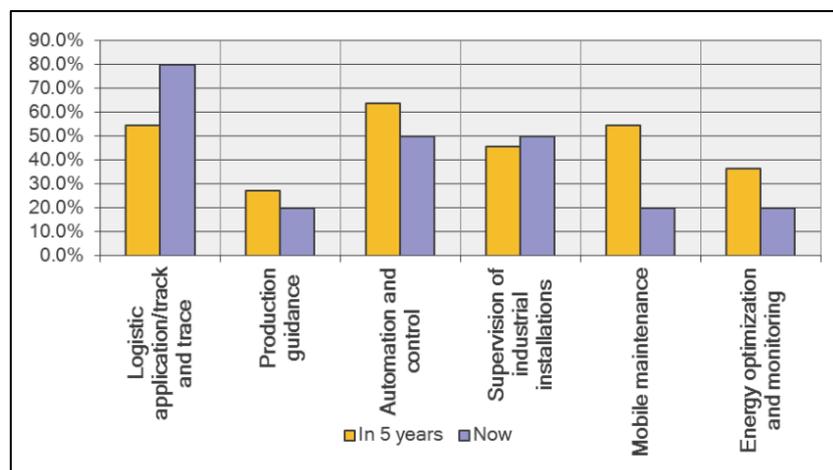


Figure 7 - Résultats de sondage, Fantana et al (2013)

Turner et al. (2022) ont présenté dans leur article un cas d'utilisation appliqué à l'industrie automobile, combinant la maintenance IOT-driven et l'économie circulaire pour suivre la dégradation et les pannes des équipements, fournissant ainsi un indicateur de santé gradué. En fonction du score de cet indicateur, ils réparent ou remplacent l'équipement par un neuf.

S'appuyant sur des recherches et des analyses de la littérature scientifique et technique, Kaliyannan et al. (2023) ont dressé un aperçu des préoccupations actuelles en matière de maintenance IOT-driven. Ils ont présenté une étude de cas visant à établir la viabilité de la technologie proposée pour permettre l'évaluation continue des machines via des équipements IOT alimentés par batterie. Pour ce faire, les auteurs ont conçu un banc d'essai pour tester 33 équipements IOT alimentés par batterie, qui transmettent des données de vibration et de température. Ils ont testé la viabilité d'un plan de maintenance basé sur les données rapportées par l'IOT, en tenant compte du temps nécessaire à leur remontée. Des prévisions de durée de vie sont ensuite établies en étudiant la consommation énergétique et les paramètres physiques.

Sawang Sri et Prasithmett (2023) ont présenté un algorithme appelé « Demo-Site », qui devrait permettre la maintenance électronique au sein d'une usine intelligente en exploitant les données remontées par l'IOT. Sur une plateforme de test reproduisant la conception d'un convoyeur industriel, six capteurs ont été installés et connectés à une base de données via un

routeur Wi-Fi. Les paramètres collectés ont été traités par un Arduino embarquant un système d'apprentissage automatique conçu pour étudier les données. Selon les auteurs, les résultats de ces tests ont été remarquables et ont permis de signaler des taux de détérioration des équipements de convoyage.

Une étude de Sugandha et al. (2023) a examiné comment le secteur pharmaceutique peut bénéficier de la transformation numérique et comment les technologies numériques telles que l'IA (intelligence artificielle), l'IOT, la blockchain et l'apprentissage automatique peuvent contribuer à accroître l'efficacité des processus. Les auteurs ont également souligné les opportunités et les défis auxquels ce secteur est confronté. Ces domaines d'application incluent la découverte de nouvelles molécules d'intérêt pour la création de médicaments, l'identification de cibles thérapeutiques et l'assistance à la conception de nouvelles molécules grâce à des approches basées sur l'IA, ainsi que l'utilisation de techniques d'apprentissage automatique dans les essais cliniques pour améliorer l'efficacité des médicaments et réduire les coûts. À la fin de leur étude, les auteurs ont abordé l'utilisation de l'IOT pour optimiser la gestion des stocks, le suivi des patients et la collecte de données d'essais cliniques, ainsi que son utilisation plus large dans la production de médicaments, permettant aux fabricants d'améliorer les processus de production et de maintenance. Cet article s'est attaché à décrire les améliorations possibles des processus de recherche et développement.

Dans les articles précédents, nous avons constaté une évolution du paradigme concernant l'utilisation de l'IOT au fil des ans depuis l'émergence des concepts de l'Industrie 4.0. Les auteurs abordent les contours de l'utilisation de l'IOT dans le secteur industriel, et notamment les besoins de robustesse et de standardisation de ces équipements, ainsi que leur potentiel d'amélioration des processus industriels. Il est en effet crucial que les équipements IOT puissent être standardisés et devenir conforme pour une utilisation en environnement industriel, en particulier dans les environnements soumis à des contraintes de qualité élevées. Les publications indiquent par exemple que l'intérêt pour l'IOT s'accroît au fil du temps. Initialement conçu pour renforcer la veille industrielle, on observe l'apparition de publications traitant de l'utilisation de l'IOT combinée à l'apprentissage automatique et à l'intelligence artificielle pour l'optimisation des processus industriels. Cependant, on constate que les auteurs cités se limitent généralement à discuter de cas d'usage théoriques, de défis et d'opportunités liés à l'IOT. Rares sont les articles évaluant l'efficacité concrète de leurs méthodes à partir de retours d'expérience terrain. On constate également que les problématiques liées à l'IOT ne se limitent pas à son utilisation, mais concernent également les architectures réseau à mettre en œuvre.

Dans la section suivante, nous nous concentrerons plus particulièrement sur certaines catégories d'équipements soumis à des modèles de maintenance pilotés par l'IOT dans l'industrie.

2.4. Modèles pour la maintenance prévisionnelle

Il existe dans la littérature scientifique de nombreux modèles de maintenance prévisionnelle permettant d'évaluer le comportement des équipements et de prédire leur durée de vie résiduelle. Ces modèles comprennent des modèles établis par les données, des modèles basés sur des lois physiques, des modèles basés sur des retours d'expérience, ainsi que des modèles hybrides. Ils sont définis comme suit :

- Le modèle établi par les données permet de modéliser le comportement en récupérant des données reflétant le comportement de l'équipement.

- Les modèles basés sur des lois physiques reposent généralement sur une loi physique connue et appliquée pour estimer un RUL, notamment dans le cas des batteries au lithium, en utilisant les lois d'Ohm et de Kirchhoff (Zwingelstein, 2020). Les lois utilisées permettent de modéliser le comportement de l'équipement étudié en conditions nominales de fonctionnement.
- Les modèles basés sur des retours d'expérience s'appuient sur la connaissance humaine des caractéristiques et des modes de défaillance de l'équipement pour estimer son RUL.
- Les modèles hybrides (Amasyali et El-Gohary, 2022) combinent deux approches, par exemple physique et basé sur les données. Cela peut être utile pour modéliser le comportement des systèmes avec des données, complétées par des connaissances physiques pour un affinage du modèle final.

En utilisant la méthode de recherche bibliographique décrite dans la section 2, cette partie présente un état de l'art des modèles recensés dans la littérature scientifique. Cette section est structurée selon le type d'équipement sujet aux modèles de maintenance. Chaque sous-section répertorie les références citées par ordre chronologique. Nous avons ainsi analysé des études portant sur des équipements mécaniques, électriques, robotiques et fluidiques.

Notre analyse de la littérature existante a révélé que les méthodes proposées dans ces études étaient basées sur des domaines d'application liés à des équipements particuliers. Nous présenterons ci-après les cas d'application les plus significatifs. Nous tenons également à préciser que pour toute référence à la RUL, nous utilisons indifféremment les termes « estimation » et « prévisions » (Lei et al., 2018).

2.4.1. Modèles de maintenance prévisionnelle pour les équipements mécaniques

Les équipements mécaniques font l'objet de nombreuses publications d'applications de modèles de maintenance prévisionnelle.

2.4.1.1. Le cas des roulements

Dans le domaine de la fabrication, l'un des cas d'utilisation qui revient régulièrement est celui des roulements.

Le premier article traite de l'estimation de la durée de vie utile des roulements à l'aide d'études acoustiques. Aye et Heyns (2015) ont proposé, à partir de données extraites d'un banc d'essai : premièrement, la création d'un indice de santé des roulements en combinant les algorithmes PKPCA (Polynomial Kernel Principal Component Analysis), GMM (Gaussian Mixture Model) et EWMA (Exponentially Weighted Moving Average), puis la proposition d'une durée de vie utile par l'utilisation séparée des algorithmes MLP (Multi-Layer Perceptron), RBF (Radial Basis Function), BLR (Bayesian Linear Regression), GMR (Gaussian Mixture Regression) et GPR (Gaussian Process Regression). Bien que tous les modèles soient capables de prévoir les défaillances des roulements, les auteurs ont soutenu que la GPR est la plus robuste et la plus précise. Le sujet des roulements est également étudié par Boukra (2015). Dans son étude, l'auteur a utilisé l'algorithme PSO (Particle Swarm Optimization) pour identifier les caractéristiques utilisées pour le pronostic avancé des défaillances. Le modèle de détection des défaillances était basé sur l'utilisation d'un PF (Particle Filters) associé à un NFS (Neuro-Fuzzy System). Les prévisions étaient basées sur des données collectées lors d'expériences réalisées sur un banc d'essai. Cependant, les prévisions faites par ces modèles ne sont pas comparées à d'autres modèles. Dans sa publication, Zheng (2019) a présenté une

méthode de prévision du RUL des roulements en utilisant l'entropie d'une transformation de Hilbert-Huang pour obtenir un signal de vibration horizontale. Ensuite, une régression linéaire est appliquée à ce signal pour obtenir une estimation du RUL. Les résultats obtenus par cette méthode étaient proches de ceux des meilleures méthodes présentées lors du IEEE PHM Data Challenge 2012 (Nectoux et al., 2012). Dans une autre étude, Y. F. Ma et al. (2020) ont proposé une méthode de détection et de pronostic des défaillances de roulements basée sur la SR (Sparse Representation) et le HSMM (Hidden Semi-Markov Model). Les données ont ensuite été regroupées. Un ensemble de 8 roulements a été testé pour la prédiction. Selon les chercheurs, cette méthode était moins précise que d'autres, comme les réseaux de neurones ou les filtres à particules. Cependant, ils ont déclaré que cette méthode peut toujours être utilisée car elle est capable de fournir de bonnes prévisions à long terme pour les applications industrielles. Nistane (2021) a proposé une application aux roulements basée d'une part, sur la création d'un indice de santé des roulements avec filtrage du signal à l'aide d'une méthode CWT (Continuous Wavelet Transform) et d'autre part sur l'estimation du RUL à l'aide d'un NAR (Nonlinear Autoregressive network). Les données utilisées pour cette méthode ont été obtenues à partir d'un banc d'essai. L'auteur a proposé une application efficace pour la détection précoce des défaillances de roulements et, plus généralement, pour les données de séries temporelles non linéaires. Cependant, les performances du modèle n'ont pas été comparées à celles d'autres modèles de prévision.

Ünal et al. (2022) ont introduit un système optique de capture de roulements qui utilisait une caméra pour surveiller physiquement les roulements. Les images étaient transmises pour traitement à un algorithme CNN (Convolutional Neural Network) capable de détecter les changements physiques dans les séquences d'images. Pour classer les caractéristiques, les méthodes SVM (Support Vector Machine) et KNN (K-Nearest Neighbors) ont ensuite été utilisées. Cependant, le modèle n'a pas été comparé à d'autres modèles existants ou à d'autres algorithmes de détection de défauts. Les auteurs ont revendiqué un score de prédiction précis de presque 100 %. Les chercheurs ont prévu de modifier le modèle et de l'adapter aux contrôles d'équipements automatisés. Lee et al. (2022) ont également proposé dans leur publication un modèle pour surveiller les roulements et estimer le RUL afin de proposer un indicateur de santé. Cette approche utilisait des méthodes EML (Extreme Machine Learning) et SFE (Systematic Feature Engineering) et proposait un modèle qui incluait non seulement l'usure physique des roulements, mais aussi un indice de comparaison avec l'usure connue d'autres roulements. Le RUL était donc estimée sur la base de deux paramètres, le premier dérivé de l'usure du roulement étudié, le second des RUL connus des autres roulements. Les auteurs ont rapporté que, par rapport aux données de Nectoux et al. (2012), pour sept des onze roulements, la méthode proposée réduisait l'erreur absolue moyenne (MAE), l'erreur quadratique moyenne (RMSE) et l'erreur absolue moyenne en pourcentage (MAPE) dans l'estimation du RUL de plus de 50 %. Juodelyte et al. (2022) ont proposé un modèle pour surveiller les roulements dans des conditions pharmaceutiques. Ils ont souligné le besoin des autorités réglementaires de surveiller la maintenance dans l'industrie pharmaceutique. Le modèle présenté était appliqué à la surveillance des sous-parties d'équipements en utilisant une segmentation de la vie du roulement basée sur des vibrations à haute fréquence, le signal étant traité par une méthode k-means et un autoencodeur. Un classificateur multiclasse a ensuite été utilisé pour prévoir les étapes de dégradation. Ils ont également déclaré que les résultats de l'expérience et les prévisions obtenues sont prometteurs, bien que des développements supplémentaires soient nécessaires avant un éventuel déploiement dans un plan de maintenance dans un environnement industriel

pharmaceutique. Une autre étude par Li et al. (2023) a proposé une méthode basée sur les données pour prévoir le RUL des roulements à travers la reconstruction de signaux, puis en utilisant une fusion de deux réseaux neuronaux, un TCN (Temporal Convolutional Network) et un CNN respectivement. Une comparaison a été faite entre la méthode proposée et d'autres méthodes impliquant une combinaison d'algorithmes. Cette opération a été réalisée sur 6 roulements différents. Dans ce cas, l'erreur quadratique moyenne et l'erreur absolue moyenne étaient plus faibles pour la méthode proposée que pour les autres modèles. Une autre étude sur les roulements a été proposée par Pandit et Xie (2023) et appliquée aux éoliennes. Le problème de la maintenance de tels équipements nécessite une connaissance précise de l'état de santé des roulements. À cet effet, les auteurs ont proposé une méthode basée sur les données pour prédire le RUL des roulements à "haute vitesse" en impliquant un algorithme SSA, combiné avec deux modèles de régression, à savoir un Support Vector Regressor (SVR) et un Random Forest (RF). La méthode proposée était basée sur l'étude des signaux de vibration à partir de données d'une éolienne réelle. Une valeur d'erreur absolue moyenne a été utilisée pour évaluer l'entraînement de l'algorithme. Selon les auteurs, le SVR, entraîné par le SSA, a donné de bons résultats en termes de prévision de RUL. Cependant, les auteurs sont limités par la disponibilité des données, n'ayant à disposition que les données d'une seule éolienne pour mener l'étude.

2.4.1.2. Le cas des boîtes de vitesse

Yubin et al. (2020) ont décrit une méthode d'estimation du RUL impliquant un DBN (Deep Belief Network) et un SOM (Self Organizing Map) applicable aux réducteurs d'éoliennes. Ces deux méthodes ont été utilisées pour réduire le bruit des signaux de vibration afin de créer un indicateur de santé pour l'équipement étudié. Cet indicateur a ensuite été traité à l'aide d'un FP optimisé par un FOA (Fly Optimization Algorithm) pour prédire le RUL. Pour prendre en compte la nature dynamique du comportement des éoliennes, un modèle de dégradation basé sur le Wiener Process a été utilisé afin d'améliorer les prévisions du RUL. Le modèle a été comparé à d'autres algorithmes de prévisions tels que le GP (Gamma-Process), le PE (Paris-Erdogan) et le LSSVM (Least Square Support Vector Machine). Les résultats ont montré que la méthode proposée basée sur un FOA permettait non seulement une prévision plus précise du RUL, mais réduisait également son incertitude. Toutefois, il apparaît qu'il n'existe toujours pas de critères clairement définis pour évaluer l'état des réducteurs d'éoliennes, ce qui devrait faire l'objet de futures recherches.

2.4.1.3. Le cas des actionneurs

Un article de Matteo et al. (2018) a discuté d'une application du ML pour la maintenance des actionneurs pneumatiques d'avions. Deux algorithmes (GA & PSO) ont été utilisés à la fois pour l'identification des pannes et pour l'optimisation. Ces algorithmes ont été comparés dans le but d'identifier 5 types de défaillances à partir des signaux électromécaniques des actionneurs. Les auteurs ont présenté les méthodes utilisées pour calibrer les algorithmes afin de les adapter à ce cas d'usage. Ils ont conclu en rapportant les résultats des deux algorithmes et ont constaté que PSO obtenait de meilleures performances que GA. Cependant, les auteurs ont souligné qu'un algorithme d'optimisation adapté à un cas d'usage peut ne pas être efficace pour un autre. De futures recherches devraient porter sur d'autres techniques d'optimisation.

2.4.1.4. Le cas des équipements industriels

La maintenance des équipements industriels est également l'une des cibles des modèles de maintenance prévisionnelle. Chia-Yu et al. (2019) ont proposé une méthode basée sur les

données pour estimer le RUL des machines-outils, à travers l'étude des accéléromètres et des capteurs de température. Le signal collecté a été prétraité par une transformée de Fourier, afin de transposer le signal brut de vibration dans le domaine fréquentiel. Les données obtenues ont ensuite été traitées par un RNN (Recurrent Neural Network). Le modèle a permis une reconstruction claire du signal et l'ajout de seuils d'alerte. Cependant, les auteurs n'ont pas effectué de comparaison avec d'autres modèles et ont simplement affirmé avoir obtenu un taux de détection de 100 % des défauts des équipements.

Une autre application concerne les équipements de fraisage, publiée par Liu et al. (2020). Cette étude a introduit une méthode basée sur les données pour étudier les signaux de vibration provenant de plusieurs capteurs sur des équipements de fraisage. Ces signaux ont d'abord été traités et décomposés en utilisant un CEEMD (Complementary Ensemble Empirical Mode Decomposition). Le signal décomposé a servi de base pour extraire des caractéristiques statistiques. Il a ensuite été optimisé par un Light-GBM (Gradient Boosting Machine) puis retraité à l'aide d'un algorithme IABC-BPNN (Improved Artificial Bee Colony – Back Propagation Neural Network) pour modéliser et prédire le RUL. Les résultats de cette méthode ont été comparés à ceux d'autres modèles : NFIS (Neuro-Fuzzy Inference System), RBFN (Radial Basis Function Network) et BPNN (Back Propagation Neural Network). D'après les résultats des expérimentations, les auteurs ont conclu que le modèle proposé était plus précis et plus efficace que ceux auxquels il a été comparé.

Filiou et al. (2020) et Nchekwube et al. (2022) ont présenté des cas d'usage sur une machine d'emballage industrielle et un convoyeur industriel, respectivement. Dans le premier cas, les auteurs ont décrit une méthode impliquant le nettoyage des données extraites de l'équipement, un prétraitement, puis l'application de deux algorithmes de ML (Machine Learning) : ARIMA (AutoRegressive Integrated Moving Average) et « Prophet », pour l'entraînement, les tests et la prévision des pannes. La performance des deux modèles a été discutée et il a été conclu que des efforts supplémentaires de recherche et de traitement devaient être menés dans le cadre d'une future étude. Dans le second article, l'accent était mis sur les CNN pour l'étude et l'estimation du RUL d'un convoyeur. Les données brutes ont également été prétraitées et normalisées avant d'être soumises au CNN. Finalement, les résultats ont été statistiquement comparés à d'autres types de réseaux de neurones décrits dans l'article. Il a été constaté que, dans ce cas, le CNN fournissait les meilleures prévisions du RUL.

Les travaux de Quatrini et al. (2022) ont porté sur un cas appliqué à un granulateur pharmaceutique. L'objectif était d'établir un pronostic de RUL en utilisant un algorithme CVA (Canonical Variate Analysis) pour faciliter la maintenance et garantir l'intégrité de l'équipement. Treize paramètres de fonctionnement du granulateur en situation de défaillance ont été utilisés dans l'étude. Le modèle obtenu a permis d'établir un RUL prenant également en compte les pannes soudaines. Les données provenaient d'un granulateur réel. Toutefois, le modèle n'a pas été comparé à d'autres modèles et l'étude repose sur un seul équipement.

Un autre exemple, présenté par Zhang et al. (2023), concernait une machine tournante industrielle. Ils ont présenté un modèle hybride combinant des approches basées sur les données, la notion d'un indicateur d'assurance qualité pour évaluer le modèle, et l'intégration d'un modèle physique pour évaluer la capacité du modèle à estimer le RUL des outils de production, l'usure de l'équipement étant estimée par un modèle physique FEM (Finite Element Method). Le modèle proposé a été comparé à trois autres modèles, à savoir : le

Wiener process, le Gaussian process et le BPNN. Les résultats ont montré que le modèle proposé permettait de meilleures prédictions pour ce cas d'usage. Cependant, la performance du modèle dépend fortement de la qualité des caractéristiques et nécessite des travaux complémentaires pour établir des performances reproductibles.

Dans un autre article, Lu et al. (2023) ont présenté un modèle de prédiction de pannes permettant d'intégrer d'autres sources d'information, en plus de celles des paramètres étudiés. Cette étude repose sur le principe qu'une grande quantité d'informations pertinentes est cachée dans des données de types multiples. À cette fin, les auteurs ont pris comme cas d'étude une machine tournante dans une raffinerie de pétrole et ont proposé un DSTN (Deep Spatio-Temporal Network), une combinaison d'un CNN et d'un LSTM (Long-Short Term Memory), ainsi qu'un mécanisme d'attention interactive pour fusionner les profils d'ondes vibratoires et les signaux de durée de vie variables dans le temps de l'équipement étudié. De cette manière, le modèle proposé a permis de traiter à la fois des données spécifiques recueillies sur le long terme et des données brutes multi-sources, qui ont ensuite été pondérées et analysées statistiquement. L'expérimentation a révélé que le modèle présentait une réduction des erreurs de prédiction tout en maintenant de bonnes performances. Les résultats de ce modèle ont été comparés à ceux d'autres modèles. La comparaison a été testée sur des données publiques fournies par la NASA (Saxena and Goebel, 2008). Ils ont conclu que leur modèle était plus performant que ceux décrits dans l'étude comparative. Toutefois, la caractérisation des caractéristiques doit être renforcée pour accroître la fiabilité du modèle.

Enfin, un autre article présenté par Do et al. (2023) a décrit un modèle adapté aux carrousels de stockage industriels verticaux, utilisant un coefficient de corrélation pour détecter les anomalies liées à ce type d'équipement. Le modèle était basé sur un LSTM-Autoencoder pour l'apprentissage et le test de la détection d'anomalies. Les données, provenant de plusieurs capteurs répartis autour de l'équipement, ont été collectées sous forme de valeurs de vibration qui, une fois traitées par l'algorithme et après affinement des hyperparamètres, ont permis d'obtenir une précision de plus de 97 %. Il convient toutefois de noter que de futurs tests doivent comparer la performance de cette méthode avec d'autres algorithmes. Par ailleurs, les LSTM étant sensibles à la qualité des données, il est nécessaire de résoudre les problèmes de déséquilibre des données.

2.4.1.5. Le cas des turboréacteurs

Un cas d'usage récurrent, appliqué à l'aéronautique, est l'étude du RUL des moteurs à double flux, à partir de données fournies par la NASA (Saxena et Goebel, 2008). Les auteurs suivants ont mené leurs travaux sur ces données. Dans leur article, Shanmugasivam et Prahlad (2021) ont présenté un modèle combinant un MLE (Multi-Loss Encoder) et un CCF (Convolutional Composite Features), dans le but d'améliorer la détection des caractéristiques à l'aide d'une approche en deux étapes, séparant la génération des caractéristiques et la prévision du RUL. Leurs résultats se sont révélés supérieurs aux méthodes de référence décrites dans l'article. Cependant, les chercheurs ont suggéré une amélioration du modèle en ajustant le nombre de paramètres.

Dans le cas de Ellefsen et al., (2019), la combinaison d'apprentissage supervisé et non supervisé a été testée pour estimer le RUL. Leur idée était d'obtenir une bonne précision de prédiction tout en réduisant la quantité de données annotées nécessaires à l'entraînement. Une approche basée sur des RBM (Restricted Boltzmann Machines) et un GA (Genetic

Algorithm) a alors été appliquée pour calibrer les hyperparamètres du modèle. Cette approche semi-supervisée a ensuite été comparée à un modèle entièrement supervisé. Les résultats ont révélé une amélioration de la précision du RUL, tout en nécessitant moins d'étiquettes. Bala et al., (2020) ont présenté un autre exemple d'application. Ils ont suggéré que les outils courants de ML pour la maintenance, tels que les RNN, peuvent être efficaces pour une application en maintenance prévisionnelle, en particulier sur les séries temporelles. Toutefois, ils ont été confrontés à un problème lié à l'instabilité des gradients. Ils ont considéré que les ESN (Echo State Networks) étaient plus à même de résoudre ce problème en n'entraînant que les poids de sortie via des régressions linéaires simples. Le modèle proposé intégrait un GOA (Grasshopper Optimization Algorithm) basé sur une ESN. Cette méthode, inspirée du comportement social des sauterelles, est utilisée pour trouver la meilleure solution à un problème dans un espace donné ou pour optimiser des systèmes. Ils ont comparé le modèle à un ensemble de méthodes appliquées à des cas similaires et ont rapporté que, dans trois cas sur quatre, le GOA basé sur ESN surpassait les autres méthodes en termes de MSE moyen. Dans le quatrième cas, les performances du modèle étaient proches de celles du meilleur algorithme. Cependant, le modèle n'était pas le plus rapide. Ils ont conclu en expliquant que les ESN peuvent être instables et nécessitent un coût de calcul élevé. Ces deux inconvénients sont liés au volume de données utilisées. Dans un autre article, Al-Dulaimi et al., (2020) ont proposé une estimation du RUL des turboréacteurs à l'aide d'un modèle hybride composé d'une architecture CNN et de plusieurs couches bruitées afin de rendre le modèle de prévision plus robuste. Le modèle a été comparé à un ensemble d'autres modèles décrits dans l'article ; il s'est révélé plus stable et robuste, notamment face au bruit de fond. Cependant, l'incertitude liée à la quantification demeure un problème dans l'utilisation des réseaux neuronaux. Les auteurs proposent que les études futures s'attaquent à cette question. Enfin, dans leur article, Mazzuto et al. (2021) ont présenté une méthode de calcul d'un indice de santé de ces mêmes équipements à l'aide d'une approche supervisée basée sur un algorithme k-means, un GWO (Grey Wolf Optimization), un SWA (Swarm Intelligence Algorithm) et un FCM (Fuzzy Cognitive Maps) pour le calcul de l'indicateur de santé de l'équipement cible et la prévision du RUL. Les performances du modèle proposé ont été comparées à celles obtenues avec des réseaux de neurones simples. Avec la méthode présentée, les chercheurs ont démontré une distinction plus précise des signaux de défauts et une estimation plus fiable du RUL. Cependant, le modèle proposé nécessite de nombreuses variables pour son application, et le processus d'extraction des caractéristiques est long. Des améliorations supplémentaires sont nécessaires pour rendre le modèle plus pratique et plus convivial.

2.4.2. Modèles de maintenance prévisionnelle pour les équipements électriques

Les équipements électriques font également l'objet de nombreuses publications concernant l'utilisation de modèles de maintenance prévisionnelle.

2.4.2.1. Le cas des transformateurs

Une étude menée par Dong et al. (2019) présente un exemple appliqué à l'industrie pétrolière et gazière. Dans le but de réduire les coûts de maintenance, ces auteurs ont proposé un algorithme de diagnostic des réseaux électriques. L'objectif spécifique de cet algorithme était de fournir un pronostic de l'état des transformateurs immergés dans l'huile. Ils ont utilisé un algorithme Bat-BPNN combiné à un ensemble de modèles de diagnostic pertinents, rendant l'approche fortement dépendante de la qualité des données. Les résultats ont montré que la précision de la caractérisation des défauts était améliorée par rapport aux autres modèles

PHM décrits dans l'article. Il est cependant important de noter que ce modèle n'était pas capable d'identifier d'autres types de défauts.

2.4.2.2. Le cas des piles à combustible

Plusieurs articles de Liu et al. (2019), Chen et al. (2020), Mingzhang et al. (2023) et Luo et al. (2023) décrivent un cas d'usage appliqué aux piles à combustible à membrane échangeuse de protons. Dans le premier article, les auteurs proposent un modèle hybride pour estimer la dégradation et l'état de vieillissement de ces piles. La première étape consiste à étudier divers modèles de dégradation existants afin d'estimer un indice de dégradation. La seconde étape applique un algorithme de ML basé sur un ANFIS optimisé par un PSO aux indices de dégradation afin de réaliser des tests de prédiction du RUL des membranes. Les données utilisées pour les tests proviennent d'un jeu de données commun fourni par Gouriveau et al. (2014). Il convient toutefois de noter que l'étude a été réalisée uniquement sur ce jeu de données commun.

Le deuxième article porte sur un modèle de pronostic du RUL construit avec un BPNN et optimisé par un algorithme évolutif, lui-même composé d'un MEA, d'un PSO et d'un GA. Les données utilisées pour créer ce modèle proviennent d'un banc d'essai conçu par les auteurs, permettant de simuler les conditions opérationnelles auxquelles les membranes étaient soumises. Ce modèle a ensuite été comparé à d'autres algorithmes similaires, notamment des classificateurs SVM et RVM (Relevance Vector Machine), et les auteurs rapportent de meilleures performances prédictives pour leur propre modèle en termes de RUL.

Mingzhang et al. proposent une méthode basée sur les données pour la prédiction à long terme du RUL. L'objectif est d'identifier les défaillances de ces équipements en exploitant le retour des paramètres électriques, tels que la tension et la résistance de polarité, puis en les traitant avec un TCN. Ces deux paramètres sont combinés pour générer un indicateur unique de défaillance. Pour améliorer les performances du modèle, des prédictions directes et récursives sont utilisées. Le modèle est comparé à quatre autres modèles décrits dans l'article. Les auteurs observent que les erreurs de prédiction sont plus faibles que celles des autres modèles, avec une meilleure estimation du RUL. L'étude n'a cependant été réalisée que sur un jeu de données commun.

Enfin, dans la quatrième étude, les chercheurs utilisent une combinaison de CNN et de BiRNN (Bidirectionnal-RNN) pour estimer le RUL de ces dispositifs. Cette combinaison permet une prédiction plus précise du RUL des piles à combustible par rapport à d'autres modèles tels que le BPNN, le LSTM ou le S-LSTM (sentence-state LSTM). Il convient toutefois de noter que l'étude a été réalisée uniquement sur un jeu de données commun.

2.4.2.3. Le cas des convertisseurs

Dans le cas des convertisseurs de puissance industriels, Wang et al. (2017) ont souligné qu'une altération d'un paramètre entraînait une baisse de la qualité de l'alimentation électrique. Les paramètres étudiés étaient les valeurs d'ondulation et les tensions en sortie du circuit, afin de calculer des distances de Mahalanobis, lesquelles, selon les auteurs, reflètent l'état de santé des circuits. L'utilisation d'un SVR a permis d'effectuer une prédiction, qui a ensuite été améliorée par un algorithme GWO pour estimer le RUL. Les résultats des simulations comparatives ont démontré que les prédictions du RUL via cette méthode étaient plus fiables et comportaient moins d'erreurs dans ce cas d'usage.

2.4.2.4. Le cas des alimentations et des générateurs

Les travaux de Kareem et al. (2022) se sont concentrés sur les alimentations électriques, et plus particulièrement sur les condensateurs électrolytiques en aluminium à découpage (SMPS-AEC). Une expérience a été menée afin de mesurer certaines caractéristiques typiques des anomalies de ces condensateurs. Les paramètres étudiés étaient l'impédance, l'ESR (Equivalent Series Resistance), le facteur de dissipation et la capacité. Pour prédire les défaillances, plusieurs algorithmes ont été testés sur les jeux de données obtenus lors des tests. Les auteurs indiquent que, dans leur cas d'usage, l'algorithme KNN surpasse les autres, en offrant de bonnes prédictions tout en minimisant le coût de calcul. Il a été noté que l'ESR était le paramètre le plus représentatif pour la détection des défauts, permettant ainsi de différencier les condensateurs. Cependant, des recherches supplémentaires sont nécessaires afin de vérifier le comportement des algorithmes en incluant une approche dépendante de la température dans le cadre de l'étude.

Zadiran et Shcherbakov (2023) ont publié un article sur les générateurs à turbine à gaz, dans lequel ils présentent une méthode de maintenance pilotée par les TIC, reposant sur un modèle non supervisé d'identification des défauts, utilisant des fenêtres glissantes et un test de Kruskal-Wallis, ainsi qu'un CNN en tant que modèle de référence. Dans cette étude, le test de Kruskal-Wallis est utilisé pour effectuer des comparaisons statistiques entre les fenêtres et détecter l'apparition d'évolutions négatives. Les résultats ont révélé une réduction des erreurs de prévision entre le modèle proposé et ceux décrits dans leur article. Il convient toutefois de noter que l'étude a été réalisée uniquement sur un jeu de données commun.

2.4.3. Modèles de maintenance prévisionnelle pour les équipements robotiques

Dans leur étude, Cardenas-Lizana et al. (2022) ont proposé une méthode de ML pour prédire la dégradation d'un dispositif médical utilisé en chirurgie de précision, à savoir un micro-robot. Ils ont classé trois états de fonctionnement distincts. Les données étudiées étaient générées par des simulations fondées sur les recommandations du fabricant et traitées par un algorithme FE afin de créer un vecteur de dégradation. Ce vecteur a ensuite servi à entraîner des algorithmes RF et VAE (Variational Auto Encoder), dont les sorties ont constitué la base de l'entraînement d'un algorithme classifieur XGBoost. Selon les auteurs, la méthode permettait de réduire significativement le bruit dans le signal de dégradation. Toutefois, des études complémentaires sont nécessaires pour explorer le potentiel de cette méthode, notamment en ce qui concerne la correction des biais et la prise en compte d'un contexte intégrant plusieurs signaux à traiter.

Une autre approche d'estimation du RUL intégrant la réalité augmentée a été proposée par Wang et al. (2022). Cette approche reposait sur l'idée de combiner une estimation précise du RUL des robots et de leurs outillages avec la facilité d'accès à un expert via des dispositifs de réalité augmentée permettant de partager des informations visuelles à distance avec un tiers et d'en déduire des instructions pour réparer rapidement les pièces ou sous-ensembles défaillants. Le RUL était estimé dans ce cas à l'aide d'un PSO-CNN, permettant d'identifier des caractéristiques à partir des données de production, et d'un GRU-Attention (Gated Recurrent Unit) pour prévoir le RUL lui-même. Cette approche était accompagnée d'une assistance à distance par un expert de l'équipement concerné, de sorte que cette combinaison pourrait, selon les auteurs, permettre un gain de temps significatif pour la réparation de l'équipement, contrairement aux approches de maintenance traditionnelles.

2.4.4. Modèles de maintenance prévisionnelle pour les équipements fluidiques

Cette section est consacrée aux équipements de traitement de liquides (industrie alimentaire, biopharmaceutique, raffineries, etc.).

2.4.4.1. Le cas des centrifugeuses industrielles

Le cas des centrifugeuses industrielles a été traité par Li et al., (2019), qui ont proposé un modèle hybride pour prévoir le RUL en intégrant les données issues de plusieurs capteurs. L'objectif du modèle était de déterminer plusieurs pronostics liés aux différentes phases de vie de l'équipement, afin de déduire une prévision précise du RUL. Les données générées par les capteurs ont été traitées à l'aide d'une méthode CVA, qui a fourni un point de départ à la prédiction et à la construction ultérieure d'un indicateur de santé. Un algorithme PF a été appliqué pour optimiser un Metabolism Grey Forecasting Model (MGMF) afin d'effectuer des prévisions à la fois sur le RUL et sur les incertitudes. Les résultats rapportés par les auteurs ont montré une efficacité de prédiction plus élevée du modèle combiné PF-MGMF par rapport à d'autres méthodes telles que le MGMF seul, LSTM ou AR.

2.4.4.2. Le cas des systèmes de filtration industriels

Skaf et al., (2015) ont créé un banc d'essai reproduisant un système de circulation de carburant afin de mesurer l'encrassement des filtres à l'aide de débitmètres. Après avoir collecté les données nécessaires à l'étude, ils ont proposé et appliqué une méthode appelée Simple State-Based Prognostics with Duration (SBPD) en trois étapes. La première étape consistait à regrouper les échantillons de données à l'aide d'une méthode de type k-means. La deuxième étape consistait à appliquer un indice de type Calinski-Harabasz aux clusters afin de déterminer le nombre optimal de clusters nécessaires à l'étude. La troisième étape consistait à calculer le RUL en sommant les durées totales attendues dans l'état de santé actuel du système et dans chaque état de santé futur avant que le système ne devienne inefficace. Cependant, les performances obtenues n'ont pas été comparées à celles d'autres modèles de ML.

Dans leur étude, Omer et al., (2016) ont étudié le phénomène d'encrassement dans les filtres de procédés industriels. Ils ont proposé un modèle physique permettant de représenter les trois phases de l'encrassement. Le RUL a ensuite été estimé via une PF, calculée à partir de données issues d'une plateforme expérimentale représentant les caractéristiques communes d'un système de filtration. Les résultats obtenus ont été validés par un système de capture optique qui surveille en continu le filtre étudié. Cela a confirmé que l'état d'encrassement estimé par le modèle correspondait bien à l'encrassement physique du filtre. Beirami et al., (2020) ont proposé un autre modèle ciblant les systèmes de filtration. Ils ont concentré leurs travaux sur une plateforme expérimentale reproduisant un système composé d'un filtre, d'une pompe péristaltique et d'un ensemble de capteurs. L'ensemble du système était placé sous l'eau. Ils ont choisi une approche combinant l'utilisation d'un classificateur SVM et d'un RFE (Recursive Feature Elimination) pour extraire et sélectionner les caractéristiques d'intérêt. Cependant, les auteurs n'ont pas comparé leur modèle à d'autres et ont conclu que leur modèle pourrait être amélioré en tenant compte d'éléments physiques et d'une meilleure paramétrisation pour caractériser les pressions dans le système. Tian et al., (2023) ont présenté une application à un système de filtration hydraulique. L'étude portait sur un ensemble de données théoriques fournissant un RUL représentatif des équipements de filtration. Ils ont proposé une approche basée sur la divergence Kullback-Leibler pour mesurer les différences dans les distributions de données. Cette méthode a été appliquée à des séries

temporelles à court et long terme en utilisant des méthodes de Transfer Ensemble Learning (TEL) et LSTM. Les résultats ont démontré que ce modèle augmentait la précision des prédictions par rapport à toutes les techniques d'analyse décrites dans l'article.

2.4.4.3. Le cas des équipements pharmaceutiques et biopharmaceutiques

Nous avons regroupé les travaux traitant de l'utilisation de modèles de maintenance dans les industries pharmaceutique et biopharmaceutique, qui constituent le domaine de notre étude. La publication la plus ancienne que nous avons trouvée, proposée par Benazzouz et al., (2011), décrivait, dans une approche d'analyse des risques, un cas d'application aux isolateurs pharmaceutiques. Le modèle proposé était basé sur un ANN (Artificial Neural Network) pour détecter et localiser les défauts en temps réel, tout en prenant en compte les contraintes réglementaires appliquées aux environnements de distribution des produits pharmaceutiques. Les paramètres étudiés étaient les pressions des différentes chambres de l'équipement et les débits d'air moyens. Après ajustement des hyperparamètres spécifiques aux réseaux de neurones décrits dans l'article, les auteurs ont rapporté un résultat simulé de détection de pannes proche de 100 % pour l'équipement étudié. Cependant, comme les auteurs revendiquent des résultats parfaits, les performances du modèle ne sont pas comparées à d'autres modèles ML pour vérifier si des résultats similaires peuvent être obtenus dans des conditions équivalentes.

Tulsyan et al., (2020) ont étudié la défaillance d'un système de capture de température utilisé dans les systèmes fluidiques de production biopharmaceutique. Ils ont introduit leur sujet en décrivant le contexte de production spécifique de l'industrie biopharmaceutique et la nécessité de respecter un système qualité strict. Les capteurs de température étudiés étaient soumis à un nettoyage à la vapeur pour garantir que l'équipement ne soit pas contaminé. Le résultat serait, au minimum, une dérive dans la capture voire des pannes. Dans cette étude, le RUL a été estimé d'abord en utilisant SFA (Slow Feature Analysis) pour diagnostiquer l'état de santé du capteur, puis en utilisant un GP pour prévoir le RUL lui-même. Le modèle a été appliqué à un capteur industriel réel. Cependant, il n'y a pas eu de comparaison des performances de ce modèle avec d'autres modèles ML.

Calzavara et al., (2021) ont abordé une approche temporelle de ML pour la maintenance prévisionnelle dans le secteur pharmaceutique en utilisant l'outil DBSCAN (Density-Based Spatial Clustering of Applications with Noise). Ils ont proposé de distinguer des indicateurs de santé à partir de projections d'eau d'un lyophilisateur des stations CIP (Cleaning In Place). L'objectif était de développer une méthode semi-automatique d'analyse des flux d'eau et de faire des prédictions de pannes à l'aide d'un volume limité de données émis par un seul capteur. Cette méthode a été utilisée pour analyser deux paramètres : le nombre de cycles de l'équipement et le débit moyen de l'eau le traversant. Les chercheurs ont ainsi pu caractériser les phases de vie de l'équipement. L'utilisation du clustering combinée à la connaissance des mainteneurs de l'équipement a permis de caractériser les signaux de défaut. Cependant, cette méthode nécessite un nettoyage de données approfondi avant d'être utilisée, et ses performances n'ont pas été comparées à d'autres modèles ML.

Une autre approche a été présentée par Kavasidis et al., (2023), où ils ont proposé un modèle basé sur des deep transformers pour estimer l'état de santé d'une ligne de production pharmaceutique. Ils ont présenté l'outil ManuTrans, basé sur des données brutes en temps réel issues d'une ligne de production pour prédire la prochaine apparition d'une panne et sa gravité. Contrairement à Calzavara et al., (2021), qui ont utilisé des données de long terme et

se sont concentrés sur un seul équipement, cette méthode utilise également des données de court terme et cible l'ensemble de la ligne de production pour permettre la prise de décision. Les données ont été obtenues à partir de capteurs de pression et de température. L'outil reposait sur les capacités des deep transformers à modéliser les relations entre les différentes corrélations dans les données rapportées. Les données ont été prétraitées puisque les auteurs ont opté pour une normalisation des données afin de faire fonctionner le modèle. Les données normalisées ont été utilisées comme base pour tester différentes techniques de classificateurs machine learning afin de les comparer aux performances de ManuTrans. Ils ont conclu que le modèle présentait théoriquement de meilleures performances que les autres classificateurs utilisés en ML.

Menanno et Salsano, (2023) ont proposé une estimation du RUL d'une pompe hydraulique basée sur les signaux de vibrations et les signaux acoustiques émis par l'équipement, en utilisant un algorithme SVM pour les prédictions. Les données ont été obtenues à partir d'une plateforme d'essai. Les résultats obtenus par le SVR ont ensuite été comparés à d'autres algorithmes comme le KNN, le DA (Discriminant Analysis), le NBC (Naïve Bayesian Classifier) et le DT (Decision Trees), le SVM montrant les meilleures performances pour cet équipement. Cependant, il est nécessaire d'optimiser les performances de prédiction qui présentent de faibles précisions.

Un autre modèle pour la surveillance des défauts de capture de température dans les systèmes fluidiques a été proposé par Kontogiannis et al., (2023). Ils ont utilisé des données issues d'un ensemble de données en ligne dans le but d'identifier un comportement anormal des capteurs. Ils ont employé une méthode d'apprentissage profond (deep learning) et ont proposé un modèle de réseau neuronal avec modulation dépendante des entrées du nombre de couches profondes pour classer les différents types de pannes potentielles. La comparaison du modèle proposé avec d'autres algorithmes comme le MLP et le LSTM a démontré son efficacité supérieure. Toutefois, les auteurs précisent que les variations dans la précision des modèles LSTM peuvent avoir un impact sur la robustesse de la solution.

2.5. Discussion : Analyse des modèles de maintenance prévisionnelle

Toutes les méthodes abordées dans la section 2.4 sont listées dans le Tableau 2 ci-dessous, et leurs acronymes sont détaillés dans la liste des acronymes (Tableau 3) retrouvée en annexe. Nous avons constaté dans la section précédente que la plupart des modèles de ML pour la maintenance prévisionnelle sont basés sur des approches data-driven. Toutefois, nous avons identifié trois modèles dans notre étude qui ne reposent pas uniquement sur des approches data-driven, mais également sur des méthodes hybrides, dont deux combinées avec un modèle physique, et un troisième employant l'expertise de l'entreprise chargée de la maintenance de l'équipement étudié.

Le tableau 2 suivant fournit un résumé des travaux identifiés par type d'équipement. Les deux premières lignes répertorient les références aux jeux de données communs décrits dans les publications analysées.

Tableau 3 - Compilation des méthodes vues en section 2.4

| Application | Type de modèle | Méthode d'apprentissage | Méthode d'optimisation | Origine des données | Auteurs | Date |
|------------------------|-------------------------------|-----------------------------|------------------------|---------------------|---------------------------|------|
| Datasets | - | - | - | - | Saxena and Goebel | 2008 |
| | - | - | - | - | Nectoux et al. | 2012 |
| | - | - | - | - | Gouriveau et al. | 2014 |
| Equipements mécaniques | Data-driven | MLP / RBF / BLR / GMR / GPR | - | Banc de test | Aye and Heyns | 2015 |
| | Data-driven | PF / NFS | PSO | Banc de test | Boukra | 2015 |
| | Data-driven | GA / PSO | GA / PSO | Banc de test | Matteo et al. | 2018 |
| | Data-driven | HHE / LR | - | Dataset commun | Zheng, Y | 2019 |
| | Data-driven | FFT / RNN | - | Banc de test | Chia-Yu et al. | 2019 |
| | Data-driven | RBM / LSTM | GA | Dataset commun | Ellefsen et al. | 2019 |
| | Data-driven | SR / HSMM | GMM/EM | Dataset commun | Y. F. Ma et al. | 2020 |
| | Data-driven | DBN/SOM | PF / FOA | Banc de test | Pan et al. | 2020 |
| | Data-driven | ARIMA / Prophet | - | Donnée réelle | Filius et al. | 2020 |
| | Data-driven | CEEMD / Light-GBM / HHT | IABC / BPNN | Banc de test | Liu et al. | 2020 |
| | Data-driven | ESN | GOA | Dataset commun | Bala et al. | 2020 |
| | Data-driven | NPBGRU | NFC | Dataset commun | Al-Dulaimi et al. | 2020 |
| | Data-driven | NAR | - | Banc de test | Nistane | 2021 |
| | Data-driven | MLE | RTC | Dataset commun | Shanmugasivam and Prahlad | 2021 |
| | Data-driven | SWA/FCM/KNN | GWO | Dataset commun | Mazzuto et al. | 2021 |
| | Data-driven | CNN / SVM / KNN | - | Banc de test | Ünal et al. | 2022 |
| | Data-driven | EML | SFE | Dataset commun | Lee et al. | 2022 |
| | Data-driven | KNN/AE | - | Dataset commun | Juodelyte et al. | 2022 |
| | Data-driven | CVA | - | Donnée réelle | Quatrini et al. | 2022 |
| | Data / Physic (hybride) model | EMF | - | Simulation | Zhang et al. | 2023 |
| | Data-driven | CNN | - | Dataset commun | Nchekwube et al. | 2022 |
| | Data-driven | TCN / CNN | SSA | Banc de test | Bin et al. | 2023 |
| Data-driven | RFR / SVR / GPR | SSA | Dataset commun | Pandit and Xie | 2023 | |
| Data-driven | CNN / LSTM / DSTN | - | Dataset commun | Lu et al. | 2023 | |
| Data-driven | LSTM-AE | - | Donnée réelle | Do et al. | 2023 | |

| | | | | | | |
|-------------------------|-----------------------------------|-----------------------|------------|----------------|-------------------------|------|
| Equipements électriques | Data-driven | SVR | GWO | Banc de test | Wang et al. | 2017 |
| | Data-driven | BPNN | BA | Donnée réelle | Dong et al. | 2019 |
| | Data / Physic (hybride) model | ANFIS | PSO | Dataset commun | Liu et al. | 2019 |
| | Data-driven | BPNN | MEA/PSO/GA | Banc de test | Chen et al. | 2020 |
| | Data-driven | KNN / PCA | - | Donnée réelle | Kareem et al. | 2022 |
| | Data-driven | TCN | - | Dataset commun | Mingzhang et al. | 2023 |
| | Data-driven | CNN | BiRNN | Dataset commun | Luo et al. | 2023 |
| | Data-driven | CNN | - | Dataset commun | Zadiran and Shcherbakov | 2023 |
| Equipements robotiques | Data-driven | FE / RF / AE / E-BIKE | XGBoost | Simulation | Cardenas-Lizana et al. | 2022 |
| | Data / Expérience (hybride) model | CNN/GRU | PSO | Banc de test | Wang et al. | 2022 |
| Equipements fluidiques | Data-driven | ANN | RPA | Donnée réelle | Benazzouz et al. | 2011 |
| | Data-driven | KNN | - | Banc de test | Skaf et al. | 2015 |
| | Data / Physic (hybride) model | PF | - | Banc de test | Omer et al. | 2016 |
| | Data-driven | CVA / MGMF | PF | Donnée réelle | Li et al. | 2019 |
| | Data-driven | SVM | RFE | Dataset commun | Beirami et al. | 2020 |
| | Data-driven | SFA | GP | Donnée réelle | Tulsyan et al. | 2020 |
| | Data-driven | DBSCAN | - | Donnée réelle | Calzavara et al. | 2021 |
| | Data-driven | Deep transformers | CEC | Donnée réelle | Kavasidis et al. | 2023 |
| | Data-driven | Deep learning | - | Dataset commun | Kontogiannis et al. | 2023 |
| | Data-driven | SVM | - | Banc de test | Menanno, and Salsano | 2023 |
| | Data-driven | Bi-LSTM | - | Dataset commun | Tian et al. | 2023 |

2.5.1. Equipements ciblés

Un nombre significatif de publications est consacré aux équipements mécaniques. Il s'agit notamment de roulements, de pièces ou de sous-ensembles d'équipements que l'on retrouve fréquemment dans les environnements industriels et dont les RUL sont estimés par des mesures de vibration et des analyses de fréquence. Les équipements électriques sont la cible des modèles d'estimation du RUL, qui mesurent les valeurs de résistance, d'ampérage et de tension. Ces paramètres sont les plus représentatifs pour caractériser les pannes futures.

Nous avons également appris que la maintenance IOT-driven peut s'appliquer à la robotique, notamment en utilisant des simulations de panne pour optimiser la prévision du RUL. Enfin, la

conception des équipements fluidiques se prête à la prévision du RUL par l'analyse des valeurs rapportées par les capteurs dont ils sont souvent équipés. Pression, température, débit et ampérage sont autant de paramètres permettant de comprendre les pannes en adoptant une approche multifactorielle, en adéquation avec la complexité du système étudié.

Une sous-catégorie des équipements fluidiques regroupe les équipements utilisés dans le domaine pharmaceutique. Les études rassemblées dans cette sous-section révèlent que les modèles de ML pour la maintenance IOT-driven se limitent à des cas d'usage ciblant des systèmes spécifiques, incluant un lyophilisateur, un compresseur d'air d'isolateur pharmaceutique, ainsi qu'une ligne de production surveillée du point de vue de capteurs physiques.

Une seule publication dans notre enquête cible spécifiquement le domaine biopharmaceutique et traite de l'utilisation de capteurs de température pour la maintenance d'une ligne de production fluidique. La Figure 8 nous permet de visualiser la répartition des différents secteurs dont sont issus les modèles identifiés.

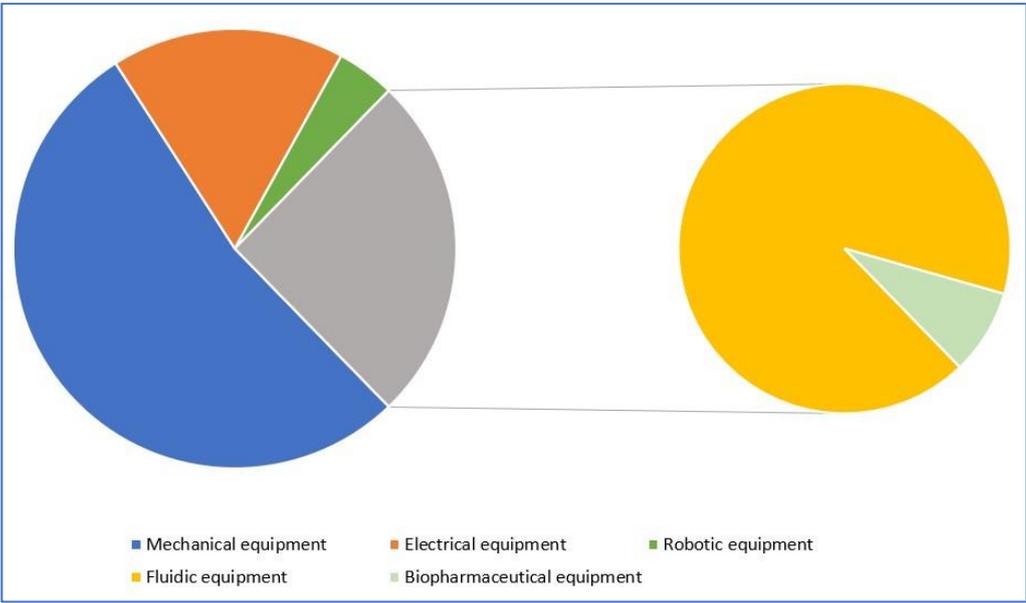


Figure 8 - Distribution des modèles selon le type d'équipement

2.5.2. Types de modèles

Près de la moitié des articles examinés dans cette étude concernent des modèles data-driven développés et appliqués soit sur des données génériques fournies lors de challenges organisés, soit par des entreprises emblématiques telles que la NASA pour l'entraînement de modèles de maintenance. D'un autre côté, d'autres types de recherches tentent de simuler ou de reproduire des systèmes sur de petites plateformes de test censées être représentatives des équipements étudiés, pour ensuite appliquer un modèle de ML afin d'estimer les RUL.

Les chercheurs présentent rarement des travaux ciblant des équipements industriels in situ, ou même des données fournies par des entreprises pour un besoin particulier. Il est également important de noter que, dans de nombreuses études, les méthodes de ML utilisées sont optimisées par un autre algorithme conçu et développé dans ce but, afin d'améliorer les prévisions de RUL. De plus, les méthodes d'apprentissage les plus couramment utilisées sont basées sur les réseaux de neurones, notamment ANN, CNN et KNN. Des méthodes d'optimisation sont ensuite mises en œuvre pour améliorer l'estimation du RUL. Pour générer les modèles de ML, nous avons noté que la majorité des chercheurs ont adopté une structure de développement schématisée dans la Figure 9 :

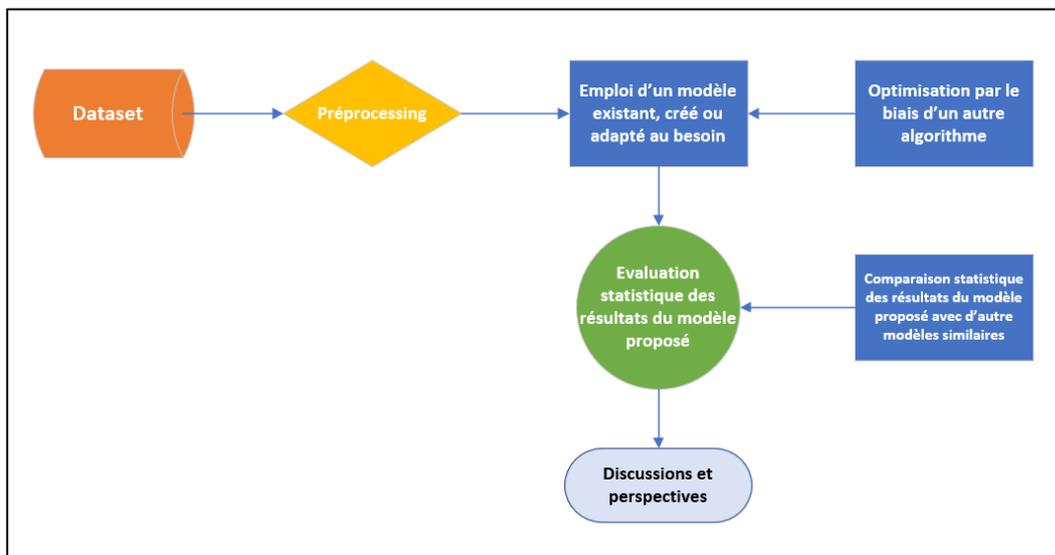


Figure 9 - Schéma de développement d'un modèle ML pour la maintenance

2.6. Conclusion du chapitre 2

Nous avons présenté dans ce chapitre, un état de l'art sur l'utilisation de modèles IOT-driven, développé dans le cadre de cette thèse. Cela nous permet de positionner notre travail en établissant les développements actuels des pratiques de maintenance IOT-driven et des modèles de ML associés dans l'industrie biopharmaceutique.

Les principales contributions de cet article se concentrent sur les points suivants :

- L'évolution de la maintenance industrielle vers des pratiques pilotées par l'IoT :

De nombreux articles nous ont permis d'observer que, dans l'industrie manufacturière, la fiabilité des IOT est une source d'intérêt et d'inquiétude. Par exemple, il a été noté que le manque de standardisation des IOT et leur maintenance sont des problématiques soulevées par les chercheurs. Cependant, nous souhaitons souligner qu'il existe une réelle attente de la

part des industriels concernant le développement des processus de production et de maintenance intégrant l'utilisation de l'IOT. Dans l'industrie pharmaceutique, l'évolution vers la maintenance prédictive constitue une préoccupation majeure pour les fabricants. Les développements dans ce domaine, cependant, décrivent principalement des cas appliqués à la logistique, à la R&D et aux utilités.

- L'état de l'art des modèles de ML pour la maintenance dans l'industrie biopharmaceutique :

Nous avons identifié un certain nombre d'articles traitant de l'utilisation des IOT pour la maintenance dans l'industrie manufacturière en général, et dans l'industrie biopharmaceutique en particulier. Cela nous a permis d'observer que l'industrie biopharmaceutique est peu représentée dans les publications scientifiques. Ces articles décrivent des cas d'usage appliqués à des systèmes spécifiques, ainsi qu'à une ligne de production surveillée à partir des données physiques de capteurs fournies par une entreprise. Nous avons constaté que la littérature ne contient pas de contributions significatives concernant des cas d'usage ciblant les membranes de vannes pneumatiques dans les systèmes fluidiques.

- Les limites des modèles de machine learning :

Comme nous l'avons vu, la majorité des modèles proposés pour la maintenance prévisionnelle ont été développés à partir de données disponibles sur Internet, fournies par des entreprises emblématiques pour traiter des cas génériques, ou fournies par des entreprises pour des cas particuliers. De plus, nous avons noté que la majorité des modèles sont basés sur les données, ce qui implique que la modélisation physique n'est pas requise dans tous les cas présentés. Enfin, la littérature examinée a révélé que plus de la moitié des modèles décrits nécessitent l'utilisation d'un autre algorithme en tant qu'optimiseur, ce qui indique que l'adaptation d'un seul modèle de machine learning peut ne pas suffire à fournir une précision acceptable du RUL de l'équipement concerné.

Sur la base des résultats discutés dans ce chapitre, la suite de ces travaux portera sur la proposition d'un modèle IOT-driven traitant le cas des vannes à membrane pneumatique, couramment présentes sur les lignes de production biopharmaceutiques. L'objectif de notre modèle est d'assurer une maintenance prévisionnelle fiable en estimant le RUL de cet équipement, en tenant compte de toutes les contraintes inhérentes à l'industrie biopharmaceutique.

Chapitre 3

3. Du modèle de dégradation de l'équipement

Comme discuté au chapitre précédent, la maintenance prévisionnelle des équipements de production connaît un gain d'intérêt croissant dans l'industrie biopharmaceutique, où la fiabilité des installations est cruciale pour garantir la qualité des produits et la disponibilité des équipements tout en respectant des normes contraignantes. Dans ce domaine, les vannes à membranes pneumatiques, utilisées pour contrôler les flux (fluides et gaz), jouent un rôle central dans les systèmes de production fluidiques. La membrane, pièce d'usure principale de ces vannes, est soumise à des contraintes mécaniques pouvant entraîner des déformations ou des ruptures. Ces défaillances peuvent perturber le processus de production et nuire à la qualité des produits.

Dans ce contexte, le Prognostics Health Management (PHM) émerge comme une méthode intéressante pour apprécier l'état de la membrane. Par l'emploi de données pertinentes et représentatives, il devient possible de modéliser l'évolution de l'état de la membrane et établir un pronostic d'apparition de défaillance. Cette approche prévisionnelle, axée sur l'analyse des phases de vie de la membrane et des tendances, peut permettre de planifier des interventions de maintenance ciblées avant qu'une défaillance ne survienne. Cet article explore l'intégration du PHM pour la maintenance des membranes de vannes pneumatiques en milieu biopharmaceutique. Nous exposons ici une méthode permettant l'identification et la récupération de données représentatives, une proposition de critères permettant d'identifier le cycle de vie de ces membranes basés sur les données collectées, une analyse de ces critères, et une mise en perspective par-rapport à des pratiques de maintenance systématique ayant cours dans une industrie biopharmaceutique. Nous opérerons également une analyse et une comparaison des signaux. L'hypothèse que nous souhaitons vérifier ici est que l'état de la membrane de vanne peut se caractériser par l'étude de la dérive des valeurs de potentiométrie, ceci pour nous permettre à termes le développement d'une maintenance prédictive liée à cet équipement. L'ensemble des éléments constituant ce chapitre ont fait l'objet d'une soumission dans un journal indexé.

3.1. Introduction et contexte industriel

Pour assurer la fiabilité des équipements et anticiper les défaillances des membranes de vannes pneumatiques, nous apportons dans ce chapitre une réflexion sur la nécessité de passer au-delà des méthodes de maintenance systématique traditionnelles pour ce cas d'usage.

Ce chapitre va présenter une problématique industrielle soumise par notre partenaire industriel. Le sujet sera positionné par le biais d'un état de l'art en section 3.2, traitant des pratiques de maintenance ciblant les vannes industrielles. Nous présenterons ensuite en section 3.3 un cas d'usage basé sur des données issues d'une plateforme expérimentale, permettant de modéliser le comportement des membranes de vannes. En section 3.4, nous analyserons les résultats obtenus et proposerons des indicateurs de vie et de dégradation de l'équipement, basés sur les données récoltées. En section 3.5, nous opérerons un comparatif physique d'une membrane neuve en rapport avec la membrane dégradée de notre étude. Enfin en section 3.6, nous conclurons cet article et proposerons des perspectives à ces travaux.

L'émergence des technologies de l'Industrie 4.0 nous permet de ne plus considérer les vannes à membrane pneumatique comme de simples vannes améliorées, mais comme des équipements se comportant comme des dispositifs IOT. Cela permet la possibilité d'évaluer les conditions fonctionnelles à l'aide d'une gamme de données collectées.

Les contributions proposées dans ce chapitre sont :

- La caractérisation des cycles de vie et la détection des défauts dans le comportement d'un équipement appliqué aux membranes de vannes pneumatiques pendant leurs phases de rodage, de maturité et de vieillesse.
- La proposition et l'analyse de critères permettant de prédire l'état de fonctionnement de cet équipement à travers la mesure des signaux captés pendant le fonctionnement du système. L'objectif de cette proposition est d'initier une méthode de maintenance prévisionnelle.
- L'application de critères de défaillance dans l'environnement biopharmaceutique, en particulier à la lumière des contraintes qualités inhérentes à cette industrie.

L'industrie biopharmaceutique fait face à des défis importants pour optimiser la maintenance des équipements, notamment en passant de stratégies systématiques à des stratégies prédictives. Les plans de maintenance systématique actuels, largement adoptés dans le secteur, reposent sur des calendriers prédéfinis pour le remplacement des composants critiques, tels que les pièces d'usure, sur la base des recommandations du fabricant. Cependant, cette approche conduit souvent à des remplacements prématurés, où des composants encore fonctionnels sont mis au rebut, entraînant un gaspillage de consommables. De plus, la maintenance systématique ne prend pas en compte les défaillances résiduelles, nécessitant des interventions curatives qui perturbent la production, compromettent la qualité du produit, et entraînent des coûts excessifs en raison des arrêts imprévus et des problèmes de qualité.

La production d'un biomédicament est caractérisée par des exigences strictes, car elle fonctionne dans des environnements contrôlés pour garantir la sécurité des produits et la conformité aux normes réglementaires. Les processus de production fluidiques impliquent la régulation précise des flux de liquides et de gaz, souvent dans des conditions rigoureuses comme le nettoyage à la vapeur ou l'exposition à des agents acides ou basiques. Ces processus exigent un équipement à la fois inerte vis-à-vis du produit et résistant aux protocoles de nettoyage, garantissant l'absence de contamination et le maintien de l'intégrité de la production.

L'avènement de l'Industrie 4.0 et de l'Internet des objets (IOT) offre des perspectives intéressantes pour l'optimisation. Ces technologies permettent la surveillance en temps réel et la prise de décisions basées sur les données, facilitant ainsi une transition vers la maintenance conditionnelle, voire prévisionnelle. Cependant, cette transition nécessite la capture et le traitement efficaces des données opérationnelles afin de surveiller l'état des équipements et de prévoir avec précision les défaillances. En l'absence de telles données, l'industrie peine à mettre en œuvre des stratégies de maintenance « juste-à-temps » alignant les interventions sur l'état réel des équipements, réduisant ainsi le gaspillage et renforçant la fiabilité.

Dans ce contexte, l'équipement sélectionné pour cette étude est la vanne à membrane pneumatique, un composant critique des systèmes de production fluidique. Cette vanne, largement utilisée dans les environnements biopharmaceutiques, assure la régulation du flux tout en maintenant une inertie en regard du produit traité. Toutefois, sa membrane est sujette à l'usure, à la déformation ou à la rupture, ce qui pose des risques pour la continuité de la production et la qualité du produit.

3.2. État de l'art

3.2.1. Maintenance prévisionnelle IOT-driven

La maintenance prévisionnelle basée sur l'IOT est apparue comme un levier d'intérêt dans l'industrie biopharmaceutique (Tulsyan et al., 2020 ; Carvalho et al., 2024), bien que son adoption reste limitée par rapport à d'autres secteurs comme l'industrie automobile (Chinta et al., 2023 ; Mohanraj, 2024), l'industrie de l'énergie (Atassi & Alhosban, 2023 ; Jambol et al., 2024), ou plus largement l'industrie manufacturière (Ahn et al., 2023 ; Hosseinzadeh et al., 2023 ; Meddaoui et al., 2023 ; Rusiński et al., 2016)). En permettant la collecte de données en temps réel ou quasi-réel, l'Internet des objets offre des opportunités sans précédent pour surveiller efficacement les équipements et anticiper les défaillances. Dans plusieurs cas, les technologies basées sur des algorithmes d'apprentissage automatique ont démontré leur efficacité en permettant un diagnostic précis et rapide des défaillances (Zadiran & Shcherbakov, 2023 ; Turner et al., 2022). Toutefois, dans les environnements biopharmaceutiques, des exigences réglementaires strictes, notamment en ce qui concerne les normes de qualité discutées à l'item 1.1.1, compliquent l'intégration de ces outils dans les processus réels. Ces contraintes nécessitent des solutions spécifiques capables de garantir la conformité aux processus de maintenance tout en exploitant pleinement le potentiel des données issues de l'IOT (Fantana et al., 2013 ; Reinhardt et al., 2020). Malgré ces défis, les applications de l'IOT dans ce secteur révèlent un potentiel d'innovation significatif, notamment en termes d'optimisation des processus et de réduction des coûts associés aux interruptions imprévues.

La méthodologie Prognostics and Health Management (figure 10) fournit un cadre méthodologique structuré permettant d'anticiper les défaillances, de prolonger la durée de vie des équipements et d'optimiser la maintenance. De nombreux exemples dans la littérature scientifique décrivent l'élaboration de plans de maintenance prévisionnelle dans l'industrie manufacturière, alignés sur les principes généraux de cette méthode (Xia et al., 2018 ; Zio, 2022).

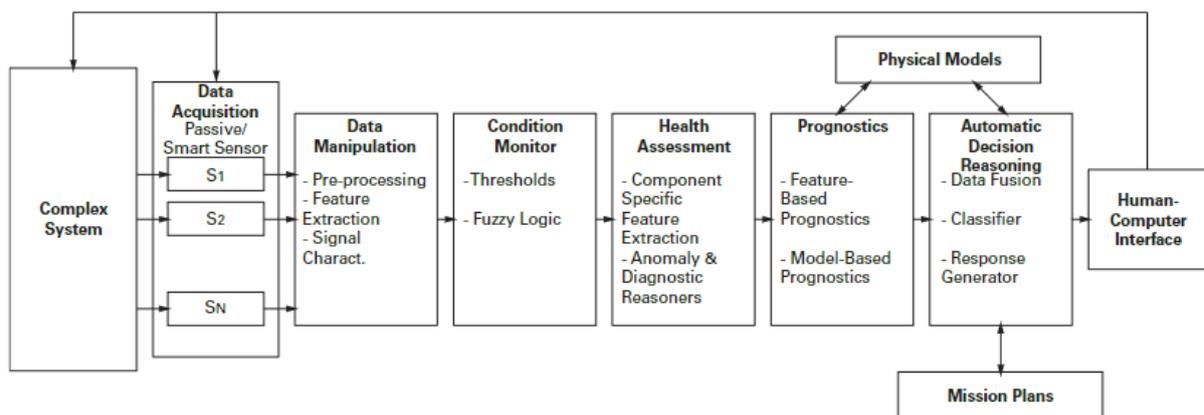


Figure 10 - Illustration des étapes de la méthode PHM selon Gouriveau et al., (2013)

Comme illustré dans la figure 10, ce paradigme repose sur un ensemble d'étapes clés : l'acquisition de données, le raffinement des données, la détection d'anomalies, le diagnostic, et l'estimation de la durée de vie restante des équipements ciblés.

Ces approches peuvent être basées sur des modèles physiques, des méthodes guidées par les données, ou des combinaisons hybrides des deux. Appliqué aux équipements fluidiques, le PHM s'avère efficace pour prévenir les défaillances, réduire les coûts d'exploitation et

minimiser les temps d'arrêt non planifiés. Dans des environnements critiques tels que l'industrie biopharmaceutique, où la continuité opérationnelle est essentielle, l'intégration du PHM peut aligner les besoins opérationnels avec les exigences réglementaires strictes. Cette méthodologie offre également une flexibilité significative, permettant son application à une variété d'équipements et de scénarios, en particulier ceux impliquant des processus fluidiques complexes. Une adoption plus large de ce cadre pourrait transformer les pratiques industrielles en combinant performance, fiabilité et conformité.

Les vannes à membrane pneumatiques présentent des défis spécifiques liés à la dégradation de la membrane. Leurs défaillances, indétectables par les méthodes traditionnelles, peuvent entraîner des arrêts non planifiés et compromettre la qualité du produit. Les avancées de l'IOT permettent désormais à ces vannes d'être équipées de systèmes de positionnement capables de collecter des données représentatives de leur état de fonctionnement. Ces données, pouvant être indépendantes du système de commande, offrent une visibilité sur des phénomènes tels que la dégradation progressive de la membrane, permettant ainsi la prévision et une planification des interventions de maintenance basée sur l'état réel de l'équipement. Dans la section suivante, nous nous concentrons sur les travaux portant sur cette vanne.

3.2.2. Détection des défaillances des vannes

En raison de son rôle critique dans la gestion des flux, la vanne est un élément central dans tout processus de fabrication impliquant des fluides. Elle peut être configurée sous diverses formes (Nesbitt, 2011). Les vannes industrielles se composent généralement d'un corps de vanne et d'un actionneur permettant leur ouverture et fermeture. Elles peuvent être manuelles, automatiques (contrôlées par un automate programmable (PLC)), l'actionneur peut être équipé d'un ressort et d'un piston. Le contrôle peut être électrique ou pneumatique. Une variante utilisée dans les processus fluidiques intègre une membrane entre l'actionneur et le siège de la vanne, dont les défaillances sont difficiles à détecter car non visibles. À l'ère des principes de l'Industrie 4.0, où les pratiques de maintenance évoluent vers des méthodes dites prédictives, la question se pose de l'applicabilité de ces pratiques aux vannes équipées de membranes. Dans cette section, nous examinons plusieurs études notables pouvant appuyer la maintenance prédictive des vannes industrielles, présentées par ordre chronologique.

Une première étude de T. Matsui et al. (1995) décrit l'utilisation d'un Réseau de Neurones Récurrent (RNN) pour détecter des défauts sur des vannes de régulation pneumatiques sans membrane. L'expérience proposée consiste à analyser le déplacement du positionneur de vanne et à comparer le signal 4-20 mA collecté d'une vanne sans défaut avec des signaux issus de pannes induites. Les auteurs relèvent des variations dans les signaux électriques pouvant être corrélées à des états de dérive, cohérents avec des classifications visuelles établies par des experts. Un réseau de neurones est utilisé pour traiter les signaux et les classer sur la base des caractéristiques et des types de défauts connus. Ensuite, Daigle et al. (2011) proposent une approche pronostique basée sur un modèle pour caractériser des défauts sur des vannes pneumatiques sans membrane dans un système de transfert d'hydrogène liquide. L'expérience consiste à modéliser des paramètres physiques et des séries temporelles telles que l'air injecté dans l'actionneur pneumatique, les positions de la vanne et les temps d'ouverture et de fermeture. D'un côté, les paramètres physiques d'une vanne saine sont présentés, et de l'autre, ceux d'une vanne dégradée. L'expérience repose à la fois sur des données réelles et simulées pour établir une durée de vie résiduelle (RUL).

Nous avons également trouvé deux travaux sur la détection et le diagnostic de défauts dans un actionneur pneumatique. Dans Subbaraj et al. (2014), les auteurs proposent un modèle ANFIS (Adaptive Neuro-Fuzzy Inference System) pour la détection et le diagnostic de défauts sur une vanne à actionneur pneumatique sans membrane dans l'industrie du ciment. Ils se concentrent dans cette publication sur l'actionneur pneumatique. Dix-neuf défauts liés à l'actionneur sont caractérisés. Le modèle proposé repose sur des données d'apprentissage générées, dont 75 % sont utilisées pour l'apprentissage et 25 % pour les tests. Ces données prennent en compte cinq variables : le mouvement du potentiomètre, la puissance d'entrée et de sortie de la pompe, la pression d'entrée, la température de sortie de l'eau, et la sortie du transmetteur de pression différentielle, créant une sortie étiquetée indiquant la présence (et le type, le cas échéant) d'un défaut ou son absence. Cependant, l'étude de Santhosh et al. (2016) utilise une approche de détection de contours pour identifier les défauts dans l'actionneur pneumatique d'une vanne sans membrane. Pour ce faire, des images du système sont collectées, une région d'intérêt est identifiée dans chaque photo, et les bords de cette région sont extraits. Les changements de position des éléments dans la région d'intérêt sont ensuite calculés. Si un changement anormal de position spatiale est détecté, un défaut est enregistré.

D'un autre côté, Sano et al. (2019) comparent divers algorithmes d'apprentissage automatique pour prédire les défauts sur les vannes de coupure à pression régulée (PRSOV). Ces vannes de régulation pneumatiques se trouvent notamment dans les circuits d'alimentation en carburant des avions. Les auteurs ont simulé des défauts tels que des fuites, des dysfonctionnements de ressort et des frictions afin de générer des données d'apprentissage. Six algorithmes (SVM, kNN, NB, CART, MLR, MLP) ont été entraînés et comparés pour le diagnostic des défauts.

En 2021, deux travaux concernant des problèmes similaires ont été publiés. Andrade et al. (2021) ont présenté une méthode de diagnostic de défauts sur des vannes pneumatiques sans membranes utilisant des techniques de détection et de diagnostic de défauts (FDD) et des arbres de décision. Une plateforme de test appelée AUTHOMATHIKA, équipée d'un actionneur pneumatique, d'une vanne de régulation et d'un positionneur, a été utilisée pour l'expérience. Les auteurs ont développé une méthodologie exploitant les données des capteurs pour caractériser les défauts. Sur la base des motifs de défaillance obtenus, une classification par arbre de décision a été réalisée pour diagnostiquer les défauts. Et Han et al. (2021) ont permis la détection de 14 défauts sur 19 liés à des vannes pneumatiques simulées sans membranes en utilisant le benchmark DAMADICS, la détection ayant été réalisée à l'aide de la méthode CVA-SMD (Canonical Variate Analysis on Square of the Mahalanobis Distance).

Diverses catégories de défauts ont été abordées, y compris celles des vannes elles-mêmes, ainsi que celles du servomoteur et du positionneur. D'autres défauts plus généraux ont également été étudiés, l'un d'entre eux portant notamment sur des défauts dus à la fatigue des membranes internes du servomoteur. Les auteurs ont testé la méthode CVA-SMD pour la détection des défauts et l'ont comparée à des méthodes statistiques plus courantes telles que PCA-T² (Principal Component Analysis) et SVA (Surrogate Variable Analysis). Selon les résultats de l'étude, CVA-SMD présentait le taux de fausses alertes le plus faible et le taux de fausses détections le plus élevé parmi les méthodes statistiques.

Enfin, Chen et al. (2024) discutent de l'application d'un système DGM (1,1) (Dynamic Grey Model type 1,1) pour la maintenance prévisionnelle de vannes pneumatiques sans membranes. L'étude porte sur des équipements utilisés dans l'industrie pétrolière. L'approche

cible les redondances dans les signaux mesurés à partir des actionneurs pneumatiques afin de modéliser les défauts, permettant leur prévision dans le futur.

Les recherches menées dans ce domaine et sur ce type d'équipement dans la littérature scientifique concernent principalement des défauts liés aux actionneurs pneumatiques des vannes. Nous n'avons identifié aucune étude ciblant spécifiquement la dégradation des membranes dans les vannes pneumatiques couramment utilisées dans les industries de procédés liquides, notamment dans le secteur biopharmaceutique.

En conclusion, nous mettons ici en évidence l'absence de méthode permettant d'évaluer l'usure des membranes de vannes. Compte tenu des contraintes réglementaires associées à la production biopharmaceutique, il est d'autant plus crucial d'établir des critères d'évaluation de l'usure des membranes, ce que nous proposons dans cet article.

3.3. Description du système étudié

3.3.1. *Vanne à membrane pneumatique*

La vanne à membrane pneumatique est utilisée dans les processus de production fluidiques. Elle est spécifiquement employée dans l'industrie du biomédicament en raison de sa compatibilité avec les environnements de production stériles et aseptiques. Comme discuté plus haut, la membrane de la vanne, au-delà de son rôle de blocage des fluides, est conçue pour être inerte vis-à-vis du produit circulant dans les lignes de production, garantissant ainsi la non-contamination du produit traité. Elle doit également résister au nettoyage des lignes de production à l'aide de vapeur ou d'autres agents, qu'ils soient acides ou alcalins. Généralement composées de matériaux en caoutchouc et/ou en plastique, ces membranes peuvent se rompre, se déformer, et impacter significativement le processus de production en réduisant la disponibilité des équipements et en introduisant un risque de contamination du produit.

Les équipements de production biopharmaceutique équipés de vannes à membrane pneumatique sont également difficilement accessibles, car ils sont généralement situés dans des environnements à atmosphère contrôlée propres aux industries produisant des biomédicaments. Le plan de maintenance le plus répandu actuellement adopté par notre partenaire industriel, et plus largement dans l'industrie biopharmaceutique, est le plan de maintenance systématique (Wadel et al., 2024). Les membranes des vannes pneumatiques, principaux composants d'usure de ces équipements, sont remplacées chez notre partenaire industriel sur la base d'un critère de durée d'utilisation dérivé des recommandations du fabricant, intégré à un plan de maintenance systématique. Bien que cette approche permette un suivi temporel des pièces d'usure, elle ne garantit pas l'absence de défaillances. Dans notre cas d'usage, l'usure prématurée des membranes due à un assemblage non conforme et les défaillances soudaines ne sont pas traitées par la maintenance systématique de façon optimale. Par ailleurs, il arrive fréquemment que les membranes remplacées l'aient été de manière prématurée, alors qu'elles étaient encore viables, posant la question du gaspillage des consommables.

Du fait que les membranes ne sont pas visibles sans démontage préalable de l'ensemble de l'équipement, il est important de disposer d'indicateurs d'état permettant une maintenance « juste-à-temps » et d'établir des horizons pertinents pour la prévision des interventions de maintenance.

La vanne à membrane pneumatique (figure 11) est constituée d'un corps de vanne à travers lequel circule le fluide, d'un actionneur pneumatique permettant l'ouverture et la fermeture de la vanne, et d'une membrane reliant le siège de la vanne à l'actionneur. L'actionneur, une fois alimenté en air comprimé, tire la membrane verticalement pour permettre la circulation du fluide à travers le siège de la vanne.

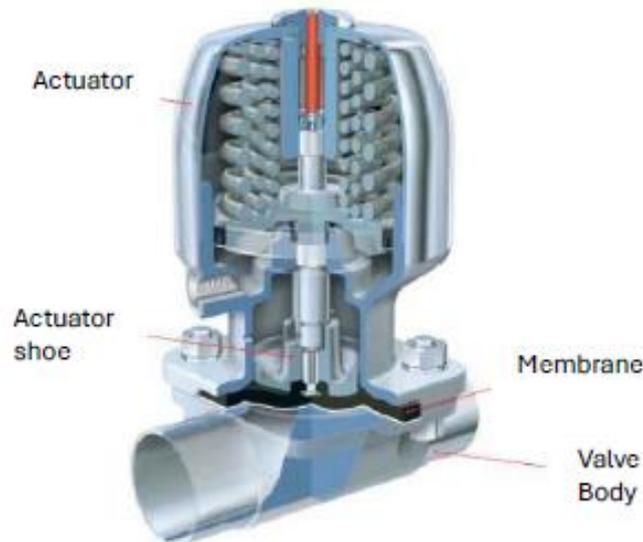


Figure 11 - Vanne à membrane pneumatique GEMÜ Biostar 650

Généralement, l'équipement peut être constitué de divers matériaux (plastique ou métal), selon le fluide qui circulera à travers la vanne et les exigences de non-contamination des produits propres au processus de production. Le corps de vanne sert également de support à l'actionneur et à la membrane.

L'actionneur pneumatique est un composant qui permet l'ouverture et la fermeture de la vanne. De l'air comprimé est injecté dans la chambre de l'actionneur, exerçant une pression sur un ressort interne, ce qui permet un déplacement vertical de la membrane et permet alors au fluide de passer à travers le corps de vanne. La vanne se ferme lorsque l'alimentation en air comprimé de la chambre de l'actionneur est arrêtée, entraînant la détente du ressort interne et le retour de la membrane à sa position initiale, elle-même limitée par le siège de vanne. Ce principe est illustré dans la Figure 3.

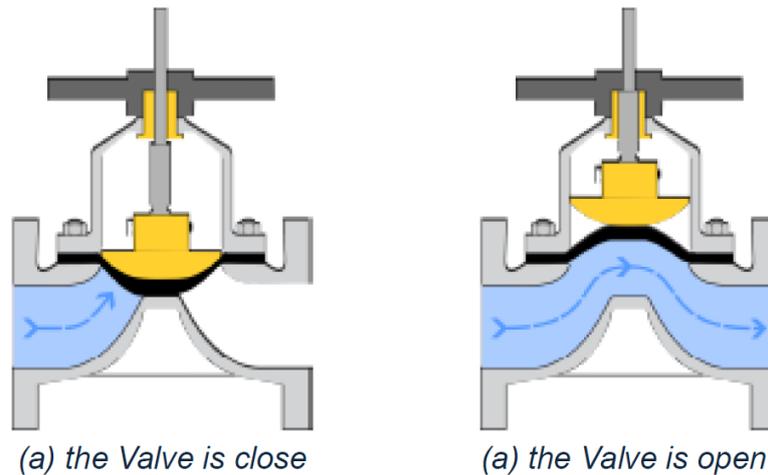


Figure 12 - Principe générique d'une vanne à membrane dérivé du schéma de Muskid, CC BY-SA 3.0, Wikimedia Commons

Fixée à l'actionneur par un pas de vis, la membrane est contrainte par l'air comprimé injecté dans l'actionneur et effectue un mouvement vertical pour permettre ou empêcher le passage du fluide. Notre étude porte spécifiquement sur les membranes EPDM/PTFE (éthylène-propylène-diène monomère / polytétrafluoroéthylène) (figure 13), couramment utilisées dans les procédés fluidiques de notre partenaire industriel, bien qu'elles puissent être fabriquées dans différents matériaux selon les contraintes du procédé.



Figure 13 - Membrane de vanne pneumatique de la marque GEMÜ

Ces vannes peuvent également être équipées d'un indicateur de position électrique (figure 14) vissé sur l'actionneur. La fonction principale de cet équipement est d'informer l'automate qui pilote les vannes de leur état d'ouverture ou de fermeture, en lui renvoyant une information basée sur une valeur potentiométrique représentative de sa position.



Figure 14 - Indicateur de position électrique de la marque GEMÜ

3.3.2. Modes de défaillance

Le principal mode de défaillance identifié pour ces membranes est la déformation de la couche de PTFE, laquelle va entraîner des fuites dans le système fluide. Le PTFE est un polymère fluoré synthétique connu pour sa résistance chimique et son faible coefficient de frottement (Feng et al., 2018). Cependant, l'une de ses principales faiblesses est sa faible résistance mécanique (Conte & Igartua, 2012). Comparé à d'autres plastiques industriels, il est plus sujet à la déformation ou à l'écrasement sous de fortes charges mécaniques ou un stress prolongé, en particulier à basse température.

La membrane de vanne étudiée comprend un composant en PTFE, qui est mobile, en contact direct avec les produits, et soumis à des contraintes mécaniques à chaque actionnement de la vanne. En raison des propriétés du matériau plastique, le mouvement d'ouverture-fermeture entraîne une déformation progressive dans les zones de contrainte de la membrane. Cette déformation peut créer une zone de rétention dans le système fluide, difficilement accessible lors des processus de nettoyage des lignes de production, pouvant potentiellement entraîner une contamination du produit. Cette déformation peut également empêcher la membrane d'assurer correctement son rôle d'étanchéité dans la ligne. Dans les cas extrêmes, la déformation peut entraîner la rupture soudaine de cette section de la membrane. Une telle rupture provoquerait une fuite importante de produit par les orifices de drainage de l'actionneur.

Bien que l'apparition de la déformation de la membrane soit inévitable en raison des propriétés du matériau et des conditions de fonctionnement, il convient de noter que le phénomène peut être aggravé par un montage non conforme de la membrane sur l'actionneur (par exemple, un serrage non conforme).

3.4. Expérimentation sur banc d'essai

L'expérience décrite ici consiste à reproduire le contrôle d'une vanne à membrane pneumatique tel qu'il serait réalisé dans le processus de production, avec la mise en œuvre d'un système permettant le suivi de certaines données liées à la sollicitation de la membrane. Les équipements et l'expertise opérationnelle utilisés pour réaliser cette expérience ont été fournis par notre partenaire industriel. L'ensemble des équipements utilisés est identique à celui que l'on trouve sur les lignes de production.

3.4.1. Matériel

Pour collecter des données représentatives de la dégradation de la membrane dans des conditions expérimentales, un banc d'essai (figure 15) a été mis en place. Celui-ci a été conçu

pour simuler les conditions rencontrées dans le processus de production fluide de notre partenaire industriel. Le banc d'essai a été équipé d'une vanne à membrane pneumatique, d'une pompe, ainsi que d'un système automatisé de contrôle et de surveillance. Le diamètre nominal (DN) de la vanne à membrane pneumatique utilisée pour l'expérimentation a été sélectionné selon un critère de représentativité, correspondant à la taille la plus couramment utilisée sur les installations de notre partenaire industriel. Enfin, la vanne étudiée a été montée et préparée par un expert de l'entreprise en question, garantissant un montage conforme.

Le banc d'essai se compose d'une boucle fermée remplie d'eau purifiée, comprenant :

- Une vanne à membrane pneumatique (1)
- Un capteur de température (2)
- Un manomètre (3)
- Une bride mécanique (4) pour garantir une pression stable dans le circuit en amont de la vanne
- Un surpresseur équipé d'un détecteur de fuites pour d'éventuelles fuites dans le circuit alimenté (5)
- Un réservoir de collecte (6)

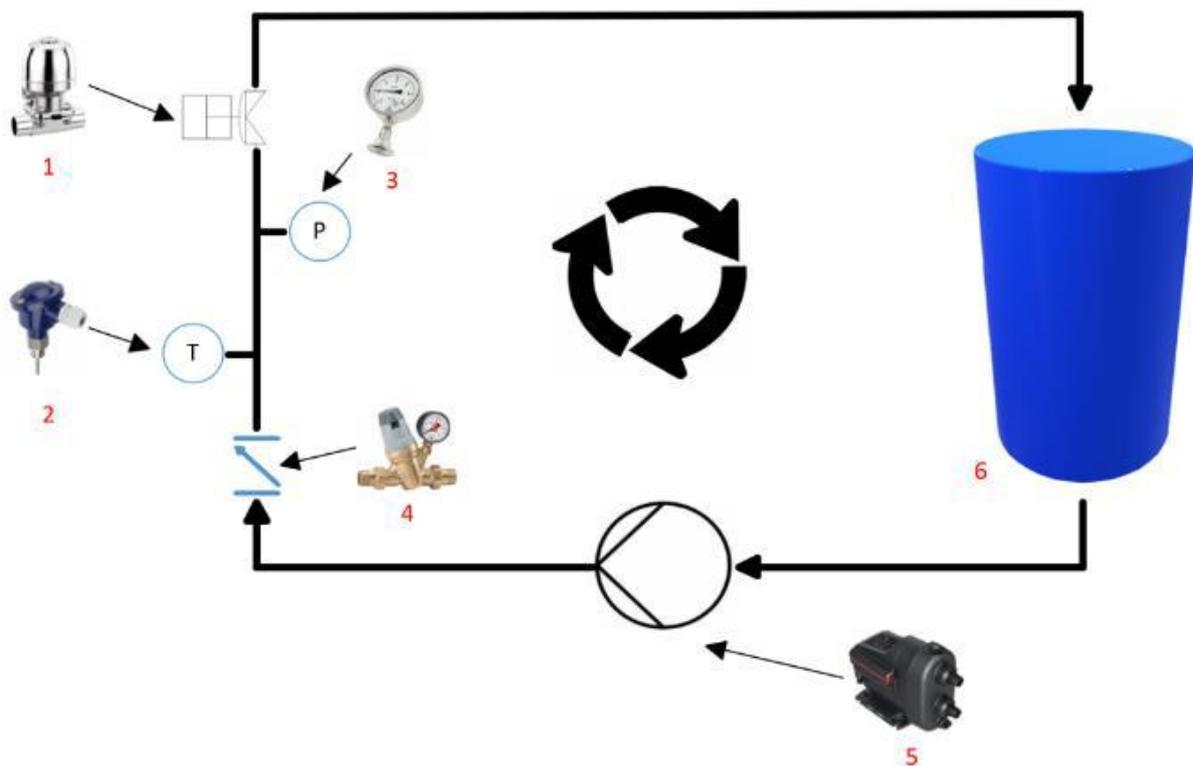


Figure 15 - Schéma de la plateforme de test

3.4.1.1. Pilotage du banc d'essai

Une architecture d'automatisation IO-Link (figure 16) a été mise en place pour contrôler le banc de test et collecter les données pertinentes à transférer dans une base de données de type Data Lake, spécifiquement créée à cet effet.

IO-Link est un protocole de communication point-à-point utilisé dans l'automatisation industrielle. Ce système se compose, au minimum, d'un PLC, d'un module « maître » auquel les capteurs et actionneurs, appelés « esclaves », sont directement connectés. Plusieurs

publications dans la littérature scientifique décrivent des plateformes de test utilisant IO-Link (Heynicke et al., 2018 ; Liebl et al., 2023).

Dans notre cas d'étude, un avantage d'IO-Link réside dans la capacité à communiquer avec les équipements « esclaves » sans impliquer le PLC et sans affecter le flux des trames de commande adressées par le PLC aux équipements.

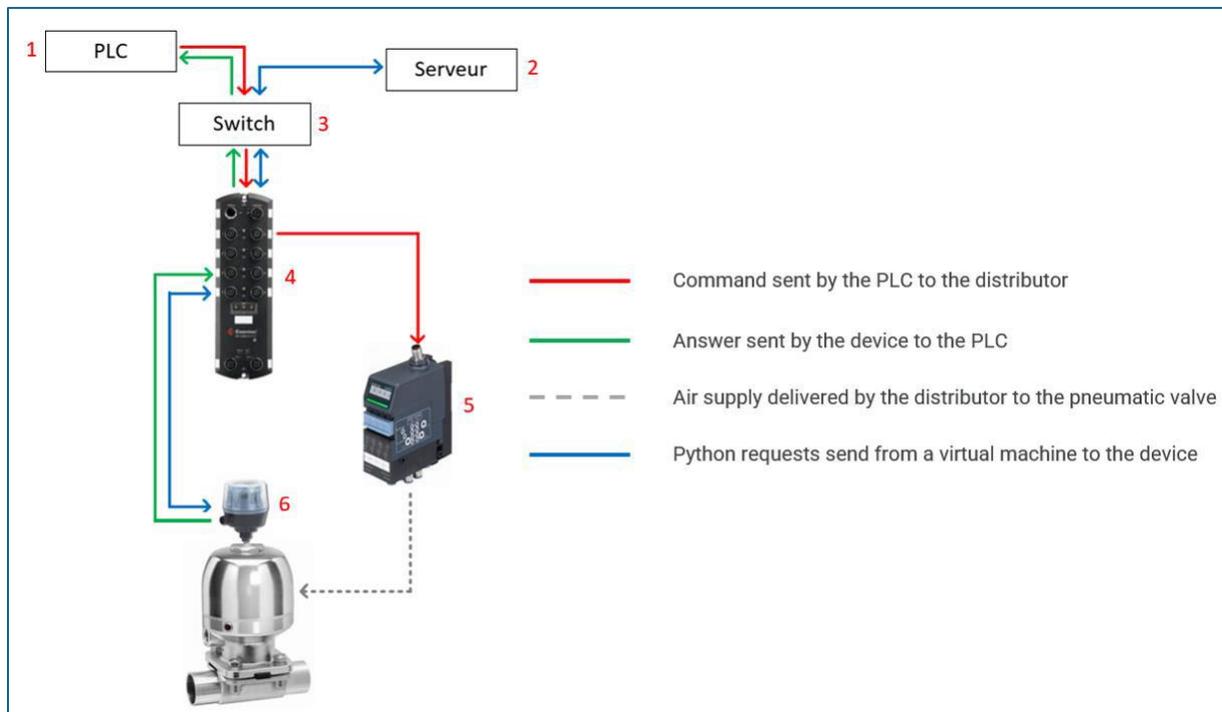


Figure 16 - Architecture IO-Link pour le pilotage de la plateforme de test + Système de collecte de données Python

L'architecture du banc d'essai se compose de :

- Un PLC (1)
- Un serveur hébergeant une base de données (2)
- Un switch non manageable (3)
- Un maître IO-Link (4)
- Un îlot de distribution pneumatique pour le contrôle de la vanne via l'actionneur (5)
- Un indicateur de position électrique à course de 30 mm (6)

3.4.2. Méthode

Les données utilisées pour notre cas d'usage proviennent de l'indicateur de position électrique de la vanne (figure 14). Elles sont collectées en continu à l'aide d'un script d'acquisition écrit en Python, permettant la communication avec l'indicateur de position électrique via le maître IO-Link, sans impliquer l'API. Ces données sont qualifiées d'« acycliques ». Elles ne sont pas spécifiquement requises par l'API pour le pilotage de l'équipement mais sont exploitées par l'équipement pour son propre fonctionnement. Dans le cadre de notre étude, nous récupérons des valeurs potentiométriques permettant à la vanne d'indiquer si elle est ouverte ou fermée, selon son diagramme de commutation (figure 17).

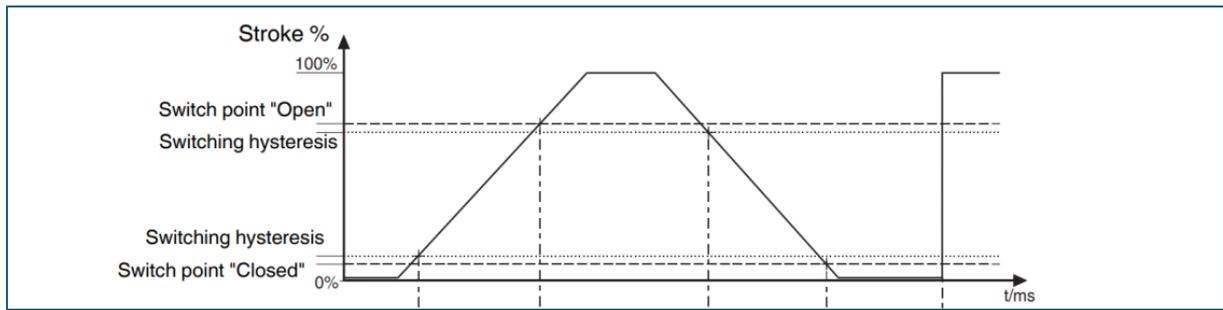


Figure 17 - Diagramme de commutation extrait de la documentation technique de fabricant

3.4.2.1. Déroulement de l'expérimentation

Une membrane EPDM/PTFE a été installée de manière conforme sur l'actionneur pneumatique de la vanne DN25 par un expert technique employé par notre partenaire industriel. Cette précaution est nécessaire pour éviter de rencontrer des défaillances liées à un montage non conforme, comme décrit dans la section « Mode de défaillance ».

3.4.2.2. Conditions de contrôle du banc d'essai

Le liquide circulant dans la boucle est de l'eau purifiée, avec les paramètres physiques du circuit suivants :

- ≈ 2 bar de pression dans le collecteur
- Liquide à température ambiante à ≈ 20 °C

L'expérimentation s'est déroulée en continu pendant douze mois. Des cycles d'ouverture/fermeture, d'une durée totale d'une heure et vingt minutes chacun, ont été réalisés et pilotés par un automate programmable industriel (API).

L'ouverture de la vanne dure cinq minutes, suivie d'une période de fermeture d'une heure et quinze minutes. Cela permet de réaliser 20 cycles par jour, simulant des conditions de procédé accélérées.

3.4.2.3. Vérification de l'état de la membrane

Étant donné que le démontage et le remontage d'une vanne provoque des perturbations dans les signaux enregistrés, nous avons décidé de planifier un seul démontage pour vérifier l'état de la membrane. Ce démontage a été programmé lorsque 3 500 cycles de commande ont été atteints. Cet horizon a été choisi en fonction des pratiques de maintenance actuellement en place chez notre partenaire industriel. Il convient de noter que dans un contexte industriel, cet horizon peut varier selon les équipements et leur position sur la ligne de production.

3.4.2.4. Caractérisation du signal

Dans cette section, nous allons discuter du signal brut ainsi que des méthodes employées pour le traiter. Nous avons suivi les étapes présentées dans la figure 18.

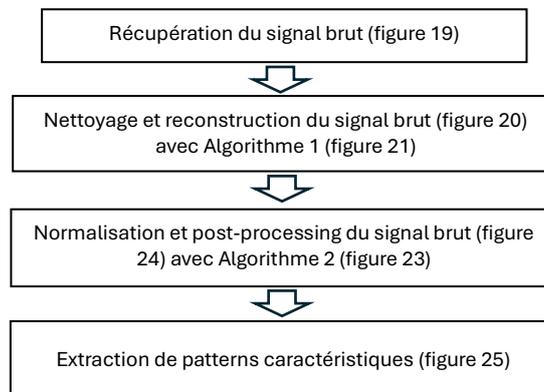


Figure 18 - Etapes de traitement du signal

Le signal brut collecté dans cette étude consiste à l'origine en des valeurs potentiométriques retournées par l'indicateur de position lors de la fermeture de la vanne. Plus précisément, il inclut toutes les valeurs enregistrées une fois que l'indicateur de position dépasse un seuil prédéfini, établi automatiquement par le système, lui permettant de signaler au PLC que la vanne est fermée. Ces valeurs sont récupérées au format numérique, pour refléter la position de la membrane une fois que l'indicateur a basculé en fermeture.

Le signal comprend un ensemble de profils de "stabilisation" de la membrane contre le siège de la vanne une fois que l'état fermé est atteint, cycle après cycle d'ouverture et de fermeture de la vanne. La figure 19 présente le signal pour un ensemble de 7 cycles d'ouverture/fermeture, mettant en évidence l'émergence d'un décalage dans les positions.

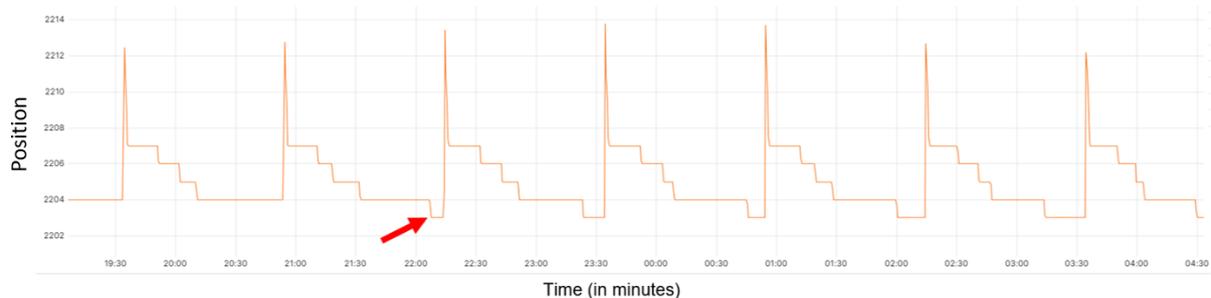


Figure 19 - Signal de position extrait de l'indicateur de position pour 7 cycles avec mise en évidence d'une dérive

Le signal étudié couvre l'ensemble du cycle de vie d'une membrane, depuis son installation sur l'actionneur jusqu'à son remplacement, après avoir atteint une condition jugée préoccupante pour le composant PTFE en contact avec le produit. Le signal subira un post-traitement pour mettre en évidence les principales phases du cycle de vie de la membrane, à savoir, le rodage, la maturité et la vieillesse, conformément à la distribution de Weibull.

Cette loi est une distribution de probabilité utilisée pour modéliser la fiabilité et la durée de vie des matériaux et des systèmes. Introduite par Waloddi Weibull en 1951 (Weibull, W., 1951), elle est particulièrement utile pour analyser la probabilité de survie et d'échec des systèmes. Elle est définie par trois paramètres : la position (γ), l'échelle (η) et la forme (β). Le paramètre de forme détermine la distribution des taux de défaillance. Si $\beta < 1$, le taux de défaillance est

relativement élevé mais décroissant, correspondant à l'élimination des défauts précoces et à la période de rodage. Si $\beta = 1$, la phase de maturité est caractérisée par un taux de défaillance faible et constant, les composants ayant prouvé leur résilience aux défauts précoces. Si $\beta > 1$, la phase de vieillesse est marquée par un taux de défaillance croissant avec le temps, à mesure que les composants subissent une usure. Cette flexibilité permet à la distribution de Weibull de modéliser la durée de vie de divers composants, tels que les composants électroniques (Méndez-González et al., 2017 ; Shuja et al., 2025), les systèmes mécaniques (Zhang et al., 2014), ou les vannes pneumatiques (Mathur et al., 2021).

Pendant la phase de rodage, les membranes sont généralement robustes et subissent peu de défaillances. La phase de maturité est caractérisée par une performance stable et une fiabilité maximale. En revanche, pendant la phase de vieillesse, les défaillances deviennent plus fréquentes en raison de la dégradation du matériau. Cependant, au-delà des changements normaux associés aux phases du cycle de vie, les capteurs peuvent détecter des signaux indiquant des perturbations anormales. Ces perturbations peuvent résulter de facteurs externes tels que des variations de pression, des fluctuations de température, des impuretés dans le fluide passant à travers la vanne, ou des dysfonctionnements mécaniques.

La caractérisation des signaux provenant des capteurs installés sur les vannes permet une détermination précise de la phase du cycle de vie d'une membrane et la prévision de son comportement futur. En analysant les données de position, qui mesurent les variations de résistance ou de tension en réponse aux conditions de fonctionnement, il est possible d'identifier des signatures distinctes associées à chaque phase du cycle de vie : rodage, maturité et vieillesse. Pendant la phase de rodage, les valeurs de position présentent généralement des variations significatives. Lorsque la membrane entre dans la phase de maturité, ces valeurs restent relativement constantes mais peuvent commencer à montrer des signes de fatigue à travers de légères fluctuations. Enfin, pendant la phase de vieillesse, les valeurs de position affichent des variations plus prononcées, indiquant un risque accru de survenue de défaillance. En surveillant ces variations de position, des seuils précis peuvent être définis pour signaler les transitions entre les phases du cycle de vie. Par exemple, un niveau de position spécifique peut indiquer le passage de la membrane du rodage à la maturité ou de la maturité à la vieillesse. Cette approche pourrait être intégrée dans une stratégie de maintenance prévisionnelle, minimisant ainsi les temps d'arrêt et les coûts associés aux défaillances inattendues.

3.4.2.5. Nettoyage du signal brut original

Comme montré dans la figure 19, les profils étudiés extraits du signal brut sont caractérisés par une succession de plateaux plus ou moins prolongés. Nous avons observé que les changements dans le profil proviennent principalement de l'apparition de plateaux supplémentaires et/ou de la modulation des plateaux existants dans le signal brut.

Nous avons initialement nettoyé et reconstruit le signal original en utilisant l'Algorithme 1 (figure 20) pour n'afficher qu'une moyenne mobile des valeurs de position minimales dans les profils, centrée par rapport à la valeur de référence obtenue lorsqu'une nouvelle membrane est montée sur l'actionneur (figure 21). Il est important de noter qu'avec cet algorithme, nous regroupons les valeurs d'origines qui sont des séries temporelles, par cycles. En effet, le signal brut n'est pas directement utilisable en raison de son bruit inhérent et des fluctuations découlant des variations opérationnelles qui masquent les tendances sous-jacentes de

dégradation de la membrane, rendant son interprétation difficile et peu fiable. Par conséquent, conformément aux principes du PHM, il est essentiel de procéder à un nettoyage du signal.

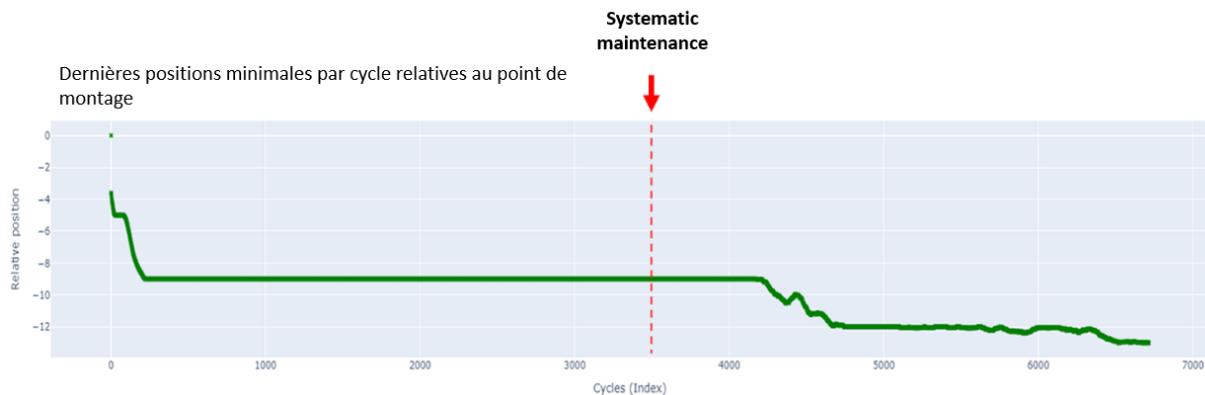


Figure 20 - Signal brut reconstruit avec les valeurs minimales

```

LOAD DATASET_PATH INTO df

ADD COLUMN cycle = ROUND(df["Valve cycles user"])
DROP COLUMNS ["Timestamp", "Valve cycles user"] FROM df
GROUP BY "cycle" INTO df_cycles:
    CALCULATE ["mean", "min", "std", "median"]

COMPUTE df_cycles["centered_closed_pos"] =
    df_cycles["Pos. Ferme Last"]["min"] - initial_closed_pos

SET WINDOW_SIZE = 5

COMPUTE df_cycles["moving_average"] =
    MOVING_AVERAGE(df_cycles["centered_closed_pos"], WINDOW_SIZE)

```

Figure 21 - Algorithme 1, pseudo-code utilisé pour la reconstruction du signal et l'établissement de la moyenne mobile

Les variables « Timestamp », « Valve cycles user » et « Pos. Ferme Last » sont les indices collectés directement à partir de l'indicateur de position électronique GEMÜ. Pour plus de clarté, nous avons choisi de ne pas modifier leurs noms dans le code.

Dans notre cas, les moyennes mobiles sont obtenues à l'aide de la formule (1) :

$$moving_average_i = \frac{1}{W} \sum_{j=i}^{i+W-1} centered_closed_pos_j \quad (1)$$

Avec les paramètres suivants :

- i : indice actuel dans la série temporelle
- w : fenêtre de lissage (taille de la moyenne mobile)
- j : l'indice de sommation qui itère sur les W points à partir de i .
- $centered_closed_pos_j$: les valeurs du signal à lisser.

Comme montré dans la figure 20, cette reconstruction initiale révèle une visualisation d'une première phase du signal jusqu'à 219 cycles. Cette phase est caractérisée par une valeur de

signal élevée qui diminue régulièrement, ce qui peut ainsi être assimilé à une période de rodage.

Une deuxième phase, s'étendant de 220 à 4212 cycles, est caractérisée par un signal constant, qui peut représenter la phase de maturité de la membrane.

Enfin, une troisième phase commence à 4213 cycles, marquée par une diminution du signal par rapport à la phase précédente. C'est dans cette troisième phase que nous pouvons anticiper l'apparition d'une défaillance, ce qui en fait la phase de vieillesse.

À des fins illustratives, nous avons indiqué sur la figure l'horizon de maintenance préventive appliqué par notre partenaire industriel, estimé à 3500 cycles.

Nous avons initialement nettoyé et reconstruit le signal original en utilisant l'Algorithme 1 (figure 21) pour n'afficher qu'une moyenne mobile des valeurs de position minimales dans les profils, par cycle, centrée par rapport à la valeur de référence obtenue lorsqu'une nouvelle membrane est montée sur l'actionneur (figure 20).

Nous nous sommes concentrés sur les plateaux intermédiaires au sein des profils du signal. Identifier les irrégularités dans les plateaux pourrait révéler des tendances de dégradation précoces, que nous pourrions ne pas être en mesure de voir dans le signal brut.

3.4.2.6. Traitement du signal brut original

De nombreux exemples dans la littérature scientifique abordent le traitement de signaux pour les équipements industriels. Par exemple, nous trouvons des cas impliquant des machines rotatives (Rajabi et al., 2022) et des équipements électriques (Cabal-Yepez et al., 2012 ; Slimane et al., 2018), où les signaux sont analysés à l'aide de transformées de Fourier et d'analyse par ondelettes. Cheng et al. (2016), en revanche, post-traite un signal audio en utilisant une Support Vector Machine (SVM) et une fenêtre de filtrage pour extraire des motifs d'intérêt. En réalité, il existe de nombreuses approches génériques qui doivent être adaptées en fonction du signal à traiter. Étant donné que nous traitons un signal spécifique, régulier et de type série temporelle, nous avons choisi de traiter notre signal à l'aide d'un algorithme de régression polynomiale (en utilisant la bibliothèque PolyFit de Python) pour capturer les événements liés à la modulation des plateaux intermédiaires dans les profils à analyser (figure 19).

Nous avons développé un algorithme qui récupère d'abord, pour chaque profil :

- Le nombre de plateaux détectés par profil.
- Leur amplitude temporelle, avec 1 point sur l'axe des y correspondant à une ½ seconde.

Toutes ces caractéristiques sont compilées dans des histogrammes, à partir desquels nous extrayons, en utilisant l'Algorithme 2 (figure 23), les valeurs d'une régression linéaire et d'une régression polynomiale de second degré (figure 22) basée sur la méthode des moindres carrés.

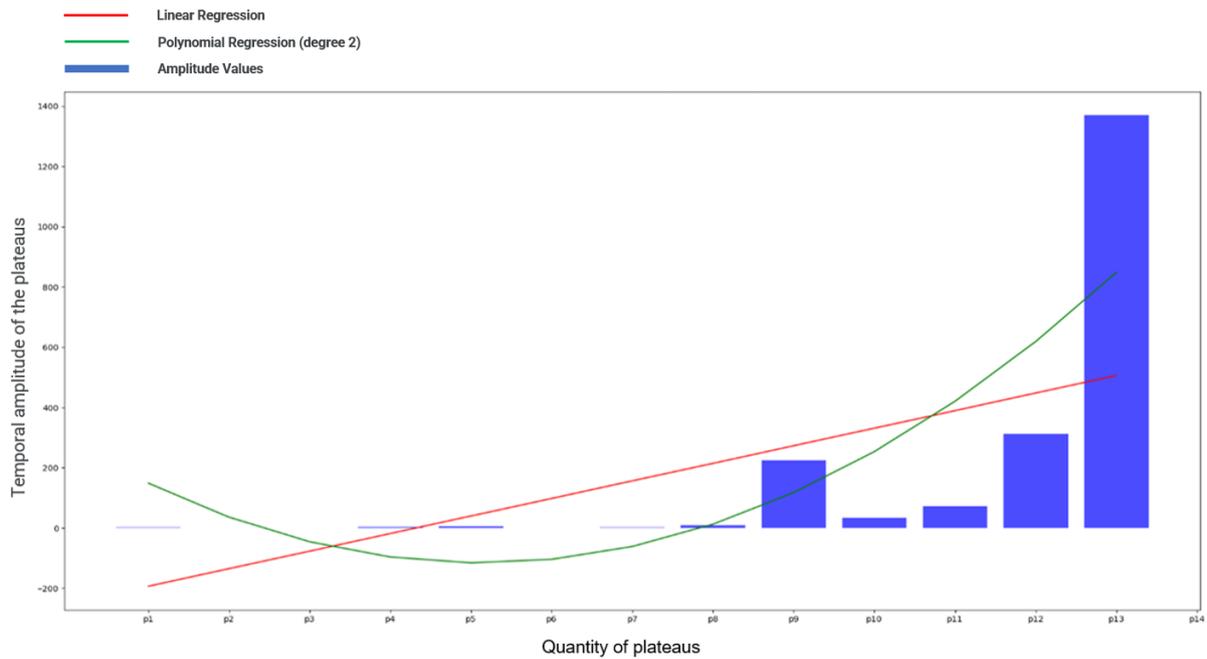


Figure 22 - Extraction des valeurs caractéristiques des cycles via PolyFit

```

LOAD DATASET_PATH INTO df
SELECT ['Timestamp', 'Pos. Ferme Last', 'Valve cycles user'] FROM df
GROUP BY "Valve cycles user" INTO df_groups

FOR cycle, group IN df_groups:
  INIT cycle_data = { 'Cycle': cycle, 'Nombre de paliers': 0 }
  palier_changes = (group['Pos. Ferme Last'] != group['Pos. Ferme
Last']).shift(1).cumsum()
  FOR palier, palier_indices IN group.groupby(palier_changes).groups:
    cycle_data["Longueur_palier_" + palier] = MAX(palier_indices) -
MIN(palier_indices)
    cycle_data['Nombre de paliers'] += 1
  APPEND cycle_data TO result_list
df = CREATE DataFrame FROM result_list

FOR cycle IN RANGE(MIN(df['Cycle']), MAX(df['Cycle'])):
  IF ALL p_columns EXIST IN df.columns:
    values = GET_VALUES(df, cycle, p_columns)
    indices = FIND_NON_NAN(values)
    IF LENGTH(indices) >= 2:
      linear_coef = POLYFIT(indices, values, DEGREE=1)
      poly_coef = POLYFIT(indices, values, DEGREE=2)
      APPEND cycle, linear_coef, poly_coef TO cycles_list

```

Figure 23 - Algorithme 2, pseudo-code utilisé pour extraire les valeurs caractéristiques des cycles

Ce système nous permet de récupérer 5 valeurs correspondant pour la régression linéaire, au coefficient de régression et à l'ordonnée à l'origine et pour la régression polynomiale, aux coefficients quadratiques et linéaires ainsi que la constante. Ces 5 valeurs permettent de caractériser et de révéler des valeurs différentielles plus ou moins marquées selon l'évolution du signal dans les différents profils étudiés. Elles sont obtenues à l'aide des formules (2) (3) (4) (5) :

Régression linéaire :

$$y = ax + b \tag{2}$$

Avec comme paramètres :

- y : Valeurs des amplitudes temporelles des paliers
- x : Indice du dernier palier détecté
- a : coefficient de régression.
- b : ordonnée à l'origine

Régression linéaire + Moindres carrés :

$$S(a, b) = \sum_{i=1}^n (y_i - (ax_i + b))^2 \tag{3}$$

Avec comme paramètres :

- $S(a, b)$: fonction de coût
- n : nombre d'observations
- x_i, y_i : valeurs des points de données (entrée/sortie)
- a, b : coefficients à optimiser pour minimiser l'erreur

Régression quadratique (polynomiale) :

$$y = ax^2 + bx + c \tag{4}$$

Avec comme paramètres :

- y : Valeurs des amplitudes temporelles des paliers
- x : Indice du dernier palier détecté
- a : Coefficient quadratique
- b : Coefficient linéaire
- c : Constante (ordonnée à l'origine)

Régression quadratique (polynomiale) + Moindres carrés :

$$S(a, b, c) = \sum_{i=1}^n (y_i - (ax_i^2 + bx + c))^2 \tag{5}$$

Avec comme paramètres :

- $S(a, b, c)$: fonction de coût

- n : nombre d'observations
- x_i, y_i : valeurs des points de données.
- a, b, c : coefficients du polynôme à ajuster.

Nous nous concentrons spécifiquement sur les valeurs polynomiales correspondant aux coefficients quadratiques car elles permettent de mettre en évidence la modulation des plateaux intermédiaires apparaissant dans les profils. Après extraction et normalisation, le signal post-traité suivant apparaît (figure 24). Nous avons à nouveau matérialisé sur la figure l'horizon de remplacement des membranes actuellement en vigueur chez notre partenaire industriel correspondant à la date de la maintenance préventive systématique. Contrairement à la figure 20, nous voyons un signal beaucoup moins stable, faisant apparaître des événements de dérives ponctuels qui correspondent à des plages de temps où des maintenances induisant des pertes de signaux ont été effectuées. Ces maintenances incluent des mises à jour de serveurs, remplacement d'équipements et démontage pour contrôle de membrane.

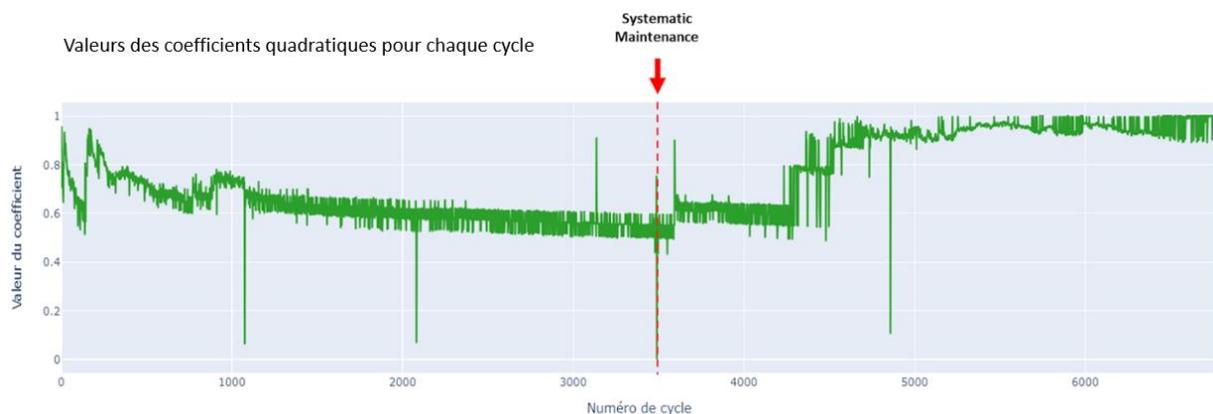


Figure 24 - Signal normalisé et post-traité représentant les coefficients quadratiques par cycles

3.4.2.7. Transformation et extraction des patterns de phase de vie

Pour approfondir notre investigation du signal brut, nous avons choisi d'appliquer une technique de transformation visant à vérifier la présence de motifs caractéristiques du cycle de vie au sein des données. Cette transformation était essentielle pour compléter notre étude, car elle nous a permis de mettre à jour autant de motifs explicites que possible qui pourraient refléter le comportement de la membrane au fil du temps. De nombreuses méthodes documentées dans la littérature scientifique se concentrent sur le retraitement des signaux pour mettre en évidence des séquences anormales, et nous avons sélectionné une approche cohérente avec ces pratiques établies. Ce faisant, nous avons cherché à détecter des anomalies ou des tendances potentielles qui pourraient ne pas être immédiatement observables dans le signal original. Cette étape garantit une analyse approfondie, offrant des perspectives plus étendues sur l'état de la membrane.

Thakur et al. (2023) proposent une classification des erreurs dans un signal électrique en utilisant PCA (Principal Component Analysis). Il s'agit d'une méthode linéaire qui transforme les variables originales en nouvelles variables orthogonales (composantes principales) qui maximisent la variance. Elle est notamment utilisée pour la compression de données et la réduction du bruit.

Une autre méthode, non linéaire cette fois, employant t-SNE, est décrite par Zhao & Zhu (2025) pour capturer les interdépendances au sein d'un signal EEG (Electroencephalography). Le principe de SNE (Distributed Stochastic Neighbor Embedding) consiste à réduire la

dimensionnalité des points de données du signal pour visualiser des données complexes en 2D ou 3D tout en préservant au maximum la structure locale des points.

Une troisième méthode de transformation fréquemment rencontrée dans la littérature, utilisant cette fois une méthode d'apprentissage profond, implique l'utilisation d'AutoEncoders. Gibb et al. (2024) appliquent cette technique pour étudier les variations acoustiques dans la biodiversité des environnements tempérés et tropicaux. L'objectif ici est d'amplifier les signaux périodiques et de réduire les biais. Cette méthode repose sur des réseaux neuronaux pour encoder les données dans un espace latent de dimension inférieure.

Pour notre cas d'application, et compte tenu de la nature des données, nous avons choisi d'appliquer une méthode t-SNE à l'ensemble des valeurs de statistiques descriptives extraites par l'Algorithme 1. Nous visons à vérifier la présence de motifs non visibles dans les signaux obtenus jusqu'à présent.

t-SNE est une méthode de réduction de dimensionnalité proposée par Maaten & Hinton (2008), utilisée pour visualiser des données multidimensionnelles en 2D ou 3D. Par conséquent, les axes d'un graphique t-SNE n'ont pas de signification physique ou mesurable. Les coordonnées résultantes ne correspondent pas à des unités spécifiques mais reflètent la structure locale des données (Wattenberg et al., 2016). t-SNE préserve les relations de voisinage, ce qui signifie que les points similaires dans l'espace original resteront proches dans l'espace réduit, bien que les distances entre les points n'aient pas de valeur absolue. Les groupements ou clusters observés indiquent des relations similaires, mais les distances exactes entre les points ne sont pas préservées. L'objectif principal de t-SNE est de révéler les structures des données, et non de maintenir des distances ou des échelles précises.

Cette méthode implique l'application d'une transformation non linéaire aux données d'entrée pour les projeter dans un espace de dimension inférieure. Un calcul de probabilité de similarité dans l'espace original est ensuite effectué sur les données d'entrée. Cela est suivi par la projection des valeurs dans un espace de basse dimension, la définition d'une distribution de similarité dans cet espace, et enfin la minimisation de la fonction de coût en utilisant la divergence de Kullback-Leibler via une descente de gradient. Appliquée sur des données se trouvant déjà un espace 2D, t-SNE permet de reprojeter ces données dans un nouvel espace 2D et minimiser la divergence entre les distributions de l'espace original et de l'espace cible, de sorte à révéler des clusters ou des motifs non visibles dans la première représentation 2D.

La méthode décrite ici, ainsi que les équations associées, est tirée de l'article de Maaten & Hinton (2008). Ci-dessous se trouvent les résultats des signaux décomposés par t-SNE, que nous considérons comme les plus pertinents, illustrés par la figure 25.

Cette figure regroupe quatre signaux distincts, étiquetés comme suit :

- a) Valeurs moyennes, représentant le comportement moyen des signaux par cycle ;
- b) Valeurs minimales, indiquant les limites inférieures de la variation du signal par cycle ;
- c) Écart-type, reflétant la variabilité des signaux au sein de chaque cycle ;
- d) Valeurs médianes, fournissant une mesure robuste de la tendance centrale par cycle.

Ces mesures statistiques offrent collectivement une vue complète des caractéristiques du signal à travers les cycles analysés.

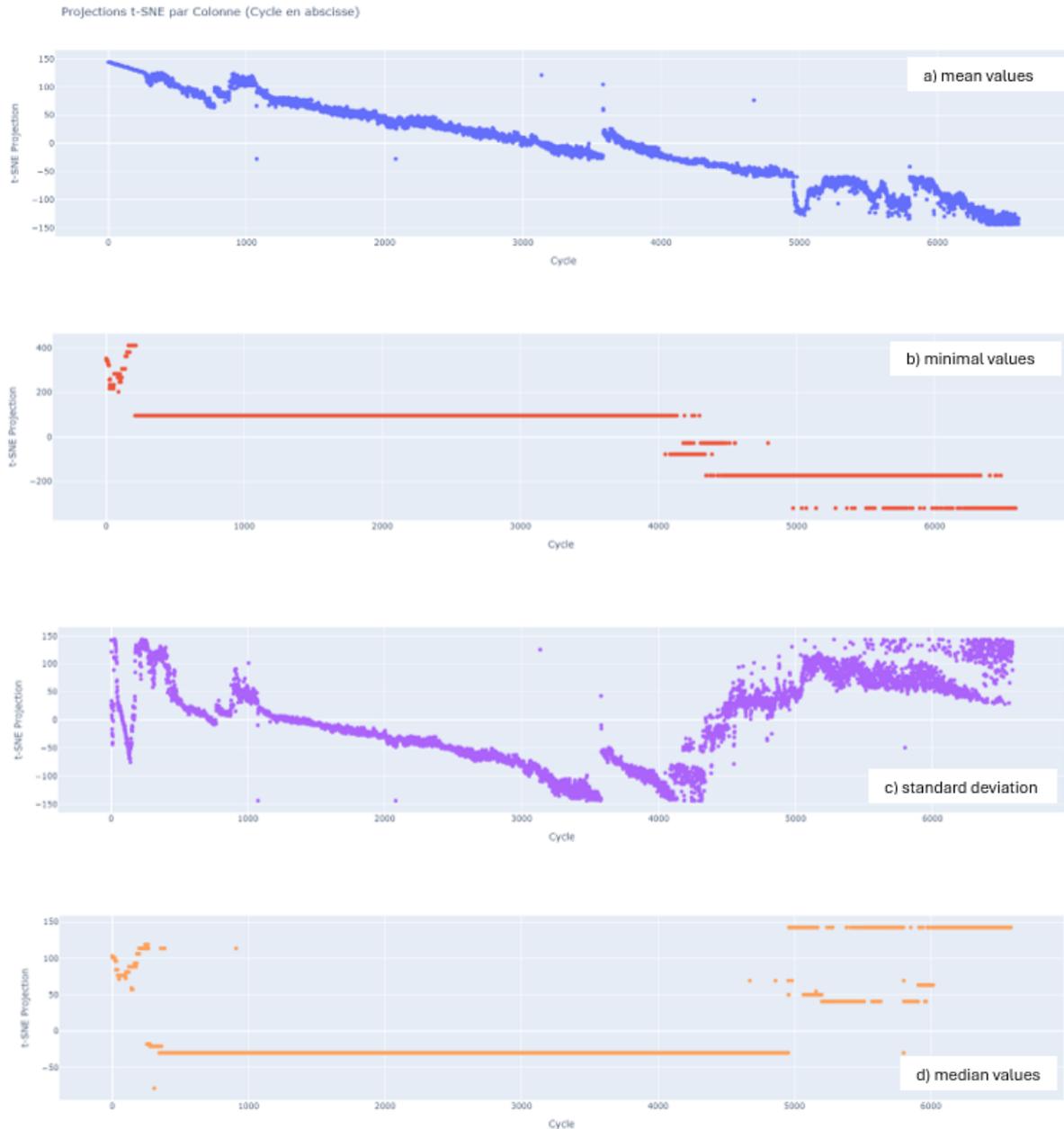


Figure 25 - Décomposition t-SNE des valeurs statistiques du signal brut par cycle

Les résultats de cette décomposition peuvent ensuite constituer une base de travail pour le traitement du signal destiné à détecter les différentes phases de la vie de l'équipement.

3.5. Analyse des résultats expérimentaux

L'analyse porte sur les données présentées dans les Figures 20 et 24. Nous avons choisi de nous concentrer sur ces résultats car :

- Dans le premier cas (Figure 20) : les valeurs sont directement issues du signal brut d'origine.
- Dans le second cas (Figure 24) : les valeurs résultent d'un post-traitement visant à mettre en évidence les modulations internes des plateaux pour chaque cycle.

La Figure 25 joue un rôle complémentaire en consolidant les premières observations issues des Figures 20 et 24. Elle offre une perspective plus large sur le comportement statistique du signal, perspective qui pourra être exploitée lors de futurs travaux.

Comme nous étudions un phénomène de dérive progressive, nous avons établi une comparaison entre un signal stable et un signal plus variable, bruité. Cela permet de vérifier si les deux signaux suivent des tendances cohérentes, en notant que le premier renvoie à la dérive du signal brut cycle par cycle, tandis que le second se concentre sur le nombre, la distribution et l'amplitude des plateaux composant le signal de chaque cycle.

Comme expliqué dans la section 3.4.2.4, nous organiserons l'analyse du signal autour des caractéristiques physiques liées à la dégradation du matériau, à savoir : le rodage, la maturité et la vieillesse.

3.5.1. Rodage

Le rodage de l'équipement commence dans notre cas au premier cycle et s'étend sur un intervalle allant du cycle 219 au cycle 277. Nous fondons cette affirmation sur les caractéristiques des deux signaux :

- Signal brut : Le signal brut reconstruit présente une chute continue du cycle 1 au cycle 219. La potentiométrie aura perdu 9 points par rapport au premier cycle avant de se stabiliser.
- Signal post-traité : Le signal post-traité présente une certaine amplitude en valeurs normalisées de 0,44 point, avant de se stabiliser à environ 0,71 point à partir du cycle 277.

3.5.2. Maturité

La maturité de l'équipement étudié se manifeste par un signal plus stable que celui de la phase de rodage, se poursuivant jusqu'à un intervalle allant du cycle 219 au cycle 4 268. La maturité est caractérisée par :

- Signal brut : Stabilité stricte du signal brut reconstruit du cycle 219 au cycle 4 212. La potentiométrie affichée dans cette plage est également de -9 points par rapport à la valeur du cycle 1.
- Signal post-traité : Courbes stables et légèrement descendantes dans le signal post-traité. Au cycle 4 268, la valeur du coefficient normalisé descend à environ 0,59, représentant une perte d'environ 0,12 depuis le début de la phase de maturité.

3.5.3. Vieillesse

La phase de vieillesse, au contraire, commence de manière relativement abrupte. À ce stade, elle s'exprime par une modulation différentielle entre les deux signaux :

- Signal brut : Au-delà de 4 219 cycles, elle apparaît comme une baisse continue du signal brut reconstruit jusqu'au cycle 4 743. À ce moment, 3 points supplémentaires auront été perdus, avec une potentiométrie se stabilisant à -12 points par rapport au cycle 1. Cette stabilité perdure jusqu'au cycle 5 103, après quoi la potentiométrie dérive lentement avec une tendance légèrement négative jusqu'à la fin de l'essai à 6 715 cycles.
- Signal post-traité : Au-delà de 4 268 cycles, le signal post-traité normalisé inverse sa courbe par rapport au signal brut. Jusqu'au cycle 5 139, le signal post-traité augmente,

passant d'environ 0,55 à environ 0,90. Au-delà, le signal se stabilise et amorce une dérive légèrement positive. Cette phase finale de dérive est également ponctuée d'événements aigus (pics à un seul cycle) amenant le signal normalisé à 1.

Dans l'ensemble, nous pouvons expliquer les écarts observés entre les signaux brut et post-traité par le fait que le premier ne représente que les positions minimales, tandis que le second implique une analyse plus fine du signal complet. Au-delà de cet écart, nous pouvons établir que les variations du signal post-traité sont corrélées à celles du signal brut.

3.5.4. Défaillance matérialisée

Comme mentionné précédemment dans la section « Mode de défaillance », les membranes de vanne sont sujettes à deux types de défaillance :

- Déformation progressive entraînant une rupture de l'étanchéité du système et la création de zones de rétention.
- Rupture de la couche de PTFE.

Dans notre étude de cas, une fuite interne du système est apparue au cycle 5 172. Celle-ci a été détectée par le booster et confirmée physiquement par un léger perlage dans le réservoir. Cela implique qu'avec une pression de 2 bars dans le système, la membrane s'était suffisamment déformée pour ne plus assurer l'étanchéité interne.

Comme la fuite ne mettait pas en danger le montage ou l'expérimentation elle-même, nous avons décidé de poursuivre l'essai afin de déterminer la dérive maximale du signal, laquelle s'est stabilisée à 1 sur une échelle normalisée. Cette dérive maximale a été observée dans le signal post-traité, maintenant une valeur de 1 du cycle 6 667 au cycle 6 715, marquant la fin de l'expérimentation.

Dans la figure 26, nous voyons cette fuite au sein d'une zone stable de la phase de vieillesse, juste avant une nouvelle phase de dérive à tendance négative.

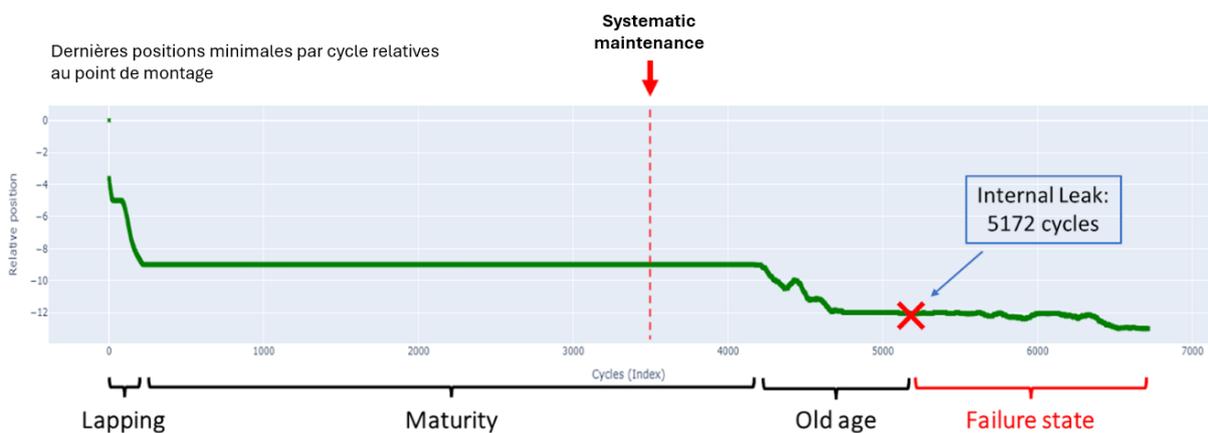


Figure 26 – Positions minimales du signal brut reconstruit annoté avec indication de la défaillance

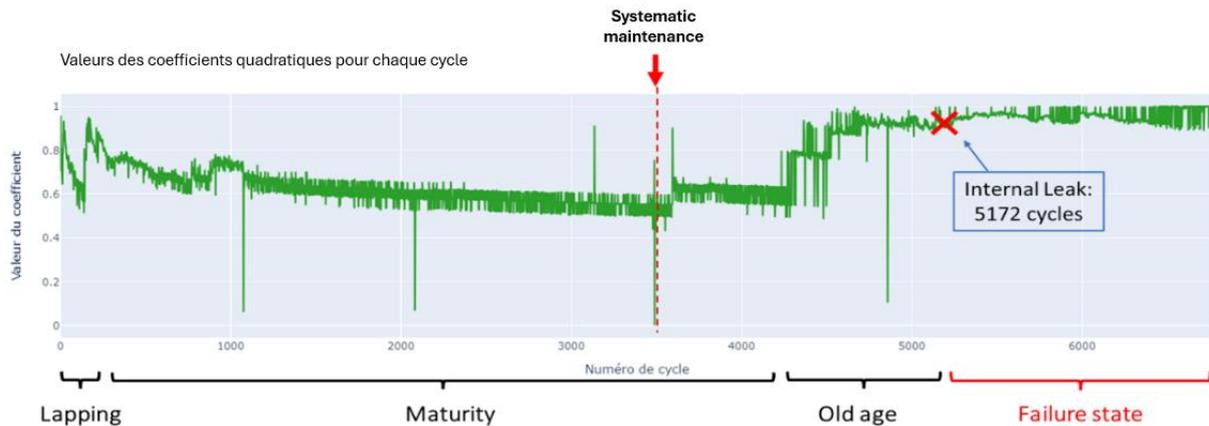


Figure 27 - Signal normalisé, post-processé et annoté avec indication de la défaillance

Comme montré dans la Figure 17, la fuite se situe cette fois dans une zone d'instabilité de la phase de vieillesse.

Afin d'établir un lien entre les dérives de signal observées et les caractéristiques physiques de la membrane, nous avons réalisé un test d'élasticité sur l'équipement. Ce test consiste en une comparaison entre l'élasticité de la membrane usagée et celle d'une membrane neuve.

3.5.5. Discussion sur la caractérisation physique de la dégradation de la membrane

Pour estimer cette dérive d'élasticité, une plateforme de test universitaire (figure 28) a été utilisée, permettant la compression des membranes et la mesure de la force (N) nécessaire pour atteindre une course donnée.

Chaque membrane a subi 200 cycles de compression. Les paramètres du test sont les suivants :

- 2 échantillons :
 - Une membrane neuve de référence
 - La membrane dégradée issue de notre étude, de la même référence
- Type de contrainte : Compression
- Déplacement : 2 mm
- Vitesse de compression : 50 mm/min

Les résultats sont présentés dans la figure 29. En bleu, la trace de la membrane neuve ; en orange, la trace de la membrane usée.



Figure 28 - Plateforme d'essai universaire



Figure 29 - Résultats des tests de compression comparatifs de la membrane usée versus une membrane neuve

Il est montré ici que la membrane neuve nécessite plus de force pour atteindre les déplacements cibles que la membrane usée. Les courbes de la membrane neuve (en bleu) présentent des valeurs de force plus élevées, exprimées en N, pour le même niveau de déplacement.

- Interprétation : la membrane neuve présente une rigidité importante, ce qui indique qu'elle résiste efficacement à la compression.

Les courbes de la membrane usée (en orange) montrent une force plus faible requise pour le même déplacement par rapport à la membrane neuve.

- Interprétation : la dégradation de la membrane a entraîné une perte de rigidité et une altération de ses propriétés mécaniques (fatigue du matériau, relâchement des composants internes). Une force requise plus faible suggère une détérioration de la membrane usée, pouvant accroître les risques de fuites ou de défaillances.

Dès le début du test, la membrane neuve nécessite une force plus importante (environ 6,47 N) pour être comprimée, alors que la membrane usée ne requiert qu'environ 0,81 N. Cela traduit une variation significative (moins de rigidité et plus d'élasticité) après usage.

Concernant la membrane neuve, le graphique indique que pour les passages successifs et dès le début de la compression, la force diminue légèrement mais reste dans une plage élevée (autour de 6,43 N). Cela suggère une certaine constance dans sa résistance à la compression. Pour la membrane usée, la force requise est faible et reste stable autour de 0,80 N, indiquant qu'elle a perdu une grande partie de sa capacité à résister à la compression. Cela implique que la membrane usée est significativement plus élastique que la membrane neuve.

Cette expérimentation met en évidence une évolution physique tangible des caractéristiques du matériau étudié. Toutefois, elle ne répond pas à certaines questions relatives aux sollicitations subies par l'équipement.

3.6. Conclusion du chapitre 3

Ce chapitre présente un cadre d'évaluation de l'état évolutif d'un équipement industriel, dans l'objectif de passer d'une maintenance systématique à une maintenance prévisionnelle. Nous avons identifié des données de fonctionnement pertinentes, en particulier des signaux indicateurs de l'état de santé de l'équipement, et avons tiré parti des technologies de l'Internet des Objets (IOT) pour capturer ces données en temps réel. Plusieurs étapes de post-traitement du signal ont ensuite été appliquées, incluant le filtrage et l'extraction de caractéristiques, afin de dégager des informations pertinentes sur l'état de l'équipement. Nous avons présenté l'application potentielle du cadre PHM (Prognostics and Health Management) à un composant d'usure couramment rencontré dans les processus de production biopharmaceutiques : la membrane de vanne pneumatique. Pour que ce cadre soit applicable, il nécessite au préalable la disponibilité de données représentatives pour l'équipement étudié, permettant ensuite l'établissement de seuils et des traitements de données plus avancés pour déterminer un horizon de maintenance prédictive.

Dans ce chapitre, nous proposons :

- Une méthode de récupération de données représentatives de l'état des membranes de vannes via des valeurs de position.
- Des critères de durée de vie de l'équipement, fondés sur une analyse des signaux extraits dans cette étude.
- Une comparaison des signaux extraits et des critères de durée de vie de l'équipement par rapport aux pratiques de maintenance systématique actuelles dans une industrie biopharmaceutique.

Cependant, bien que les données collectées aient permis de fournir des éléments intéressants, plusieurs limites demeurent, notamment en ce qui concerne la représentativité des tests réalisés. En effet, les conditions expérimentales et le type de vanne testé ne peuvent

être considérés comme pleinement représentatifs de toutes les configurations utilisées dans les environnements de production biopharmaceutique. La variabilité des conditions d'exploitation, comme la nature des fluides, les températures et pressions variables, ainsi que les spécificités propres à chaque installation de production, peuvent influencer le comportement des membranes de valves pneumatiques. Par conséquent, les résultats obtenus à partir de tests effectués sur une vanne spécifique ne peuvent pas être généralisés à toutes les valves présentes dans divers processus industriels, bien qu'ils puissent être applicables à des valves à commande pneumatique fonctionnant dans des conditions similaires. Ainsi, les travaux futurs pourront porter sur la proposition d'un cadre générique qui pourrait être appliqué à différents types d'équipements dans plusieurs domaines industriels.

En outre, l'étude a été fondée sur un seul essai. Il serait nécessaire de répliquer cette étude pour parvenir à une représentativité statistique et garantir que les signaux et les indicateurs de durée de vie des membranes soient reproductibles.

La suite de ces travaux va se porter sur une méthode permettant d'établir un pronostic d'occurrence de défaillance basé sur le modèle comportemental proposé dans cet article. Pour ce faire, nous exploiterons l'ensemble des profils présentés dans la section « Post-traitement du signal ».

Chapitre 4

4. De l'établissement du pronostic

Dans ce chapitre, nous allons appliquer aux signaux obtenus précédemment (chapitre 3) la trame de la 6ème brique du PHM, qui concerne l'établissement du pronostic. Pour rappel, le PHM fournit un cadre méthodologique sans détailler précisément la mise en œuvre de chaque étape. Ainsi, à partir des données traitées et des anomalies identifiées dans le chapitre 2, l'objectif de ce chapitre est de développer un modèle de pronostic permettant d'estimer le RUL des équipements.

Dans la section « Transformation et extraction des patterns de phase de vie » (chapitre 3), nous avons évoqué plusieurs méthodes permettant d'extraire des informations pertinentes afin de modéliser le comportement d'un équipement. Les données exposées dans les figures 20 et 24 nous ont permis de déduire un modèle de dégradation de la membrane, mettant en évidence les différentes phases de vie de l'équipement. Ces phases, établies en conformité avec le 3ème et 4èmes briques du PHM, ont constitué une base pour la mise en place d'un modèle de maintenance prévisionnelle intégrant des techniques de Machine Learning. De plus, la figure 25 illustre des données intéressantes pour construire un tel modèle, en représentant l'évolution des paramètres de l'équipement jusqu'à l'apparition d'une défaillance caractérisée physiquement par une fuite du système.

L'objectif de cette 6ème brique est spécifiquement de classifier les anomalies détectées et d'établir une estimation fiable du RUL. Pour ce faire, il est indispensable de recourir à des méthodes adaptées au pronostic. Plusieurs approches peuvent être employées, selon la nature des données disponibles et le modèle de dégradation sous-jacent et seront discutées plus bas.

Ce chapitre s'articule autour de plusieurs étapes. Nous débutons par une analyse des approches du pronostic en section 4.1, en explorant les méthodes data-driven, les techniques d'apprentissage supervisé et non-supervisé, ainsi que le Deep Learning, avant de conclure par un choix raisonné de la méthode retenue. Nous poursuivons sur l'apprentissage supervisé en section 4.2, en détaillant les algorithmes disponibles et leur fonctionnement. Nous présentons ensuite des méthodes de validation, en distinguant les approches de classification et de régression adaptées au pronostic. Par la suite, une section dédiée à l'optimisation examine les tests réalisés sur les algorithmes, incluant la préparation des données et l'analyse des résultats obtenus. La section 4.3 traite du test des algorithmes retenus tandis qu'en section 4.4, nous abordons ensuite la méthode d'entraînement de l'algorithme, en décrivant le fonctionnement des algorithmes RandomForestClassifier et XGBRegressor, ainsi que la création du modèle de dégradation en section 4.5. Enfin, une conclusion en section 4.6 synthétise les résultats obtenus dans ce chapitre.

4.1. Approches du pronostic

Le pronostic repose sur trois grandes catégories de modèles : les modèles physiques, les approches basées sur les données et les approches hybrides combinant les deux précédentes présentées succinctement dans le chapitre 1. Les modèles physiques s'appuient sur des équations mathématiques et des lois physiques décrivant la dégradation des composants, tandis que les approches data-driven exploitent des algorithmes d'apprentissage automatique pour établir des prévisions à partir des tendances observées dans les données historiques. Les approches hybrides visent à tirer parti des avantages de chacune de ces méthodes en combinant la robustesse des modèles physiques et la flexibilité des approches basées sur les

données. Ces différentes approches ont été détaillées dans le chapitre 1 consacré à l'état de l'art et ne seront donc pas redéveloppées ici.

Pour notre cas d'usage et au vu du type de modèle de dégradation dont nous disposons, nous avons choisi une approche data-driven. Ceci du fait que le modèle proposé en chapitre 2 est déjà basé sur des données pertinentes extraites du système étudié. Ce choix se justifie également par la capacité des méthodes d'apprentissage automatique à identifier des tendances complexes dans les signaux et à fournir des estimations précises du RUL, même dans des conditions de fonctionnement relativement variables. L'intégration de ces modèles et méthodes au sein du processus de maintenance prévisionnelle permettra ainsi d'optimiser les interventions et de minimiser les risques de défaillance non anticipée.

4.1.1. Approche data-driven

Les approches data-driven en pronostic reposent exclusivement sur l'exploitation des données expérimentales collectées pour estimer la durée de vie résiduelle d'un système. Contrairement aux modèles physiques qui nécessitent une compréhension fine des mécanismes de dégradation, ces méthodes s'appuient sur des techniques statistiques et d'apprentissage automatique pour modéliser les tendances de défaillance.

4.1.1.1. Apprentissage supervisé

L'apprentissage supervisé est l'une des méthodes employables pour estimer la durée de vie résiduelle (RUL), particulièrement adaptée lorsque l'on dispose de données historiques étiquetées. Par données étiquetées, nous entendons des ensembles de données pour lesquels chaque observation est associée à une valeur cible connue, comme le temps restant avant une défaillance dans le cas du RUL, permettant ainsi d'entraîner un modèle à prévoir cette valeur. À l'inverse, les données non étiquetées ne comportent pas de telles annotations, rendant leur utilisation plus adaptée à des approches non supervisées visant à identifier des motifs ou des regroupements sans cible prédéfinie (Zhou, 2018).

Une étude de Losi et al. (2021) a démontré l'efficacité des forêts aléatoires (Random Forests) pour prédire le RUL de turbines à gaz à partir de données simulées, soulignant leur capacité à gérer des relations non linéaires complexes entre les signaux et les défaillances recensées. De même, les machines à vecteurs de support (SVM) ont été utilisées par Yan et al. (2020) pour le pronostic de dégradation des roulements industriels, montrant une robustesse notable face aux données bruitées, un cas fréquent dans les systèmes industriels. Ces méthodes nécessitent cependant un volume important de données étiquetées, comme le notent Li et al. (2018), qui ont évalué des algorithmes de régression supervisée pour des systèmes aéronautiques, montrant que la précision dépend fortement de la qualité et de la représentativité des données d'entraînement. Dans notre cas, les données extraites du système étudié (chapitre 2) rendent cette approche adaptée pour établir des prévisions fiables du RUL.

4.1.1.2. Apprentissage non-supervisé

L'apprentissage non-supervisé est applicable lorsque les données ne sont pas étiquetées ou lorsque les mécanismes de dégradation sont peu connus. Une étude de Lei et al. (2016) concerne une analyse en composantes principales (PCA) pour réduire la dimensionnalité des signaux vibratoires dans le diagnostic de machines, permettant d'identifier des tendances de dégradation sans étiquettes préalables. Une autre méthode généralement employée, le clustering, a été appliquée par Choi & Kim (2024) pour détecter des anomalies dans des systèmes industriels, offrant une segmentation initiale des données qui peut précéder une analyse supervisée. Cependant, comme le soulignent Lei et al. (2020) dans leur revue sur les

méthodes non-supervisées en PHM, ces techniques sont moins précises pour estimer directement le RUL, mais elles restent intéressantes pour explorer des données complexes ou évolutives. Pour notre cas d'usage et au vu de la particularité du signal que nous traitons, nous avons employé une méthode non-supervisée discutée au chapitre 2 nommée t-SNE pour extraire des patterns de dégradation dans nos données.

4.1.1.3. Deep Learning

Enfin, le Deep Learning est particulièrement applicable dans le traitement de données volumineuses et complexes. Par exemple, Liu et al. (2022) ont proposé un modèle combinant des réseaux convolutifs (CNN) et des Long-Short Time Memory (LSTM) pour prédire le RUL de systèmes mécaniques, capturant efficacement les dépendances temporelles dans des conditions variables. De même, Li et al. (2018) ont utilisé des réseaux de neurones récurrents (RNN) pour estimer le RUL de turbines, démontrant leur capacité à modéliser des séquences temporelles sans nécessiter un prétraitement extensif des caractéristiques. Toutefois, comme le notent Fink et al. (2020) dans leur analyse des approches de Deep Learning pour le PHM, ces méthodes nécessitent une puissance de calcul élevée et un volume de données conséquent, ce qui peut poser des défis dans des environnements industriels aux ressources limitées.

4.2. Conclusion et choix

Le choix d'une approche data-driven dans notre étude s'appuie sur une combinaison pragmatique de ces techniques, étayée par la littérature scientifique. L'apprentissage supervisé est privilégié pour sa précision dans l'estimation du RUL à partir des données historiques disponibles (Sharma & al., 2020). L'apprentissage non-supervisé peut enrichir l'analyse en prétraitant les données ou en détectant des anomalies, tandis que le Deep Learning offre une perspective d'amélioration future, notamment si les données augmentent en volume ou en complexité (Shinde et Shah, 2018). Une stratégie hybride, combinant les avantages de chaque méthode utilisée, s'aligne avec les recommandations de Drissi et al. (2022), qui insistent sur l'intégration de multiples approches pour optimiser la maintenance prévisionnelle.

Ainsi, au vu de la nature et de la disponibilité des données employées pour créer le modèle de dégradation discuté au chapitre 2, notre choix se dirige sur une approche hybride employant d'une part l'apprentissage supervisé, complété par l'outil non-supervisé t-SNE décrit au chapitre 2. Cette dernière méthode a notamment été employée pour les signaux reconstruits extraits des statistiques descriptives du signal brut, facilitant la visualisation des états de dégradation.

4.2. Apprentissage supervisé

Dans cet item, nous allons passer en revue les algorithmes supervisés disponibles, en examinant leurs avantages et inconvénients pour mieux comprendre leurs spécificités.

4.2.1. Algorithmes disponibles et fonctionnement

L'apprentissage supervisé s'appuie sur un certain nombre d'algorithmes capables de modéliser la relation entre les données d'entrée et des objectifs tels que la classification d'états ou l'estimation de valeurs continues comme la durée de vie résiduelle. Ces méthodes analysées à la suite, peuvent être appliquées à des tâches combinées, comme classer les signaux en catégories (par exemple, états de dégradation) puis établir un RUL, avec des forces et faiblesses intrinsèques découlant de leur conception mathématique, leur complexité et leurs exigences en termes de données.

D'un côté, les arbres de décision (Decision Trees) segmentent les données via des seuils sur les variables d'entrée pour prédire une classe ou une valeur numérique. Kamiński et al., (2018) ont exploré leur utilisation dans des contextes de prévision, soulignant leur simplicité d'implémentation et leur faible coût computationnel, permettant un traitement rapide même sur des systèmes limités. Leur interprétabilité est un atout majeur, chaque branche offrant une règle claire. Cependant, ils sont sujets au surapprentissage sans régulation et leur sensibilité aux variations locales peut réduire leur stabilité (Amro et al., 2021).

Les forêts aléatoires (Random Forests) combinent de multiples arbres entraînés sur des sous-ensembles aléatoires, comme étudié par Biau et Scornet (2016). Elles offrent précision et robustesse grâce à l'agrégation, réduisant la variance tout en maintenant un biais faible, ce qui améliore la généralisation sur des données bruitées et non linéaires. Cependant, leur complexité computationnelle est élevée, nécessitant des ressources importantes, ce qui peut poser un problème sur des systèmes contraints. De plus, leur interprétabilité est limitée : malgré des mesures comme l'importance des variables, la prédiction finale, issue d'un vote ou d'une moyenne, perd la transparence d'un arbre unique, compliquant l'analyse des décisions dans des contextes exigeant une explicabilité claire.

Les machines à vecteurs de support (SVM) identifient une frontière optimale pour la classification ou la régression, comme discuté par Cervantes et al. (2020). Elles projettent les données dans un espace supérieur via des noyaux comme le RBF, capturant des relations non linéaires complexes. Robustes au bruit grâce aux variables d'écart et au paramètre C, elles tolèrent les outliers et lissent les perturbations. Efficaces sur de petits datasets en se focalisant sur les vecteurs de support, elles nécessitent cependant un réglage fin des hyperparamètres (C, gamma), coûteux en temps via validation croisée. De plus, leur coût computationnel croît significativement avec le volume des données.

Les réseaux de neurones artificiels (ANN) modélisent des relations complexes via des couches interconnectées, comme analysé par Abiodun et al. (2018). Leur flexibilité découle de leur architecture adaptable, capable d'extraire automatiquement des caractéristiques pertinentes sans ingénierie manuelle, ce qui leur permet d'atteindre des performances élevées sur des datasets volumineux et non structurés. Cependant, les auteurs notent qu'ils nécessitent de grandes quantités de données pour éviter le surapprentissage, ainsi qu'une puissance de calcul importante pour l'entraînement, en raison de la complexité des calculs matriciels. De plus, leur opacité, due à la nature non linéaire des transformations dans les couches cachées, limite leur interprétabilité, rendant difficile l'explication des prédictions.

Une autre étude traitant cette fois-ci de la régression logistique, décrite par Hosmer et al. (2020), estime des probabilités pour la classification en utilisant une fonction logistique qui modélise la relation entre les variables explicatives et la probabilité d'appartenance à une classe. Les auteurs soulignent sa simplicité d'implémentation, nécessitant peu de ressources, et sa rapidité pour des tâches catégoriques, comme la distinction entre classes binaires ou multiclassées. De plus, ses coefficients explicites permettent une interprétabilité claire, chaque coefficient indiquant l'impact d'une variable sur la probabilité prédite. Cependant, les auteurs notent également qu'elle est limitée à la classification et ne peut pas estimer directement des valeurs continues, comme une durée ou une métrique numérique, ce qui restreint son usage pour des tâches de régression.

D'autre part, les modèles de régression bayésienne, étudiés par Van de Schoot et al. (2021), génèrent des prédictions probabilistes en intégrant des distributions a priori sur les paramètres, offrant une estimation complète des incertitudes. Les auteurs mettent en avant leur capacité à quantifier le risque via des intervalles de crédibilité, ce qui est important pour évaluer la fiabilité des prédictions dans des contextes critiques. Ils reposent sur des méthodes comme l'échantillonnage MCMC (Monte Carlo Markov Chain) pour inférer les distributions a posteriori, améliorant la robustesse face à l'incertitude des données. Cependant, l'étude

souligne que leur dépendance à des hypothèses sur les distributions a priori peut introduire des biais si ces hypothèses sont mal spécifiées. De plus, leur complexité, due aux calculs itératifs de l'échantillonnage, accroît significativement le temps de calcul. Pour ce qui est des techniques de clustering, une étude de Zhang et al. (2022) discute des k-nearest neighbors (k-NN) dont la fonction est de prédire en fonction des points les plus similaires. Les auteurs mettent en avant leur simplicité, car ils reposent sur une distance (ex. euclidienne) pour classer ou régresser sans nécessiter d'entraînement explicite, ce qui accélère leur déploiement. L'étude met en avant que cette absence de phase d'entraînement permette une mise en œuvre rapide, même sur des systèmes légers. Cependant, leur recherche exhaustive, nécessitant de calculer les distances à tous les points, augmente les coûts en mémoire et en calcul, particulièrement pour de grands datasets. De plus, leur sensibilité aux outliers, qui peuvent fausser les distances et biaiser les prédictions, dégrade les performances dans des scénarios bruités.

Nous nous intéressons également aux modèles de boosting (Gradient Boosting, XGBoost), sont décrits par Chen et Guestrin (2016) et Bentéjac et al. (2020). Chen et Guestrin expliquent que XGBoost optimise une fonction de perte en ajoutant des arbres qui corrigent les erreurs des précédents, atteignant une précision élevée sur des données hétérogènes grâce à cette correction adaptative. Bentéjac et al. ajoutent que cette approche excelle dans la gestion de données complexes via des techniques comme la régularisation L1/L2. Cependant, les auteurs notent qu'un réglage minutieux des hyperparamètres (ex. taux d'apprentissage, profondeur des arbres) est requis, un processus coûteux en temps. Leur coût computationnel, dû à l'entraînement séquentiel, et leur faible interprétabilité, liée à la complexité ensembliste, limitent leur usage dans des contextes nécessitant explicabilité.

Enfin, les processus gaussiens (GPR) sont étudiés par Schulz et al. (2018). Ils modélisent les sorties avec des intervalles de confiance, les auteurs soulignant leur précision sur de petits datasets. Ils s'appuient sur une covariance définie par un noyau (ex. Radiale Base Function) pour capturer les relations entre points, offrant des prédictions probabilistes robustes. Schulz et al. valorisent leur approche probabiliste, qui quantifie l'incertitude via des distributions prédictives, utile pour des analyses fiables. Cependant, leur coût computationnel est élevé, ce qui les rend peu pratiques pour de grands volumes de données. Cette limitation restreint leur scalabilité malgré leur efficacité sur des ensembles restreints.

Dans notre étude, les données exposées au chapitre 2, issues du suivi de l'état de la membrane jusqu'à une défaillance, seront utilisées pour une double tâche : classer les signaux en état de vieillesse (basés sur le modèle de dégradation) et estimer un RUL mis à jour cycle après cycle.

Compte tenu de la diversité des algorithmes disponibles, une approche expérimentale sera adoptée pour sélectionner une méthode : tous ces algorithmes sont testés pour évaluer leurs performances en classification (discrimination des états) et en régression (précision du RUL).

Les métriques obtenues, telles que l'exactitude pour la classification et l'erreur quadratique moyenne pour la régression, permettront une comparaison objective afin de déterminer l'algorithme le plus adapté à notre cas d'usage, en tenant compte des contraintes industrielles et des objectifs de maintenance prévisionnelle.

4.2.2. Métriques de validation

L'évaluation des modèles supervisés repose sur des métriques adaptées aux tâches de classification et de régression.

4.2.2.1. Classification

Pour la classification, l'exactitude (6) mesure la proportion de prédictions correctes, métrique privilégiée par Kocher & Kumar (2021) pour sa simplicité dans l'évaluation des états de santé des systèmes. L'aire sous la courbe ROC (Receiver Operating Characteristic) (7) (8), utilisée par Bammou et al. (2024) dans un contexte similaire, quantifie la capacité discriminante des modèles, essentielle pour distinguer les classes dans des diagnostics industriels, un AUC-ROC proche de 1 indiquant une séparation parfaite. La précision moyenne en validation (9) croisée, mise en avant par Nti et al. (2021), évalue la robustesse des modèles face à la variabilité des données, importante pour les applications critiques comme les turbines. Le détail des métriques est donné dans les items suivants.

$$Exactitude = \frac{VP + VN}{VP + VN + FP + FN} \quad (6)$$

Où :

- *VP* (vrais positifs) : cas positifs correctement prédits.
- *VN* (vrais négatifs) : cas négatifs correctement prédits.
- *FP* (faux positifs) : cas négatifs prédits comme positifs.
- *FN* (faux négatifs) : cas positifs prédits comme négatifs.

L'exactitude indique ici la performance de la classification avec des valeurs allant de 0 à 1, où 1 indiquant une classification parfaite.

Pour ce qui est de l'AUC-ROC, il s'obtient par le calcul de l'intégrale du taux de vrais positifs (TVP) avec la formule suivante :

$$TVP = \frac{VP}{VP + FN} \quad (7)$$

En fonction du taux de faux positifs (TFP) :

$$TFP = \frac{FP}{FP + VN} \quad (8)$$

Les retours varient de 0 à 1. L'AUC est l'intégrale de la courbe ROC, qui trace TVP en fonction de TFP pour différents seuils. Une AUC = 1 indique une séparation parfaite ; AUC = 0,5 indique une classification aléatoire ; AUC < 0,5 indique que les résultats de classification sont moins performants encore que l'aléatoire. Cette valeur mesure la capacité d'un modèle à discriminer entre les classes positives et négatives.

Pour ce qui est de la précision moyenne, elle s'obtient grâce à la formule suivante :

$$Précision\ moyenne = \frac{1}{k} \sum_{i=1}^k Exactitude_i \quad (9)$$

où $Exactitude_i$ est l'exactitude calculée sur le i -ème pli test. Le jeu de données est divisé en k plis, le modèle est entraîné sur $k - 1$ plis et testé sur le pli restant, répété k fois. Fournit une estimation des performances sur des données non vues, appréciant par conséquent le risque de surapprentissage.

4.2.2.2. Régression

Pour la régression, l'erreur quadratique moyenne (MSE) (10) et sa racine (RMSE) (11) mesurent les écarts entre prédictions et valeurs réelles. Dans leur étude, Kocher & Kumar (2021) favorisent leur sensibilité aux grandes erreurs pour éviter les sous-estimations du RUL dans les systèmes industriels. L'erreur absolue moyenne (MAE) (12), utilisée par Bammou et al. (2024), offre une interprétation intuitive en unités de temps dans un contexte de maintenance prévisionnelle pour des machines rotatives. Le coefficient de détermination (R^2) (13) évalue la variance expliquée. Patra et al. (2024) l'emploient notamment pour valider la précision des prédictions, un R^2 proche de 1 signifiant une forte adéquation.

L'erreur quadratique moyenne (MSE) s'obtient avec la formule suivante :

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (10)$$

où y_i est la valeur réelle, \hat{y}_i est la valeur prédite et n est le nombre d'observations. Une MSE faible indique des prédictions proches des valeurs réelles.

La racine de l'erreur quadratique moyenne quant à elle s'établit avec la formule suivante :

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (11)$$

Une racine carrée de la MSE ou une RMSE de 0 signifie une prédiction parfaite.

Pour ce qui est de la MAE, la formule est la suivante :

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (12)$$

La MAE est la moyenne des valeurs absolues des écarts entre prédictions et valeurs réelles, offrant une mesure linéaire de l'erreur.

Enfin, pour ce qui est du coefficient de détermination, la formule applicable est la suivante :

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (13)$$

où \bar{y}_i est la moyenne des valeurs réelles, $\sum_{i=1}^n (y_i - \hat{y}_i)^2$ est la somme des carrés des résidus et $\sum_{i=1}^n (y_i - \bar{y}_i)^2$ est la variance totale. Le R^2 représente la proportion de la variance totale des

données expliquée par le modèle, mesurant son adéquation aux observations. Il prend des valeurs comprises entre $-\infty$ et 1. Pour un $R^2 = 1$, le modèle explique toute la variance ; Pour un $R^2 = 0$, le modèle n'explique pas la variance mieux qu'une moyenne ; Pour un $R^2 < 0$, le modèle est moins performant que ce que pourrait faire une prévision constante.

Pour notre cas d'usage :

- Pour ce qui est de la classification, l'exactitude et l'AUC-ROC seront utilisés pour évaluer la précision et la discrimination sur les données, tandis que la précision moyenne en validation croisée évaluera la généralisation.
- Pour ce qui est de la régression, la MSE, la MAE, et le R^2 quantifieront la précision globale du modèle.

Une phase d'optimisation peut être nécessaire afin d'affiner les performances du modèle de pronostic, pour identifier les paramètres optimaux des algorithmes dans notre cas d'usage. En effet, les résultats initiaux des métriques telles que le R^2 , la MSE ou la MAE peuvent révéler des écarts entre les prédictions et les valeurs réelles, notamment en raison de la variabilité des données ou des hyperparamètres sous-optimaux.

4.2.3. Optimisation

Une optimisation systématique est importante pour adapter les modèles aux exigences particulières de chaque problème. En fait, les hyperparamètres génériques des algorithmes d'apprentissage automatique constituent des leviers pour configurer leur comportement avant l'entraînement, par opposition aux paramètres appris directement à partir des données, tels que les poids d'un modèle. Ces hyperparamètres, définis a priori par l'utilisateur, influencent la capacité, la complexité et la généralisation des modèles supervisés (Weerts et al., 2020).

Par exemple, dans les arbres de décision et leurs dérivés comme les forêts aléatoires ou le gradient boosting, des hyperparamètres tels que la profondeur maximale (`max_depth`) contrôle la complexité pour prévenir le surajustement en limitant la profondeur des arbres, tandis que le nombre d'estimateurs (`n_estimators`) renforce la robustesse grâce à l'agrégation de multiples arbres, réduisant ainsi la variance des prédictions. Pour les machines à vecteurs de support (SVM), le paramètre de régularisation (`C`) ajuste l'équilibre entre une marge de séparation large et la tolérance aux erreurs, influençant la capacité de généralisation, et le paramètre `gamma` détermine l'étendue de l'influence des points d'entraînement, modifiant la flexibilité des frontières de décision selon le noyau utilisé (par exemple, RBF, linéaire ou polynomial). Dans les réseaux neuronaux, comme les perceptrons multicouches (MLP), la taille des couches cachées (`hidden_layer_sizes`) régit la capacité du modèle à capturer des relations complexes en augmentant le nombre de neurones, tandis que le taux d'apprentissage (`learning_rate`) pilote la vitesse et la stabilité de la convergence lors de l'optimisation par descente de gradient, variant selon les stratégies d'ajustement. Selon Hastie et al. (2009), ces hyperparamètres, lorsqu'ils sont laissés à leurs valeurs par défaut, offrent souvent une performance raisonnable mais rarement optimale, car ils ne tiennent pas pleinement compte de la diversité des jeux de données ou des spécificités des tâches, qu'il s'agisse de classification ou de régression. Ainsi, une optimisation systématique devient importante pour personnaliser ces modèles en fonction des exigences particulières de chaque problème.

Au test d'un algorithme sur un dataset particulier comme celui de notre cas d'usage, la question se pose de l'emploi du système avec ses fonctions et paramètres fournis par défaut VS des fonctions et paramètres personnalisés. Ceci pouvant alourdir le système, le rendant plus complexe, nécessitant davantage de puissance de calcul, etc ... Pour l'étude à suivre, nous souhaitons effectuer un complément d'étude sur chacun des algorithmes discutés de sorte à pouvoir conclure sur le fait que les systèmes testés doivent être optimisés ou non et

dans quelle mesure et avec quelle fonction, dans l'optique de pouvoir offrir les meilleures prévisions pour notre cas d'usage.

Dans cette optique, nous avons testé exhaustivement toutes les variantes accessibles de ces modèles : arbres de décision avec différents critères de perte (par exemple, erreur quadratique ou absolue), forêts aléatoires avec divers critères d'évaluation, gradient boosting avec plusieurs fonctions de perte (comme l'erreur quadratique, absolue ou Huber), SVM avec différents noyaux (RBF, linéaire, polynomial), perceptrons multicouches avec diverses fonctions d'activation (ReLU, tanh), ainsi que des approches bayésiennes et des k-plus proches voisins avec leurs propres configurations. Cette démarche permet de comparer les performances des configurations par défaut à celles optimisées pour chaque variante, en évaluant des métriques telles que le RMSE, la précision temporelle ($\pm 10\%$) et le R^2 , afin d'identifier les combinaisons les plus performantes pour prédire les cycles restants et de détecter les cas où l'optimisation pourrait ne pas surpasser les réglages initiaux.

Pour réaliser cette optimisation de manière systématique, nous avons utilisé Python, qui met à disposition un large ensemble d'outils et de bibliothèques, tels que Scikit-Learn, permettant la mise en œuvre de méthodes avancées pour l'ajustement des hyperparamètres et l'évaluation des modèles (Bisong, 2019).

La bibliothèque Python de Scikit-Learn fournit notamment à l'utilisateur un régleur automatique d'hyperparamètres, appelable en tant que classe dans un code. Son utilisation dans les modèles supervisés repose sur son approche systématique pour identifier les paramètres optimaux, pouvant surpasser les paramètres par défaut souvent sous-optimaux des algorithmes d'apprentissage automatique. Contrairement à une recherche manuelle subjective, chronophage et potentiellement inefficace ou aux valeurs par défaut qui ne garantissent pas des performances maximales, GridSearchCV explore exhaustivement un espace de paramètres prédéfini, couplé à une validation croisée, pour ajuster les hyperparamètres dans le but d'optimiser une métrique cible spécifiée.

Son objectif est de jouer sur ces paramètres pour minimiser les erreurs de prédiction, comme l'erreur quadratique moyenne (MSE) ou l'erreur absolue moyenne (MAE) en régression, afin de réduire les écarts entre les valeurs prédites et réelles, ou encore pour maximiser des métriques comme l'exactitude et l'aire sous la courbe ROC (AUC-ROC) en classification, de sorte à augmenter la proportion de prédictions justes et la capacité discriminante. Bergstra et Bengio (2012) soulignent que, dans des espaces de paramètres restreints où chaque valeur a un impact significatif, une recherche en grille est efficace pour cibler ces optima, par rapport à des alternatives aléatoires comme RandomizedSearchCV qui risquent de manquer des combinaisons clés. Pedregosa et al. (2011) mettent en avant dans la documentation de scikit-learn que GridSearchCV permet de régler précisément les configurations pour atteindre ces objectifs, un avantage pour des modèles sensibles aux hyperparamètres tels que les machines à vecteurs de support ou les réseaux neuronaux, où les paramètres par défaut peuvent entraîner des prédictions imprécises ou une faible généralisation. Comparée à des méthodes bayésiennes, qui nécessitent des hypothèses probabilistes complexes, GridSearchCV offre une simplicité et une transparence prisées en contexte scientifique, notamment pour des études visant à comparer rigoureusement les performances par défaut et optimisées. Ainsi, cet optimiseur est privilégié dans notre étude pour sa capacité à identifier de manière fiable et systématique les paramètres optimaux, ajustant les modèles afin d'abaisser les erreurs et d'améliorer les métriques de performance clés selon les exigences de la tâche.

Tenant compte de cette approche méthodologique, nous proposons à la section suivante une évaluation concrète des algorithmes supervisés dans un cadre expérimental, afin de valider leur efficacité et de guider le choix des modèles les plus adaptés à notre problématique.

4.3. Test des algorithmes

Pour mettre en œuvre cette évaluation, nous avons conçu une série d'expériences visant à comparer les performances des algorithmes avec et sans optimisation des hyperparamètres. Avant d'exposer les résultats, il est essentiel de détailler la préparation des données, qui constitue une étape préalable critique pour garantir la fiabilité et la comparabilité des performances observées.

4.3.1. Préparation des données

La préparation des données est une étape importante pour structurer et enrichir le jeu de données initial afin d'orienter les modèles supervisés dans les tâches de classification et de régression.

Nos données issues d'une analyse t-SNE et représentant des signaux temporels (Pos. Ferme Last_mean, _min, _std, _median), ont été chargées dans un DataFrame pandas, avec un index temporel cycle inséré comme première colonne pour refléter la progression des observations. Il est important de noter que cet index pseudo-temporel (car basé sur des cycles successifs et non des timestamps) ne fait pas partie des features retenues pour entraîner les modèles.

Une annotation manuelle, conforme aux méthodologies de labelling basées sur les phases de vie des systèmes industriels décrites par Lei et al. (2016), a été appliquée pour catégoriser les cycles en quatre étapes :

- Rodage (cycles 0-248, étiquette 1)
- Maturité (249-4240, étiquette 2)
- Vieillesse (4241-5121, étiquette 3)
- Défaillance (5122-6737, étiquette 4)

Les ranges d'annotation ont été déterminés sur la base des critères proposés dans le chapitre 2, item "Analyse des résultats expérimentaux". Les horizons ont été déterminés en prenant pour chaque phase de vie proposée, une valeur à mi-chemin entre l'horizon fixé sur la base du signal brut, et l'horizon fixé sur la base du signal post-traité.

L'annotation permet de créer une variable cible Y pour la classification, mappée en valeurs numériques (1 à 4). Cette approche d'annotation, validée pour le pronostic des défaillances dans des systèmes mécaniques complexes et pour l'étiquetage des données de dégradation dans la maintenance prévisionnelle (Lei et al., 2016), reflète les transitions typiques de dégradation.

Pour pallier la limitation potentielle du volume de données et accroître la robustesse des modèles, une augmentation méthodique a été réalisée sur les signaux, s'inspirant des techniques d'enrichissement de données. Dans notre cas, un bruit aléatoire gaussien (moyenne 0, écart-type 10) a été ajouté aux signaux moyens, minimaux, standards et médians pour simuler des variations "naturelles" (Jittered), stratégie décrite par Wen et al. (2021) pour améliorer la généralisation des modèles supervisés face aux perturbations dans les séries temporelles industrielles. Par ailleurs, un décalage temporel de 20 cycles, suivi d'un remplissage rétroactif a été appliqué pour introduire des dynamiques temporelles (Shifted), une méthode décrite par les auteurs précédents, pour capturer les dépendances temporelles dans des données à destination d'une maintenance prévisionnelle. Ces transformations ont généré huit nouvelles colonnes (ex. : Pos. Ferme Last_mean Jittered, Pos. Ferme Last_mean Shifted), enrichissant le jeu de données initial.

Enfin, une visualisation interactive (figure 30) a permis de valider l'augmentation en projetant les signaux transformés par rapport aux cycles, confirmant la diversité introduite.



Figure 30 - Signaux augmentés et bruités

Ces étapes, réalisées avec numpy et pandas (Sapre & Vartak, 2020), visent à fournir un jeu de données annoté et diversifié, essentiel pour entraîner et tester efficacement les algorithmes supervisés dans notre contexte de pronostic industriel. Pour plus de clarté, les valeurs sont fusionnées avec une moyenne des valeurs des colonnes par cycle, comme montré sur la Figure 31.

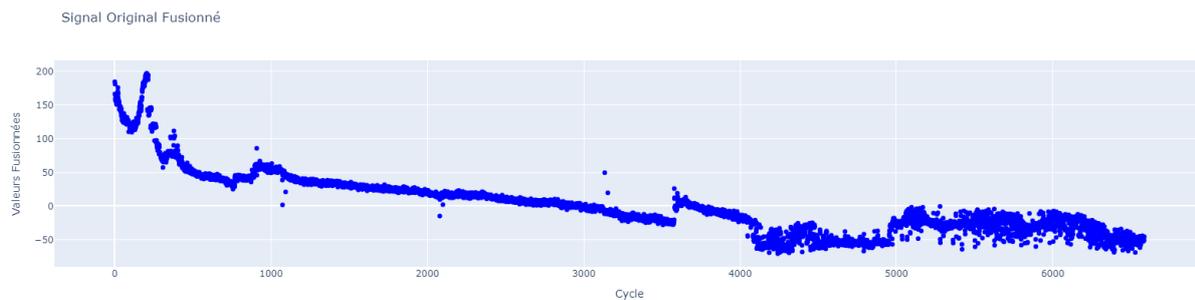


Figure 31 - Visualisation obtenue par fusion des valeurs

Forts de ce jeu de données préparé et mis en forme, nous passons maintenant à l'analyse des performances des algorithmes supervisés.

4.3.2. Résultat test algorithmes

Afin de choisir une combinaison d'algorithmes de classification et de régression qui soit optimale pour notre cas d'usage, nous exposons dans les tableaux 4 et 5 les résultats en termes de métriques discutées précédemment dans la section "Métriques de validation". Pour la classification, un algorithme est considéré comme optimal lorsque les métriques telles que l'exactitude (accuracy) et l'AUC-ROC se rapprochent le plus possible de 1, indiquant respectivement une proportion maximale de prédictions correctes et une bonne capacité de discrimination entre les classes. Pour la régression, l'optimalité est atteinte lorsque le RMSE (Root Mean Squared Error) est le plus petit, reflétant une erreur moyenne minimale entre les valeurs prédites et réelles, et lorsque le R^2 est proche de 1, signifiant que le modèle explique une grande proportion de la variance des données. Le RMSE doit être minimal, car il mesure l'écart quadratique moyen entre les prédictions et les observations, et une valeur faible indique une meilleure précision du modèle, réduisant ainsi les erreurs significatives dans l'estimation du RUL.

Dans ce test, nous comparons une gamme de modèles standards (arbres de décision, forêts aléatoires, machines à vecteurs de support, réseaux neuronaux, et autres) en examinant leurs métriques clés avec et sans optimisation des hyperparamètres via GridSearchCV. L'objectif est d'identifier les configurations les plus performantes pour chaque tâche, en mesurant la

précision et la capacité discriminante en classification (via Exactitude, AUC-ROC, Mean_CV_Accuracy), et l'exactitude des prédictions en régression (via MSE, RMSE, MAE, R², Temporal_Accuracy). Cette analyse vise à éclairer les choix algorithmiques et l'impact de l'optimisation dans notre contexte.

Tableau 4 - Récapitulatif des performances en classification

| Modèle | Variant | GridSearchCV | Accuracy | AUC-ROC | Mean CV Accuracy |
|-----------------------------|-------------------------------------|--------------|----------|---------|------------------|
| DecisionTree Classifier | criterion='gini' | Non | 0.9810 | 0.9846 | 0.8984 |
| RandomForest Classifier | criterion='gini' | Oui | 0.9947 | 0.9998 | 0.9220 |
| SVC | kernel='rbf', probability=True | Oui | 0.9689 | 0.9986 | 0.9387 |
| MLPClassifier | activation='tanh', max_iter=2000 | Oui | 0.9780 | 0.9987 | 0.8958 |
| Logistic Regression | penalty='l2' | Oui | 0.9598 | 0.9901 | 0.9326 |
| GaussianNB | - | Non | 0.9431 | 0.9831 | 0.9016 |
| GradientBoosting Classifier | loss='log_loss' | Oui | 0.9901 | 0.9996 | 0.9180 |
| XGB Classifier | objective='multi:softp rob' | Oui | 0.9932 | 0.9998 | 0.9245 |
| KNeighbors Classifier | metric='manhattan' | Oui | 0.9879 | 0.9975 | 0.9262 |

Les résultats des tests révèlent que les modèles d'ensemble, tels que RandomForestClassifier, XGBClassifier, et GradientBoostingClassifier, dominent en classification multiclasse, affichant des performances intéressantes avec des Accuracy supérieures à 0.99 et des AUC-ROC proches de 1.0, même sans optimisation.

En classification, RandomForestClassifier (variant gini, optimisé) atteint une Accuracy de 0.9947 et un AUC-ROC de 0.9998, surpassant légèrement XGBClassifier (Accuracy = 0.9932, AUC-ROC = 0.9998), grâce à une précision comparable et une robustesse élevée (Mean CV Accuracy ~0.9220 vs 0.9245). Ces performances permettent une segmentation fiable des phases de dégradation (Rodage, Maturité, Vieillesse, Défaillance), avec une capacité particulière de RandomForestClassifier à capturer efficacement les frontières entre les classes, notamment la phase critique de Vieillesse, sans nécessiter d'ajustements complexes des hyperparamètres.

Dans l'ensemble, l'optimisation via GridSearchCV améliore légèrement RandomForestClassifier (Accuracy de 0.9939 à 0.9947, AUC-ROC de 0.9996 à 0.9998), stabilise XGBClassifier (Accuracy inchangée à 0.9932, AUC-ROC de 0.9999 à 0.9998), et renforce GradientBoostingClassifier (Accuracy de 0.9894 à 0.9901). Les gains restent modestes, les performances de base étant déjà proche de 1. Pour ces modèles, l'optimisation affine surtout la robustesse sans bouleverser les résultats.

La figure 32 expose les valeurs classifiées par l'algorithme RandomForestClassifier, tandis que la figure 33 met en évidence les erreurs de classification pour ce même algorithme.

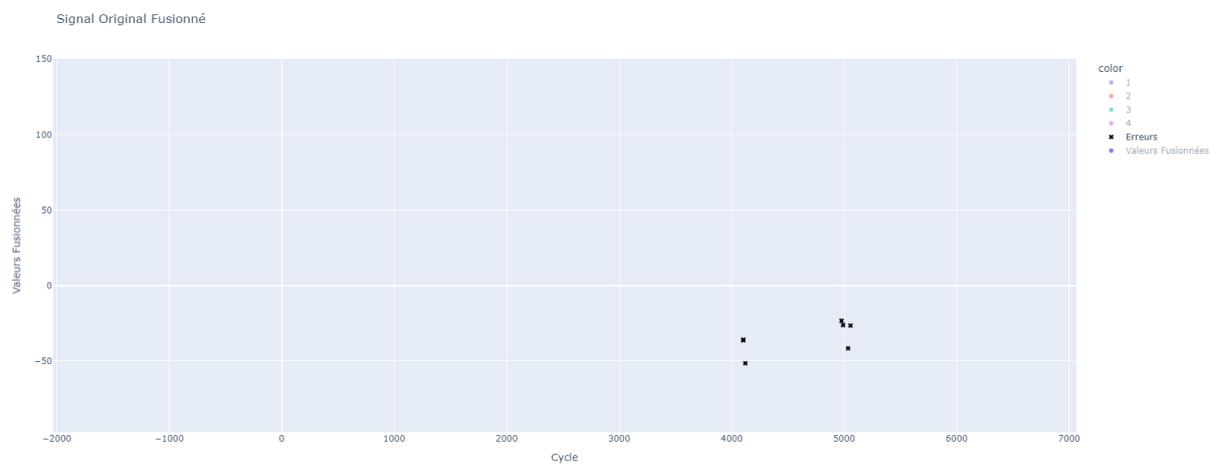
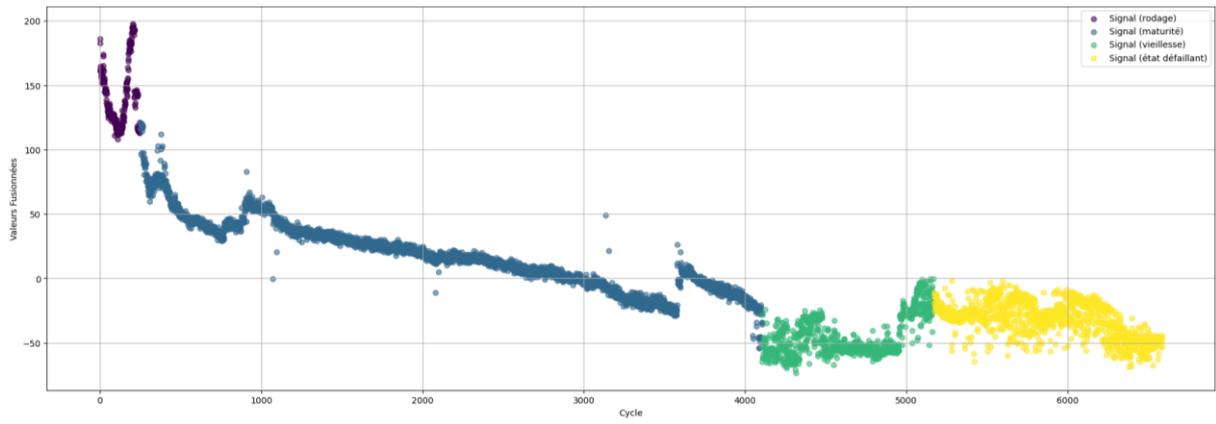


Tableau 5 - Récapitulatif des performances en régression

| Modèle | Variant | GridSearch CV | MSE | RMSE | MAE | R ² | Mean CV MSE |
|----------------------------|---------------------------------|---------------|---------|-------|-------|----------------|-------------|
| DecisionTree Regressor | criterion='squared_error' | Oui | 4724.59 | 68.74 | 44.59 | 0.9457 | 5860.31 |
| RandomForest Regressor | criterion='squared_error' | Oui | 2710.97 | 52.07 | 37.91 | 0.9688 | 3394.33 |
| SVR | kernel='linear' | Oui | 8920.32 | 94.45 | 75.39 | 0.8975 | 10040.91 |
| MLP Regressor | activation='tanh' | Oui | 3122.07 | 55.88 | 40.62 | 0.9641 | 3946.78 |
| Bayesian Ridge | compute_score=False | Non | 8868.01 | 94.17 | 77.08 | 0.8981 | 9995.61 |
| GradientBoosting Regressor | loss='huber' | Oui | 2635.97 | 51.34 | 36.34 | 0.9697 | 3331.60 |
| XGB Regressor | objective='reg:squaredlogerror' | Oui | 2572.42 | 50.71 | 37.57 | 0.9704 | 3339.18 |
| KNeighbors Regressor | metric='manhattan' | Non | 5385.67 | 73.39 | 55.20 | 0.9381 | 5984.19 |

En régression, XGBRegressor (optimisé) surpasse RandomForestRegressor en offrant les meilleures prédictions, avec un RMSE minimal de 50.72 contre 52.07 et un R² supérieur (0.9704 vs 0.9688), démontrant une capacité accrue à estimer précisément la durée de vie restante (RUL) à partir des signaux fusionnés. Cette tâche, centrale pour anticiper la défaillance à partir de la phase de Vieillesse détectée en classification, bénéficie de la précision et de la flexibilité de XGBRegressor, optimisé via GridSearchCV, qui ajuste ses hyperparamètres pour minimiser l'erreur sur les cycles restants. En comparaison, RandomForestRegressor, bien que robuste avec un RMSE de 52.07 et un R² de 0.9688, reste légèrement inférieur, tandis que GradientBoostingRegressor (RMSE 51.34, R² 0.9697) se positionne comme un compromis efficace, tous deux bénéficiant de l'optimisation mais sans atteindre la précision de XGBRegressor.

La figure 34 schématise le décompte du RUL cycle après cycle, dès la détection du début de la phase de Vieillesse.

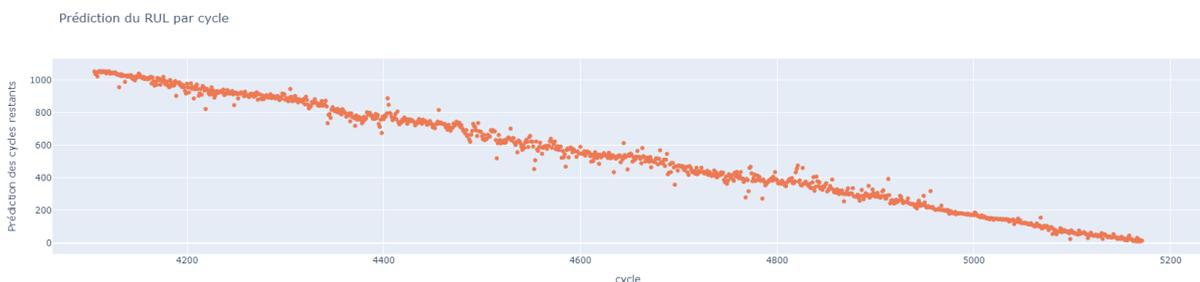


Figure 34 - Décompte progressif du RUL

L'optimisation via GridSearchCV renforce XGBRegressor, qui atteint un RMSE minimal de 50.38 et un R² de 0.9708, surpassant RandomForestRegressor (RMSE 52.07, R² 0.9688), grâce à un réglage des hyperparamètres pour estimer la durée de vie restante (RUL). GradientBoostingRegressor (RMSE 51.34, R² 0.9697) bénéficie également de l'optimisation, mais avec des gains plus modestes. Pour ces modèles, l'optimisation affine la précision, mais XGBRegressor reste particulièrement intéressant dans la prévision des cycles restants à partir de la phase de Vieillesse.

Pour notre cas d'usage, une combinaison de RandomForestClassifier et XGBRegressor optimisé apparaît optimale, équilibrant précision, généralisation et efficacité opérationnelle. Le tableau 6 suivant présente les résultats de prévisions de la durée de vie restante pour un sous-ensemble de cycles d'un système, spécifiquement ceux appartenant à la classe 3 (vieillesse) dans notre cas : 1 pour Rodage, 2 pour Maturité, 3 pour Vieillesse, 4 pour Défaillance.

Les données sont exposées dans trois colonnes :

- **Cycle** : le numéro du cycle dans la séquence temporelle, allant de 4108 à 5168. Les sauts sont dus à la sélection des cycles de la classe 3 et à un échantillonnage aléatoire de 20 % des données (213 cycles sur 1066).
- **RUL réel** : la durée de vie restante réelle, exprimée en nombre de cycles restants jusqu'à l'état de défaillance (classe 4), variant progressivement de 1064 à 4. Le RUL réel est mesuré en calculant la différence entre le cycle courant et le cycle où l'état de défaillance (classe 4) est atteint pour la première fois dans les données, c'est-à-dire le cycle où la classe passe de 3 à 4 dans la séquence temporelle.
- **RUL Prédit** : la durée de vie restante estimée par notre modèle de régression.

Tableau 6 - Comparaison du RUL réel VS RUL prédit

| Cycle | RUL réel | RUL prédit | Cycle | RUL réel | RUL prédit | Cycle | RUL réel | RUL prédit | Cycle | RUL réel | RUL prédit |
|-------|----------|------------|-------------|----------|------------|-------------|----------|------------|-------------|----------|------------|
| 4108 | 1064 | 1042 | 4371 | 801 | 718 | 4610 | 562 | 533 | 4856 | 316 | 389 |
| 4109 | 1063 | 1021 | 4379 | 793 | 744 | 4613 | 559 | 502 | 4867 | 305 | 301 |
| 4116 | 1056 | 1047 | 4381 | 791 | 757 | 4619 | 553 | 485 | 4868 | 304 | 255 |
| 4118 | 1054 | 1040 | 4386 | 786 | 797 | 4628 | 544 | 521 | 4870 | 302 | 299 |
| 4129 | 1043 | 1031 | 4392 | 780 | 791 | 4631 | 541 | 519 | 4884 | 288 | 259 |
| 4131 | 1041 | 955 | 4395 | 777 | 729 | 4634 | 538 | 434 | 4888 | 284 | 268 |
| 4136 | 1036 | 1025 | 4396 | 776 | 712 | 4639 | 533 | 535 | 4890 | 282 | 305 |
| 4137 | 1035 | 988 | 4398 | 774 | 674 | 4641 | 531 | 495 | 4893 | 279 | 350 |
| 4145 | 1027 | 1006 | 4400 | 772 | 759 | 4644 | 528 | 612 | 4898 | 274 | 306 |
| 4155 | 1017 | 1010 | 4402 | 770 | 775 | 4648 | 524 | 451 | 4902 | 270 | 307 |
| 4160 | 1012 | 1004 | 4404 | 768 | 887 | 4649 | 523 | 541 | 4904 | 268 | 287 |
| 4161 | 1011 | 1012 | 4405 | 767 | 847 | 4658 | 514 | 564 | 4907 | 265 | 316 |
| 4164 | 1008 | 978 | 4412 | 760 | 791 | 4663 | 509 | 582 | 4913 | 259 | 392 |
| 4165 | 1007 | 1015 | 4418 | 754 | 738 | 4676 | 496 | 494 | 4916 | 256 | 260 |
| 4169 | 1003 | 970 | 4425 | 747 | 758 | 4681 | 491 | 568 | 4925 | 247 | 289 |
| 4172 | 1000 | 1007 | 4427 | 744 | 757 | 4687 | 485 | 510 | 4929 | 243 | 252 |
| 4173 | 999 | 970 | 4429 | 743 | 753 | 4688 | 484 | 546 | 4931 | 241 | 250 |
| 4176 | 996 | 1018 | 4433 | 739 | 738 | 4689 | 483 | 421 | 4949 | 223 | 289 |
| 4178 | 994 | 973 | 4434 | 738 | 733 | 4690 | 482 | 431 | 4950 | 222 | 248 |
| 4182 | 990 | 955 | 4439 | 733 | 731 | 4695 | 477 | 463 | 4954 | 218 | 225 |
| 4184 | 988 | 977 | 4440 | 732 | 756 | 4696 | 476 | 357 | 4956 | 216 | 318 |
| 4189 | 983 | 902 | 4442 | 730 | 748 | 4702 | 470 | 468 | 4965 | 207 | 206 |
| 4192 | 979 | 985 | 4448 | 723 | 736 | 4705 | 467 | 454 | 4967 | 205 | 229 |
| 4194 | 978 | 974 | 4450 | 722 | 709 | 4710 | 462 | 417 | 4975 | 197 | 179 |
| 4202 | 970 | 923 | 4456 | 716 | 815 | 4719 | 453 | 421 | 4983 | 189 | 179 |
| 4206 | 966 | 956 | 4460 | 712 | 729 | 4721 | 451 | 436 | 4994 | 178 | 179 |

| | | | | | | | | | | | |
|------|-----|-----|-------------|-----|-----|-------------|-----|-----|-------------|-----|-----|
| 4207 | 965 | 886 | 4461 | 711 | 730 | 4724 | 448 | 435 | 5001 | 171 | 187 |
| 4212 | 960 | 903 | 4465 | 707 | 694 | 4739 | 433 | 402 | 5002 | 170 | 180 |
| 4213 | 958 | 917 | 4468 | 704 | 688 | 4740 | 432 | 421 | 5008 | 164 | 169 |
| 4215 | 957 | 909 | 4471 | 701 | 731 | 4749 | 423 | 363 | 5013 | 159 | 150 |
| 4219 | 953 | 821 | 4483 | 689 | 671 | 4755 | 417 | 390 | 5015 | 157 | 151 |
| 4226 | 946 | 913 | 4487 | 685 | 634 | 4756 | 416 | 381 | 5016 | 156 | 145 |
| 4242 | 930 | 918 | 4494 | 678 | 690 | 4759 | 413 | 435 | 5020 | 152 | 142 |
| 4245 | 927 | 916 | 4496 | 676 | 735 | 4763 | 409 | 435 | 5030 | 142 | 146 |
| 4247 | 925 | 911 | 4515 | 657 | 519 | 4766 | 406 | 374 | 5040 | 132 | 144 |
| 4264 | 908 | 909 | 4517 | 655 | 662 | 4768 | 404 | 279 | 5041 | 131 | 116 |
| 4274 | 898 | 912 | 4519 | 653 | 611 | 4771 | 401 | 317 | 5045 | 127 | 149 |
| 4280 | 892 | 898 | 4536 | 636 | 556 | 4772 | 400 | 460 | 5064 | 108 | 119 |
| 4284 | 888 | 892 | 4551 | 621 | 620 | 4773 | 399 | 467 | 5068 | 104 | 155 |
| 4290 | 882 | 886 | 4553 | 619 | 453 | 4776 | 396 | 398 | 5069 | 103 | 90 |
| 4298 | 874 | 907 | 4554 | 618 | 507 | 4779 | 393 | 384 | 5076 | 96 | 79 |
| 4304 | 868 | 889 | 4557 | 615 | 565 | 4785 | 387 | 272 | 5080 | 92 | 55 |
| 4305 | 867 | 944 | 4568 | 604 | 596 | 4795 | 377 | 382 | 5084 | 88 | 53 |
| 4314 | 858 | 881 | 4570 | 602 | 639 | 4798 | 374 | 434 | 5098 | 74 | 24 |
| 4316 | 856 | 870 | 4571 | 601 | 583 | 4809 | 363 | 348 | 5100 | 72 | 64 |
| 4319 | 853 | 883 | 4581 | 591 | 522 | 4819 | 353 | 433 | 5114 | 58 | 54 |
| 4321 | 851 | 896 | 4583 | 589 | 525 | 4820 | 352 | 455 | 5116 | 56 | 30 |
| 4324 | 848 | 825 | 4584 | 588 | 544 | 4821 | 351 | 474 | 5131 | 41 | 25 |
| 4337 | 835 | 870 | 4585 | 587 | 468 | 4822 | 350 | 376 | 5132 | 40 | 53 |
| 4343 | 829 | 734 | 4588 | 584 | 525 | 4826 | 346 | 460 | 5142 | 30 | 54 |
| 4350 | 822 | 809 | 4592 | 580 | 556 | 4836 | 336 | 372 | 5166 | 6 | 31 |
| 4353 | 819 | 822 | 4596 | 576 | 528 | 4837 | 335 | 404 | 5168 | 4 | 21 |
| 4362 | 810 | 774 | 4600 | 572 | 589 | 4838 | 334 | 320 | | | |
| 4366 | 806 | 746 | 4603 | 569 | 546 | 4840 | 332 | 407 | | | |

Pour évaluer la qualité des prévisions, une comparaison entre le RUL réel et le RUL prédit montre des écarts variables. Par exemple, pour le cycle 4108, le RUL réel est de 1064 cycles, tandis que le RUL prédit est de 1042 cycles, soit un écart de 22 cycles, témoignant d'une précision proche en phase initiale. À l'inverse, pour le cycle 4913, le RUL réel est de 259 cycles, mais le RUL prédit est de 392 cycles, soit un écart plus marqué de 133 cycles. Les prédictions sont généralement proches du RUL réel pour les cycles initiaux et bien que des écarts plus importants puissent apparaître pour des cycles intermédiaires, on observe une amélioration notable sur les derniers cycles. Par exemple, pour le cycle 5131, l'écart est de 16 cycles, et pour le cycle 5168, il est de 17 cycles, montrant que les écarts se resserrent dans les phases proches de la défaillance. Cela indique que malgré une sensibilité aux variations des données dans certaines phases intermédiaires, le modèle affiche une précision intéressante dans les derniers cycles.

4.4. Méthode d'entraînement de l'algorithme

La sélection de ces deux algorithmes (RandomForestClassifier et XGBRegressor) repose sur leurs méthodes d'entraînement et leurs principes algorithmiques bien établis, adaptés à notre

problématique de segmentation des phases de dégradation et de prédiction du RUL. Une approche d'ensemble combine les prévisions de multiples modèles individuels, appelés apprenants faibles (ici, des arbres de décision), pour produire une prédiction globale plus précise et moins sensible aux erreurs d'un seul modèle, exploitant ainsi la "sagesse des foules" (Hastie et al. (2009) Le principe de la "sagesse des foules" repose sur l'idée qu'un groupe d'apprenants faibles, bien que chacun puisse être imparfait, peut collectivement aboutir à une décision précise en agrégeant leurs prédictions. Dans ce contexte, chaque arbre de décision explore une perspective différente des données (via des sous-ensembles aléatoires ou des pondérations spécifiques), ce qui permet de réduire le biais et la variance globale du modèle. Cette diversité des apprenants compense les erreurs individuelles, rendant les algorithmes comme RandomForestClassifier et XGBRegressor intéressants pour capturer les patterns liés aux phases de dégradation et à l'estimation du RUL. Ces principes de fonctionnement se déclinent en deux stratégies principales :

- Le bagging (agrégation parallèle)
- Le boosting (ajustement séquentiel)

4.4.1. RandomForestClassifier : Fonctionnement

RandomForestClassifier utilise une approche d'ensemble basée sur le bagging (Bootstrap Aggregating), introduite par Breiman (2001). Chaque arbre de décision est entraîné indépendamment sur un sous-ensemble aléatoire des données, obtenu par échantillonnage avec remplacement (bootstrap), et un sous-ensemble aléatoire des caractéristiques à chaque nœud, réduisant ainsi la corrélation entre les arbres et le risque de surajustement. Pour une observation x , la prévision finale est déterminée par un vote majoritaire (14) parmi T arbres :

$$\hat{y}(x) = \text{mode}\{h_t(x)\}_{t=1}^T \quad (14)$$

où $\hat{y}(x)$ est la classe prédite, $h_t(x)$ représente la prédiction de classe (dans notre cas : 1 pour Rodage, 2 pour Maturité, 3 pour Vieillesse, 4 pour Défaillance) du t -ième arbre, et T est le nombre total d'arbres (par défaut 100 dans sklearn (Fai et al., 2023), mais ajustable).

L'entraînement minimise une mesure d'impureté, comme l'indice Gini (Akırmak et Altan, 2023), calculé dans chaque nœud (15) pour évaluer la pureté des classes :

$$Gini = 1 - \sum_{k=1}^K p_k^2 \quad (15)$$

où K est le nombre de classes (ici, $K=4$), et p_k est la proportion d'échantillons de la classe k dans le nœud. Par exemple : si 60 % des échantillons sont de classe 3, alors $p_3 = 0.6$. Un $Gini$ de 0 indique une pureté parfaite (tous les échantillons d'une même classe), tandis qu'un $Gini$ élevé reflète une mixité. Dans notre cas d'usage, l'indice $Gini$ est celui qui nous a retourné les métriques les plus performantes.

Outre l'indice $Gini$, l'entropie (16) peut être utilisée pour mesurer l'impureté, offrant une sensibilité différente aux distributions de classes et dont la formule est la suivante :

$$- \sum_{k=1}^K p_k \log_2(p_k) \quad (16)$$

L'entropie pénalise plus fortement les distributions inégales que Gini, car la fonction logarithmique amplifie les petites probabilités. Elle mesure le désordre ou l'incertitude dans un nœud. Une entropie de 0 signifie une totale absence de mixité (une seule classe), tandis qu'une entropie maximale (ex. : $\log_2 K$ pour des classes équiprobables) indique une mixité totale.

Le bagging réduit la variance des prédictions. Pour T arbres indépendants $h_t(x)$, chacun prédisant une classe pour une observation x , l'erreur moyenne d'un arbre individuel est atténuée par le vote majoritaire. Breiman (1996) montre que, sous l'hypothèse d'arbres faiblement corrélés, l'erreur agrégée diminue approximativement comme $\frac{1}{T}$ fois l'erreur moyenne d'un arbre, améliorant ainsi la stabilité des prédictions. Dans notre étude, cette approche permet à RandomForestClassifier d'atteindre une précision intéressante (Accuracy = 0.9947) sans optimisation, en segmentant efficacement les phases, et démontre sa capacité à capturer les frontières complexes entre les classes avec une bonne stabilité.

4.4.2. XGBRegressor : Fonctionnement

XGBRegressor, basé sur le gradient boosting (Chen & Guestrin, 2016), adopte une approche d'ensemble séquentielle où les arbres sont construits pour corriger les erreurs des prédictions précédentes. Contrairement au bagging, qui agrège indépendamment, le boosting ajuste les arbres itérativement en minimisant une fonction de perte cumulative, ici l'erreur quadratique pour la régression (17) est donnée par l'équation :

$$L(y, \hat{y}) = \frac{1}{2} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (17)$$

où n est le nombre d'observations (dans notre cas, nombre de cycles dans la phase de Vieillesse), y_i est le RUL réel (cycles restants), et \hat{y}_i est la prévision pour l'observation i . La prévision finale est une somme des contributions des T arbres, ajustée par un taux d'apprentissage (18) (ex. : 0.1) :

$$\hat{y}_i(x) = \sum_{t=1}^T f_t(x) \quad (18)$$

où $f_t(x)$ est la fonction prédictive du t -ième arbre (une valeur continue), T est le nombre d'arbres (ex. : 200 dans notre optimisation, paramètre optimal identifié par l'optimiseur GridSearchCV). Ce taux d'apprentissage contrôle la contribution de chaque arbre pour éviter les ajustements trop rapides. L'entraînement optimise une fonction *Objectif* (19) régularisée :

$$\text{Objectif} = L(y, \hat{y}) + \sum_{t=1}^T \Omega(f_t) \quad (19)$$

où $\Omega(f_t) = \gamma L_t + \frac{1}{2} \lambda \|w\|^2$ est un terme de régularisation. Ω est une notation introduite par les auteurs de XGBoost pour représenter la pénalité. Ici, γ est un paramètre de pénalité sur le nombre de feuilles L_t de l'arbre f_t , tandis que λ applique une régularisation L2 sur les poids des feuilles λ (valeurs prédites aux terminaisons), et $\|w\|^2$ est la norme carrée des poids, limitant les prévisions extrêmes. Cette régularisation, combinée à l'optimisation par descente de gradient, réduit le biais en ajustant les erreurs résiduelles et contrôle la variance en évitant le surajustement. Dans notre cas, XGBRegressor (optimisé avec GridSearchCV, ex. : max_depth=5, n_estimators=200) offre de bonnes performances pour prédire le RUL avec un

RMSE de 50.71, surpassant RandomForestRegressor grâce à cet ajustement précis des hyperparamètres.

Globalement, la fonction *Objectif* joue le rôle d'une fonction de perte. Elle est programmée par défaut pour mesurer l'erreur quadratique. Les auteurs de l'algorithme ont prévu des alternatives à ce paramétrage pour traiter des cas de figures spécifiques. L'ensemble de ces fonctions alternatives ont été testées pour vérifier si elles permettaient une amélioration des prévisions du RUL. L'une d'entre elle s'est avérée supérieure en performance et est détaillée comme suit :

L'erreur quadratique logarithmique (20), accessible via un paramétrage explicite, mesure l'erreur quadratique entre les logarithmes des valeurs prédites et réelles. La formule est la suivante :

$$L = \frac{1}{2} [\log(1 + y_{pred}) - \log(1 + y_{true})]^2 \quad (20)$$

Cette fonction est adaptée aux cibles positives continues et traite plus spécifiquement les erreurs relatives plutôt que les écarts absolus, par le biais de la transformation logarithmique. Cela a pour effet de réduire les grandes différences sur de larges échelles et donc de réduire la sensibilité aux valeurs aberrantes.

Du fait de sa meilleure performance, limitée mais existante, nous avons choisi d'employer une configuration avec fonction quadratique logarithmique pour la création de notre modèle.

4.5. Création du modèle de dégradation

Pour ce qui est de l'entraînement du modèle de dégradation, le processus commence par l'entraînement d'un RandomForestClassifier, avec une optimisation des hyperparamètres via GridSearchCV, choisi pour sa capacité à détecter avec précision la phase de Vieillesse (classe 3) sans nécessiter d'optimisation complexe, offrant ainsi une base fiable pour la suite. Ensuite, un XGBRegressor est employé pour prédire le RUL à partir des cycles de cette phase, avec une optimisation des hyperparamètres via GridSearchCV. Le pseudo-code ci-dessous (figure 35) détaille ces étapes de manière structurée, illustrant la création et la sauvegarde des deux modèles.

Algorithme: ModeleHybridePourPronostic

Entrées:

- fichier_csv: Chemin vers les données (augmentées, data.csv)
- signal_columns: Liste des colonnes de caractéristiques sauvegardées

Sorties:

- clf_model: Modèle de classification sauvegardé
- rul_model: Modèle de régression RUL sauvegardé

Début

```
df_mean ← LireCSV(fichier_csv)
```

```
SupprimerLigne(df_mean, 'cycle', ignorer_entrée=True)
```

```
AjouterColonne(df_mean, 'cycle', index_des_cycles)
```

```
X ← Normaliser(df_mean[signal_columns])
```

```
y ← df_mean[Y] - 1
```

```
(X_train, X_test, y_train, y_test) ← DiviserDonnées(X, y, test_size=0.2, random_state=42)
```

```
param_grid_clf ← {
```

```
  'n_estimators': [100, 200],
```

```
  'max_depth': [None, 10],
```

```
  'criterion': ['gini']
```

```
}
```

```
clf_base ← InitialiserRandomForestClassifier(random_state=42)
```

```
grid_search_clf ← GridSearchCV(clf_base, param_grid_clf, cv=5, scoring='accuracy')
```

```
Entraîner(grid_search_clf, X_train, y_train)
```

```
clf ← MeilleurEstimateur(grid_search_clf)
```

```
Sauvegarder(clf, 'random_forest_model.pkl')
```

```
df_class_3 ← Filtrer(df_mean, Y == 3)
```

```
Pour chaque c dans df_class_3['cycle']:
```

```
  df_class_4['cycle'] ← Minimum(df_mean['cycle'] où cycle > c et Y == 4)
```

```
SupprimerLignesNaN(df_class_3, 'Next_Class_4_Cycle')
```

```
df_class_3['Cycles Restants'] ← df_class_3['Next_Class_4_Cycle'] - df_class_3['cycle']
```

```
X_rul ← Normaliser(df_class_3[signal_columns])
```

```
y_rul ← df_class_3['Cycles Restants']
```

```
(X_train_rul, X_test_rul, y_train_rul, y_test_rul) ← DiviserDonnées(X_rul, y_rul, test_size=0.2, random_state=42)
```

```
param_grid_rul ← {
```

```
  'n_estimators': [100, 200, 300],
```

```
  'max_depth': [5, 7],
```

```
  'learning_rate': [0.01, 0.1, 0.3],
```

```
  'subsample': [0.8, 1.0],
```

```
  'colsample_bytree': [0.8, 1.0],
```

```
  'objective': ['reg:squarederror']
```

```
}
```

```
xgb_reg ← InitialiserXGBRegressor(random_state=42)
```

```
grid_search_rul ← GridSearchCV(xgb_reg, param_grid_rul, cv=5, scoring='neg_mean_squared_error')
```

```
Entraîner(grid_search_rul, X_train_rul, y_train_rul)
```

```
rul_model ← MeilleurEstimateur(grid_search_rul)
```

```
Sauvegarder(rul_model, 'xgb_rul_model.pkl')
```

Fin

Figure 35 - Pseudo-code de création du modèle d'estimation du RUL

Les modèles ainsi créés deviennent appelables, pour nous permettre de leur soumettre un certain nombre de valeurs, de sorte à pouvoir classifier et prédire le cas échéant, dès la détection du début de la phase de vieillesse.

4.6. Conclusion du chapitre 4

Pour répondre aux besoins de notre cas d'usage en pronostic pour un équipement industriel, une approche hybride combinant classification et régression est proposée comme une solution la plus adaptée.

En classification, l'utilisation de `RandomForestClassifier`, optimisé via `GridSearchCV`, permet une détection précise et robuste de la phase de Vieillesse (classe 3), une étape déterminante pour initier le suivi de la durée de vie restante (RUL) avant l'entrée en Défaillance (classe 4). Ce modèle, dans sa configuration non-optimisée, atteint une précision très satisfaisante et une capacité discriminante quasi-parfaite, garantissant une identification fiable des cycles critiques.

Une fois cette phase détectée, `XGBRegressor`, optimisé via `GridSearchCV`, prend le relais pour prédire progressivement le RUL avec une erreur minimale et une forte adéquation aux données.

Cette combinaison hybride est efficace dans notre contexte en offrant une transition fluide entre la segmentation des phases de dégradation et l'estimation continue des cycles restants, comme illustré par les prévisions visualisées dans la figure 32.

Contrairement à des modèles alternatifs décrits plus haut comme SVR ou MLP, qui nécessitent une optimisation lourde pour des performances inférieures, cette approche tire parti de la robustesse naturelle des modèles d'ensemble. Ainsi, le choix de `RandomForestClassifier` pour la classification des phases, suivi de `XGBRegressor` optimisé pour le calcul du RUL, constitue une stratégie intéressante, alliant précision dans la détection de la "Vieillesse" et fiabilité dans la prévision de la durée de vie restante (figure 34), répondant ainsi aux besoins opérationnels de notre étude.

Enfin, ces informations seront exploitées pour optimiser la maintenance en proposant des actions adaptées, facilitant ainsi la prise de décision. L'objectif est de maximiser la fiabilité des équipements tout en réduisant les coûts d'intervention et les arrêts non planifiés.

Chapitre 5

5. De la prise de décision

L'aide à la décision constitue la dernière brique essentielle de tout plan de maintenance prévisionnel, valorisant le modèle de dégradation et transformant les prévisions issues des étapes précédentes en actions opérationnelles concrètes. Comme souligné par Jardine et al. (2006), la valeur d'un système de maintenance prévisionnelle réside dans sa capacité à réduire les incertitudes et à optimiser les stratégies de maintenance, un enjeu important dans les environnements industriels où la fiabilité des équipements est primordiale. Dans notre cas d'usage, les prévisions dérivées du modèle de dégradation développées dans les chapitres 2 et 3, vont être exploitées pour élaborer un système d'aide à la décision pour proposer un plan de maintenance hybride combinant les approches systématique et prévisionnelle pour des actions adaptées aux environnements de production contraints.

Dans ce chapitre, nous nous concentrons sur un scénario de validation de cette approche hybride, en confrontant les performances de la maintenance systématique actuelle (en place chez OCTAPHARMA SAS) à celles de la maintenance prévisionnelle proposée par notre modèle. À travers une analyse structurée basée sur plusieurs scénarios de confrontation entre ces deux plans, nous évaluerons la faisabilité et l'efficacité potentielle de cette combinaison pour répondre aux objectifs opérationnels clés : maximiser la fiabilité des équipements tout en réduisant les coûts d'intervention et les arrêts non planifiés. Aussi, nous discuterons de la manière dont ces résultats pourraient constituer une véritable aide à la décision, en posant les bases pour une intégration opérationnelle future.

5.1. Introduction

La méthodologie PHM repose sur une séquence structurée de briques visant à surveiller et gérer la santé des systèmes industriels. Les étapes précédentes, de l'acquisition des données au pronostic de la durée de vie restante, ont permis d'établir un modèle pour estimer la durée de vie d'une membrane industrielle. Dans ce cadre, l'aide à la décision, considérée comme la septième brique du PHM (Vachtsevanos et al., 2006), joue un rôle important. Elle valorise les informations issues des étapes précédentes en les transformant en recommandations pratiques, permettant une gestion proactive des actifs industriels.

Dans le contexte de l'industrie du biomédicament, l'aide à la décision prend une importance particulière en raison des contraintes spécifiques liées à la qualité des produits et aux exigences réglementaires strictes, telles que celles imposées par les bonnes pratiques de fabrication (BPF/GMP) discutées notamment en introduction. Pour rappel, comme vu dans le chapitre 1, les exigences réglementaires poussent naturellement les industriels du biomédicament à adopter un plan systématique, de sorte à démontrer que les opérations de maintenance sont sous contrôle. L'objectif principal est d'exploiter les prévisions de phase de vieillissement de la membrane, illustrées par le RUL, pour optimiser les stratégies de maintenance, tout en garantissant la conformité des processus de production et les exigences qualité. Une maintenance optimisée dans notre contexte doit répondre à un triple impératif :

- Maximiser la fiabilité des équipements de sorte à assurer la continuité des opérations de production,

- Minimiser les coûts d'intervention et réduire les arrêts non planifiés qui pourraient compromettre la qualité des lots produits,
- Réduire les incertitudes liées à la défaillance des équipements, en s'appuyant sur des prévisions fiables pour planifier les interventions au moment le plus opportun.

Ce chapitre explore comment ces prévisions peuvent être exploitées pour concevoir un plan de maintenance hybride, en traduisant les sorties techniques en recommandations concrètes adaptées aux exigences industrielles.

5.2. Contexte opérationnel et maintenance actuelle

Pour aborder la maintenance spécifique des membranes de vannes, il convient d'abord de décrire le cadre général dans lequel s'inscrit la politique de maintenance de notre partenaire industriel, afin de mieux appréhender ses pratiques actuelles et leurs effets sur la performance opérationnelle.

5.2.1. Contexte de maintenance globale actuel

Notre partenaire industriel adopte une approche de maintenance structurée, planifiée sur un cycle annuel, appliquée à tous les équipements de production. Ce système vise à assurer une disponibilité optimale des installations tout en limitant les arrêts imprévus. La figure 36 présente cette organisation.

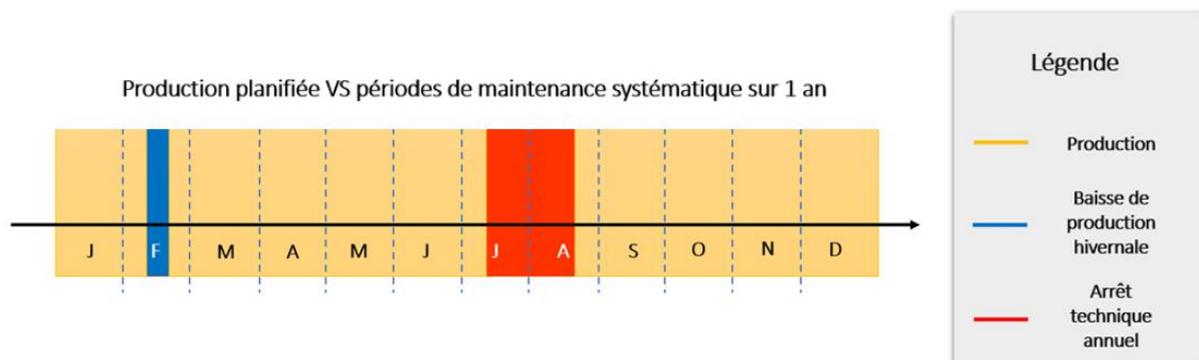


Figure 36 - Schématisation de la maintenance systématique chez OCTAPHARMA SAS

La planification annuelle est structurée autour de deux périodes distinctes d'arrêt technique, programmées pour minimiser l'impact sur la production.

Une première intervention, qualifiée de "baisse de production hivernale", est prévue au courant du mois de février. Cette période d'une durée de quelques semaines, permet de réaliser des opérations de maintenance préventive, des inspections et des ajustements mineurs, sans nécessairement arrêter la production.

Une seconde période, plus longue, est planifiée au courant de l'été. Elle correspond à un arrêt technique annuel et est dédiée à des interventions plus lourdes incluant réparations et remplacements d'équipements. L'entreprise profite également de l'arrêt de la production pour réaliser un certain nombre de projets tels que les extensions de zones de production par exemple, ou toute autre intervention lourde ne pouvant se faire dans un environnement à atmosphère contrôlée.

Les périodes de production, représentées en jaune sur le schéma, couvrent les mois restants, permettant à notre partenaire industriel de maintenir une activité soutenue en dehors des fenêtres de maintenance.

Cette approche de maintenance systématique, qui vise généralement à maintenir les équipements dans un état proche du neuf, engendre des coûts significatifs, tant en termes de ressources matérielles que de main-d'œuvre, sans pour autant garantir l'absence de défaillances imprévues. De plus, la grande majorité de ces équipements sont qualifiés de "critiques", ce qui implique d'un point de vue du système qualité de l'entreprise, que ces équipements sont en contact direct avec le produit. Dans un cas urgent, le remplacement de cet équipement va entraîner à minima un traçage dans le système GMAO (Maintenance Assistée par ordinateur), voire un "événement" ou une "déviation", c'est-à-dire un écart à la norme. Cela va potentiellement déclencher une investigation de la cause et en fonction des cas, des actions correctives et préventives ainsi qu'une évaluation de l'impact. Ce système qualité est décrit succinctement dans l'item suivant.

5.2.2. Cadre qualité pharmaceutique

Face à la criticité des équipements et aux exigences strictes qui en découlent, la gestion des anomalies repose sur un processus rigoureux que nous allons maintenant aborder, en commençant par la définition et le traitement des événements.

5.2.2.1. Événements

Dans le contexte pharmaceutique, un événement désigne toute anomalie observée lors des processus de production ou de maintenance, qu'elle respecte ou non les procédures établies. Ces événements s'inscrivent dans le cadre du système qualité de l'entreprise, conformément aux Bonnes Pratiques de Fabrication (BPF / GMP) et aux réglementations internationales, telles que celles de l'Agence Européenne des Médicaments (EMA) ou de la Food and Drug Administration (FDA). À la différence d'une déviation, un événement ne représente pas forcément une non-conformité : il peut s'agir d'une observation sans effet sur la qualité du produit. Cependant, une analyse est nécessaire pour éviter qu'il n'évolue en problème plus significatif.

La gestion des événements commence par leur enregistrement dans un journal d'opérations ou un système électronique, garantissant une traçabilité initiale. Une évaluation préliminaire est ensuite menée pour déterminer la gravité de l'événement et décider s'il nécessite une requalification en déviation. Cette étape peut inclure une revue des données opérationnelles ou une inspection des équipements concernés. Si l'événement est jugé sans conséquence sur la qualité, il est clos après documentation.

La documentation des événements n'en reste pas moins essentielle. Elle comprend une description de l'occurrence, les conclusions de l'analyse préliminaire et les actions éventuelles, validées en interne par un responsable opérationnel. Ce suivi permet de maintenir une vigilance sur les opérations tout en évitant des interventions superflues.

5.2.2.2. Déviations

Une déviation, dans le secteur pharmaceutique, se définit comme tout écart par rapport aux procédures, spécifications ou normes établies au sein du système qualité, en conformité avec les BPF/GMP et les exigences des autorités comme l'EMA ou la FDA. Ces écarts, susceptibles

d'affecter la qualité, la sécurité ou l'efficacité du produit, nécessitent une gestion stricte pour garantir la conformité réglementaire.

Le processus débute par la détection et le signalement de l'écart. Une investigation approfondie suit pour identifier la cause racine, en examinant les conditions opérationnelles et les données de production. Des actions correctives et préventives (CAPA) sont ensuite mises en œuvre : les actions correctives corrigent l'incident immédiat (par exemple, en remédiant à une anomalie technique), tandis que les actions préventives visent à prévenir sa récurrence, via des ajustements de procédures ou des contrôles renforcés. Une évaluation de l'impact sur la qualité du produit est également réalisée, pouvant conduire dans les pires scénarios à des décisions comme le rejet d'un lot, le rappel d'un lot, ceci pouvant impliquer la santé et la sécurité des patients récipiendaires des biomédicaments.

La documentation des déviations est exhaustive, incluant le rapport initial, les résultats de l'investigation, les CAPA et l'évaluation de l'impact, le tout approuvé par le service qualité. Une fois ces étapes terminées, la déviation est clôturée, avec un suivi pour vérifier l'efficacité des mesures préventives. Ce processus mobilise des ressources importantes, surtout en cas de récurrence.

5.2.3. Focus sur la maintenance appliquée aux membranes de vannes

Dans le cadre spécifique de la maintenance des membranes de vannes, l'entreprise a adopté une stratégie de remplacement systématique tous les trois ans, ce qui correspond à environ 3500 cycles dans notre cas d'usage, conformément aux recommandations des constructeurs. Cette périodicité est établie sur les données de durabilité fournies par le fabricant, qui estiment que les membranes atteignent leur limite d'usure après une période d'utilisation définie, dans des conditions opérationnelles standards. Ainsi, à la fin de chaque cycle triennal, le remplacement des membranes s'effectue généralement lors de l'arrêt technique, afin de minimiser les perturbations sur les périodes de production. Cette approche, bien que conforme aux préconisations industrielles, implique que toute action corrective non anticipée par le calendrier de maintenance systématique, telle que le remplacement imprévu d'une membrane défaillante, doit nécessairement être effectuée pendant les plages de production. Ces interventions, réalisées en dehors des arrêts techniques planifiés, peuvent entraîner des interruptions non prévues de la production et augmentant les risques opérationnels et de contamination.

Le remplacement des membranes des vannes pneumatiques en dehors des plages d'arrêts techniques s'organise selon deux scénarios distincts et en fonction de l'urgence de l'intervention :

- Dans le premier cas, décrit en interne comme « correctif planifié », le remplacement s'opère lorsqu'aucune activité de production n'est en cours, afin d'éviter toute perturbation. Pour donner un ordre de grandeur concret, sur une période de trois ans (2022-2025), 127 interventions de ce type ont été recensées.
- Dans le second cas, décrit en interne comme « correctif urgent », l'intervention est de nature corrective curative, motivée par une défaillance avérée de la membrane, telle qu'une rupture ou une déformation critique. Ces remplacements, par leur caractère imprévu, doivent être réalisés dans les plus brefs délais, en dehors des plages d'arrêts techniques, ce qui implique une intervention durant les périodes de production. Sur la période de trois ans évoqués plus haut, 86 interventions urgentes ont été enregistrées.

5.2.4. Cas concret OCTAPHARMA

Pour ce qui est des cas de “correctifs urgents”, les interventions suivent le processus décrit en figure 37 tandis que les “correctifs planifiés” suivent le processus décrit en figure 38 :

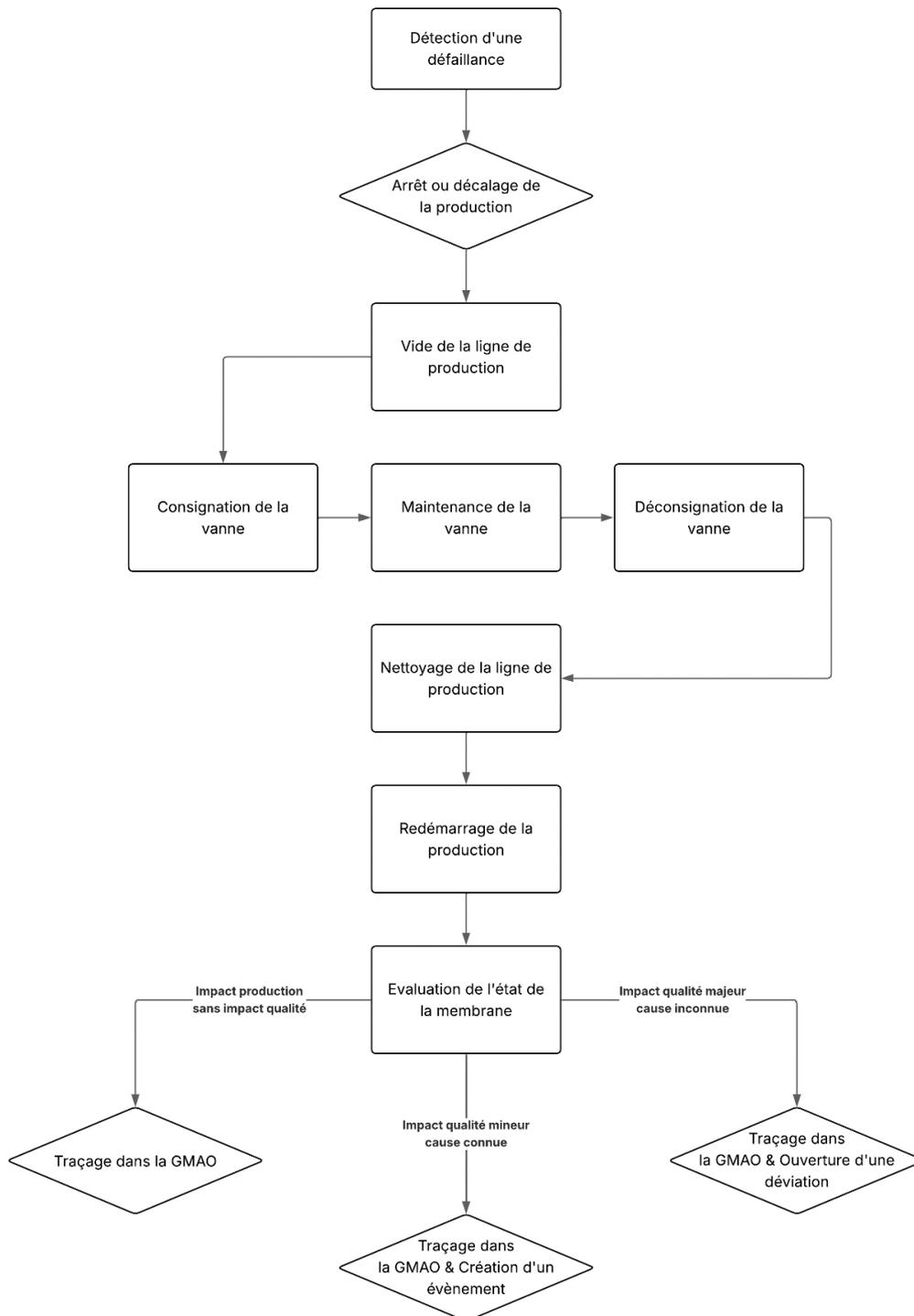


Figure 37 - Processus de remplacement et de suivi d'une membrane sous “correctif urgent”

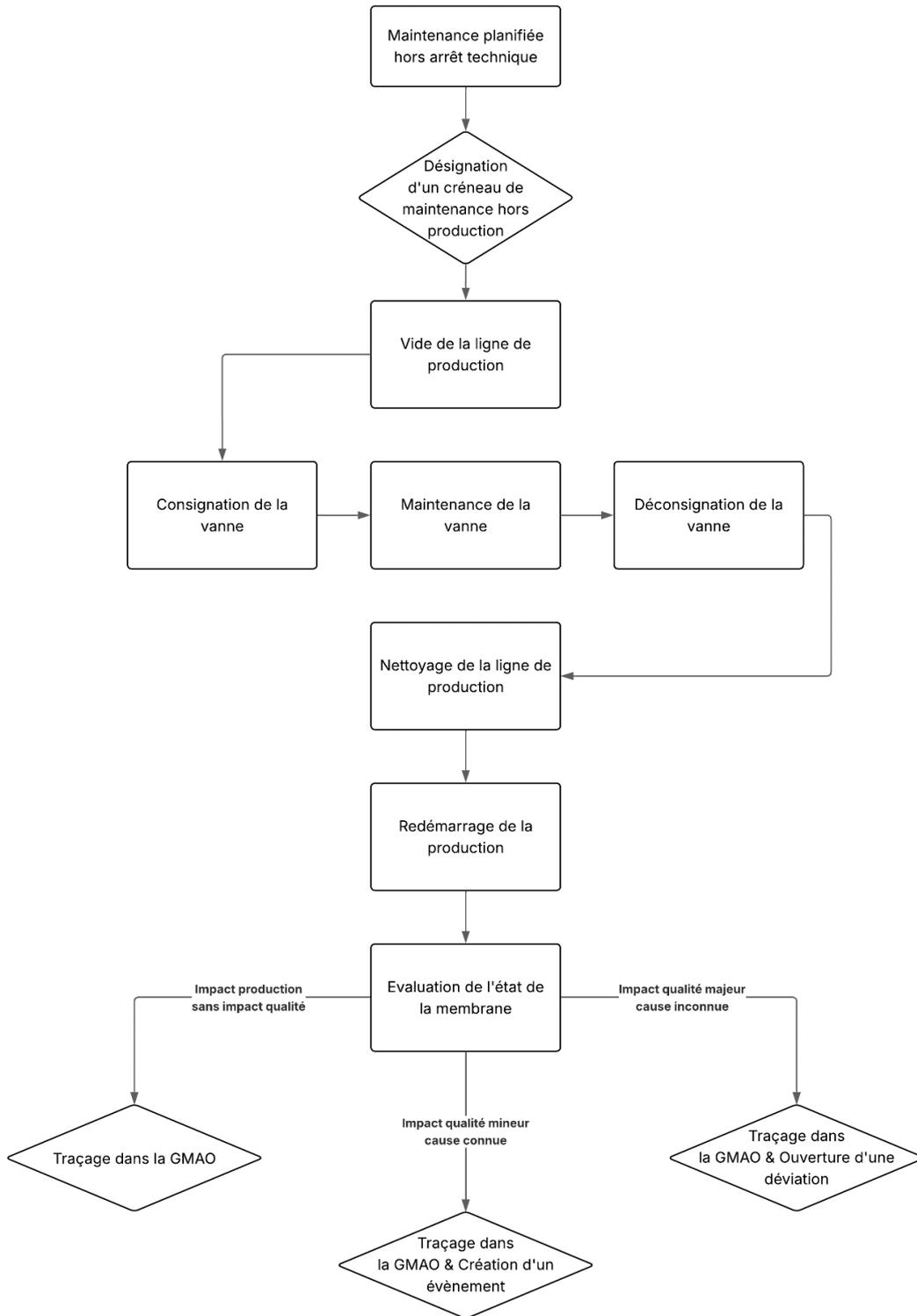


Figure 38 - Processus de remplacement et de suivi d'une membrane sous "correctif planifié"

A titre indicatif, chaque occurrence de "correctif urgent" entraîne un report d'une heure de production, avec un coût estimé à 13 000 € par heure. Sur les 86 cas observés, cela représente une perte totale de 1 180 000 € sur trois ans. Ces arrêts imprévus ne se limitent pas à un impact financier, ils augmentent les risques de non-conformité, chaque incident pouvant générer une déviation nécessitant une investigation et une documentation supplémentaires.

Cette charge additionnelle mobilise des ressources humaines et complexifie la gestion des opérations.

5.3. Intégration des prévisions dans la planification

Pour dépasser les limites de la maintenance systématique actuelle et exploiter pleinement les prévisions issues de notre modèle, il est nécessaire de traduire ces données en un calendrier opérationnel concret.

5.3.1. Exploitation des prévisions pour la prise de décision

Les résultats du modèle ne constituent qu'un premier pas : ils doivent être traduits en décisions concrètes pour produire un réel impact. C'est pourquoi il est essentiel de comprendre la signification et la portée des sorties générées par le modèle.

5.3.1.1. Lecture opérationnelle des sorties du modèle de dégradation

Actuellement, la maintenance combinée (systématique/corrective) entraîne des inefficacités notables propres aux approches systématiques. Celles-ci, souvent opérées prématurément, mobilisent des ressources inutiles, tandis que les défaillances imprévues entre deux cycles planifiés provoquent des arrêts impactant les cadences de production.

En comparaison, une maintenance prévisionnelle basée sur notre modèle peut offrir une alternative intéressante. La classification des phases de vieillesse des équipements identifie une dégradation avancée, tandis que l'estimation du RUL définit une fenêtre d'intervention relativement précise. Cette stratégie surmonte les limites de la maintenance systématique en ciblant les actions sur l'état réel des membranes de vannes, évitant les interventions inutiles, précoces ou tardives.

Pour ce qui est des vannes industrielles, les bénéfices potentiels incluent une réduction des coûts liés aux ressources gaspillées, une diminution des arrêts non planifiés affectant la production et une meilleure garantie de conformité aux exigences GMP.

Dans notre étude, le modèle de dégradation élaboré au chapitre 2 montre un potentiel pour améliorer la gestion des membranes de vannes. Nos analyses révèlent qu'il pourrait prolonger la durée de vie d'environ 33 % de ces composants, en se basant sur une évaluation technique pure, sans prendre en compte jusqu'ici les aspects liés à la prise de décision. Cette estimation souligne l'intérêt d'une approche prévisionnelle pour optimiser l'utilisation des membranes.

Si les prévisions issues du modèle permettent d'identifier les phases de dégradation, encore faut-il pouvoir les exploiter de manière concrète sur le terrain. Cela suppose pour commencer, la conversion des cycles en unités temporelles, mieux adaptées aux logiques de planification terrain.

5.3.1.2. De la prévision au calendrier d'intervention :

Pour que cette amélioration soit opérationnelle, il est essentiel de traduire ces prédictions en termes pratiques, ce qui nous conduit à proposer une méthode de conversion des cycles en unités de temps.

5.3.1.2.1. Conversion des cycles en temps

La transformation de la durée de vie résiduelle (RUL) des membranes, exprimée initialement en cycles, vers des unités de temps comme des heures ou des jours, repose sur l'introduction d'une fonction spécifique. Une fonction proposée dans l'équation (21) utilise la fréquence

d'utilisation de l'équipement, c'est-à-dire un taux de cyclage (équation 22), qui peut être déterminé à partir de données historiques ou opérationnelles. Par exemple, en supposant un taux de 10 cycles par heure, un RUL de 50 cycles équivaut à 5 heures de fonctionnement restant. Cette approche nécessite une connaissance précise du taux de cycles par unité de temps, qui peut être obtenue via des enregistrements passés, d'informations extraites des équipements directement, voire des timestamps d'enregistrement des valeurs dans la base de données. Cette conversion permet de passer d'une prévision théorique à une planification pratique. La formule générale est simple à déduire :

$$RUL \text{ en temps} = \frac{RUL \text{ en cycles}}{\text{Taux de cyclage}} \quad (21)$$

Dans notre cas, l'équipement sollicité pour établir le modèle de dégradation des membranes est en mesure de nous fournir une valeur incrémentale correspondant aux cycles d'ouverture/fermeture successifs. Ceci couplé aux timestamps relevés à chaque incrémentation permet de déterminer la durée d'un cycle. Considérant que nous avons à faire à une production cadencée, avec un nombre de lots à produire défini à l'avance, nous pouvons estimer avec une relative précision la cadence d'utilisation des vannes à membranes.

Ici, le RUL en temps est exprimé en minutes, car notre essai a été cadencé à 1 cycle / 80 minutes, de sorte à représenter une dégradation accélérée. Si le modèle, à détection du démarrage de la phase de vieillesse, détermine un RUL en cycles à 1038 par exemple, ce qui est l'horizon fourni par notre modèle dans le cas d'usage décrit au chapitre 3. Cela nous donnerait :

- RUL en cycles = 1038 cycles
- Taux de cyclage (T_c) = nombre de cycles par unité de temps. Ici,

$$T_c = \frac{1}{80} \quad (22)$$

Dans notre exemple, en appliquant la formule (21) proposée plus haut, cela donne :

$$RUL \text{ en temps} = \frac{1038}{\frac{1}{80}}$$

Avec pour résultat :

$$RUL \text{ en temps} = 1038 \times 80$$

$$RUL \text{ en temps} = 83040 \text{ minutes}$$

Le résultat de cette conversion affiche donc, dans notre cas d'usage, un RUL prédit à 83040 minutes, soit environ 57 jours restant à partir du moment où nous avons réalisé la prévision. Cela nous donne un autre angle de prévision pour donner aux équipes de maintenance un horizon raisonnable pour planifier une maintenance.

Cette première mise en forme de la sortie de notre modèle nous permet de fournir à l'utilisateur deux variables pour établir un horizon de planification que sont : le RUL en cycles et le RUL en temps.

5.3.1.3. Seuils de décision

Une fois le RUL converti en unités de temps, la prochaine étape consiste à établir des seuils de décision pour déclencher les actions de maintenance. Ces seuils définissent le moment où le temps restant devient nécessaire pour justifier une intervention. Il est essentiel de situer cette approche dans le cadre opérationnel existant de notre partenaire. En effet, la définition de ces seuils ne peut être établie sans une compréhension approfondie de la stratégie de maintenance systématique actuelle. À ce stade, notre analyse se concentre sur la faisabilité technique de cette approche, mais à terme, ces seuils devront intégrer des facteurs supplémentaires tels que : les coûts de maintenance, les besoins de production, les exigences réglementaires, les procédures internes de l'entreprise. Cette conversion en temps et l'établissement de seuils en fonction du contexte industriel pose les bases d'une planification.

5.3.2. Scénarios de Confrontation des Approches

Pour valider l'approche hybride et évaluer son efficacité par rapport à la maintenance systématique actuelle, nous proposons une analyse comparative à travers trois scénarios génériques, confrontant les horizons d'intervention prévisionnels aux horizons systématiques.

La méthode prévisionnelle que nous proposons vise à éliminer ces 86 interventions urgentes en anticipant les défaillances des membranes grâce à l'estimation de leur durée de vie résiduelle décrite dans les chapitres précédents. Cette approche permettrait de programmer les remplacements avant que les défaillances ne surviennent, évitant ainsi tout impact sur les cadences de production. Cependant, pour définir ces seuils de manière précise et adaptée, il est nécessaire de confronter les prévisions du modèle aux pratiques actuelles de maintenance systématique.

Ci-dessous, nous décrivons 3 scénarios génériques plausibles comparant les performances de la maintenance prévisionnelle telle que proposée avec les paramètres suivants :

- T_{RUL} : Horizon calculé à partir de la détection du démarrage de la phase de vieillesse
- T_{vieil} : Détection du démarrage de la phase de vieillesse
- T_{sys} : Horizon déterminé par le plan de maintenance systématique
- Δ_{plan} : Quantification de la différence temporelle $T_{sys} - T_{RUL}$

Scénario 1 : La maintenance prévisionnelle intervient plus tôt que la systématique, figure 39.

- **Description** : La phase de vieillesse est détectée à T_{vieil} , et le RUL est calculé, définissant T_{RUL} comme la date d'intervention prévisionnelle. Dans ce scénario, T_{RUL} est inférieur à T_{sys} , indiquant que la maintenance prévisionnelle prévoit une intervention avant l'échéance systématique. L'écart Δ_{plan} est positif, montrant que T_{sys} est plus tardif.
- **Interprétation** : Ce scénario suggère que la maintenance prévisionnelle anticipe une dégradation potentielle plus tôt que le plan systématique. Si aucune défaillance ne survient avant T_{RUL} , cela peut indiquer une prédiction juste, mais une évaluation au démontage est nécessaire pour confirmer la nécessité de l'intervention.

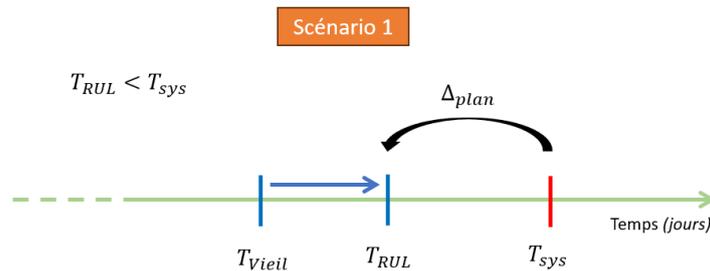


Figure 39 - Schéma du scénario 1 ($T_{RUL} < T_{sys}$)

Scénario 2 : Les maintenances prévisionnelle et systématique affichent des horizons similaires, figure 40.

- **Description** : La phase de vieillesse est détectée à T_{vieil} , et le RUL est calculé, définissant T_{RUL} . Dans ce scénario, T_{RUL} est approximativement égal à T_{sys} , indiquant que les deux approches prévoient une intervention au même moment. L'écart Δ_{plan} est proche de zéro. La figure montre une chronologie où T_{vieil} précède T_{RUL} et T_{sys} , qui coïncident.
- **Interprétation** : Ce scénario indique une convergence entre les deux approches, suggérant que la prédiction prévisionnelle est alignée avec l'intervalle systématique.

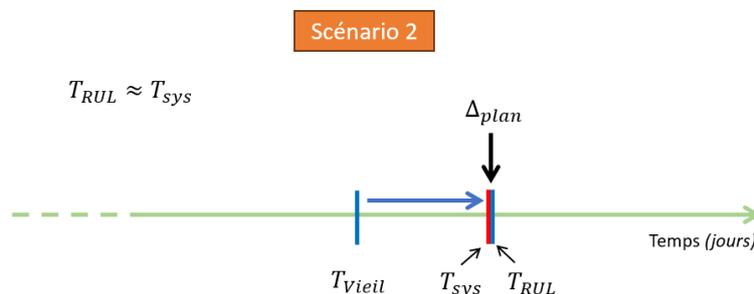


Figure 40 - Schéma du scénario 2 ($T_{RUL} \approx T_{sys}$)

Scénario 3 : La maintenance systématique intervient plus tôt que la prévisionnelle, figure 41.

- **Description** : La phase de vieillesse est détectée à T_{vieil} , et le RUL est calculé, définissant T_{RUL} . Dans ce scénario, T_{sys} est inférieur à T_{RUL} , indiquant que la

maintenance systématique impose une intervention avant la date prédite par la maintenance prévisionnelle. L'écart Δ_{plan} est négatif, montrant que T_{sys} est plus précoce. La figure illustre une chronologie où T_{vieil} précède T_{sys} , qui lui-même précède T_{RUL} .

- **Interprétation** : Ce scénario suggère que la maintenance systématique est plus conservatrice, intervenant avant la prédiction prévisionnelle. Si aucune défaillance ne survient avant T_{sys} , cela peut indiquer que l'intervention systématique était prématurée, mais une défaillance avant T_{sys} révélerait une sous-estimation par les deux approches.

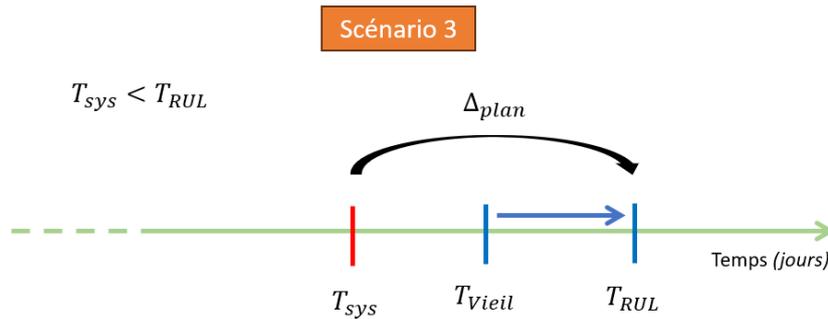


Figure 41 - Schéma du scénario 3 ($T_{sys} < T_{RUL}$)

Ces scénarios peuvent servir de base pour décrire un système d'optimisation du plan de maintenance prévisionnel proposé.

5.3.3. Validation de l'Approche Hybride

Ces scénarios, en évaluant l'écart temporel entre les deux approches, permettent d'identifier les conditions dans lesquelles une intervention prévisionnelle devient pertinente, fournissant ainsi une base concrète pour valider l'approche hybride proposée.

5.3.3.1 Fiabilité et validation du système :

La fiabilité d'un nouveau système de maintenance, tel que celui proposé pour les membranes de vannes dans un contexte biopharmaceutique, repose sur sa capacité à prévoir avec précision le RUL des composants, tout en minimisant les défaillances imprévues. Dans notre cas, le système étant basé sur des données expérimentales, certaines incertitudes vont peser sur sa performance initiale. Pour évaluer sa fiabilité, une approche de déploiement parallèle est proposée, où le système prévisionnel coexiste avec le plan de maintenance systématique existant lors d'un premier cycle complet d'utilisation des équipements cibles.

Le plan systématique impose des limites fixes (ex. : remplacement tous les 3 ans), tandis que le plan prévisionnel utilise un modèle de dégradation pour prédire le RUL de manière dynamique (ex. : ~57 jours restants selon les conditions de procédé). En déployant les deux plans en parallèle, les prédictions du modèle prévisionnel peuvent être confrontées aux conditions réelles, permettant d'identifier les biais et d'évaluer la fiabilité du système prévisionnel.

5.3.3.2. Stratégie de validation de maintenance proposée

Cette stratégie hybride s'appuie à la fois sur les horizons déterminés par le plan systématique et sur la détection précoce de la phase de vieillesse des membranes, pour déclencher

l'émission d'un pronostic RUL. Pour valider concrètement cette approche chez OCTAPHARMA SAS tout en respectant les contraintes opérationnelles et réglementaires, une méthodologie différenciée est proposée selon la criticité des équipements :

Équipements critiques :

La maintenance systématique reste la doctrine de référence. Ces vannes, dont les membranes sont en contact direct avec le produit, sont remplacées tous les trois ans selon le plan actuellement en vigueur. Cela permet à l'entreprise, durant la phase de transition vers une maintenance prévisionnelle, de maintenir ses standards qualité.

Le recours à un modèle prévisionnel reste complémentaire : il estime un RUL à partir de la détection de la phase de vieillesse (T_{vieil}), mais sa pertinence est jugée a posteriori, au moment du remplacement imposé par le plan systématique. C'est la confrontation entre l'horizon prédit (T_{RUL}) et le remplacement effectif (T_{sys}) qui permet d'évaluer la pertinence du modèle.

- Si T_{RUL} est inférieur à T_{sys} (scénario 1), une dégradation marquée observée sur l'équipement lors du démontage vient conforter la prédiction.
- Si T_{RUL} est proche de T_{sys} (scénario 2), et que l'inspection révèle une usure avancée, cela indique que la planification actuelle recoupe les prévisions, renforçant l'intérêt du modèle sans modifier la stratégie en place.
- Si T_{RUL} est supérieur à T_{sys} (scénario 3) et que l'état de l'équipement s'avère encore satisfaisant, cela suggère que la maintenance systématique pourrait être optimisée, bien que, dans le cas des critiques, toute modification reste conditionnée à des exigences de sûreté et de qualité strictes.

Ainsi, dans ce cadre, c'est bien la stratégie systématique qui prévaut, et la modélisation est analysée à travers elle.

Équipements non critiques :

Ici, la maintenance prévisionnelle devient la doctrine principale. Le modèle guide la prise de décision, et le remplacement est déclenché lorsque le RUL estimé arrive à échéance. Ce remplacement est idéalement intégré à une fenêtre d'arrêt planifié, ou effectué en-dehors des périodes de production sous forme de correctif planifié lorsque nécessaire. Dans ce cadre, la pertinence du modèle est évaluée à partir de la confrontation entre la prévision et l'état réel de l'équipement au moment du remplacement.

Si l'analyse post-intervention révèle une dégradation conforme à celle attendue (usure visible, membranes altérées, tests mécaniques ou visuels en accord avec une fin de vie), alors le modèle est validé dans sa capacité à anticiper correctement l'évolution du composant. Cette convergence entre le modèle et la réalité observée conforte la fiabilité du dispositif prévisionnel, justifie la décision de remplacement, et alimente la confiance dans la stratégie de maintenance conditionnée par le RUL. Elle permet également de renforcer l'acceptabilité terrain de ces approches, en montrant que la modélisation s'inscrit dans une logique opérationnelle tangible.

Si, en revanche, l'équipement démontre un état de santé divergent, comme :

- Soit en meilleur état que prévu avec une usure minimale malgré un RUL jugé critique,
- Soit à l'inverse en état de dégradation avancée alors que le modèle estimait une durée de vie résiduelle encore significative,

Cela met en évidence une inadéquation entre la prédiction et le comportement réel. Un recalibrage du modèle devient alors nécessaire. Celui-ci peut passer par une reconsidération des signaux utilisés, un ajustement des seuils de déclenchement, ou une meilleure prise en compte des contextes d'utilisation spécifiques (par exemple, des cycles de production atypiques ou des conditions environnementales particulières). Cette étape de réajustement peut s'avérer essentielle dans une logique d'amélioration continue du dispositif. Ce retour d'expérience permet d'affiner progressivement la précision des prévisions. Ainsi, pour ces équipements, la maintenance prévisionnelle prévaut et le retour terrain, via les observations effectuées a posteriori, joue un rôle de validation et d'amélioration continue du dispositif prédictif.

L'ensemble de ces éléments est récapitulé dans la figure 42. Ce logigramme expose concrètement quelle forme pourrait prendre ce processus de validation du modèle de maintenance prévisionnelle proposé.

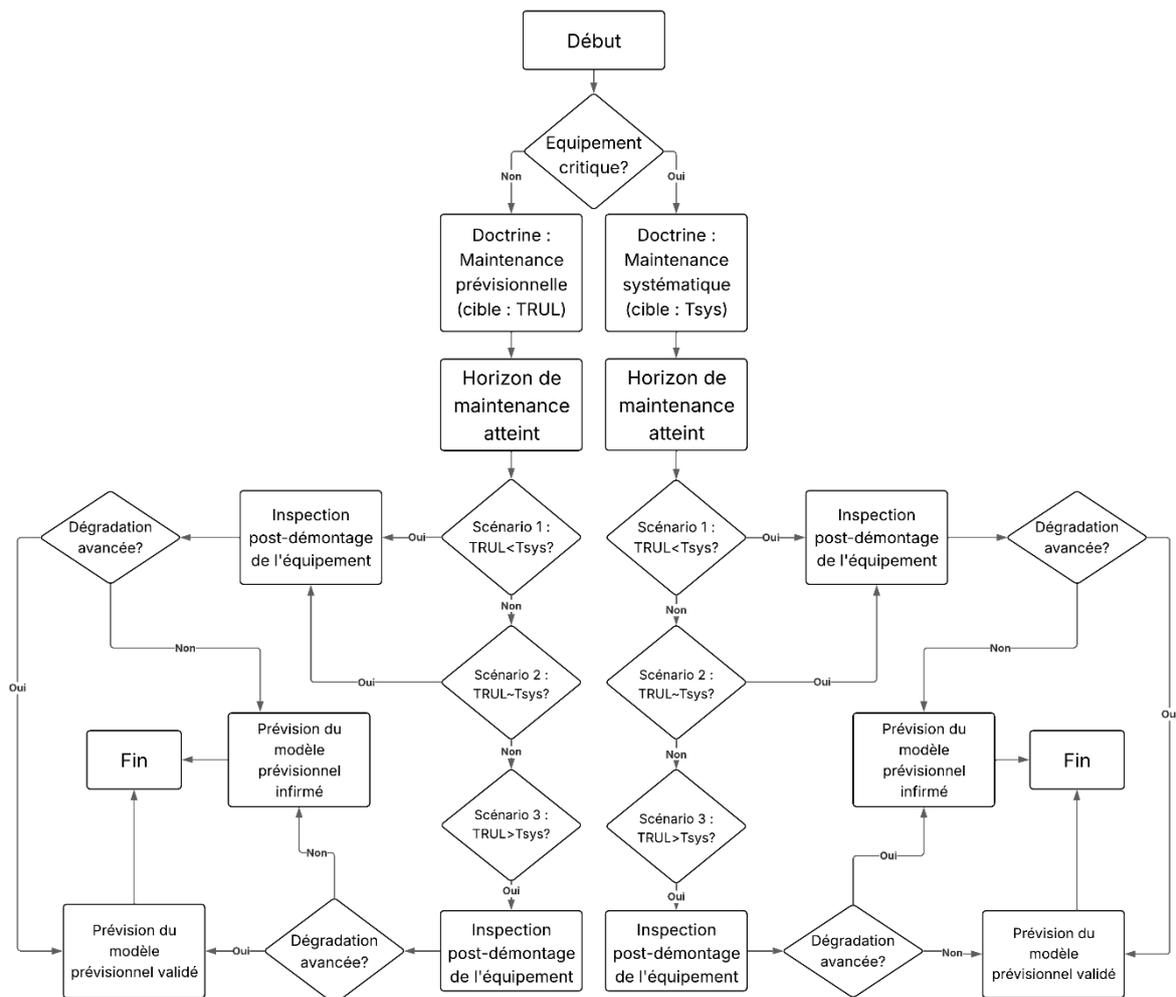


Figure 42 - Stratégie de maintenance hybride orientée en fonction de la criticité des équipements

Dans la section suivante, nous allons proposer un système d'aide à la décision tenant compte à la fois des éléments discutés pour la validation et des contraintes réglementaires et économiques propres à l'entreprise.

5.4. Aide à la décision

Dans cette section, nous proposons une première application concrète d'aide à la décision qui serait une perspective de déploiement d'un plan prévisionnel suffisamment fiable pour s'affranchir des horizons de maintenance systématique.

5.4.1. Algorithme d'affectation des interventions

Une fois l'étape de validation du plan de maintenance prévisionnel et dans l'hypothèse où son efficacité est effectivement constatée, nous proposons une méthode de prise à la décision. L'algorithme développé et proposé en figure 43 vise à optimiser la planification des interventions de maintenance des membranes de vannes en prenant en compte à la fois leur criticité et leur impact business. Ce dernier, dans notre cas d'usage, correspond aux conséquences économiques d'une interruption ou d'une défaillance à un moment donné du cycle de fabrication. Plus le produit progresse dans le processus et gagne en valeur ajoutée (matières premières transformées, main d'œuvre mobilisée, énergie consommée, etc.), plus une défaillance ou un arrêt à ce stade engendre des pertes financières importantes. Ainsi, l'impact business augmente avec la valorisation progressive du produit. Une interruption en fin de chaîne de production est généralement bien plus coûteuse qu'en amont, car elle compromet un produit déjà presque finalisé, avec un investissement cumulé bien plus élevé.

Le processus débute par l'obtention des RUL des membranes de vannes, comme décrit dans les chapitres précédents. Cette estimation permet d'identifier les composants dont la probabilité de défaillance est forte.

Ensuite, l'algorithme identifie les créneaux de maintenance disponibles dans le calendrier d'exploitation. Ces créneaux, qui peuvent correspondre à des périodes d'arrêt programmé ou de moindre charge, sont triés par ordre chronologique afin de garantir une intervention aussi anticipée que possible. L'existence de créneaux exploitables est alors vérifiée. Si aucun créneau n'est disponible, l'algorithme interrompt l'exécution et reporte la décision à une itération ultérieure. Il s'agira dans ce cas de figure d'émettre une alerte et de revoir potentiellement les ordonnancements des opérations de maintenance.

Dans le cas contraire, les membranes à risque sont classées selon deux critères principaux : leur criticité et leur impact business. Ce classement permet de prioriser les interventions sur les éléments dont la défaillance entraînerait les conséquences les plus graves, que ce soit en termes de sécurité, de conformité réglementaire ou de pertes financières. Le premier créneau disponible est alors attribué à la membrane présentant la plus forte combinaison de criticité et d'impact business. La maintenance est programmée en conséquence.

Après chaque attribution, l'algorithme vérifie s'il reste encore des membranes à traiter et des créneaux disponibles. Tant que ces deux conditions sont réunies, le processus se répète de manière itérative. Cette boucle permet de traiter l'ensemble des priorités identifiées dans la limite des ressources temporelles disponibles. Si l'une de ces conditions n'est plus remplie, absence de membranes critiques restantes ou saturation du planning, les interventions non planifiées sont reportées pour une future itération.

Ainsi, ce processus permet d'assurer une allocation dynamique des ressources de maintenance, en privilégiant la prévention des défaillances les plus impactantes dans notre cadre opérationnel.

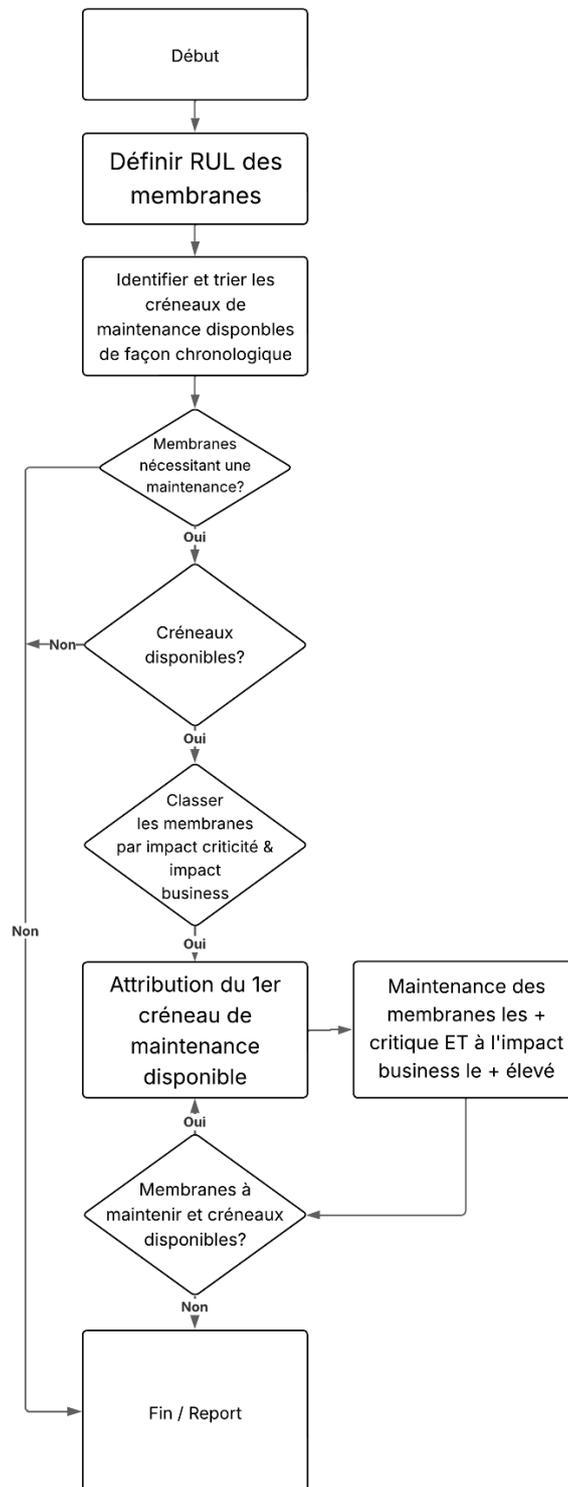


Figure 43 - Proposition de système d'aide à la décision basé sur impacts criticité & business

Ce premier algorithme, figure 43, conçu pour les membranes de vannes, illustre l'intérêt d'une approche structurée et priorisée de la maintenance prévisionnelle dans un environnement

critique. Fort de cette démonstration, il devient pertinent de réfléchir à élargir la portée de la méthode.

5.5. Généralisation du modèle

Dans une perspective de généralisation de l'approche développée précédemment pour la maintenance prévisionnelle chez notre partenaire et appliquée sur les membranes, nous proposons ici un modèle générique extensible à d'autres équipements, critiques ou non, d'un procédé industriel fortement contraint. Ce modèle a pour objectif de guider la planification dynamique des interventions de maintenance en s'appuyant sur une logique adaptable, tenant compte des spécificités de chaque équipement, de leur rôle dans la chaîne de production et de leur impact économique potentiel en cas de défaillance.

1. La première étape de cette approche consiste à identifier un paramètre significatif de dégradation, possiblement propre à chaque type d'équipement. Contrairement à une approche fondée sur des critères standards tels que le temps de fonctionnement ou le nombre de cycles, cette stratégie repose sur un suivi contextualisé de l'état réel de l'équipement, à travers un indicateur représentatif de son usure ou de son vieillissement. Ce paramètre peut varier considérablement d'un équipement à un autre : pour une pompe, il pourra s'agir de la consommation énergétique ou de la température d'échauffement ; pour un moteur, de la vitesse de réponse ou des vibrations mécaniques ; pour une vanne, d'un signal de position. Ce choix repose sur une analyse croisée entre l'expertise terrain, les retours d'expérience, et les historiques de données disponibles.
2. Une fois ce paramètre déterminé, une modélisation de la dégradation est mise en œuvre. Elle vise à décrire l'évolution temporelle ou fonctionnelle du paramètre en lien avec l'état de santé de l'équipement. Cette modélisation peut reposer sur différents niveaux de sophistication, allant de méthodes empiriques basées sur des seuils à des approches probabilistes ou d'apprentissage automatique. L'objectif est d'aboutir à une estimation fiable du RUL, c'est-à-dire la durée, en unités de temps, de cycles ou d'usage, avant que l'équipement ne franchisse un seuil de dégradation critique au-delà duquel le risque de défaillance devient inacceptable. Le RUL devient ainsi le pivot du pilotage anticipé de la maintenance.
3. Sur la base de cette information, l'étape suivante consiste à intégrer l'estimation du RUL dans une stratégie de priorisation des interventions de maintenance, en mobilisant des critères techniques, économiques et organisationnels. L'un des axes structurants de cette priorisation est l'évaluation de la criticité fonctionnelle de l'équipement : cette notion renvoie à son rôle dans le procédé (redondance, impossibilité de contournement, liens de causalité avec d'autres équipements) et à la gravité des conséquences de sa défaillance. Parallèlement, le modèle intègre un impact business dynamique, c'est-à-dire une mesure des pertes potentielles liées à une interruption ou une défaillance en fonction du moment où celle-ci intervient dans le cycle de production. Dans les environnements à forte valeur ajoutée, la valeur du produit tend à croître à mesure qu'il progresse dans le processus : ainsi, une défaillance en fin de chaîne engendre des coûts généralement plus élevés qu'en début de production. Cette variabilité de l'impact business est un autre élément central du modèle, car elle justifie une approche différenciée et contextualisée de la maintenance. Tout ceci nous amène

à un 1^{er} processus de validation générique de l'approche prévisionnelle décrit dans la figure 44.

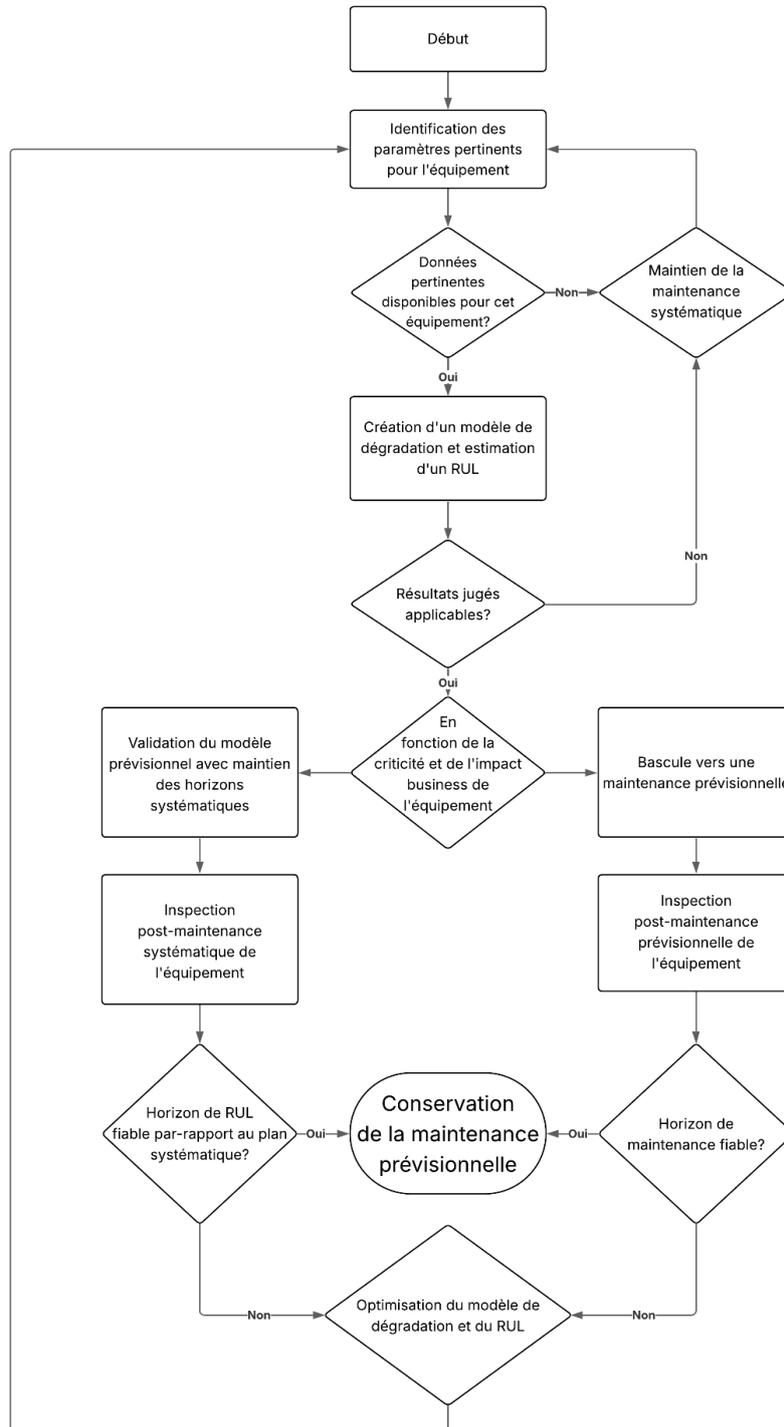


Figure 44 - Processus de validation générique de l'approche prévisionnelle

Tenant compte de ces éléments, une logique de décision visant à recommander les interventions les plus pertinentes est établie (figure 45) en tenant compte du RUL, de la criticité, de l'impact économique estimé et, si besoin, des contraintes de disponibilité des ressources (équipes, créneaux d'intervention, pièces de rechange). L'aide à la décision devient ainsi défini pour fonctionner de manière itérative, en réattribuant en continu les priorités

à partir des nouvelles données de surveillance collectées sur le terrain. Il constitue un outil d'aide à la décision robuste et adaptable, permettant de dépasser les approches de maintenance préventive systématique, au profit d'un pilotage dynamique, rationalisé, et économiquement optimisé de la maintenance.

Tenant compte de ces éléments, une logique de décision visant à prioriser les interventions est établie (figure 45) en tenant compte du RUL, de la criticité, de l'impact économique estimé et, si besoin, des contraintes de disponibilité des ressources (équipes, créneaux d'intervention, pièces de rechange). L'aide à la décision devient ainsi défini pour fonctionner de manière itérative, en réattribuant en continu les priorités à partir des nouvelles données de surveillance collectées sur le terrain. Il constitue un outil d'aide à la décision robuste et adaptable, permettant de dépasser les approches de maintenance préventive systématique, au profit d'un pilotage dynamique, rationalisé, et économiquement optimisé de la maintenance.

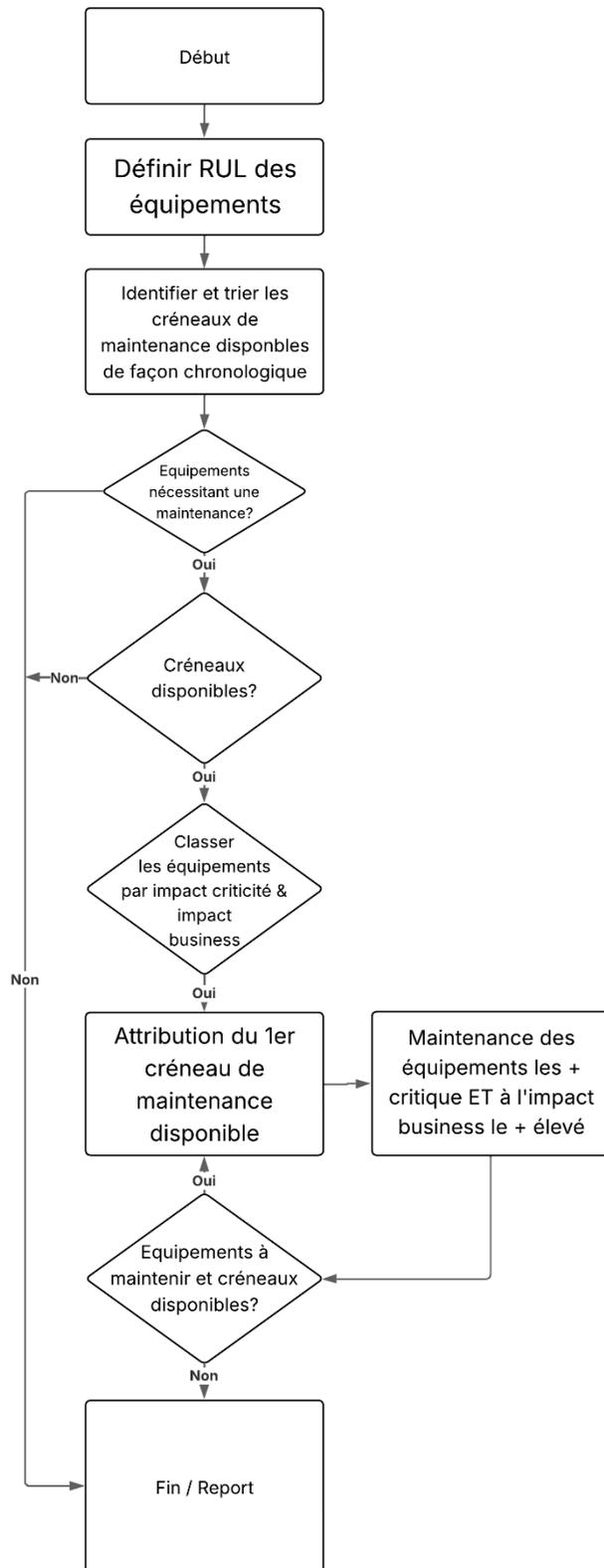


Figure 45 - Processus de système d'aide à la décision générique basé sur impacts criticité & business

5.6. Conclusion du chapitre 5

Dans ce chapitre, nous avons discuté du déploiement concret de l'approche de maintenance prévisionnelle développée dans le cadre de cette thèse, en réponse aux besoins exprimés par OCTAPHARMA SAS. En s'appuyant sur les résultats des chapitres précédents, nous avons abordé la validation de l'approche hybride et proposé une méthode d'aide à la décision, tout en explorant les perspectives de généralisation à d'autres équipements industriels.

La section 5.3, dédiée à la validation de l'approche hybride, a proposé une stratégie différenciée selon la criticité des équipements. Pour les équipements critiques, la maintenance systématique resterait la doctrine de référence, avec une évaluation a posteriori de la pertinence des prédictions du modèle prévisionnel via la confrontation des RUL estimés et des observations réelles. Pour les équipements non critiques, le modèle prévisionnel guide directement les interventions, avec une validation basée sur l'analyse post-intervention des membranes, permettant d'identifier les écarts entre prévisions et réalité. Un système de recalibrage continu du modèle peut être envisagé en perspectives. Cette méthodologie, illustrée par le logigramme de la figure 42, propose une transition progressive vers la maintenance prévisionnelle tout en respectant les exigences réglementaires et opérationnelles.

L'aide à la décision, détaillée dans la section 5.4, se matérialise par un algorithme d'affectation des interventions (figure 43) qui priorise les remplacements des membranes en fonction de leur criticité et de leur impact économique. En intégrant les RUL estimés, les créneaux d'arrêt planifiés, et l'impact business croissant avec la valeur ajoutée du produit, cet algorithme propose une planification dynamique.

Enfin, la section 5.5 propose un élargissement de la portée de ces travaux en proposant un modèle générique de maintenance prévisionnelle, extensible à d'autres équipements industriels. En identifiant des paramètres d'intérêt, en modélisant leur évolution, et en intégrant des critères de criticité et d'impact économique, ce modèle (illustré par les figures 44 et 45) offre une logique itérative d'aide à la décision.

En conclusion, ce chapitre étend les contributions de la thèse en contextualisant l'approche hybride proposée, en opérationnalisant l'aide à la décision, et en posant les bases d'une généralisation à d'autres contextes industriels.

6. Conclusion générale

Cette thèse s'est attachée à répondre à trois questions scientifiques pour accompagner l'évolution des pratiques de maintenance dans l'industrie biopharmaceutique, en ciblant comme cas d'usage les vannes à membranes pneumatiques et en s'alignant sur les principes de l'Industrie 4.0. Ces questions, liées aux objectifs industriels d'OCTAPHARMA SAS, ont structuré une démarche scientifique, dont les résultats sont présentés au travers des chapitres de ce manuscrit.

La première question scientifique visait à caractériser les pratiques de maintenance dans le secteur du biomédicament et à évaluer l'adoption des principes de l'Industrie 4.0, notamment la maintenance prévisionnelle. Le premier chapitre a répondu à cette question en établissant un état de l'art basé sur une enquête auprès d'industriels du biomédicament et une revue de littérature scientifique. Ces travaux ont révélé un retard significatif pour ce qui est de la maintenance prévisionnelle, IOT-driven, dans ce secteur et notamment par rapport à l'industrie manufacturière générale. Ce constat a non seulement positionné OCTAPHARMA SAS parmi d'autres acteurs confrontés à des contraintes similaires, mais a également mis en évidence une lacune dans la recherche, offrant une opportunité pour une contribution scientifique. Ainsi, ce chapitre a répondu au premier objectif industriel en proposant un cadre pour l'évolution des pratiques de maintenance.

La deuxième question scientifique portait sur l'établissement d'un modèle de dégradation des membranes de vannes pneumatiques, en s'appuyant sur la méthode Prognostics Health Management (PHM) pour permettre une maintenance prévisionnelle. Le deuxième chapitre a abordé cette question en développant une approche basée sur l'analyse de données représentatives, issues d'une architecture fonctionnelle d'acquisition de données d'OCTAPHARMA SAS. En étudiant la dérive des valeurs de position, un modèle de dégradation a été construit, mettant en évidence les phases de vie des membranes et leur corrélation avec les déformations physiques observées. Cette contribution, conforme aux quatre premières briques du PHM, a répondu au deuxième objectif industriel en fournissant un modèle comportemental et une architecture technique concrète pour l'intégration d'un plan de maintenance prévisionnelle.

La troisième question scientifique concernait les pronostics de défaillance à partir du modèle de dégradation, pour en déduire une méthode d'aide à la décision pour transformer ces prévisions en actions opérationnelles. Le troisième chapitre a répondu à cette question en exploitant les données et les anomalies identifiées précédemment. Un modèle de pronostic a été développé à l'aide de techniques de Machine Learning, notamment RandomForestClassifier et XGBRegressor, pour estimer la durée de vie résiduelle des membranes. Ce chapitre a également exploré des approches data-driven, supervisées et non supervisées, pour classifier les anomalies et optimiser les performances du modèle. Enfin, le quatrième chapitre a complété cette réponse en proposant un plan de maintenance hybride, combinant approches systématique et prévisionnelle. En confrontant ce plan aux pratiques actuelles d'OCTAPHARMA SAS, il a démontré une opportunité d'amélioration de la fiabilité des équipements et une réduction des coûts et des arrêts non planifiés, répondant ainsi au troisième objectif industriel.

Enfin, nos travaux proposent des possibilités d'extension de ces approches à d'autres secteurs fluidiques, tout en consolidant l'alignement du secteur du biomédicament avec les paradigmes

de l'Industrie 4.0. Cette généralisation repose sur plusieurs leviers méthodologiques développés tout au long de cette recherche : la structuration des données représentatives de paramètres significatifs, la modélisation des signatures de dégradation spécifiques à des composants industriels, et l'implémentation d'architectures modulaires pouvant être adaptées à différentes topologies d'équipements. En particulier, la flexibilité des modèles data-driven employés permet d'envisager leur requalification sur d'autres organes critiques soumis à des contraintes similaires. Cette démarche ouvre la voie à une approche standardisée de la maintenance prévisionnelle dans les environnements contraints, tout en favorisant l'essaimage des pratiques vers d'autres industries régulées ou sensibles aux défaillances, telles que l'agroalimentaire, la chimie fine ou encore les raffineries par exemple.

7. Perspectives

Les perspectives de cette thèse sont multiples et sont proposées structurées par chapitre, hors état de l'art, de sorte à consolider l'ensemble des éléments apportés par ces travaux.

Le chapitre 3 a établi un modèle de dégradation des membranes de vannes pneumatiques basé sur l'analyse de la dérive des valeurs de position, en s'appuyant sur la méthode Prognostics Health Management (PHM). Bien que ce modèle ait démontré sa pertinence dans les conditions expérimentales décrites dans ce chapitre, son applicabilité reste limitée aux vannes testées et aux contextes de pilotage étudiés. Une perspective consiste à consolider ce modèle en l'étendant à d'autres types de vannes pneumatiques. Cette extension nécessiterait un contingent d'étude plus large, incluant des vannes de diverses conceptions (matériaux, tailles, mécanismes) et soumises à des conditions de pilotage variées, telles que des pressions, températures ou compositions des fluides distinctes. De plus, l'étude de cadences de pilotage différenciées (par exemple, cycles rapides vs. cycles lents, fonctionnement intermittent vs. continu) permettrait d'enrichir le modèle en capturant une gamme plus large de comportements de dégradation. Ces travaux renforceraient son potentiel d'intégration dans des environnements industriels diversifiés, tout en consolidant son alignement avec les exigences de l'Industrie 4.0.

Le chapitre 4 propose un modèle de pronostic pour estimer la durée de vie résiduelle des membranes de vannes pneumatiques, en exploitant des techniques de Machine Learning, notamment RandomForestClassifier et XGBRegressor. Si les performances de classification des anomalies et de prévision du RUL ont été validées dans les conditions étudiées, la présence d'événements ponctuels non impactant, tels que des arrêts temporaires de pilotage, des maintenances de la ligne de production, ou des variations environnementales, peut introduire des signaux non-conformes faussement interprétés. Une perspective consiste à étendre la pertinence des sorties des algorithmes en intégrant ces événements dans les données d'entraînement. Cela impliquerait de collecter et d'annoter des ensembles de données incluant ces perturbations, afin d'entraîner les modèles à distinguer les signaux liés à des dégradations réelles de ceux induits par des conditions opérationnelles transitoires. Une telle approche renforcerait la confiance dans les pronostics et faciliterait leur adoption dans des environnements industriels où les interruptions sont fréquentes. Cette optimisation pourrait ouvrir la voie à une application plus large dans des contextes industriels variés.

Le chapitre 5 a proposé une approche hybride de maintenance et décrit un algorithme d'aide à la décision basé sur la criticité des équipements et leur impact économique, tout en esquissant un modèle générique extensible à d'autres équipements. Pour consolider ces avancées, une perspective à ces travaux consiste à opérer un retour d'expérience systématique sur les modèles d'aide à la décision déployés chez OCTAPHARMA SAS. Ce retour d'expérience permettrait d'évaluer la pertinence des priorisations effectuées (par exemple, l'adéquation des interventions planifiées avec les états réels des membranes) et d'identifier leurs éventuelles limites. En complément, une méthode d'optimisation de la prise de décision pourrait être développée par l'établissement d'un indicateur fiable de l'état des membranes, relevé au démontage post-maintenance. Cet indicateur permettrait de quantifier l'adéquation entre les prédictions du modèle et l'état réel des composants. En intégrant ces données dans une boucle d'amélioration continue, l'algorithme d'aide à la décision pourrait être affiné pour optimiser la planification des interventions, réduire les coûts, et renforcer la fiabilité des équipements, notamment critiques. Cette approche, en s'appuyant sur des retours

terrain concrets, favoriserait également l'acceptabilité des solutions prévisionnelles par les opérateurs et les décideurs. Enfin, une analyse multicritère pourrait être intégrée à l'algorithme afin d'affiner la hiérarchisation des actions de maintenance. Cette approche permettrait d'identifier, de formaliser et de pondérer les différents critères influençant la décision, tels que la fréquence des anomalies, la durée de vie résiduelle estimée, les coûts d'arrêt, la disponibilité des pièces, ou encore les enjeux sanitaires ou réglementaires. Une telle analyse enrichirait la logique décisionnelle du modèle en y intégrant la complexité réelle des arbitrages industriels.

Ces perspectives, en prolongeant les contributions des chapitres 3, 4 et 5, visent à renforcer l'ensemble des travaux déroulés dans cette thèse. En consolidant le modèle de dégradation pour d'autres vannes à membranes et en optimisant l'aide à la décision grâce à des retours d'expérience et des indicateurs fiables, nous ambitionnons d'ouvrir la voie à une adoption élargie des principes de l'Industrie 4.0 dans l'industrie biopharmaceutique et au-delà. Ces pistes, en phase avec les objectifs industriels d'OCTAPHARMA SAS, pourraient également bénéficier à d'autres secteurs fluidiques, contribuant à une gestion de la maintenance plus dynamique, économique et durable.

8. Références

Ahn, J.; Lee, Y.; Kim, N.; Park, C.; Jeong, J. Federated Learning for Predictive Maintenance and Anomaly Detection Using Time Series Data Distribution Shifts in Manufacturing Processes. *Sensors* 2023, 23, 7331.

<https://doi.org/10.3390/s23177331>

Akırmak, O. O., & Altan, A. (2023). Estimation of extrusion process parameters in tire manufacturing industry using random forest classifier. *Balkan Journal of Electrical and Computer Engineering*, 11(2), 138-143.

<https://doi.org/10.17694/bajece.1232811>

Amro, A., Al-Akhras, M., Hindi, K., Habib, M., & Shawar, B. (2021). Instance Reduction for Avoiding Overfitting in Decision Trees. *Journal of Intelligent Systems*, 30(1), 438-459.

<https://doi.org/10.1515/jisys-2020-0061>

Andrade, A.; Lopes, K.; Lima, B.; Maitelli, A. Development of a Methodology Using Artificial Neural Network in the Detection and Diagnosis of Faults for Pneumatic Control Valves. *Sensors* 2021, 21, 853.

<https://doi.org/10.3390/s21030853>

Atassi, R., & Alhosban, F. (2023). Predictive Maintenance in IoT: Early Fault Detection and Failure Prediction in Industrial Equipment. *Journal of Intelligent Systems & Internet of Things*, 9(2).

<https://doi.org/10.54216/JISIoT.090217>

Bammou, Y., Benzougagh, B., Abdessalam, O., Brahim, I., Kader, S., Spalevic, V., ... & Ercişli, S. (2024). Machine learning models for gully erosion susceptibility assessment in the Tensift catchment, Haouz Plain, Morocco for sustainable development. *Journal of African Earth Sciences*, 213, 105229.

<https://doi.org/10.1016/j.jafrearsci.2024.105229>

Bentéjac, C., Csörgő, A. & Martínez-Muñoz, G. A comparative analysis of gradient boosting algorithms. *Artif Intell Rev* 54, 1937–1967 (2021).

<https://doi.org/10.1007/s10462-020-09896-5>

Bergstra, J., & Bengio, Y. (2012). Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*, 13, 281-305, ISSN: 1532-4435

Biau, G., Scornet, E. A random forest guided tour. *TEST* 25, 197–227 (2016).

<https://doi.org/10.1007/s11749-016-0481-7>

Bisong, E. (2019). Introduction to Scikit-learn. In *Building machine learning and deep learning models on google cloud platform: a comprehensive guide for beginners* (pp. 215-229). Berkeley, CA: Apress.

https://doi.org/10.1007/978-1-4842-4470-8_18

Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32.

<https://doi.org/10.1023/A:1010933404324>

Breiman, L. Bagging predictors. *Mach Learn* 24, 123–140 (1996).
<https://doi.org/10.1007/BF00058655>

Cabal-Yepez, E., Garcia-Ramirez, A. G., Romero-Troncoso, R. J., Garcia-Perez, A., & Osornio-Rios, R. A. (2012). Reconfigurable monitoring system for time-frequency analysis on industrial equipment through STFT and DWT. *IEEE transactions on industrial informatics*, 9(2), 760-771.
<https://doi.org/10.1109/TII.2012.2221131>

Carvalho JN, Silva FRD, Nascimento EGS. Challenges of the Biopharmaceutical Industry in the Application of Prescriptive Maintenance in the Industry 4.0 Context: A Comprehensive Literature Review. *Sensors (Basel)*. 2024 Nov 7;24(22):7163.
<https://doi.org/10.3390/s24227163>

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, 408, 189-215.
<https://doi.org/10.1016/j.neucom.2019.10.118>

Chen, Tianqi and Guestrin, Carlos, XGBoost: A Scalable Tree Boosting System, 2016, Association for Computing Machinery, New York, NY,
<https://doi.org/10.1145/2939672.2939785>

Chen, Y., Qiu, J., Wang, M., Rao, J., Xia, T., & Yang, Y. (2024). Fault prediction of pneumatic valves in an LNG plant by the DGM (1, 1) model. *International Journal of System Assurance Engineering and Management*, 15(3), 775-785.
<https://doi.org/10.1007/s13198-023-02130-9>

Cheng, C. F., Rashidi, A., Davenport, M. A., & Anderson, D. (2016). Audio signal processing for activity recognition of construction heavy equipment. In *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction* (Vol. 33, p. 1). IAARC Publications.
<https://doi.org/10.22260/ISARC2016/0078>

Chinta, V. S., Reddi, S. K., & Yarramsetty, N. (2023). Optimal feature selection on Serial Cascaded deep learning for predictive maintenance system in automotive industry with fused optimization algorithm. *Advanced Engineering Informatics*, 57, 102105.
<https://doi.org/10.1016/j.aei.2023.102105>

Choi, W.-H., & Kim, J. (2024). Unsupervised Learning Approach for Anomaly Detection in Industrial Control Systems. *Applied System Innovation*, 7(2), 18.
<https://doi.org/10.3390/asi7020018>

Conte, M., Igartua, A. Study of PTFE composites tribological behavior, *Wear*, Volume 296, Issues 1–2, 2012, Pages 568-574, ISSN 0043-1648,
<https://doi.org/10.1016/j.wear.2012.08.015>

Czmochowski, J., Moczko, P., & Pietrusiak, D. Challenges and Strategies of Long-life Operation and Maintenance of Technical Objects.

<https://doi.org/10.5937/fmet1603219R>

Daigle, M. J., & Goebel, K. (2011). A model-based prognostics approach applied to pneumatic valves. *International Journal of Prognostics and Health Management* Volume 2 (color), 84.

<https://doi.org/10.36001/ijphm.2011.v2i2.1359>

Dhanumalayan, E., Joshi, G.M. Performance properties and applications of polytetrafluoroethylene (PTFE)—a review. *Adv Compos Hybrid Mater* 1, 247–268 (2018).

<https://doi.org/10.1007/s42114-018-0023-8>

Drissi, A., lung, B., Voisin, A., Galinier, V. Coupling Prognostics and Decision Making (P&DM) Processes in PHM: A literature review and proposal of a Residual Performance Lifetime concept, *IFAC-PapersOnLine*, Volume 55, Issue 19, 2022, Pages 145-150, ISSN 2405-8963,

<https://doi.org/10.1016/j.ifacol.2022.09.198>

Fai, N. J., Wey, W. K., Qi, K. Y., Xian, G. J., Chun, R. J. M., & bin Abdul Salam, Z. A. (2023). Digits Classification Using Random Forest Classifier. *Journal of Applied Technology and Innovation (e-ISSN: 2600-7304)*, 7(3), 63. e-ISSN: 2600-7304

Fantana, Nicolaie & Riedel, Till & Schlick, Jochen & Ferber, Stefan & Hupp, Jürgen & Miles, Stephen & Michahelles, Florian & Svensson, Stefan. (2013). Internet of Things - Converging Technologies for Smart Environments and Integrated Ecosystems.

<https://doi.org/10.1201/9781003338659>

Feng, S., Zhong, Z., Wang, Y., Xing, W., Drioli, E. Progress and perspectives in PTFE membrane: Preparation, modification, and applications, *Journal of Membrane Science*,

<https://doi.org/10.1016/j.memsci.2017.12.032>

Fink, O., Wang, Q., Svensén, M., Dersin, P., Lee, W.-J., Ducoffe, M. Potential, challenges and future directions for deep learning in prognostics and health management applications, *Engineering Applications of Artificial Intelligence*, Volume 92, 2020, 103678, ISSN 0952-1976,

<https://doi.org/10.1016/j.engappai.2020.103678>

Gibb, K. A., Eldridge, A., Sandom, C. J., & Simpson, I. J. (2024). Towards interpretable learned representations for ecoacoustics using variational auto-encoding. *Ecological Informatics*, 80, 102449.

<https://doi.org/10.1016/j.ecoinf.2023.102449>

Gouriveau. (2013). PHM–Prognostics and health management-De la surveillance au pronostic de défaillances de systèmes complexes. *Editions TI*.

<https://doi.org/10.51257/a-v1-mt9570>

Han, X., Jiang, J., Xu, A., Huang, X., Pei, C., Sun, Y. "Fault Detection of Pneumatic Control Valves Based on Canonical Variate Analysis," in *IEEE Sensors Journal*, vol. 21, no. 12, pp. 13603-13615, 15 June 15, 2021,

<https://doi.org/10.1109/JSEN.2021.3070035>

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

<https://doi.org/10.1007/978-0-387-84858-7>

Heynicke, R., Krush, D., Cammin, C., Scholl, G., Kaercher, B., Ritter, J., Gaggero, P., and Rentschler, M.: IO-Link Wireless enhanced factory automation communication for Industry 4.0 applications, *J. Sens. Syst.* 7, 131–142,

<https://doi.org/10.5194/jsss-7-131-2018>, 2018

Hosseinzadeh, A., Chen, F. F., Shahin, M., & Bouzary, H. (2023). A predictive maintenance approach in manufacturing systems via AI-based early failure detection. *Manufacturing Letters*, 35, 1179-1186.

<https://doi.org/10.1016/j.mfglet.2023.08.125>

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2020). *Applied Logistic Regression* (3th ed.). Wiley. ISBN: 978-1-119-52481-6

Jambol, D. D., Sofoluwe, O. O., Ukato, A., & Ochulor, O. J. (2024). Transforming equipment management in oil and gas with AI-Driven predictive maintenance. *Computer Science & IT Research Journal*, 5(5), 1090-1112.

<https://doi.org/10.51594/csitrj.v5i5.1117>

Jardine, A. K. S., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*.

<https://doi.org/10.1016/j.ymsp.2005.09.012>

Kamiński, B., Jakubczyk, M., & Szufel, P. (2018). A framework for sensitivity analysis of decision trees. *Central European Journal of Operations Research*, 26(1), 135-159.

<https://doi.org/10.1007/s10100-017-0479-6>

Kocher, G., & Kumar, G. (2021). Analysis of machine learning algorithms with feature selection for intrusion detection using UNSW-NB15 dataset. *Available at SSRN 3784406*.

Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L., & Siegel, D. (2014). Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications. *Reliability Engineering & System Safety*, 130, 1-14.

<https://doi.org/10.1016/j.res.2014.04.013>

Lei, Y., Jia, F., Lin, J., Xing, S., Ding, S. X. "An Intelligent Fault Diagnosis Method Using Unsupervised Feature Learning Towards Mechanical Big Data," in *IEEE Transactions on Industrial Electronics*, vol. 63, no. 5, pp. 3137-3147, May 2016,

<https://doi.org/10.1109/TIE.2016.2519325>

Lei, Y., Li, N., Gontarz, S., Lin, J., Radkowski, S., Dybala, J. "A Model-Based Method for Remaining Useful Life Prediction of Machinery," in *IEEE Transactions on Reliability*, vol. 65, no. 3, pp. 1314-1326, Sept. 2016,
<https://doi.org/10.1109/TR.2016.2570568>

Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., Nandi, A. K. Applications of machine learning to machine fault diagnosis: A review and roadmap, *Mechanical Systems and Signal Processing*, Volume 138, 2020, 106587, ISSN 0888-3270,
<https://doi.org/10.1016/j.ymssp.2019.106587>

Li, X., Zhang, W., Ding, Q. Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction, *Reliability Engineering & System Safety*, Volume 182, 2019, Pages 208-218, ISSN 0951-8320,
<https://doi.org/10.1016/j.ress.2018.11.011>

Li, Z., Goebel, K., & Wu, D. (2018). Degradation Modeling and Remaining Useful Life Prediction of Aircraft Engines Using Ensemble Learning. *Journal of Engineering for Gas Turbines and Power*.
<https://doi.org/10.1115/1.4041674>

Liebl, M., Holder, J., Mohr, T., Dorneich, A., Liebgott, F., Middendorf, P. (2023). Development, Implementation and Evaluation of a Prototype System for Data-Driven Optimization of a Preforming Process. In: Kiefl, N., Wulle, F., Ackermann, C., Holder, D. (eds) *Advances in Automotive Production Technology – Towards Software-Defined Manufacturing and Resilient Supply Chains. SCAP 2022. ARENA2036*. Springer, Cham.
https://doi.org/10.1007/978-3-031-27933-1_27

Liu, H., Chen, W., Chen, W., Gu, Y. A CNN-LSTM-based domain adaptation model for remaining useful life prediction, *Meas. Sci. Technol.* 33 115118,
<https://doi.org/10.1088/1361-6501/ac7f7f>

Losi, E., Venturini, M., Manservigi, L., Ceschini, G. F., Bechini, G., Cota, G., and Riguzzi, F. (January 21, 2022). "Prediction of Gas Turbine Trip: A Novel Methodology Based on Random Forest Models." *ASME. J. Eng. Gas Turbines Power*. March 2022; 144(3): 031025.
<https://doi.org/10.1115/1.4053194>

Maaten, L. V. D., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of machine learning research*, 9(Nov), 2579-2605.

Mathur, N., Asirvadam, V. S., & Aziz, A. A. (2021). Mechanical Damage Assessment for Pneumatic Control Valves Based on a Statistical Reliability Model. *Sensors*, 21(10), 3307.
<https://doi.org/10.3390/s21103307>

Matsui, T., Nogami, T., Yokoi, Y. "Failure diagnosis for pneumatic control valves by neural network," *Proceedings of ICNN'95 - International Conference on Neural Networks*, Perth, WA, Australia, 1995, pp. 724-729 vol.2,
<https://doi.org/10.1109/ICNN.1995.487506>

Meddaoui, A., Hain, M. & Hachmoud, A. The benefits of predictive maintenance in manufacturing excellence: a case study to establish reliable methods for predicting failures. *Int J Adv Manuf Technol* 128, 3685–3690 (2023).

<https://doi.org/10.1007/s00170-023-12086-6>

Méndez-González, L. C., Rodríguez-Picón, L. A., Valles-Rosales, D. J., Romero-López, R., & Quezada-Carreón, A. E. (2017). Reliability analysis for electronic devices using beta-Weibull distribution. *Quality and reliability engineering international*, 33(8), 2521-2530.

<https://doi.org/10.1002/qre.2214>

Mohanraj, E., N, E., S, S., S, S. "Digital Twins for Automotive Predictive Maintenance," 2024 *International Conference on Inventive Computation Technologies (ICICT)*, Lalitpur, Nepal, 2024, pp. 1579-1584,

<https://doi.org/10.1109/ICICT60155.2024.10544392>

Nesbitt, B. (2011). *Handbook of valves and actuators: valves manual international*. ISBN: 9781493303885, Hardback ISBN: 9781856174947, eBook ISBN: 9780080549286, Elsevier.

Nti, I. K., Nyarko-Boateng, O., & Aning, J. (2021). Performance of machine learning algorithms with different K values in K-fold cross-validation. *International Journal of Information Technology and Computer Science*, 13(6), 61-71.

<https://doi.org/10.5815/ijitcs.2021.06.05>

Patra, G. K., Kuraku, C., Konkimalla, S., Boddapati, V. N., Sarisa, M., & Reddy, M. S. (2024). An Analysis and Prediction of Health Insurance Costs Using Machine Learning-Based Regressor Techniques. *Journal of Data Analysis and Information Processing*, 12(4), 581-596.

<https://doi.org/10.4236/jdaip.2024.124031>

Pedregosa, Fabian and Varoquaux and Gramfort, Alexandre and Michel, Vincent and Thirion, Bertrand and Grisel, Olivier and Blondel, Mathieu and Prettenhofer, Peter and Weiss, Ron and Dubourg, Vincent and Vanderplas, Jake and Passos, Alexandre and Cournapeau, David and Brucher, Matthieu and Perrot, Matthieu and Duchesnay, Edouard}, Scikit-learn: Machine Learning in Python, 2011, *JMLR.org*, issn: 1532-4435,

<https://doi.org/10.5555/1953048.2078195>

Rajabi, S., Azari, M. S., Santini, S., & Flammini, F. (2022). Fault diagnosis in industrial rotating equipment based on permutation entropy, signal processing and multi-output neuro-fuzzy classifier. *Expert systems with applications*, 206, 117754.

<https://doi.org/10.1016/j.eswa.2022.117754>

Reinhardt, I. C., Oliveira, J. C., Ring, D. T. (2020). Current Perspectives on the Development of Industry 4.0 in the Pharmaceutical Sector, *Journal of Industrial Information Integration*, Volume 18, 100131, ISSN 2452-414X,

<https://doi.org/10.1016/j.jii.2020.100131>

Sano, H., Malere, J., Berton, L. "Single and Multiple Failures Diagnostics of Pneumatic Valves using Machine Learning", in *Anais do XVI Encontro Nacional de Inteligência Artificial e Computacional*, Salvador, 2019, pp. 202-213,
<https://doi.org/10.5753/eniac.2019.9284>

Santhosh, K. V., Shenoy, V., Navada, B. R. "Fault identification and isolation in pneumatic valve using image processing," *2016 IEEE Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER)*, Mangalore, India, 2016, pp. 205-210,
<https://doi.org/10.1109/DISCOVER.2016.7806219>

Sapre, A., & Vartak, S. (2020). Scientific computing and data analysis using NumPy and Pandas. *International Research Journal of Engineering and Technology*, 7, 1334-1346. e-ISSN: 2395-0056

Schulz, E., Speekenbrink, M., Krause, A. A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions, *Journal of Mathematical Psychology*, Volume 85, 2018, Pages 1-16, ISSN 0022-2496,
<https://doi.org/10.1016/j.jmp.2018.03.001>

Sharma, R., Sharma, K., & Khanna, A. (2020). Study of supervised learning and unsupervised learning. *International Journal for Research in Applied Science and Engineering Technology*, 8(6), 588-593.
<http://doi.org/10.22214/ijraset.2020.6095>

Shinde, P. P., & Shah, S. (2018, August). A review of machine learning and deep learning applications. In *2018 Fourth international conference on computing communication control and automation (ICCUBEA)* (pp. 1-6). IEEE.
<http://doi.org/10.1109/ICCUBEA.2018.8697857>

Shuja, S. Z., Alotaibi, M., & Rehman, S. (2025). Evaluation of Onshore Wind Power Potential to Maximize Renewable Energy-Supported Transportation Electrification–Applied Along the Red Sea Coast of Saudi Arabia. *FME Transactions*, 53(1), 85-104.
<https://doi.org/10.5937/fme2501085S>

Slimane, A., Kebdani, S., Bouchouicha, B., Benguediab, M., Slimane, S., Bahram, K., ... & Sardi, N. (2018). An interactive method for predicting industrial equipment defects. *The International Journal of Advanced Manufacturing Technology*, 95, 4341-4351.
<https://doi.org/10.1007/s00170-017-1416-5>

Speiser, J. L., Miller, M. E., Tooze, J., Ip, E. A comparison of random forest variable selection methods for classification prediction modeling, *Expert Systems with Applications*, Volume 134, 2019, Pages 93-101, ISSN 0957-4174,
<https://doi.org/10.1016/j.eswa.2019.05.028>

Subbaraj, P., Kannapiran, B. Fault detection and diagnosis of pneumatic valve using Adaptive Neuro-Fuzzy Inference System approach, *Applied Soft Computing*, Volume 19, 2014, Pages 362-371, ISSN 1568-4946,

<https://doi.org/10.1016/j.asoc.2014.02.008>

Thakur, A. K., Mukherjee, A., Kundu, P. K. et al. Assessment of Different Signal Processing Techniques for Classifying Induction Motor Faults Using PCA Features: A Comparative Analysis. *J. Inst. Eng. India Ser. B* 105, 93–108 (2024).

<https://doi.org/10.1007/s40031-023-00948-2>

Tulsyan, A., Garvin, C., & Undey, C. (2020). Condition-based sensor-health monitoring and maintenance in biomanufacturing. *IFAC PapersOnLine*, 53(2), 170-175.

<https://doi.org/10.1016/j.ifacol.2020.12.116>

Turner, C., Okorie, O., Emmanouilidis, C., & Oyekan, J. (2022). Circular production and maintenance of automotive parts: An Internet of Things (IoT) data framework and practice review. *Computers in Industry*, 136, 103593.

<https://doi.org/10.1016/j.compind.2021.103593>

Vachtsevanos, G., Lewis, F., Roemer, M., Hess, A., & Wu, B. (2006). *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*. Wiley.

<https://doi.org/10.1002/9780470117842>

Van de Schoot, R., Depaoli, S., King, R. et al. Publisher Correction: Bayesian statistics and modelling. *Nat Rev Methods Primers* 1, 16 (2021).

<https://doi.org/10.1038/s43586-021-00017-2>

Wadel, F., Houssin, R., Coulibaly, A., & Tighazoui, A. (2024). Analysis of models for IoT-driven predictive maintenance under constraints in the case of the biopharmaceutical industry. *Journal of Intelligent Manufacturing*, 1-23.

<https://doi.org/10.1007/s10845-024-02542-9>

Wattenberg, M., Viégas, F., & Johnson, I. (2016). How to use t-SNE effectively. *Distill*, 1(10), e2.

<http://doi.org/10.23915/distill.00002>

Weerts, H. J., Mueller, A. C., & Vanschoren, J. (2020). Importance of tuning hyperparameters of machine learning algorithms. *arXiv preprint arXiv:2007.07588*.

<https://doi.org/10.48550/arXiv.2007.07588>

Weibull, W. (1951). A statistical distribution function of wide applicability. *Journal of applied mechanics*.

<https://hal.science/hal-03112318v1>

Wen, Q., Sun, L., Yang, F., Song, X., Gao, J., Wang, X., & Xu, H. (2021). Time Series Data Augmentation for Deep Learning: A Survey. 4653-4660.

<https://doi.org/10.24963/ijcai.2021/631>

Xia, T., Dong, Y., Xiao, L., Du, S., Pan, E., & Xi, L. (2018). Recent advances in prognostics and health management for advanced manufacturing paradigms. *Reliability Engineering & System Safety*, 178, 255-268.

<https://doi.org/10.1016/j.ress.2018.06.021>

Yan, M., Wang, X., Wang, B., Chang, M., Muhammad, I. Bearing remaining useful life prediction using support vector machine and hybrid degradation tracking model. *ISA Trans.* 2020 Mar;98:471-482.

<https://doi.org/10.1016/j.isatra.2019.08.058>

Zadiran, K., & Shcherbakov, M. (2023). New Method of Degradation Process Identification for Reliability-Centered Maintenance of Energy Equipment. *Energies*, 16(2), 575.

<https://doi.org/10.3390/en16020575>

Zhang, Q., Hua, C., & Xu, G. (2014). A mixture Weibull proportional hazard model for mechanical system failure prediction utilising lifetime and monitoring data. *Mechanical Systems and Signal Processing*, 43(1-2), 103-112.

<https://doi.org/10.1016/j.ymssp.2013.10.013>

Zhang, S., Li, X., Zong, M., Zhu, X., Wang, R. "Efficient kNN Classification With Different Numbers of Nearest Neighbors," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 5, pp. 1774-1785, May 2018,

<https://doi.org/10.1109/TNNLS.2017.2673241>

Zhao, Q., & Zhu, W. (2025). TMSA-Net: A novel attention mechanism for improved motor imagery EEG signal processing. *Biomedical Signal Processing and Control*, 102, 107189.

<https://doi.org/10.1016/j.bspc.2024.107189>

Zhou, Z.-H. A brief introduction to weakly supervised learning, *National Science Review*, Volume 5, Issue 1, January 2018, Pages 44–53.

<https://doi.org/10.1093/nsr/nwx106>

Zio, E. (2022). Prognostics and Health Management (PHM): Where are we and where do we (need to) go in theory and practice. *Reliability Engineering & System Safety*, 218, 108119.

<https://doi.org/10.1016/j.ress.2021.108119>