

THÈSE DE DOCTORAT DE

L'ÉCOLE NATIONALE
D'INGÉNIEURS DE BREST

ÉCOLE DOCTORALE N° 644

*Mathématiques et Sciences et Technologies
de l'Information et de la Communication*

Spécialité : « *Informatique et Architectures numériques* »

Par

Jean-Victor AUTRAN

DADDY : Détection par Anticipation De DYsfonctionnement dans des systèmes embarqués

Thèse présentée et soutenue à Plouzané, le 14 novembre 2025

Unité de recherche : IRL CROSSING 2010

Thèse N° : 2025ENIB004

Rapporteurs avant soutenance :

Nicolas BOUSQUET Professeur associé, Sorbonne Université
Patrick REIGNIER Professeur des Universités, Grenoble INP

Composition du Jury :

| | | |
|--------------------|----------------------|---|
| Examinateurs : | Angélique DRÉMEAU | Professeure, ENSTA Bretagne |
| Dir. de thèse : | Cédric BUCHE | Professeur des Universités, CNRS IRL CROSSING (Australie) |
| Co-dir. de thèse : | Jean-Philippe DIGUET | Directeur de recherche, CNRS IRL CROSSING (Australie) |
| Co-enc. de thèse : | Véronique KUHN | Manager, Ariane Group |

REMERCIEMENTS

Mes tout premiers remerciements vont sans aucun doute à Véronique Kuhn, mon encadrante au sein d'ArianeGroup. Je lui suis profondément reconnaissant pour son accompagnement, son soutien indéfectible, sa patience et ses encouragements tout au long de ces années. J'ai particulièrement apprécié sa disponibilité malgré un emploi du temps toujours très chargé, ainsi que par l'efficacité et la rigueur qu'elle m'a transmises. Ses relectures attentives et ses corrections précises de mes articles et du manuscrit de ma thèse ont été pour moi d'une aide inestimable. Plus qu'une encadrante, Véronique a été une véritable guide pendant toute la durée de ma thèse. Elle a cru en moi dès le départ et m'a donné l'opportunité de réaliser cette thèse, ce pour quoi je lui en serai toujours reconnaissant.

Je tiens également à exprimer ma profonde gratitude à mes directeurs de thèse, Cédric Buche et Jean-Philippe Diguet. Leur expertise et leurs conseils avisés ont été essentiels à l'avancée de ce travail. Je leur suis particulièrement reconnaissant d'avoir su rendre nos échanges aussi fluides que fructueux, et ce, malgré l'éloignement géographique et le décalage horaire.

J'associe à ces remerciements le laboratoire CROSSING pour avoir fourni le cadre institutionnel nécessaire à cette collaboration internationale entre la France et l'Australie.

Je souhaite par ailleurs exprimer toute ma gratitude aux rapporteurs de mon manuscrit, Nicolas Bousquet et Patrick Reignier, pour avoir pris le temps d'évaluer ce travail avec rigueur et bienveillance, et pour leurs remarques constructives qui ont contribué à l'améliorer. Je remercie aussi Angélique Drémeau de m'avoir fait l'honneur de présider ce jury.

Je remercie également Guy Gogniat et Dominique Faudreydaz d'avoir accepté de faire partie de mon comité de suivi de thèse. Leurs conseils et leurs retours m'ont permis de prendre du recul et d'améliorer la qualité de mon travail de recherche.

Je tiens par ailleurs à remercier l'AID et l'ANRT, dont le financement via une convention CIFRE a rendu possible la réalisation de cette thèse.

Une pensée particulière pour l'ensemble des membres d'ArianeGroup, qui m'ont accueilli avec chaleur et bienveillance. Je pense notamment à Jean-Pierre, responsable de

l'équipe, pour son soutien constant, ainsi qu'à Matthias et Sébastien, dont l'aide et le suivi tout au long de la thèse ont été déterminants pour mener à bien ce travail. Je remercie aussi mes collègues du bâtiment : Aude, Clément, Davy, Elodie, Erwan, Fabien, Lionel, Nicolas et tous les autres, pour la bonne ambiance quotidienne, les rires partagés lors du déjeuner et les parties de baby-foot à la pause du midi. Au fil de ces années, ils sont devenus bien plus que de simples collègues : de véritables amis.

Je souhaite aussi remercier mes amis, qui ont rendu ces années plus légères et joyeuses :

À Brest, merci à Anaïs, Coraline, Corentin, Hugo, Jeremy, Lucas, Nicolas, Pierre, Sébastien, Yann, ainsi qu'à tous les membres du CERV pour leur amitié fidèle et leur soutien, mais aussi pour toutes nos discussions passionnées autour de nos péripéties de thèse.

À mes amis d'enfance, Dorian, Vincent, Félix, Sharon, Alban, Jean-Gabriel, merci d'avoir toujours été là, que ce soit de l'autre bout de la France ou de l'autre côté de la Terre. Malgré la distance, vos encouragements, votre écoute et vos mots bienveillants ont été pour moi d'un grand réconfort.

Enfin, un merci tout particulier à ma famille, sans qui rien de tout cela n'aurait été possible. Merci à ma mère et à mon père pour leur patience et leur fierté qui m'ont toujours donné la force d'avancer. Merci aussi à mes frères, pour leur présence constante et pour être une source d'inspiration à mes yeux. Merci enfin à tout le reste de ma famille, pour leur affection et leurs encouragements au fil des années.

À toutes et à tous, je vous adresse ma profonde gratitude.

TABLE DES MATIÈRES

| | |
|---|-----------|
| Introduction | 1 |
| 1 Maintenance Prévisionnelle : Fondements, Enjeux et Applications Industrielles | 7 |
| 1.1 La maintenance prévisionnelle dans le domaine industriel | 8 |
| 1.1.1 Évolution des pratiques de maintenance industrielle | 8 |
| 1.1.2 Les étapes clés de la maintenance prévisionnelle | 12 |
| 1.1.3 Bénéfices de la maintenance prévisionnelle | 16 |
| 1.2 Systèmes industriels complexes et maintenance prévisionnelle | 19 |
| 1.2.1 La transition vers la maintenance prévisionnelle dans l'industrie : approches et facteurs facilitants | 19 |
| 1.2.2 Complexité des systèmes industriels | 23 |
| 1.2.3 Défis et problématiques de la maintenance prévisionnelle | 24 |
| 1.2.4 Cas particulier de l'étude | 27 |
| 2 Crédit d'un Indicateur d'État de Santé par Discréétisation : la Méthode MSLD | 33 |
| 2.1 Contexte et méthodologies de création d'un état de santé | 34 |
| 2.1.1 Définition et rôle d'un état de santé | 34 |
| 2.1.2 Méthodes de création d'un état de santé | 35 |
| 2.1.3 Limites des approches existantes pour la création d'état de santé . | 38 |
| 2.2 Proposition d'une méthode de création d'état de santé basée sur la discréétisation | 42 |
| 2.2.1 Présentation et Formalisation de la méthode MSLD | 42 |
| 2.2.2 Rôle de la discréétisation | 43 |
| 2.2.3 Processus de création d'un état de santé avec MSLD | 46 |
| 2.2.4 Résultats | 47 |
| 2.2.5 Discussion et limitations | 51 |

TABLE DES MATIÈRES

| | |
|---|-----------|
| 3 Modélisation d'État de Santé et Diagnostic par Réseau Bayésien | 55 |
| 3.1 Utilisation des bases de données publiques en maintenance prévisionnelle | 56 |
| 3.1.1 Motivations et contraintes | 56 |
| 3.1.2 Sélection de bases de données | 57 |
| 3.1.3 Prétraitement des bases de données | 62 |
| 3.2 Choix et justification du modèle de diagnostic | 66 |
| 3.2.1 Critères de sélection des algorithmes | 66 |
| 3.2.2 Comparaison des approches de diagnostic | 67 |
| 3.2.3 Choix des réseaux bayésiens | 71 |
| 3.3 Modélisation et mise en œuvre du réseau bayésien | 74 |
| 3.3.1 Fondements des réseaux bayésiens | 74 |
| 3.3.2 Apprentissage de la structure et des paramètres | 77 |
| 3.3.3 Inférence | 85 |
| 3.4 Résultats et analyses | 87 |
| 3.4.1 Évaluation des performances | 87 |
| 3.4.2 Analyse des résultats | 89 |
| 4 Modélisation Temporelle de la Dégradation de Systèmes Industriels par Réseaux Bayésiens Dynamiques | 95 |
| 4.1 Construction d'un modèle dynamique à partir des réseaux bayésiens | 96 |
| 4.1.1 Du diagnostic à la prédition : principe des réseaux bayésiens dynamiques | 96 |
| 4.1.2 Méthodologie de création d'un RBD pour la maintenance prévisionnelle industrielle | 102 |
| 4.2 Application des RBD à des jeux de données industriels | 107 |
| 4.2.1 Construction de jeux de données temporels à partir des données transformées | 107 |
| 4.2.2 Résultats expérimentaux et analyse des performances | 109 |
| 4.3 Discussion approfondie, analyse méthodologique et perspectives | 113 |
| 4.3.1 Analyse rétrospective de la méthodologie de thèse | 113 |
| 4.3.2 Adéquation des modèles RBD aux conditions industrielles : limites et enjeux | 115 |
| 4.3.3 Limitations spécifiques du travail de thèse et recommandations | 120 |
| 4.3.4 Perspectives d'évolution et pistes de recherche futures | 127 |

TABLE DES MATIÈRES

| | |
|---------------------|-----|
| Conclusion Générale | 133 |
| Bibliographie | 141 |

INTRODUCTION

La performance et la compétitivité des systèmes industriels modernes reposent sur une maintenance des équipements à la fois efficace et optimisée. Or, la survenue de défaillances et les arrêts de production imprévus qui en découlent représentent une menace constante pour la continuité des opérations et engendrent une charge économique considérable. Pour les 500 plus grandes entreprises mondiales, cette charge est estimée à environ 1,5 billion de dollars annuellement, soit près de 11% de leurs revenus (SIEMENS, 2024). Plus largement, le coût de la maintenance peut représenter entre 15 et 60% du coût de production total selon le secteur industriel (MOBLEY, 2002). Face à cette réalité économique, les stratégies de maintenance ont évolué d'approches purement réactives vers des méthodologies proactives, parmi lesquelles la maintenance prévisionnelle (PdM) occupe aujourd’hui une place centrale.

L'évolution des stratégies de maintenance, détaillée au Chapitre 1, a débuté avec la maintenance corrective, où les interventions n'ont lieu qu'après la défaillance, conformément à la norme AFNOR NF-X 60 000 (AFNOR, 2016). Pour pallier les coûts élevés et les arrêts de production non maîtrisés inhérents à cette approche, la maintenance préventive a été introduite. Celle-ci, qu'elle soit systématique ou conditionnelle, vise à réduire la probabilité de défaillance par des actions anticipées. La maintenance prévisionnelle constitue l'étape évolutive suivante, s'appuyant sur l'analyse de données et des modèles pour anticiper les défaillances avant leur survenue. Cette transition est une réponse logique aux limitations des méthodes précédentes et aux défis posés par les systèmes industriels complexes, caractérisés par de nombreuses interactions entre composants et des dynamiques non linéaires.

L'intérêt pour la maintenance prévisionnelle a connu une croissance exponentielle au cours des deux dernières décennies, comme l'illustre la Figure 1 qui montre une augmentation spectaculaire du nombre de publications scientifiques sur le sujet, passant de 1500 en 1981 à environ 8 000 en 2004 et à plus de 48 000 en 2024. Cette dynamique témoigne d'une industrialisation avancée, contexte favorable au déploiement de la PdM. Le contexte actuel est également marqué par la quatrième révolution industrielle, ou Industrie 4.0, qui fournit les leviers technologiques, notamment l'Internet des Objets (IoT) et l'analyse de

TABLE DES MATIÈRES

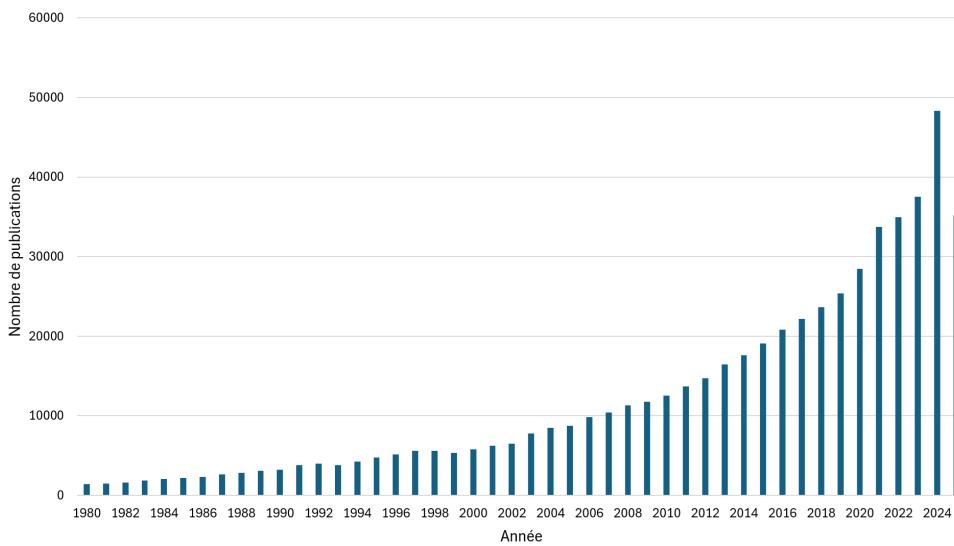


FIGURE 1 – Évolution du nombre d'articles scientifiques annuels sur le thème de la maintenance prévisionnelle dans la base de données d'Elsevier.

données massives (Big Data), indispensables à l'essor de la PdM.

La mise en œuvre de la PdM s'articule autour d'étapes clés décrites au Chapitre 1, à savoir l'acquisition et le prétraitement des données, le diagnostic, le pronostic qui inclut l'estimation de la durée de vie utile restante (RUL) (NEAGU ; SOLEIMANI et CAMPEAN, 2021), la prise de décision et l'interface homme-machine (IHM). Plusieurs approches méthodologiques coexistent, telles que les approches basées sur la physique, qui emploient par exemple les filtres de Kalman, ou celles basées sur les données, qui exploitent des algorithmes comme les SVM et les LSTM, sans oublier les approches fondées sur les connaissances et les méthodes hybrides.

Un élément fondamental de la PdM est la définition d'un indicateur de santé pertinent. Face aux limitations des indicateurs binaires souvent rencontrés, la méthode MSLD (Multivariable System Labeling through Discretization), développée au Chapitre 2 de cette thèse, a été proposée. Elle permet de créer des états de santé discrets, multi-niveaux et sémantiquement riches, à savoir : *Optimal*, *Stable*, *Prudent*, *Vulnérable* et *Défaillance*. Cette approche par discréétisation non supervisée offre ainsi une granularité essentielle au diagnostic. Pour la modélisation, les Réseaux Bayésiens et leur extension temporelle, les Réseaux Bayésiens Dynamiques, offrent des cadres formels puissants. Les Réseaux Bayésiens, présentés au Chapitre 3, permettent de modéliser les relations probabilistes et de gérer l'incertitude, ce qui est crucial pour le diagnostic. Leur extension, les Réseaux Bayésiens Dynamiques étudiés au Chapitre 4, intègre la dimension temporelle pour le pronostic

en modélisant l'évolution des états de santé, souvent via une représentation compacte de type 2TBN (Two-Time-Slice Bayesian Network) (MURPHY, 2002).

Malgré ces avancées, la mise en œuvre de la PdM est confrontée à des défis majeurs, qui sont abordés au Chapitre 1. La gestion des données, incluant leur disponibilité, leur qualité et le bruit, est un obstacle central (COMPARE ; BARALDI et ZIO, 2019). La rareté des données de défaillance, paradoxe inhérent à une maintenance efficace, complexifie l'apprentissage supervisé. Des contraintes technologiques et organisationnelles freinent également son adoption (BOKRANTZ et al., 2017).

La présente thèse s'inscrit dans un contexte industriel spécifique, se concentrant sur des systèmes complexes où l'acquisition de données est irrégulière, partielle et non continue. Ces spécificités, liées à des mesures effectuées uniquement lors d'interventions de maintenance planifiées, exacerbent les défis existants. Elles compliquent la capture des tendances de dégradation, mènent à un volume de données d'apprentissage limité et à une très faible proportion de données de pannes réelles. La stratégie de maintenance actuelle, bien que préventive, conduit encore à des défaillances inattendues et à des opérations curatives coûteuses.

De ces constats découlent plusieurs questions de recherche centrales qui animent ce travail. Il s'agit premièrement d'étudier comment mettre en œuvre efficacement des approches de pronostic basées sur les données dans un contexte de faible quantité de données, et de définir le lien quantitatif entre la taille des données et la performance. La thèse explore également comment l'augmentation de données appliquée aux séries temporelles peut être exploitée et optimisée pour améliorer le pronostic. De plus, nous examinons de quelle manière la connaissance experte peut être intégrée à ces méthodologies pour en améliorer la fiabilité dans un cadre industriel. Enfin, un axe de recherche essentiel concerne l'évaluation de l'impact de la performance du pronostic sur la prise de décision en maintenance, afin de quantifier les gains potentiels.

Ces défis définissent la problématique centrale de nos travaux. La question de recherche principale qui guide cette thèse est donc de déterminer comment développer une méthodologie de maintenance prévisionnelle complète et robuste, du diagnostic au pronostic à court terme, pour des systèmes industriels complexes caractérisés par des données acquises de manière irrégulière, incomplète et parcimonieuse.

Ce travail repose sur plusieurs hypothèses fondamentales. Nous postulons d'abord qu'il est possible de générer des indicateurs d'état de santé discrets, multi-niveaux et sémantiquement riches à partir de données de capteurs brutes et acquises irrégulièrement, grâce

TABLE DES MATIÈRES

à la méthode MSLD. Ensuite, nous avançons que les Réseaux Bayésiens constituent un formalisme approprié pour diagnostiquer de manière robuste ces états de santé, même en présence de données manquantes. Nous émettons également l'hypothèse que leur extension, les Réseaux Bayésiens Dynamiques, est capable de modéliser la dégradation temporelle et de prédire les états de santé futurs à court terme. Finalement, nous supposons que la méthodologie globale qui en résulte permettra d'améliorer les stratégies de maintenance actuelles en fournissant des diagnostics fiables et des pronostics exploitables.

Pour répondre à ces questions, une méthodologie structurée en trois étapes a été adoptée. La première contribution, détaillée au Chapitre 2, est le développement de la méthode MSLD pour la création d'états de santé interprétables. Cette méthode transforme les variables continues en catégories discrètes en se basant sur des analyses statistiques et des seuils de défaillance, puis définit l'état de santé global par le paramètre le plus dégradé.

Pour la modélisation diagnostique, présentée au Chapitre 3, nous avons utilisé les Réseaux Bayésiens statiques. Ce travail a nécessité la transformation de bases de données publiques, AI4I devenant AI4I-PMDI, et l'utilisation d'une base de données industrielle, Genesis-PMDI, pour simuler et refléter les contraintes réelles de données manquantes et de gestion de parc. Les performances du réseau bayésien à reproduire la logique de diagnostic MSLD ont ainsi été validées.

L'extension vers le pronostic, qui fait l'objet du Chapitre 4, a été réalisée via les Réseaux Bayésiens Dynamiques. Les données ont été structurées en paires temporelles pour apprendre un modèle de transition 2TBN. La structure intra-tranche du réseau dynamique a hérité de celle du réseau statique, tandis que les dépendances inter-tranches ont été apprises. La capacité du modèle à prédire l'état de santé à l'horizon suivant a ensuite été rigoureusement évaluée sur les deux jeux de données.

L'objectif principal de cette thèse est donc de développer et d'évaluer une méthodologie complète pour la maintenance prévisionnelle au sein de systèmes industriels complexes, spécifiquement ceux confrontés à des données acquises de manière irrégulière et incomplète. Cette méthodologie se focalise sur la création d'états de santé interprétables et sur l'utilisation des réseaux bayésiens pour le diagnostic et le pronostic à court terme. L'importance de ce travail réside dans sa contribution à la résolution des limitations des stratégies de maintenance traditionnelles, avec pour ambition de réduire les coûts, d'améliorer la fiabilité et la disponibilité des systèmes, notamment dans des contextes industriels exigeants comme celui de la sidérurgie, où les données sont rares et précieuses.

Contributions scientifiques

1. Jean-Victor Autran, Véronique Kuhn, Jean-Philippe Diguet, Matthias Dubois, et Cédric Buche. **AI4I-PMDI : Predictive maintenance datasets with complex industrial settings' irregularities**. Dans *Procedia Computer Science*, volume 246, pages 1201–1209. Elsevier, 2024. Actes de la *28th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2024)*.
2. Jean-Victor Autran, Véronique Kuhn, Jean-Philippe Diguet, Matthias Dubois, et Cédric Buche. **Discretization Strategies for Improved Health State Labeling in Multivariable Predictive Maintenance Systems**. Dans les actes de la *13th International Conference on Data Science, Technology and Applications (DATA 2024)*, pages 434–441. SCITEPRESS, 2024.

Organisation du manuscrit

Ce manuscrit est organisé en quatre chapitres principaux :

Le **Chapitre 1**, intitulé *Maintenance prévisionnelle : fondements, enjeux et applications industrielles*, pose les bases de la recherche. Il retrace l'évolution des pratiques de maintenance, détaille les étapes clés, les bénéfices et les défis de la PdM, et introduit le cas d'étude spécifique.

Le **Chapitre 2**, *État de santé d'un équipement : approche basée sur la discréétisation des mesures*, se consacre à la définition d'un indicateur de santé pertinent et introduit la méthode MSLD.

Le **Chapitre 3**, *Modélisation d'état de santé et diagnostic par réseau bayésien*, aborde la phase de diagnostic, justifie le choix des Réseaux Bayésiens, et décrit la construction et l'évaluation d'un modèle RB statique.

Le **Chapitre 4**, *Modélisation temporelle de la dégradation de systèmes industriels par réseaux bayésiens dynamiques*, étend le modèle au pronostic via les Réseaux Bayésiens Dynamiques, détaille leur construction et évaluation.

La **conclusion** revient sur l'ensemble des travaux présentés, synthétise les contributions majeures et propose des perspectives pour l'orientation des futures recherches dans ce domaine.

MAINTENANCE PRÉVISIONNELLE : FONDEMENTS, ENJEUX ET APPLICATIONS INDUSTRIELLES

Introduction

La maintenance des équipements industriels est un enjeu fondamental pour garantir la continuité et l'efficacité des opérations. Elle englobe l'ensemble des actions visant à assurer le bon fonctionnement des équipements, qu'il s'agisse de vérifications de routine, de réparations ponctuelles ou du remplacement d'éléments critiques. Le choix des stratégies de maintenance est crucial et dépend des exigences spécifiques de l'entreprise, des contraintes économiques et des caractéristiques des systèmes en place.

Dans un environnement industriel en constante évolution, déterminer le moment optimal pour réaliser une intervention de maintenance est un défi majeur. L'incertitude associée à l'apparition des pannes provient à la fois de la variabilité des conditions d'exploitation et des interactions complexes entre les composants des systèmes industriels. En effet, les arrêts imprévus et le coût global de la maintenance représentent des charges financières considérables pour les industries, mettant en évidence l'importance de stratégies efficaces pour optimiser la gestion des équipements et minimiser les pertes.

Historiquement, les entreprises ont adopté des approches réactives et préventives pour assurer la maintenance de leurs systèmes. La maintenance corrective, qui consiste à intervenir uniquement après une panne, entraîne des coûts élevés et des arrêts prolongés. En réponse, la maintenance préventive a été largement adoptée afin d'anticiper les pannes grâce à des inspections régulières et des remplacements planifiés. Toutefois, ces stratégies présentent des limites importantes, notamment en raison d'interventions parfois inutiles ou d'une incapacité à prévoir précisément l'évolution des dégradations.

Pour pallier ces limitations, la maintenance prévisionnelle a émergé comme une ap-

proche plus avancée, combinant surveillance des équipements et analyse de données. En exploitant des technologies comme l'Internet des objets (IoT), l'apprentissage automatique et les modèles physiques, elle vise à anticiper les défaillances avant leur survenue, permettant ainsi d'optimiser les interventions de maintenance. Cette évolution marque un tournant majeur dans la gestion des systèmes industriels, où la prise de décision repose désormais sur une analyse fine et continue de l'état des équipements.

Ce chapitre explore en détail les fondements, les enjeux et les applications de la maintenance prévisionnelle. Il retrace l'évolution des pratiques de maintenance, en examinant la transition progressive des stratégies réactives aux approches prévisionnelles basées sur l'exploitation des données. Les concepts clés, les technologies sous-jacentes et les bénéfices de cette approche seront analysés à travers divers secteurs industriels. Enfin, nous mettrons en lumière les défis spécifiques liés à la mise en œuvre de la maintenance prévisionnelle, en particulier dans le cas des systèmes complexes, où l'acquisition des données est irrégulière et contrainte. Ce dernier aspect sera illustré par une étude de cas détaillant les problématiques spécifiques à un environnement industriel particulier, servant de base aux développements méthodologiques des chapitres suivants.

1.1 La maintenance prévisionnelle dans le domaine industriel

1.1.1 Évolution des pratiques de maintenance industrielle

L'évolution des pratiques de maintenance dans le secteur industriel a été marquée par une transition progressive des approches réactives à des stratégies de plus en plus proactives et prédictives (MOBLEY, 2002). Cette évolution répond à la nécessité de garantir la fiabilité des équipements tout en réduisant les coûts d'exploitation et en maximisant les performances des systèmes.

Maintenance corrective : une approche réactive

La maintenance corrective, ou réactive, est définie par la norme AFNOR NF-X 60 000 comme une "maintenance exécutée après détection d'une panne et destinée à remettre un bien dans un état dans lequel il peut accomplir une fonction requise" (AFNOR, 2016). Il s'agit de la forme la plus basique de maintenance, consistant à intervenir uniquement après la survenue d'une défaillance pour résoudre le problème.

Cette méthode présente l'avantage de nécessiter un investissement minimal en termes de planification et de ressources initiales. En effet, les équipes de maintenance interviennent uniquement en cas de dysfonctionnement avéré. Toutefois, cette simplicité s'accompagne de nombreuses contraintes et inconvénients.

Tout d'abord, la maintenance corrective engendre généralement de coûts élevés, principalement en raison des réparations en urgence et des arrêts imprévus de production. Lorsque les pièces de rechange ne sont pas disponibles immédiatement, il est nécessaire d'engager des frais supplémentaires pour des commandes en urgence, avec parfois des délais incompatibles avec les exigences de production (MOBLEY, 2002). De plus, l'absence de planification préalable allonge la durée d'indisponibilité des équipements. Il faut en effet consacrer du temps à diagnostiquer l'origine de la panne, puis mobiliser les ressources humaines et matérielles nécessaires à la réparation, ce qui augmente considérablement le temps d'arrêt et engendre des pertes économiques.

Les conséquences de cette maintenance peuvent être particulièrement lourdes, où les défaillances peuvent avoir un effet domino et perturber l'ensemble du processus de production. Les retards accumulés entraînent des pertes de production, des coûts supplémentaires liés aux réparations d'urgence et, dans certains cas, des pénalités contractuelles en raison du non-respect des délais de livraison. Cette situation se traduit également par une insatisfaction des clients, notamment lorsque les interruptions de production retardent la disponibilité des produits.

Face à ces impacts, il est nécessaire de différencier les différentes formes d'intervention correctives, qui s'adaptent selon la nature et la gravité de la panne. On peut distinguer trois sous-catégories de maintenance corrective :

- **La maintenance palliative** : elle consiste en une remise en état provisoire pour permettre au système de poursuivre son fonctionnement temporairement, en attendant une intervention définitive.
- **La maintenance curative** : elle vise à remettre l'équipement en état de fonctionnement de manière définitive, en réparant ou en remplaçant les composants défaillants.
- **La maintenance d'amélioration** : elle inclut une modification du système pour éliminer la cause récurrente de la panne et améliorer la fiabilité globale de l'équipement.

Ainsi, si la maintenance corrective semble initialement attrayante en raison de sa simplicité et de son faible coût initial, elle présente de nombreuses limites dans un envi-

ronnement industriel moderne. L'absence d'anticipation, la difficulté à disposer de pièces de rechange en temps utile et les temps d'arrêt non maîtrisés entraînent des coûts indirects élevés. Ces inconvénients en font une stratégie peu fiable pour les équipements critiques, où les défaillances peuvent affecter durablement la production et la satisfaction client.

Face à ces limites, les industriels ont progressivement adopté des pratiques de maintenance préventive, permettant d'anticiper les pannes et de mieux planifier les interventions. Cette approche, qui sera abordée dans la partie suivante, marque une étape essentielle dans l'évolution des pratiques de maintenance.

Maintenance préventive : un pas vers l'anticipation

Face aux limites de la maintenance corrective, la maintenance préventive est apparue comme une alternative plus efficace. Elle est définie par la norme AFNOR NF-X 60 000 comme une "maintenance exécutée à des intervalles prédéterminés ou selon des critères prescrits et destinés à réduire la probabilité de défaillance ou la dégradation du fonctionnement d'un bien" (AFNOR, 2016). Son objectif principal est de limiter les risques de panne en intervenant de manière anticipée, avant qu'une défaillance ne survienne. Cette approche peut être subdivisée en deux types : la maintenance systématique et la maintenance conditionnelle.

La **maintenance systématique** repose sur des intervalles prédéterminés de temps ou d'utilisation pour planifier les interventions (ex. remplacement d'un filtre à air toutes les 500 heures de fonctionnement). Les équipements sont entretenus selon un calendrier établi à l'avance, indépendamment de leur état réel. Cela permet de mieux organiser les opérations de maintenance et d'éviter certaines pannes inattendues. Cependant, cette approche est coûteuse, car elle peut conduire à des interventions prématurées si les équipements sont encore en bon état. Par conséquent, elle entraîne des coûts supplémentaires liés aux remplacements inutiles et à une consommation excessive de pièces détachées (MOBLEY, 2002).

Afin de pallier ces inconvénients, la **maintenance conditionnelle** repose sur la surveillance en temps réel des paramètres de l'équipement via des capteurs. Les données collectées sont comparées à des seuils d'alerte prédéfinis à partir d'analyses historiques ou théoriques afin de déclencher une intervention lorsque les indicateurs de fonctionnement suggèrent une dégradation imminente. Cette approche permet d'adapter la maintenance aux besoins réels de l'équipement, évitant ainsi les interventions superflues tout en réduisant le risque de panne. Elle est particulièrement efficace pour détecter des anomalies

en amont, comme une augmentation anormale de la température d'un moteur ou une vibration excessive signalant une usure prématuée (MOBLEY, 2002).

Toutefois, l'efficacité de cette méthode dépend fortement de la précision des seuils d'alerte. Si ces seuils sont mal définis, cela peut engendrer deux types de problèmes : Un seuil trop bas entraîne des interventions trop fréquentes et inutiles, augmentant les coûts et perturbant l'exploitation des équipements. Un seuil trop élevé risque de ne pas déclencher d'intervention à temps, laissant la défaillance se produire sans avertissement, avec des conséquences parfois critiques.

De plus, même lorsque les anomalies sont détectées, le délai entre l'alerte et la panne peut être trop court pour organiser efficacement une intervention. L'inefficacité de la maintenance préventive est exacerbée dans les environnements industriels où les machines ne suivent pas un comportement linéaire et où les conditions de fonctionnement varient fréquemment.

Ainsi, bien que la maintenance conditionnelle représente une avancée significative par rapport aux approches purement systématiques, elle repose encore sur des modèles fixes ou des seuils empiriques, qui ne tiennent pas toujours compte de la complexité des systèmes industriels. Ces limites ont conduit à l'émergence d'approches plus avancées, comme les jumeaux numériques, où les modèles de données et l'analyse en temps réel sont utilisés pour simuler et anticiper la dégradation des équipements. C'est dans ce contexte que la maintenance prévisionnelle s'est imposée comme une alternative plus prometteuse. Elle exploite l'analyse avancée des données et l'intelligence artificielle pour anticiper les pannes avec une meilleure précision, en s'adaptant dynamiquement aux évolutions des équipements.

Maintenance prévisionnelle : l'optimisation basée sur les données

L'apparition de la maintenance prévisionnelle représente une avancée majeure par rapport aux pratiques précédentes. Définie par la norme AFNOR NF-X 60 000 comme une "maintenance conditionnelle exécutée en suivant les prévisions extrapolées de l'analyse et de l'évaluation de paramètres significatifs de la dégradation du bien" (AFNOR, 2016), elle se distingue par son approche proactive basée sur l'exploitation des données. Ce type de maintenance pousse la maintenance conditionnelle un pas plus loin en utilisant des modèles de données avancés pour prédire les défaillances avant qu'elles ne surviennent. Cette méthode repose sur des modèles physiques ou basé sur les données, ainsi que sur l'analyse des données historiques et en temps réel, pour suivre l'état de santé des équi-

gements et planifier les interventions au moment optimal (Figure 1.1). En réduisant les interventions inutiles, la maintenance prévisionnelle maximise la durée de vie des équipements et permet d'anticiper des défaillances que les méthodes traditionnelles ne peuvent identifier.

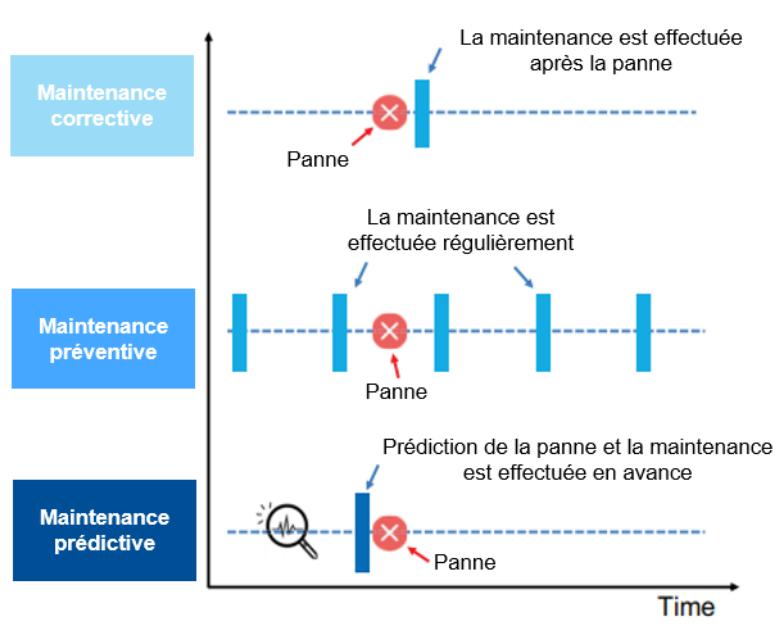


FIGURE 1.1 – Comparaison des stratégies de maintenance et de leurs temporalités d'intervention (RAN et al., 2019).

L'évolution des stratégies de maintenance a conduit à l'essor de la maintenance prévisionnelle, qui se distingue par son approche fondée sur l'analyse des données et la modélisation des défaillances. Toutefois, sa mise en œuvre repose sur un enchaînement structuré d'étapes permettant d'assurer une surveillance efficace des équipements et d'anticiper les pannes avec précision. Il est donc essentiel de détailler ces étapes méthodologiques qui conditionnent la performance des systèmes de maintenance prévisionnelle.

1.1.2 Les étapes clés de la maintenance prévisionnelle

La maintenance prévisionnelle repose sur une série d'étapes méthodologiques qui, combinées, permettent d'assurer un suivi précis des équipements industriels et d'anticiper les défaillances. Ces étapes incluent l'acquisition et le prétraitement des données, le diagnostic, le pronostic, la prise de décision, ainsi que l'interface Homme-Machine (IHM). Chacune

de ces étapes est essentielle pour garantir une mise en œuvre réussie de la maintenance prévisionnelle.

Acquisition des données

L'acquisition des données constitue la première étape et se concentre sur la collecte de données variées provenant de capteurs ou d'événements. Ces données incluent des paramètres tels que l'intensité, la tension, la température ou encore les vibrations, ainsi que des enregistrements d'événements liés aux pannes ou aux maintenances (LEUKEL ; GONZÁLEZ et RIEKERT, 2021).

Cette étape est cruciale, car toutes les phases suivantes de la maintenance prévisionnelle reposent sur la qualité et la pertinence des données collectées. Des données incomplètes ou bruitées peuvent compromettre l'efficacité des analyses et réduire la fiabilité des modèles prédictifs, limitant ainsi leur capacité à anticiper correctement les défaillances. Il est donc essentiel de garantir une acquisition de données cohérente et représentative des conditions réelles de fonctionnement des équipements.

Dans certains environnements industriels, il peut être nécessaire d'ajouter des capteurs supplémentaires pour enrichir les données disponibles et assurer un suivi plus précis des paramètres critiques. L'intégration de capteurs permet non seulement d'améliorer la couverture des données collectées, mais aussi d'assurer une meilleure détection des anomalies et des tendances de dégradation.

Selon les études, la durée d'acquisition s'étend généralement sur plusieurs mois à plusieurs années, avec des périodes d'échantillonnage variant de quelques minutes à plusieurs heures. Les bases de données publiques, comme celles utilisées pour les turboréacteurs, les machines de fraisage ou les roulements, jouent un rôle clé en fournissant des ensembles de données standardisés pour comparer les performances des méthodes de maintenance prévisionnelle (EXCELLENCE, s. d.). Cependant, pour garantir l'efficacité des modèles, ces bases de données doivent être adaptées aux spécificités des équipements et des conditions d'exploitation rencontrées.

Prétraitement des données

Le prétraitement des données est une étape cruciale pour transformer les données brutes en informations exploitables. Elle commence par un nettoyage des données pour éliminer les valeurs aberrantes et traiter les données incomplètes. Les valeurs manquantes peuvent être imputées à l'aide de moyennes, de zéros neutres, ou encore complétées par des

algorithmes plus avancés. Cette étape doit également tenir compte de la nature temporelle des données pour préserver les dynamiques de dégradation pertinentes.

Une fois nettoyées, les données sont transformées pour en extraire des caractéristiques pertinentes. Par exemple, face à de multiples signaux de capteurs corrélés (vibrations, températures), des techniques comme l'analyse en composantes principales (PCA) peuvent réduire la dimensionnalité en créant un indicateur de santé synthétique. Des algorithmes plus modernes comme les Auto-Encodeurs (AE) peuvent extraire automatiquement des caractéristiques non linéaires. Il s'agit de réseaux de neurones entraînés à reconstruire leurs propres données d'entrée, forçant l'apprentissage d'une représentation compressée et pertinente dans une couche intermédiaire. L'erreur de reconstruction peut alors servir d'indicateur de santé, particulièrement utile pour distinguer les régimes de fonctionnement normaux des dégradations fines.

Diagnostic

Le diagnostic est l'une des étapes principales de la maintenance prévisionnelle, visant à détecter, isoler et identifier les erreurs dans un système. Cela est souvent réalisé via des indicateurs de santé (IS), qui représentent l'état des équipements sous forme de valeurs numériques ou de classes. Les indicateurs de santé physiques (ISP) reposent sur des méthodes statistiques ou de traitement du signal qui synthétisent des signaux bruts complexes, comme les vibrations, en une seule métrique représentative de l'état de santé, telle que la moyenne quadratique des signaux vibratoires (HUANG et al., 2017; SOUALHI; MEDJAHER et ZERHOUNI, 2015a). En revanche, les indicateurs de santé virtuels (ISV) utilisent des algorithmes de fusion de données, tels que l'analyse en composantes principales ou des techniques de machine learning non supervisées. L'indicateur de santé peut être sous la forme d'une valeur numérique indiquant l'usure de l'équipement ou directement une catégorie explicite de son état. Le diagnostic offre une vue claire de l'état actuel des équipements et fournit des informations précieuses pour les phases suivantes.

Pronostic

Le pronostic, considéré comme l'objectif principal de la maintenance prévisionnelle, consiste à estimer l'état de santé futur ou la durée de fonctionnement avant défaillance (DFAD), souvent appelée *Remaining Useful Life* (RUL) (NEAGU ; SOLEIMANI et CAMPEAN, 2021). Cette étape repose sur la prévision de l'évolution de l'état de santé d'un équipement, en utilisant des données historiques ou des modèles physiques de dégradation. Lorsque

l'état de santé dépasse un seuil critique, l'équipement est considéré comme défaillant. Les méthodes de prévision peuvent inclure des approches basées sur les données, des modèles physiques ou une combinaison des deux, comme détaillé dans la section 1.2.1. Le pronostic permet d'anticiper les défaillances et de planifier efficacement les interventions.

Prise de décision

Une fois une panne future prédictive, la prise de décision intervient pour déterminer la meilleure réponse à celle-ci. Elle repose sur l'analyse des diagnostics et des pronostics, en tenant compte de paramètres tels que la criticité de la panne, les ressources disponibles, et les objectifs opérationnels (BOUSDEKIS et al., 2021). Les outils d'aide à la décision, tels que les arbres de décision (ARENA et al., 2022) ou les réseaux bayésiens dynamiques (SCHENKELBERG ; SEIDENBERG et ANSARI, 2020), sont souvent utilisés pour proposer des stratégies optimales. Ces méthodes permettent d'équilibrer le coût, le temps d'indisponibilité, et la fréquence des interventions de maintenance, tout en assurant la sécurité et la fiabilité des systèmes.

Interface Homme-Machine

L'interface Homme-Machine (IHM) est l'étape finale mais essentielle de la maintenance prévisionnelle. Elle assure la communication entre les systèmes automatisés et les utilisateurs en affichant les informations nécessaires à la prise de décision. Une IHM bien conçue doit être intuitive et inclure des visualisations claires de l'état des équipements, des prédictions de pannes et des recommandations d'actions. L'implication des utilisateurs finaux, tels que les techniciens de maintenance, est cruciale dès la phase de conception de ces interfaces pour s'assurer que les informations présentées sont directement exploitables. Leur expertise permet de définir les indicateurs les plus pertinents à visualiser (par ex., probabilité de panne, variables les plus influentes) et d'assurer que l'ergonomie de l'IHM est adaptée aux contraintes du terrain. Chaque étape de la maintenance prévisionnelle contribue de manière essentielle à une gestion proactive et optimisée des actifs industriels. La coordination entre ces étapes garantit des interventions ciblées et une réduction significative des temps d'arrêt, renforçant ainsi la compétitivité des entreprises. L'ensemble de ces étapes sera approfondi au fil des prochains chapitres, en détaillant les approches méthodologiques et les outils technologiques permettant leur mise en œuvre efficace. Ces développements permettront d'analyser en profondeur l'impact de chaque phase sur la performance globale de la maintenance prévisionnelle et sur son intégration

dans des environnements industriels complexes. L'ensemble de ces étapes, de l'acquisition des données à l'interface Homme-Machine, est synthétisé visuellement dans la Figure 1.2. Au-delà de son aspect méthodologique, la maintenance prévisionnelle génère des bénéfices concrets qui influencent directement la fiabilité des systèmes, les coûts d'exploitation et la durabilité des infrastructures. Il convient donc d'examiner les avantages majeurs qu'elle procure et leur influence sur la performance industrielle.

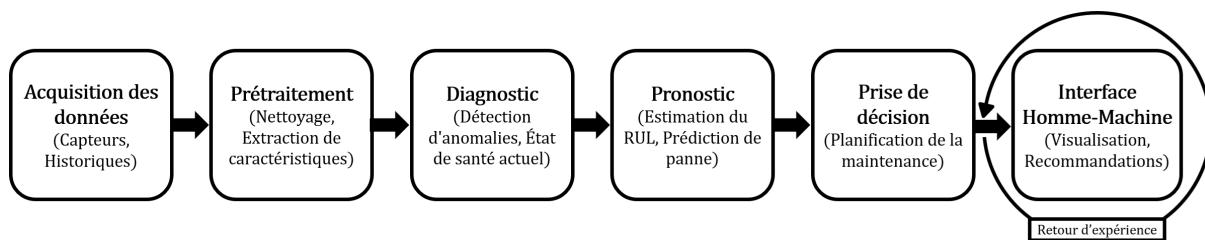


FIGURE 1.2 – Les étapes clés du processus de maintenance prévisionnelle.

1.1.3 Bénéfices de la maintenance prévisionnelle

La maintenance prévisionnelle permet d'optimiser la gestion des équipements en réduisant les interruptions non planifiées, en minimisant les coûts opérationnels et en améliorant la fiabilité des systèmes. Cette section explore ces bénéfices en détail.

Réduction des interruptions non planifiées et amélioration de la disponibilité

Les arrêts de production non planifiés représentent une source majeure de pertes économiques pour les entreprises industrielles. Une étude de Deloitte (SIEMENS, 2024) a révélé que, dans certaines industries, ces arrêts imprévus peuvent représenter jusqu'à 20% de la capacité de production annuelle, entraînant des pertes se chiffrant en milliards de dollars pour les plus grosses entreprises.

La maintenance prévisionnelle réduit significativement ces interruptions en détectant les anomalies avant qu'elles ne se transforment en pannes critiques (LEE et al., 2006). Dans l'industrie manufacturière, par exemple, des capteurs de vibration installés sur les machines-outils et les robots industriels permettent de repérer des déséquilibres ou des dégradations mécaniques à un stade précoce (KANAWADAY et SANE, 2017). Cela donne aux équipes de maintenance le temps nécessaire pour intervenir de manière ciblée, sans perturber le flux de production, réduisant ainsi les pertes de matière et les coûts liés aux interruptions.

Dans l'aéronautique, la maintenance prévisionnelle contribue à la sécurité et à la ponctualité des vols en surveillant en temps réel les moteurs et les systèmes hydrauliques des avions (LIN ; LUO et ZHONG, 2018). Les interventions peuvent être planifiées en amont, évitant ainsi des retards liés à des réparations d'urgence. De même, dans le secteur ferroviaire, les capteurs placés sur les bogies et les freins détectent l'usure prématuée des composants, permettant de planifier les opérations de maintenance hors des heures de pointe et d'améliorer la régularité des trains (DAVARI et al., 2021).

Optimisation des coûts opérationnels et gestion des ressources

L'optimisation des coûts est une priorité pour les industries cherchant à améliorer leur rentabilité tout en assurant la fiabilité de leurs équipements. La maintenance prévisionnelle contribue à cet objectif en évitant les dépenses inutiles associées aux interventions basées sur des calendriers fixes, comme c'est le cas avec la maintenance préventive (MOBLEY, 2002). Elle permet de n'intervenir que lorsque cela est nécessaire, réduisant ainsi les remplacements prématués d'équipements encore fonctionnels (FLORIAN ; SGARBOSSA et ZENNARO, 2021).

Dans le secteur maritime, où les navires opèrent souvent loin des infrastructures de maintenance, l'optimisation des interventions est un enjeu clé (ZHAO et YANG, 2018). Les systèmes embarqués surveillent les moteurs et les équipements de propulsion, permettant d'organiser les réparations durant les escales, limitant ainsi les coûts et les risques de pannes en mer.

De plus, la maintenance prévisionnelle facilite la gestion des stocks de pièces de rechange et l'affectation des ressources humaines. Dans les parcs éoliens, par exemple, l'accès difficile aux turbines offshore rend les interventions coûteuses (RINALDI et al., 2020). Grâce à une analyse continue des capteurs installés sur les pales et les générateurs, les entreprises peuvent planifier les interventions avec précision et limiter les déplacements inutiles des équipes techniques.

Amélioration de la fiabilité et de la sécurité des équipements

La fiabilité des infrastructures industrielles est un facteur essentiel pour assurer la continuité des opérations et la sécurité des employés et des usagers. Les défaillances de composants critiques peuvent avoir des conséquences graves, en particulier dans des secteurs à haut risque comme l'énergie, la pétrochimie et le transport.

Dans les centrales nucléaires et les raffineries, la maintenance prévisionnelle joue un rôle essentiel dans la prévention des accidents (MOLEDA et al., 2023). Elle permet d'identifier les signes de dégradation dans les pompes et compresseurs avant qu'ils ne provoquent des fuites ou des incendies (MOBLEY, 2002). Cette surveillance continue garantit un meilleur contrôle des infrastructures et réduit les risques pour les opérateurs.

De même, dans l'industrie agroalimentaire, la détection précoce des anomalies dans les lignes de production permet de prévenir les contaminations et d'assurer la conformité des produits aux normes sanitaires (LIU et al., 2022). Cela évite des rappels coûteux et renforce la confiance des consommateurs dans la qualité des produits.

Impact environnemental et durabilité

En optimisant la gestion des équipements, la maintenance prévisionnelle contribue également à la durabilité industrielle en réduisant le gaspillage et en prolongeant la durée de vie des machines. Elle favorise une utilisation plus rationnelle des ressources, limitant ainsi la production de déchets industriels et les coûts associés à leur traitement.

Dans le secteur des transports, la surveillance des moteurs et des systèmes de propulsion permet d'optimiser la consommation de carburant et de réduire les émissions de gaz à effet de serre (SIMÕES et al., 2017). Cette optimisation énergétique ne se traduit pas seulement par des économies financières, mais participe également à l'effort global de réduction de l'empreinte carbone des entreprises.

Enfin, dans les infrastructures énergétiques, la maintenance prévisionnelle améliore l'efficacité des installations en minimisant les pertes d'énergie (FAUSING OLESEN et SHAKER, 2020). Dans les centrales électriques et les réseaux de distribution, elle permet d'identifier les composants sous-performants et de les remplacer avant qu'ils ne deviennent une source de gaspillage.

Comme illustré dans le tableau 1.1, la maintenance prévisionnelle combine les avantages des approches corrective et préventive tout en minimisant leurs inconvénients respectifs. Elle permet une réduction significative des coûts, des interruptions et des risques liés aux défaillances, tout en améliorant la sécurité et l'efficacité énergétique des équipements industriels.

Toutefois, la mise en œuvre de la maintenance prévisionnelle présente plusieurs défis, notamment en raison de la complexité croissante des systèmes industriels et des exigences technologiques associées. La section suivante aborde ces défis en examinant les problématiques spécifiques aux environnements industriels modernes.

| Maintenance | Avantages | Inconvénients |
|-------------|--|--|
| Corrective | Interventions uniquement après défaillance Peu de planification nécessaire | Coût élevé en cas de panne Indisponibilité importante des équipements Nécessité d'une enquête approfondie pour identifier la cause Impact sur la production et la productivité |
| Préventive | Réduction des risques de pannes majeures Amélioration de la fiabilité des équipements Planification facilitée Facilite la gestion des ressources | Coût élevé dû à la fréquence des interventions Maintenances parfois inutiles sur des équipements en bon état Peut entraîner des interruptions programmées Moins réactive aux conditions réelles de fonctionnement |
| Préditive | Réduction des coûts et des interruptions Amélioration significative de la fiabilité Optimisation des interventions de maintenance Détection de défaillances avant leur apparition | Nécessite une infrastructure technologique avancée Dépendance à la qualité et à la quantité des données historiques Complexité d'implémentation Difficulté à prévoir des défaillances rares |

TABLE 1.1 – Comparaison des types de maintenance : avantages et inconvénients

1.2 Systèmes industriels complexes et maintenance prévisionnelle

1.2.1 La transition vers la maintenance prévisionnelle dans l'industrie : approches et facteurs facilitants

La transition vers la maintenance prévisionnelle dans l'industrie est le résultat d'avancées technologiques significatives et de la nécessité d'optimiser les performances des équipements. Elle repose sur quatre principales approches méthodologiques : l'approche basée sur la physique, l'approche basée sur les données, l'approche basée sur les connaissances et l'approche hybride. Ces approches, chacune ayant ses avantages et inconvénients, ont été renforcées par plusieurs facteurs facilitant leur mise en œuvre dans les environnements industriels modernes.

L'approche basée sur la physique utilise des modèles mathématiques pour représenter les mécanismes fondamentaux de dégradation des systèmes. Elle s'avère particulièrement efficace lorsque les phénomènes physiques à l'origine des défaillances sont bien compris, comme dans le cas des fissures ou des vibrations anormales. Des modèles classiques incluent l'équation de Paris-Erdogan pour la propagation des fissures (PARIS et ERDOGAN, 1963) et les filtres de Kalman pour l'estimation des paramètres système (BARALDI ; MANGILI et ZIO, 2012). Avec l'augmentation des capacités de calcul au cours des dernières décennies, l'utilisation de la méthode des éléments finis s'est généralisée pour l'étude de la propagation des dommages et la prédiction des défaillances, notamment dans les rotors (CLIMENTE-ALARCON et al., 2017) et les actionneurs électromagnétiques (LI et al., 2018a). Ces modèles permettent de réaliser des simulations numériques précises afin d'évaluer l'évolution d'un système ou d'un composant dans des conditions contrôlées. Cepen-

dant, leur précision dépend fortement de l'exhaustivité des modèles physiques utilisés et de leur capacité à prendre en compte les conditions réelles d'exploitation. Les systèmes techniques étant influencés par des facteurs environnementaux tels que la température ou la pression, la prise en compte de ces éléments reste un défi majeur (SI et al., 2011). Dans ce contexte, l'ajout de modèles complémentaires, voire l'hybridation avec d'autres approches basées sur les données, peut améliorer leur robustesse et leur applicabilité à des environnements industriels complexes.

L'approche basée sur les données exploite les volumes massifs d'informations générées par les capteurs industriels. Elle s'appuie sur des algorithmes de machine learning (ML) ou de deep learning (DL), capables de détecter des tendances et des anomalies dans les données. Les SVM (CORTES et VAPNIK, 1995), les Random Forest (BREIMAN, 2001) ou encore les réseaux neuronaux profonds comme les LSTM, spécifiquement adaptés à l'analyse des séries temporelles (YUAN ; WU et LIN, 2016), figurent parmi les outils couramment utilisés.

L'intégration des données issues des dispositifs IoT et des plateformes big data a également permis d'accroître l'efficacité et la précision de cette approche. Par exemple, des dispositifs intelligents installés sur des machines-outils à commande numérique permettent d'acquérir des données sur les conditions d'usure des outils et de prévoir leur durée de vie restante. Cependant, cette approche nécessite des volumes importants de données de qualité, ce qui peut être résolu grâce à des dispositifs IoT et des plateformes big data pour historiser une grande quantité de données. Cependant, quand ce n'est pas possible, cela peut poser des défis dans les environnements où les données historiques sont rares, bruitées ou censurées (NAGPAL et al., 2021). En dépit de ces limites, l'approche basée sur les données offre une adaptabilité et une performance accrues dans des environnements dynamiques, où les modèles traditionnels peinent à capturer les interactions complexes entre les variables.

L'approche basée sur les connaissances s'appuie sur l'expertise humaine traduite en règles, cas ou logiques floues. Cette méthodologie est particulièrement adaptée aux systèmes critiques où l'explicabilité des décisions est essentielle, comme dans les centrales électriques ou l'aéronautique (HUSSAIN et al., 2016). Par exemple, des systèmes experts basés sur le raisonnement "Si/Alors" permettent de simplifier la prise de décision dans des environnements à haut risque. Ces systèmes intègrent des bases de faits, des règles d'inférence et des moteurs d'interprétation pour générer des diagnostics précis (VINGERHOEDS et al., 1995). En outre, les approches basées sur des cas permettent de constituer des

bases de connaissances à partir d'expériences passées, facilitant la résolution rapide de problèmes similaires (SIKORSKA ; HODKIEWICZ et MA, 2011). Cependant, cette approche reste limitée dans sa capacité à effectuer des prédictions complexes, nécessitant souvent des volumes importants de connaissances initiales provenant d'experts. Son efficacité est également contrainte par la difficulté à formaliser des règles explicites dans des environnements très dynamiques ou incertains. C'est pourquoi cette approche est souvent hybridée, comme dans les modèles neuro-flous (ANFIS) (JANG, 1993) qui combinent des règles floues avec des réseaux de neurones pour améliorer la prédition tout en conservant une bonne explicabilité.

Enfin, l'approche hybride combine les forces des autres méthodes pour pallier leurs limites. Elle permet de coupler des modèles physiques avec des algorithmes basés sur les données ou les connaissances, offrant ainsi une combinaison puissante qui exploite les avantages de chaque méthodologie. Cette intégration améliore la précision des prédictions et l'interprétabilité des modèles tout en réduisant la dépendance à de grandes quantités de données (PENG et al., 2021). Par exemple, (ARIAS CHAO et al., 2022) propose une approche hybride combinant des réseaux neuronaux profonds avec des modèles physiques basés sur des filtres de Kalman pour prédire l'état de santé de moteurs d'avion. Cette approche hybride a permis d'améliorer l'horizon de prédition de près de 127% par rapport à l'approche uniquement basée sur les données tout en utilisant moins de données. De manière similaire, des modèles neuro-flous comme ANFIS, intégrant des règles floues et des réseaux neuronaux, permettent d'améliorer la prédition tout en offrant une bonne explicabilité (JANG, 1993). Cependant, la complexité accrue de ces méthodes nécessite une expertise spécialisée et des investissements importants dans la conception et la mise en œuvre des modèles.

Parallèlement à ces approches, le domaine de la fiabilité a développé des modèles statistiques spécifiques pour quantifier l'impact des interventions de maintenance. Dépassant les hypothèses simplistes d'une réparation parfaite ("Comme Neuf", AGAN) ou nulle ("Aussi Mauvais que Vieux", ABAO), des modèles de réparation imparfaite comme l'Arithmetic Reduction of Age (ARA) ou l'Arithmetic Reduction of Intensity (ARI) (DOYEN et GAUDOIN, 2004) permettent de modéliser un "âge virtuel" pour un composant après une réparation. De même, des cadres mathématiques comme les Processus Markoviens Déterministes par Morceaux (PDMP) (DAVIS, 1984) offrent une alternative robuste pour les systèmes hybrides, en modélisant une dégradation continue qui est interrompue par des événements brusques et aléatoires comme des pannes ou des interventions. Bien que ces

modèles soient puissants pour des problématiques ciblées, ils n'ont pas été retenus comme approche centrale dans cette thèse. Les modèles ARA/ARI se focalisent sur l'effet post-maintenance plutôt que sur le diagnostic prédictif, tandis que les PDMP requièrent une modélisation physique fine de la dégradation. Notre méthodologie privilégie la flexibilité et l'interprétabilité des réseaux bayésiens pour gérer l'incertitude inhérente aux données incomplètes et aux dépendances complexes, typiques de notre cas d'étude.

L'intégration de capteurs intelligents et de dispositifs IoT a joué un rôle clé dans la transition vers la maintenance prévisionnelle. Ces technologies permettent de collecter des données en temps réel, offrant une visibilité globale sur l'état des équipements et facilitant la détection précoce des anomalies. Des capteurs de vibration, de température ou de pression, connectés via des plateformes IoT, sont couramment utilisés pour surveiller en continu les systèmes industriels (ÇINAR et al., 2020). Par exemple, les dispositifs intelligents installés sur des machines-outils à commande numérique permettent d'acquérir des données sur les conditions d'usure des outils et de prévoir leur durée de vie restante (JIMENEZ-CORTADI et al., 2020). Cette anticipation permet aux équipes de maintenance d'intervenir de manière ciblée sans perturber le flux de production, réduisant ainsi les pertes de matière et les coûts liés aux interruptions. De plus, les avancées dans le traitement des données, grâce à l'essor des plateformes de big data et des systèmes d'analyse avancée, permettent de traiter des volumes massifs d'informations avec une précision accrue.

Les entreprises adoptent de plus en plus ces systèmes pour réduire les coûts de maintenance, améliorer la durabilité des équipements et optimiser leur efficacité opérationnelle. Selon (ÇINAR et al., 2020), l'utilisation de la maintenance prévisionnelle peut réduire les temps d'arrêt non planifiés de 30 à 50 %, augmentant ainsi la productivité et réduisant les coûts d'exploitation. Cependant, cette transition n'est pas sans défis. Elle nécessite des investissements significatifs pour l'acquisition de capteurs, l'installation d'infrastructures IoT et la formation des équipes aux nouvelles technologies. Les entreprises doivent également relever des défis liés à la gestion des données, notamment leur qualité, leur confidentialité et leur sécurité. Les systèmes industriels modernes doivent se conformer à des réglementations strictes en matière de cybersécurité, ce qui ajoute une couche de complexité au déploiement des infrastructures IoT (GRABOT, 2020).

En dépit de ces obstacles, la maintenance prévisionnelle représente une avancée stratégique pour les industries cherchant à rester compétitives dans un environnement de plus en plus exigeant. Les approches méthodologiques et les technologies facilitant cette

transition ouvrent la voie à une gestion proactive et optimisée des actifs industriels. Elles permettent de mieux anticiper les défaillances, de maximiser la durée de vie des équipements et de répondre aux exigences croissantes en matière de durabilité et d'efficacité énergétique.

1.2.2 Complexité des systèmes industriels

Les systèmes industriels modernes se distinguent par leur complexité croissante, un phénomène directement lié à l'évolution rapide des technologies et à la diversification des secteurs industriels. Cette complexité découle à la fois de la variété des composants présents dans ces systèmes et des interdépendances dynamiques qui les relient (BROCAL et al., 2019). Chaque machine, capteur ou système de contrôle joue un rôle crucial, mais ces éléments ne fonctionnent pas de manière isolés. Les interactions fréquentes entre ces composants forment des réseaux hautement intégrés, où une perturbation localisée peut rapidement se propager à l'ensemble du système. Ce phénomène de dépendance fonctionnelle, crée un environnement dans lequel les effets en cascade peuvent avoir des conséquences majeures sur la production et la sécurité (LADYMAN ; LAMBERT et WIESNER, 2013).

Les environnements industriels modernes sont également confrontés à une explosion des données générées par les dispositifs connectés. L'intégration de technologies comme l'Internet des Objets (IoT) et les systèmes SCADA (Supervisory Control and Data Acquisition) a permis une collecte massive d'informations en temps réel sur l'état des équipements, les conditions opérationnelles et les performances des processus (SAJID ; ABBAS et SALEEM, 2016).

Cependant, cette abondance de données pose des défis majeurs. L'hétérogénéité des sources de données et la variabilité de leur qualité compliquent considérablement leur interprétation, notamment en raison des interactions entre de nombreux sous-systèmes autonomes (HMELO-SILVER et AZEVEDO, 2006). Ces interactions à plusieurs niveaux rendent la modélisation et l'analyse plus complexes, car les systèmes industriels évoluent de manière dynamique et ne suivent pas toujours des lois de transition simples (WILENSKY et RESNICK, 1999). En effet, les systèmes industriels modernes sont souvent caractérisés par des connexions hiérarchiques en réseau et des dynamiques non linéaires qui peuvent conduire à des comportements inattendus (SAN MIGUEL et al., 2012).

Un autre facteur déterminant est l'incertitude liée aux environnements industriels. Les systèmes sont soumis à des fluctuations imprévisibles des conditions opérationnelles, telles

que les variations de température, les vibrations excessives ou encore des perturbations dans les chaînes d'approvisionnement. Ces incertitudes influencent non seulement la performance des équipements, mais introduisent également une dépendance aux conditions initiales qui complique toute tentative de prédition fiable (WILENSKY et RESNICK, 1999). Les comportements auto-organisés de certains systèmes complexes accentuent cette difficulté, rendant l'anticipation des défaillances particulièrement délicate (HMELO-SILVER et AZEVEDO, 2006).

En conséquence, la complexité croissante des systèmes industriels a conduit les entreprises à repenser leurs approches pour mieux gérer ces environnements exigeants. Cette réflexion a profondément influencé l'évolution des pratiques de maintenance, qui se sont adaptées pour répondre aux défis spécifiques posés par ces systèmes modernes.

1.2.3 Défis et problématiques de la maintenance prévisionnelle

Bien que la maintenance prévisionnelle représente une avancée majeure dans l'optimisation de la gestion des équipements industriels, son implémentation soulève des défis importants. Ces défis sont particulièrement accentués dans les environnements industriels complexes, où les interactions entre systèmes, la variabilité des conditions d'exploitation et les contraintes technologiques rendent son adoption difficile. Les principales problématiques rencontrées peuvent être regroupées en trois grandes catégories : la gestion des données, les défis technologiques et les contraintes organisationnelles.

Gestion des données et limitations des modèles prédictifs

La maintenance prévisionnelle repose sur l'exploitation de données collectées en temps réel, mais la disponibilité et la qualité de ces données constituent un frein majeur. L'efficacité des modèles d'analyse dépend de la quantité et de la pertinence des données historiques utilisées pour entraîner les algorithmes prédictifs. Or, dans de nombreux environnements industriels, les jeux de données sont souvent incomplets ou biaisés, ce qui peut nuire à la robustesse et à la fiabilité des prédictions (COMPARE ; BARALDI et ZIO, 2019). De plus, les capteurs industriels, bien que de plus en plus sophistiqués, ne fournissent pas toujours des mesures exemptes de bruit ou d'erreurs de transmission. Les interférences électromagnétiques, les pannes des capteurs eux-mêmes ou encore la mauvaise calibration des instruments de mesure peuvent entraîner des données corrompues ou incohérentes.

L'un des paradoxes de la maintenance prévisionnelle réside dans la nécessité de dispo-

ser d'exemples de défaillances pour entraîner les modèles, alors que son objectif principal est précisément de les éviter. Ainsi, dans certains secteurs où les pannes sont rares mais critiques, comme l'aéronautique ou le nucléaire, la faiblesse du nombre d'occurrences de défaillances complique l'élaboration de modèles fiables (BOKRANTZ et al., 2017). De plus, le stockage et le traitement de ces volumes de données massifs nécessitent des infrastructures adaptées, telles que des plateformes de big data et des solutions cloud, qui impliquent des investissements significatifs.

Contraintes technologiques et compatibilité avec les systèmes existants

L'intégration de la maintenance prévisionnelle dans les environnements industriels existants constitue un autre défi majeur. De nombreuses infrastructures ont été conçues bien avant l'émergence des technologies IoT et des modèles de maintenance avancés, rendant leur adaptation complexe. L'ajout de capteurs et de dispositifs de surveillance nécessite souvent une refonte des équipements et des réseaux de communication, ce qui engendre des coûts élevés et pose des problèmes de compatibilité avec les systèmes existants.

Dans les installations plus anciennes, la modernisation passe par plusieurs étapes, notamment :

- **L'installation de capteurs et de dispositifs IoT** sur des machines initialement non instrumentées pour collecter des données en continu.
- **La mise en place d'infrastructures de communication** adaptées pour assurer la transmission des données vers des plateformes d'analyse.
- **L'adoption de systèmes d'analyse avancés** capables de traiter les données en temps réel et de fournir des prédictions exploitables.
- **La migration vers des architectures cloud** pour stocker et centraliser les informations tout en garantissant leur accessibilité.

Toutefois, ces évolutions s'accompagnent de défis techniques, notamment en termes de cybersécurité et de fiabilité des infrastructures numériques. L'augmentation du volume de données échangées expose les entreprises à des risques accrus de cyberattaques, nécessitant une mise en place rigoureuse de protocoles de sécurité. Par ailleurs, les prédictions fournies par les modèles de maintenance doivent être suffisamment précises et adaptées aux dynamiques des systèmes industriels, ce qui représente un défi dans les environnements soumis à de fortes variabilités.

Enjeux organisationnels et économiques

L’adoption de la maintenance prévisionnelle ne repose pas uniquement sur des considérations technologiques, mais également sur des transformations organisationnelles profondes. Les méthodes de maintenance traditionnelles, qu’elles soient correctives ou préventives, sont bien établies et largement adoptées dans les industries. L’introduction de la maintenance prévisionnelle implique une refonte des pratiques existantes, ce qui peut se heurter à des résistances au changement. Les équipes de maintenance doivent être formées aux nouveaux outils et à l’interprétation des analyses prédictives, ce qui nécessite un investissement en formation et en accompagnement des techniciens (BOKRANTZ et al., 2017).

Un autre défi réside dans la coordination entre les différents départements d’une entreprise. L’exploitation efficace des analyses prédictives demande une collaboration étroite entre les services de maintenance, de production et d’ingénierie, ce qui peut complexifier la gestion des interventions et nécessiter des adaptations dans les processus décisionnels. Par ailleurs, l’implémentation de la maintenance prévisionnelle peut donner lieu à des interventions de maintenance anticipées qui, bien que justifiées par des modèles analytiques, peuvent être perçues comme inutiles si elles ne sont pas accompagnées de justifications claires.

De plus, d’un point de vue économique, les coûts initiaux de mise en place de la maintenance prévisionnelle peuvent être prohibitifs pour certaines entreprises, en particulier les petites et moyennes structures. L’installation de capteurs, l’achat de logiciels d’analyse et la mise en place d’infrastructures adaptées représentent des investissements conséquents (COMPARE ; BARALDI et ZIO, 2019). Même si ces coûts peuvent être amortis à long terme grâce aux économies réalisées sur la maintenance et les temps d’arrêt, ils constituent souvent un frein à l’adoption rapide de cette approche.

Enfin, les réglementations et normes en vigueur dans certaines industries peuvent également ralentir l’adoption de la maintenance prévisionnelle. Dans des secteurs où la sécurité est un enjeu primordial, comme l’industrie chimique ou le nucléaire, l’introduction de nouvelles technologies de surveillance et d’analyse peut être limitée par des exigences réglementaires strictes. L’expérimentation et la validation des modèles prédictifs doivent respecter des processus rigoureux, ce qui peut allonger le temps nécessaire à leur déploiement.

Malgré les avancées notables qu’offre la maintenance prévisionnelle, son implantation efficace reste conditionnée par la qualité et la disponibilité des données, la compati-

bilité avec les infrastructures existantes et l'adhésion des équipes aux nouvelles pratiques de maintenance. L'adaptation des environnements industriels aux exigences de la maintenance prévisionnelle implique non seulement des investissements technologiques et organisationnels, mais aussi une réévaluation des méthodologies traditionnelles pour mieux les intégrer aux nouvelles approches basées sur les données.

Les défis mentionnés prennent une ampleur encore plus marquée dans les systèmes industriels complexes, notamment lorsque l'acquisition des données ne se fait pas en continu mais de manière ponctuelle, comme c'est le cas dans notre étude. Dans de telles configurations, la collecte d'informations n'est pas systématique, rendant difficile la construction de modèles prédictifs robustes. L'absence de mesures continues, la variabilité des sous-systèmes et l'hétérogénéité des structures des équipements soulèvent des problématiques spécifiques qui nécessitent des solutions adaptées, allant au-delà des méthodologies conventionnelles de la maintenance prévisionnelle.

La section suivante explore en détail le cas particulier de notre étude, où l'acquisition des données est restreinte aux périodes de maintenance et où la structure du système introduit des défis supplémentaires liés à la disponibilité, la fiabilité et l'interprétabilité des informations collectées. Ces contraintes spécifiques influencent directement les stratégies à adopter pour mettre en place une maintenance prévisionnelle efficace dans un tel contexte.

1.2.4 Cas particulier de l'étude

Dans de nombreux environnements industriels, la maintenance prévisionnelle repose sur la surveillance continue des systèmes grâce à des capteurs mesurant divers paramètres en temps réel. Cependant, certains systèmes ne permettent pas une telle surveillance, notamment les infrastructures anciennes ou complexes qui n'ont pas été conçues pour intégrer cette approche. Notre étude se concentre sur une typologie spécifique de systèmes embarqués, dont la nature dicte les contraintes uniques d'acquisition de données qui motivent cette thèse. Dans ce contexte, le terme "embarqué" ne désigne pas le sens habituel de dispositifs connectés transmettant des données en continu, comme dans l'Internet des Objets (IoT). Il se réfère plutôt à des systèmes complexes et autonomes intégrés au sein d'une plateforme plus large (telle qu'un aéronef, un navire ou un véhicule), rendant toute télémétrie continue impossible ou impraticable durant les phases opérationnelles.

L'exemple classique est celui d'un turboréacteur sur un avion : les données de performance détaillées et les diagnostics approfondis ne peuvent être collectés que lorsque l'avion est au sol pour des opérations de maintenance. Cette réalité opérationnelle est

la source directe des défis illustrés par la Figure 1.3 : l’acquisition des données n’est ni continue, ni régulière, mais ponctuelle et opportuniste.

Même si ces interventions sont planifiées, l’acquisition de données reste structurellement irrégulière. En effet, comme cela sera détaillé plus loin, le plan de maintenance n’exige pas que l’ensemble des contrôles soit réalisé à chaque visite, entraînant ainsi une complétude variable des données d’un passage à l’autre.

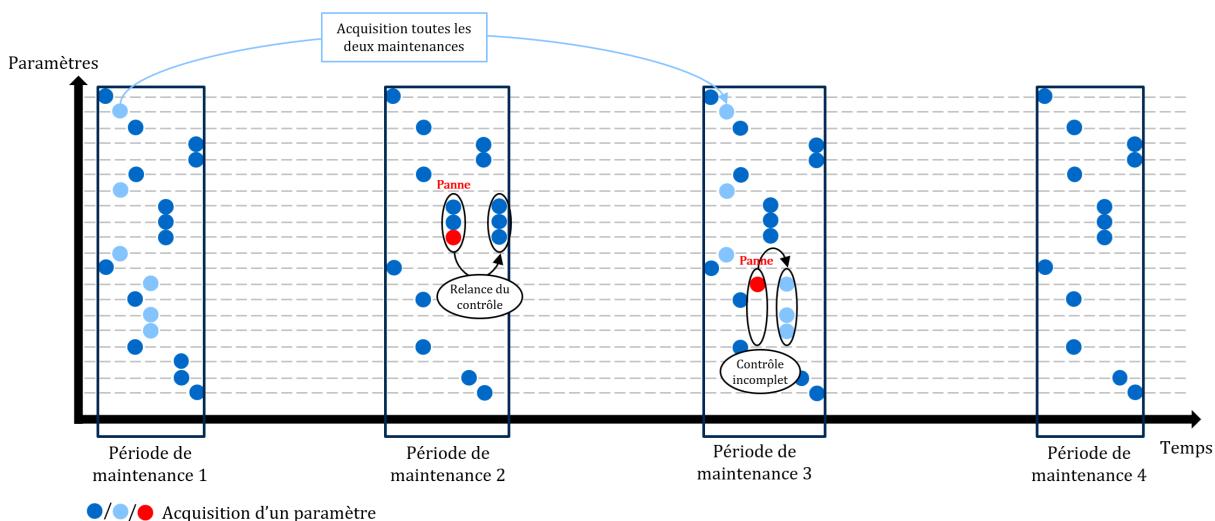


FIGURE 1.3 – Illustration des contraintes d’acquisition des données dans le cas d’étude : acquisition ponctuelle et irrégulière, données manquantes et rareté des pannes.

Contrairement aux systèmes équipés de télémétrie continue, les données de ces équipements ne peuvent être collectées que lors du retour au site de maintenance. À titre d’exemple, considérons un système aéronautique : les mesures de performance ou de dégradation ne peuvent être acquises que lors des opérations de maintenance au sol, et non pendant le vol. Pour récupérer ces mesures et valider le bon fonctionnement du système, il est nécessaire de dérouler des contrôles spécifiques qui mesurent certaines grandeurs clés. Les défis majeurs proviennent de la nature même de la collecte des données. Celle-ci est ponctuelle et structurellement irrégulière, car elle est réalisée uniquement lors des phases de maintenance, contrairement aux systèmes équipés d’une surveillance en continu. Cette acquisition intermittente complique la détection des tendances de dégradation et des signaux faibles annonciateurs de défaillances, rendant difficile l’établissement de modèles prédictifs fiables. Il en résulte des problématiques de temporalité, avec des informations fragmentées, mais aussi de complétude, avec de nombreuses données manquantes. De plus, le volume de données historiques est limité et contient une très faible proportion de pannes,

ce qui constitue un obstacle majeur pour l'entraînement des modèles. La maintenance de ces systèmes repose donc sur des interventions régulières, lors desquelles des contrôles spécifiques sont réalisés pour évaluer le bon fonctionnement et déterminer si une réparation ou un remplacement est nécessaire. Le processus n'est considéré comme achevé que lorsque toutes les chaînes d'équipements respectent les limites de fonctionnement définies. Cette méthode d'acquisition des données constitue la base de cette étude.

L'irrégularité de l'acquisition se manifeste aussi par le fait que toutes les données décrivant un équipement ne sont pas systématiquement acquises à chaque intervention de maintenance. En fonction du plan de maintenance défini par les experts, certains contrôles ne sont pas déroulés à chaque période de maintenance, ces équipements étant jugés assez fiables pour être moins vérifiés que d'autres. Cependant, cela entraîne un déséquilibre entre les équipements dans les données disponibles et une irrégularité d'acquisition entre les contrôles. Ces informations manquantes compliquent l'entraînement des modèles prédictifs, qui nécessitent généralement des bases de données complètes et homogènes pour fonctionner efficacement. Dans ce contexte, il devient indispensable d'utiliser des techniques avancées pour traiter les données manquantes. Au-delà des méthodes d'imputation classiques (statistiques, bayésiennes), des approches plus robustes comme les algorithmes de maximisation de la vraisemblance (par ex. Expectation-Maximization - EM) ou les modèles génératifs sont de plus en plus utilisées pour gérer l'incertitude inhérente à ces données incomplètes (REN et al., 2023). Un autre défi majeur réside dans le volume limité de données disponibles. Dans des systèmes où les données sont abondantes et diversifiées, il est possible de concevoir des modèles robustes, capables de généraliser efficacement les comportements des équipements. En revanche, dans le cas étudié, la rareté des données engendre un risque de sur-apprentissage, où les modèles deviennent trop spécifiques aux observations disponibles et peinent à généraliser à de nouveaux cas (HAKAMI, 2020). Ce manque de diversité empêche également d'identifier correctement l'ensemble des modes de défaillance potentiels, réduisant ainsi l'efficacité des modèles prédictifs. Sans un volume suffisant d'exemples représentatifs des différentes conditions d'exploitation et de vieillissement des équipements, les algorithmes risquent d'être inefficaces pour détecter des pannes complexes.

Enfin, la faible proportion de pannes dans les données constitue une contrainte supplémentaire. Dans la majorité des cas, les défaillances apparaissent rarement dans l'historique disponible, et certains types de pannes ne se sont jamais produits. Cette situation complique l'entraînement des modèles de machine learning, qui reposent sur des corrélations

entre les données d'entrée et les événements de défaillance. Avec un nombre insuffisant d'exemples de pannes réelles, les modèles risquent soit de ne pas reconnaître les signes avant-coureurs d'une future panne, soit de générer un grand nombre de fausses alertes, entraînant des interventions inutiles et une augmentation des coûts opérationnels. De plus, les pannes rares qui ne sont jamais survenues dans l'historique sont impossibles à anticiper avec des méthodes supervisées classiques, ce qui rend nécessaire l'utilisation d'approches alternatives comme l'apprentissage par transfert, la génération de données synthétiques ou des modèles hybrides combinant expertise métier et intelligence artificielle.

L'ensemble de ces défis a un impact direct sur la mise en place d'une stratégie de maintenance prévisionnelle efficace. L'irrégularité et l'incomplétude des données rendent plus difficile la construction d'un indicateur de santé fiable et interprétable. De plus, l'entraînement des modèles est limité par la faible représentativité des pannes, ce qui impose le développement de nouvelles approches pour pallier ces insuffisances. Dans ce contexte, l'étude de ce cas particulier représente une opportunité pour développer des solutions innovantes qui pourraient être généralisées à d'autres secteurs industriels confrontés à des contraintes similaires, comme l'automobile, l'aérospatial ou les infrastructures critiques.

L'ensemble de ces problématiques sera traité en détail dans les chapitres suivants. Différentes approches méthodologiques seront explorées afin d'atténuer les limitations liées à l'acquisition irrégulière des données, de pallier l'absence de suivi continu et d'optimiser les stratégies de modélisation pour améliorer la fiabilité des prédictions. Ces développements permettront d'adapter la maintenance prévisionnelle aux contraintes spécifiques du système étudié et d'ouvrir la voie à des solutions généralisables à d'autres environnements industriels complexes. L'objectif est de démontrer que, malgré ces défis, il est possible de tirer parti des données existantes pour améliorer la fiabilité des équipements et réduire les interventions correctives.

Conclusion

La maintenance prévisionnelle s'impose aujourd'hui comme une solution clé pour améliorer la fiabilité et la performance des systèmes industriels. Grâce aux avancées technologiques en matière de capteurs intelligents, de machine learning et de traitement des données, elle permet une surveillance proactive des équipements, réduisant ainsi les coûts et les interruptions non planifiées. Contrairement aux approches correctives et préventives, elle optimise les interventions en ciblant précisément les équipements nécessitant

une maintenance, garantissant une efficacité opérationnelle accrue.

Toutefois, sa mise en œuvre reste complexe, notamment en raison des défis liés à la gestion des données, aux coûts d'implémentation et à la nécessité d'une expertise spécialisée. L'étude de notre cas particulier met en évidence ces contraintes dans un environnement où l'acquisition des données est irrégulière et incomplète. L'absence de surveillance continue, la présence de données manquantes et la rareté des pannes observées posent des obstacles majeurs pour le développement de modèles prédictifs robustes. Ces problématiques nécessitent l'adaptation des méthodologies existantes et l'exploration de nouvelles approches pour exploiter efficacement les informations disponibles.

La maintenance prévisionnelle repose ainsi sur une compréhension approfondie du fonctionnement des systèmes industriels et sur l'exploitation intelligente des données collectées. Cette analyse passe nécessairement par la détermination d'un état de santé pertinent et représentatif des équipements, un processus que nous approfondirons dans le chapitre suivant.

CRÉATION D'UN INDICATEUR D'ÉTAT DE SANTÉ PAR DISCRÉTISATION : LA MÉTHODE MSLD

Introduction

Le Chapitre 1 a posé les fondements et les enjeux de la maintenance prévisionnelle, en soulignant les défis liés à son application en contexte industriel. Pour mettre en œuvre une telle stratégie, la définition d'un indicateur fiable et interprétable de l'état de santé est une première étape méthodologique indispensable. Ce chapitre se consacre à l'élaboration de cet indicateur, en se concentrant sur les méthodes de structuration et de discréétisation des données. Nous explorerons les méthodes de structuration des données, en mettant l'accent sur la discréétisation pour améliorer l'interprétabilité et la précision des diagnostics.

La création d'un état de santé pertinent et exploitable représente un défi majeur, notamment dans un contexte industriel réel (HENG et al., 2009). Les données industrielles sont souvent hétérogènes, incomplètes et collectées de manière irrégulière, ce qui complique l'établissement d'une représentation fidèle de la dégradation des équipements. De plus, les déséquilibres dans les jeux de données, où les occurrences de défaillance sont rares par rapport aux états normaux, rendent difficile l'entraînement de modèles de diagnostic précis et fiables. La discréétisation constitue une approche prometteuse pour transformer des variables continues en catégories discrètes, facilitant ainsi la création d'états de santé interprétables.

Ce chapitre se concentre sur les fondements théoriques et méthodologiques permettant de déterminer comment discréétiser efficacement les données pour construire un état de santé robuste. Après avoir défini l'état de santé et son rôle dans la maintenance prévisionnelle, nous explorerons différentes approches de création, en mettant en lumière leurs forces et faiblesses. Nous introduirons ensuite la méthode MSLD (Multivariable System

Labeling through Discretization), développée spécifiquement dans le cadre de cette étude pour surmonter les limites des approches existantes. Le processus complet de création d'un état de santé via MSLD sera détaillé, montrant comment cette méthode permet de structurer les données complexes pour améliorer le diagnostic et la prédition des pannes. Ce travail vise à établir une base solide pour les analyses et résultats qui seront présentés dans les chapitres suivants.

2.1 Contexte et méthodologies de création d'un état de santé

2.1.1 Définition et rôle d'un état de santé

Un état de santé, ou indicateur de santé, est une représentation synthétique et interprétable de la condition d'un équipement ou d'un système (LEI et al., 2018a). Il est utilisé pour identifier les signes précurseurs de défaillance ou de dysfonctionnement et peut prendre diverses formes, telles qu'une valeur numérique, des états discrets ou des catégories qualitatives (bon, moyen, mauvais). Le rôle opérationnel de cet indicateur est de transformer des données de capteurs, souvent complexes et multidimensionnelles, en une information synthétique qui guide directement les actions de maintenance, comme l'illustre la Figure 2.1.

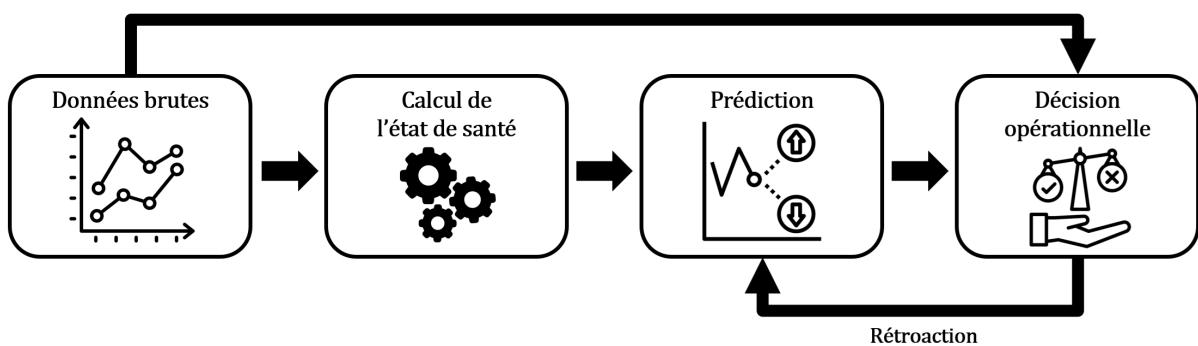


FIGURE 2.1 – Cycle opérationnel de l'état de santé : de la donnée brute à la décision de maintenance.

La mise en œuvre d'un suivi efficace des états de santé repose sur diverses approches adaptées aux spécificités des systèmes industriels. Les méthodes traditionnelles, souvent basées sur des modèles physiques, fournissent des diagnostics précis pour des systèmes

bien compris (HANACHI et al., 2015). Cependant, elles montrent leurs limites lorsque les interactions entre les paramètres deviennent complexes ou difficilement modélisables, ce qui oblige souvent à recourir à des approximations de la réalité dans les modèles physiques.

Avec l'essor des capteurs IoT et des systèmes d'acquisition avancés, les volumes massifs de données collectés permettent désormais d'utiliser des algorithmes d'apprentissage automatique et des techniques de traitement de données massives (*Big Data*) pour identifier des schémas complexes de dégradation. Ces méthodes facilitent la détection de signaux faibles, annonciateurs de défaillances potentielles, et améliorent considérablement la précision des diagnostics. Ainsi, la surveillance de l'état de santé est devenue un enjeu majeur pour garantir la fiabilité et la continuité des opérations industrielles.

Toutefois, la construction d'un état de santé pertinent repose sur des méthodologies variées, adaptées aux caractéristiques spécifiques des équipements et aux contraintes du contexte industriel. Les méthodes utilisées peuvent s'appuyer sur des modèles physiques, exploitant les lois de la mécanique ou de la thermodynamique, ou sur des approches basées sur les données, exploitant la richesse des informations issues des capteurs et des historiques de fonctionnement. La section suivante détaille ces différentes approches et analyse leurs avantages et leurs limites en fonction des applications industrielles.

2.1.2 Méthodes de création d'un état de santé

Les méthodes de création d'un état de santé peuvent être regroupées en deux grandes catégories : les approches basées sur des modèles physiques et celles fondées sur l'exploitation des données (HU et al., 2012). Chacune de ces approches présente des avantages et des limites qui influencent leur application selon le contexte industriel.

Approches basées sur des modèles physiques

Les approches basées sur des modèles physiques reposent sur la modélisation mathématique des mécanismes fondamentaux de dégradation des systèmes. Elles s'appuient sur des lois physiques connues pour établir des relations directes entre les paramètres mesurés et l'état de santé de l'équipement.

Un exemple classique est l'utilisation de l'équation de Paris-Erdogan pour modéliser la propagation des fissures dans les matériaux (PUGNO et al., 2006). De même, l'analyse vibratoire est couramment utilisée pour détecter des déséquilibres ou des anomalies mécaniques à partir des signatures fréquentielles des signaux collectés (SOUALHI ;

MEDJAHER et ZERHOUNI, 2015b). Dans certains cas, des indicateurs de santé physiques simples sont dérivés de mesures directes, comme la moyenne quadratique des signaux vibratoires (HUANG et al., 2017) ou l'évolution des paramètres de température et de pression.

Ces méthodes présentent plusieurs avantages. Elles offrent une forte interprétabilité et permettent de comprendre précisément les mécanismes de dégradation en jeu. De plus, dans les environnements critiques tels que l'aéronautique ou le nucléaire, ces modèles sont souvent privilégiés pour leur robustesse et leur fiabilité.

Cependant, leur mise en œuvre suppose une connaissance approfondie des phénomènes sous-jacents, ce qui n'est pas toujours le cas dans les systèmes complexes. La complexité croissante des équipements modernes rend également ces modèles parfois difficiles à calibrer.

Approches basées sur les données

Contrairement aux approches fondées sur des modèles physiques, les méthodes basées sur les données exploitent directement les informations collectées pour analyser l'état de santé des équipements. Avec l'essor des capteurs intelligents et des systèmes connectés, ces approches ont pris une place centrale dans la création d'états de santé. Elles permettent d'identifier des anomalies, de détecter des tendances de dégradation et d'anticiper les pannes potentielles.

Une stratégie courante consiste à générer des indicateurs de santé virtuels (*Virtual Health Indicators* - VHI) par fusion de multiples signaux ou mesures (LEI et al., 2018b). Ces indicateurs sont calculés à l'aide de techniques comme l'analyse en composantes principales (PCA), qui réduit la dimensionnalité en projetant les données sur de nouveaux axes orthogonaux capturant le maximum de variance. Des méthodes de clustering sont aussi utilisées pour réduire la dimensionnalité des données tout en conservant l'essentiel de l'information (KURREWAR et al., 2021).

Les algorithmes d'apprentissage non supervisés sont également utilisés pour segmenter les états de santé. Parmi eux, le regroupement par k-means (LLOYD, 1982) vise à partitionner un ensemble de N observations $\{x_1, \dots, x_N\}$ en k clusters $C = \{C_1, \dots, C_k\}$. L'objectif est de minimiser l'inertie intra-classe, définie par l'équation suivante :

$$\arg \min_C \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (2.1)$$

où μ_i est le centroïde du cluster C_i .

Les modèles de mélange gaussien (GMM) (REYNOLDS et al., 2009), quant à eux, modélisent les données comme une combinaison de plusieurs distributions gaussiennes, permettant une assignation probabiliste des points aux clusters. Les approches supervisées, quant à elles, exploitent des jeux de données annotés pour construire des modèles prédictifs capables d'anticiper l'évolution des équipements. Cependant, l'annotation des données constitue un défi majeur, car elle revient à définir l'état de santé de l'équipement à partir de critères souvent subjectifs ou incomplets. Dans de nombreux cas, ces annotations sont obtenues soit par expertise humaine (classification des équipements sur la base d'examens visuels ou de mesures spécifiques), soit par l'utilisation de seuils empiriques définis à partir de l'historique des défaillances connues.

La qualité des annotations influence directement la performance des modèles prédictifs (BUDACH et al., 2022). Des annotations bruitées ou imprécises entraînent des prédictions erronées et réduisent l'efficacité des stratégies de maintenance prévisionnelle. Par ailleurs, l'absence d'un consensus clair sur la définition de l'état de santé peut conduire à des incohérences dans l'apprentissage, compromettant la fiabilité des modèles.

Les approches basées sur les données offrent plusieurs avantages, notamment une capacité à capturer des interactions complexes entre variables et à s'adapter aux environnements dynamiques. Cependant, elles nécessitent une structuration rigoureuse des données et une méthodologie robuste d'annotation pour garantir la pertinence des états de santé générés.

Les approches de création d'états de santé présentent chacune des avantages et des limites selon le contexte industriel. Les méthodes basées sur les modèles physiques offrent une forte interprétabilité et sont adaptées aux environnements critiques, tandis que les approches basées sur les données apportent une flexibilité accrue et permettent d'exploiter des relations complexes entre variables.

Dans notre cas d'étude, la faible quantité de données, leur acquisition irrégulière et la complexité des interactions entre sous-systèmes limitent l'application directe de ces méthodes. Les modèles physiques deviennent difficiles à calibrer, tandis que les approches basées sur les données nécessitent des adaptations pour pallier l'absence de mesures continues et garantir l'explicabilité des indicateurs. La section suivante détaille ces limitations et analyse leurs impacts sur la création d'états de santé dans notre contexte.

2.1.3 Limites des approches existantes pour la création d'état de santé

L'application des méthodes classiques pour la création d'état de santé présente plusieurs limites lorsqu'elles sont appliquées à notre cas d'étude. Les spécificités du système analysé, telles que la faible quantité de données disponibles, l'acquisition irrégulière et partielle des mesures, la complexité structurelle des sous-systèmes, et l'exigence d'interprétabilité, soulèvent des défis majeurs pour les approches traditionnelles, qu'elles soient basées sur des modèles physiques ou des données. Cette section explore en détail ces contraintes et leurs impacts sur la création d'états de santé dans notre contexte.

Contraintes liées à la faible quantité et à l'acquisition irrégulière des données

Comme détaillé au Chapitre 1, notre cas d'étude se caractérise par une acquisition de données ponctuelle et irrégulière, réalisée uniquement lors des interventions de maintenance. Cette contrainte fondamentale, qui limite la représentation continue de l'évolution des équipements, rend difficile l'application directe des approches traditionnelles de création d'état de santé. Un défi supplémentaire provient du faible nombre de pannes observées en raison des stratégies de maintenance préventive souvent déjà en place. Ce type de maintenance, qui consiste à remplacer ou réparer les équipements avant l'apparition de défaillances pour éviter les interventions correctives coûteuses, empêche par conséquent de recueillir des données relatives aux états de panne réels. Cette situation, bien que bénéfique pour la disponibilité des installations, constitue un obstacle significatif pour l'élaboration de modèles prédictifs robustes qui nécessitent des exemples de défaillances pour s'entraîner efficacement.

Parallèlement, les approches basées sur des modèles physiques s'appuient sur des lois et principes fondamentaux pour modéliser la dégradation des composants. Elles fournissent des représentations précises lorsque les mécanismes de défaillance sont bien compris et que tous les paramètres influents peuvent être correctement identifiés et mesurés. Cependant, dans des environnements complexes, ces approches sont limitées par la difficulté de modéliser les interactions entre sous-systèmes et la présence de paramètres non observables. À mesure que le nombre de variables à prendre en compte augmente, le modèle devient plus complexe. Cela entraîne une augmentation de l'incertitude sur les prédictions, notamment en raison de risques de sur-apprentissage ou de difficultés à bien capturer les relations entre les paramètres (ATAMURADOV et al., 2017).

les méthodes basées sur les données sont également limitées par l'irrégularité des acquisitions. Les algorithmes de création d'état de santé, comme les techniques de clustering ou de réduction de dimensionnalité, nécessitent généralement des ensembles de données denses pour identifier les tendances et extraire des indicateurs pertinents (CALABRESE et al., 2021). Lorsque les données sont échantillonnées de manière irrégulière, ces méthodes peinent à construire une représentation fidèle et cohérente de l'état de l'équipement.

Données incomplètes et acquisition partielle

Un autre défi majeur concerne l'acquisition partielle des données. Tous les équipements ne sont pas systématiquement mesurés à chaque période de maintenance, et certains paramètres ne sont pas systématiquement contrôlés à chaque maintenance. Cette acquisition sélective génère des jeux de données incomplets, rendant difficile la création d'un indicateur de santé global des équipements.

Les approches basées sur les données nécessitent généralement une base de données complètes pour appliquer des techniques comme l'analyse en composantes principales (PCA) ou les algorithmes de clustering. En présence de valeurs manquantes, ces méthodes doivent recourir à des techniques d'imputation, qui peuvent introduire des biais et altérer la qualité des indicateurs générés (GINKEL, 2023). Il est néanmoins possible d'appliquer directement des méthodes comme la PCA ou des algorithmes de clustering sur des jeux de données incomplets, mais cela demeure complexe et exige souvent de faire des hypothèses sur les valeurs manquantes, ce qui peut affecter la robustesse des résultats. De plus, l'incomplétude des données limite la capacité à exploiter pleinement les relations entre variables, compliquant la création d'états de santé précis et fiables.

Complexité structurelle du système étudié

Le système analysé est composé de multiples sous-systèmes interconnectés, chacun présentant des comportements distincts et des interactions complexes. Cette architecture introduit des interdépendances difficiles à modéliser, que ce soit avec des approches physiques ou basées sur les données.

Les approches classiques de création d'état de santé sont souvent conçues pour des équipements isolés, où les relations entre les variables sont claires et bien définies. Dans un système complexe comme celui étudié, les interactions entre sous-systèmes compliquent la tâche. Les méthodes de fusion de données ou de création d'indicateurs globaux peinent

à refléter fidèlement l'état global du système, car elles doivent agréger des informations hétérogènes de multiples sous-système (ATAMURADOV et al., 2017).

De plus, l'acquisition des données se fait lors de contrôles mesurant plusieurs équipements chainés. Cette configuration complique l'analyse, car en cas de défaillance détectée lors d'un contrôle, il devient difficile d'identifier précisément l'origine de l'erreur. L'absence de mesures individuelles pour chaque composant rend l'isolation des pannes complexe, augmentant le risque de diagnostics erronés ou imprécis.

Les approches physiques nécessitent une modélisation détaillée des interactions entre les composants, ce qui devient rapidement inextricable dans un environnement complexe. Les modèles simplifiés, qui ignorent certaines dépendances, risquent de produire des états de santé peu représentatifs, tandis que les modèles exhaustifs sont souvent irréalisables en raison de la complexité (HENG et al., 2009).

Défis liés à l'explicabilité des états de santé

L'explicabilité est un critère essentiel dans la création d'états de santé, et elle revêt une importance particulière dans notre cas d'étude. Les indicateurs produits doivent être compréhensibles par les experts de terrain afin de faciliter la justification de la prise de décision en matière de maintenance. Dans des environnements industriels complexes, comme celui analysé ici, il est crucial que les états de santé puissent être explicités de manière claire et précise pour que les opérateurs puissent comprendre l'origine des dégradations et les interventions nécessaires.

Or, de nombreuses approches avancées, notamment les méthodes basées sur la fusion de données complexes ou les techniques non supervisées comme le clustering, produisent des indicateurs difficiles à expliquer à cause de leur éloignement à la physique (ATAMURADOV et al., 2017). Les approches basées sur les données, bien qu'efficaces pour extraire des schémas complexes, souffrent souvent d'un manque d'explicabilité. Les indicateurs virtuels créés à partir de techniques comme les réseaux de neurones sont des représentations abstraites qui n'ont pas nécessairement de signification physique directe, rendant leur interprétation difficile pour les utilisateurs finaux (UCAR ; KARAKOSE et KIRIMCA, 2024). De plus, les résultats générés par ces méthodes sont souvent probabilistes, introduisant une incertitude supplémentaire dans l'estimation de l'état de santé des équipements.

Pour pallier ce manque d'explicabilité, des techniques d'intelligence artificielle expliable (*Explainable AI - XAI*) sont parfois utilisées afin d'interpréter le fonctionnement des modèles complexes. Des méthodes comme SHAP (*SHapley Additive exPlanations*)

ou LIME (*Local Interpretable Model-agnostic Explanations*) permettent d'analyser l'importance des variables d'entrée (CUMMINS et al., 2024). LIME explique une prédiction individuelle en construisant un modèle linéaire simple et interprétable dans son voisinage local. SHAP, basé sur la théorie des jeux, calcule la contribution de chaque variable à la prédiction en considérant toutes les combinaisons de variables possibles, offrant des garanties de cohérence.

Cependant, ces techniques ne fournissent qu'une approximation du comportement réel des modèles et introduisent une incertitude supplémentaire liée aux sources d'erreur inhérentes aux données (variabilité des mesures), aux modèles (imperfections ou approximations) et au contexte d'application (conditions différentes de celles utilisées pour l'entraînement du modèle). Ainsi, les interprétations générées par les approches XAI doivent être considérées avec prudence et ne garantissent pas une explicabilité parfaite.

Les modèles physiques offrent généralement une meilleure explicabilité, car ils reposent sur des relations causales bien établies et des phénomènes physiques connus. Cependant, leur application est souvent limitée dans les systèmes complexes ou lorsque les données disponibles sont incomplètes, comme c'est le cas dans notre étude. La complexité du système et les lacunes dans les données rendent en effet ces approches difficiles à mettre en œuvre. Ainsi, il devient nécessaire de développer des méthodes capables de fournir des états de santé à la fois robustes et explicables, tout en tenant compte des contraintes spécifiques du système analysé.

Le tableau 2.1 résume les principales limitations des approches existantes pour la création d'états de santé dans le contexte de notre étude.

Ces limitations soulignent les défis liés à la création d'états de santé robustes et explicables dans un contexte caractérisé par des données irrégulières, incomplètes et une structure complexe du système analysé. Les approches basées sur des modèles physiques peinent à capturer la complexité du système et nécessitent des données continues difficiles à obtenir, tandis que les méthodes fondées sur l'exploitation des données souffrent de la rareté et de l'irrégularité des mesures ainsi que du manque d'explicabilité des résultats.

Face à ces défis, il est crucial de développer une approche alternative capable de surmonter ces limites. La section suivante introduit la méthode MSLD (*Multi-Scale Level Discretization*), conçue pour traiter ces contraintes et permettre la création d'états de santé robustes, explicables et adaptés aux spécificités du système étudié.

TABLE 2.1 – Synthèse des limites des approches existantes pour la création d'états de santé.

| Catégorie | Problème rencontré | Impact sur la création d'état de santé |
|-------------------------------|--|---|
| Acquisition des données | Mesures irrégulières et ponctuelles | Difficulté à représenter la dégradation et perte d'informations temporelles |
| Données incomplètes | Valeurs manquantes et acquisition partielle | Complexité dans l'application des algorithmes et qualité réduite des indicateurs |
| Structure complexe du système | Interactions multiples entre sous-systèmes | Agrégation difficile des données et complexité de la modélisation |
| Faible volume de données | Données insuffisantes pour extraire des tendances et peu de pannes observées | Difficulté à établir des modèles de dégradation et manque de données pour l'apprentissage |
| Explicabilité des indicateurs | Indicateurs abstraits ou non physiques | Difficulté d'interprétation par les opérateurs et faible confiance dans les résultats |

2.2 Proposition d'une méthode de création d'état de santé basée sur la discrétisation

2.2.1 Présentation et Formalisation de la méthode MSLD

L'étiquetage de systèmes multivariables par discrétisation (*Multivariable System Labeling through Discretization*, MSLD) est une méthode développée spécifiquement dans le cadre de cette étude pour pallier l'absence d'états de santé détaillés dans les bases de données de maintenance prévisionnelle. Elle a été conçue pour répondre aux limites identifiées dans les approches existantes, notamment face aux contraintes posées par les données incomplètes, irrégulières et par la nécessité d'une explicabilité accrue.

Dans de nombreux cas, les bases de données de maintenance prévisionnelle se limitent à des états binaires, tels que *panne* ou *non-panne*, offrant une vision simpliste de la santé des équipements. MSLD vise à dépasser ces limitations en créant des états de santé interprétables et exploitables à travers une discrétisation non supervisée des données.

Cette méthode améliore la granularité des analyses en introduisant plusieurs états de transition, permettant ainsi une évaluation plus précise des risques et des dynamiques

de dégradation. L'objectif principal de MSLD est de fournir une représentation détaillée et nuancée de la santé d'un système multivarié, facilitant ainsi une maintenance prévisionnelle plus efficace et des stratégies d'intervention mieux adaptées aux conditions spécifiques des équipements.

La conception de MSLD répond directement aux exigences d'explicabilité imposées par le contexte industriel. Elle privilégie une démarche explicite, fondée sur des principes statistiques clairs, comme la définition de la plage de fonctionnement optimale par la moyenne et l'écart-type, et sur une logique métier claire, celle du composant le plus faible. Ce choix méthodologique vise à ce que l'indicateur de santé final soit facilement interprétable par les experts de terrain. La pertinence de cette approche, notamment face aux méthodes de discréétisation plus génériques, sera confirmée par les résultats expérimentaux.

Plus formellement, la méthode définit ces états de transition en calculant une série de seuils qui s'appuient à la fois sur la distribution statistique du paramètre et sur les limites de défaillance connues. Le processus de discréétisation génère $k - 1$ seuils intermédiaires entre la zone optimale et les limites de panne, comme illustré dans la Figure 2.2. Soit un paramètre de moyenne μ et d'écart-type σ , avec des seuils de défaillance F (inférieur) et F' (supérieur), et un nombre total de k états de transition souhaités. Les seuils S_n sont définis pour chaque état intermédiaire indexé par $n \in \{1, 2, \dots, k - 1\}$.

Pour les seuils inférieurs (entre F et $\mu - \sigma$), le n -ième seuil $S_{\text{inf},n}$ est donné par :

$$S_{\text{inf},n} = (\mu - \sigma) - n \frac{(\mu - \sigma) - F}{k} \quad (2.2)$$

Pour les seuils supérieurs (entre $\mu + \sigma$ et F'), le n -ième seuil $S_{\text{sup},n}$ est donné par :

$$S_{\text{sup},n} = (\mu + \sigma) + n \frac{F' - (\mu + \sigma)}{k} \quad (2.3)$$

Où n est un index entier variant de 1 à $k - 1$ qui positionne chacun des $k - 1$ seuils intermédiaires de manière équidistante (selon le principe *Equal Width*) dans l'espace de dégradation.

Le processus complet de la méthode MSLD peut être formalisé par l'Algorithme 1.

2.2.2 Rôle de la discréétisation

La discréétisation joue un rôle clé dans la création d'un état de santé en transformant les variables continues en catégories discrètes appelées *états* ou *classes* (GARCÍA et al.,

Algorithm 1 Création d’un état de santé global avec MSLD

- 1: **Entrée** : Matrice de données brutes X (observations i , paramètres j), Vecteur des seuils de panne F_j , Nombre d’états k .
- 2: **Sortie** : Vecteur Y des états de santé globaux pour chaque observation.
- 3: Initialiser une matrice d’états discrets D de même taille que X .
- 4: **for all** paramètre j dans X **do**
- 5: Calculer la moyenne μ_j et l’écart-type σ_j de la colonne j .
- 6: Définir l’intervalle optimal $I_{opt} = [\mu_j - \sigma_j, \mu_j + \sigma_j]$.
- 7: Calculer les $k - 1$ seuils de discrétisation $S_{inf,n}$ et $S_{sup,n}$ pour $n \in \{1, \dots, k - 1\}$ avec les équations (2.2) et (2.3).
- 8: **for all** observation i **do**
- 9: Discrétiser la valeur $x_{i,j}$ en un état entier $d_{i,j}$ basé sur sa position par rapport aux seuils calculés (ex : 0 pour optimal, ..., k pour panne).
- 10: $D[i, j] \leftarrow d_{i,j}$.
- 11: **end for**
- 12: **end for**
- 13: **for all** observation i **do**
- 14: // Appliquer le principe du composant le plus faible
- 15: $Y[i] \leftarrow \max_j(D[i, j])$.
- 16: **end for**
- 17: **return** Y

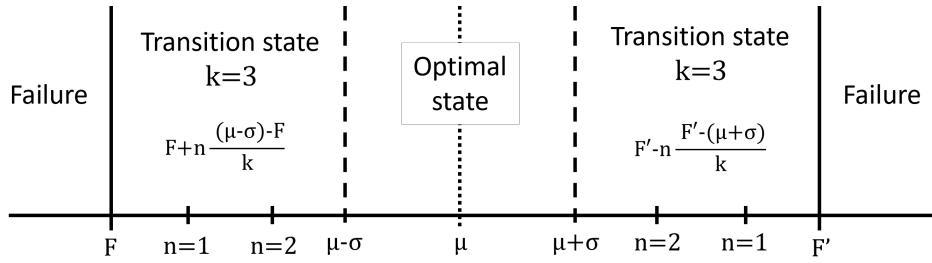


FIGURE 2.2 – Processus de répartition des étiquettes par discréétisation.

2013). Cette approche peut être notamment utile pour structurer les données complexes et les rendre plus interprétables.

Les techniques supervisées exploitent des données étiquetées pour diviser les caractéristiques continues en intervalles distincts. Les approches les plus courantes incluent ChiMerge (KERBER, 1992), le principe de longueur de description minimale (RISSANEN, 1986), et les techniques basées sur l'entropie (FAYYAD et IRANI, 1993). Ces méthodes nécessitent toutefois des étiquettes de classe préexistantes, ce qui peut limiter leur application lorsque les étiquettes sont absentes ou difficiles à définir.

Les techniques non supervisées, comme la discréétisation par largeur égale (Equal Width, EW) et par fréquence égale (Equal Frequency, EF), permettent de segmenter les données sans avoir recours à des étiquettes. La méthode EW divise les données en intervalles de taille constante, mais elle est sensible aux valeurs aberrantes (CATLETT, 1991). La méthode EF, en revanche, répartit les données en intervalles contenant un nombre similaire de points de données, offrant une meilleure gestion des distributions inégales. Cependant, EF peut produire des intervalles inadéquats pour des distributions hétérogènes.

Dans les environnements industriels, la discréétisation est fréquemment utilisée pour identifier des seuils critiques et catégoriser des paramètres tels que la température ou la pression en états *normal*, *alerte* et *panne*. Cette structuration facilite l'interprétation des résultats et améliore la communication entre les opérateurs et les experts techniques.

Le choix des seuils et du nombre d'états constitue un défi majeur dans la discréétisation. Des seuils mal définis peuvent entraîner une perte d'information ou des biais dans les analyses. Pour résoudre ce problème, des méthodes avancées basées sur l'apprentissage automatique ou des techniques d'optimisation peuvent être employées pour ajuster les paramètres de discréétisation selon les besoins spécifiques des systèmes industriels.

2.2.3 Processus de création d'un état de santé avec MSLD

Étapes du processus

Le processus de création d'un état de santé à l'aide de la méthode MSLD repose sur deux étapes clés : la discrétisation des paramètres et la catégorisation des états. Ces étapes sont présentées dans la Figure 2.3, qui illustre la transformation des données brutes en un état de santé interprétable.

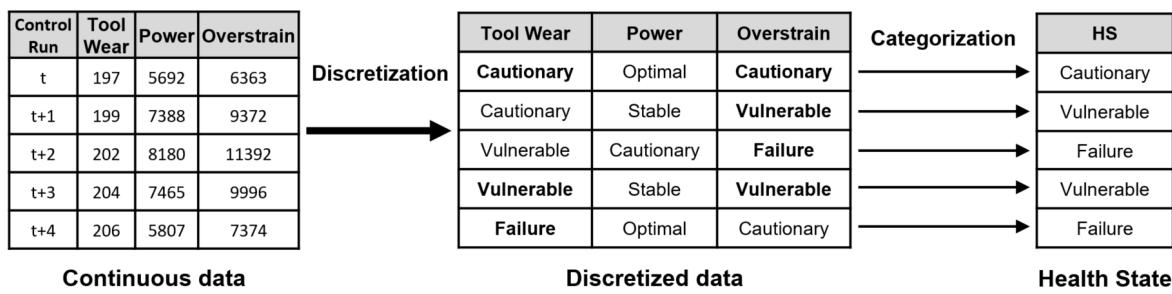


FIGURE 2.3 – Processus d'étiquetage par discrétisation dans la méthode MSLD.

Étape 1 : Discrétisation des paramètres. La première étape du processus consiste à discrétiser chaque paramètre mesuré en se basant sur son historique statistique. Cette phase est essentielle pour transformer les données continues en catégories discrètes, facilitant ainsi l'interprétation et l'analyse des tendances de dégradation.

L'écart-type (σ) autour de la moyenne (μ) est utilisé comme critère pour définir la plage de fonctionnement optimale du paramètre. Les valeurs situées entre $\mu - \sigma$ et $\mu + \sigma$ sont considérées comme reflétant un état optimal (Figure 2.4a), représentant un fonctionnement normal et sans risque. En revanche, les valeurs qui s'écartent de cette plage sont interprétées comme des signes de dégradation. Plus l'écart est grand avec la classe optimale, plus le paramètre est considéré comme dégradé.

Un aspect central de cette discrétisation est l'intégration de la limite de panne spécifique au cas d'étude. Cette limite fixe le seuil au-delà duquel un paramètre est jugé en panne, déclenchant une alerte critique. Elle permet de définir la borne supérieure des états de transition, assurant que toute valeur dépassant cette limite est automatiquement étiquetée comme critique. Cette donnée est cruciale, car elle garantit que les intervalles de transition reflètent fidèlement les risques opérationnels, tout en maintenant une distinction claire entre les états normaux, dégradés et défaillants.

Pour enrichir la granularité de l'analyse et mieux suivre les tendances de dégradation, des états intermédiaires sont introduits entre l'état optimal et la limite de panne. Ces états, tels que *vulnérable*, *instable*, et *stable*, permettent de suivre la progression graduelle vers une panne (Figure 2.4b). La méthode de discrétisation *Equal Width* (EW) est utilisée pour créer ces intervalles, divisant les données en segments de taille égale tout en tenant compte de la distribution globale des valeurs.

Étape 2 : Catégorisation des états. Une fois les paramètres discrétisés, l'étape suivante consiste à générer un état de santé global pour chaque exécution de contrôle. Ce processus repose sur l'hypothèse que la performance globale d'un système est limitée par son composant le plus faible. Ainsi, l'état global est défini par le paramètre présentant le niveau de dégradation le plus élevé. Cette logique reflète la méthodologie utilisée lors des contrôles de panne : si un seul paramètre dépasse la limite de panne, le contrôle est considéré comme un échec, même si les autres paramètres restent dans des plages acceptables. Cette approche garantit une évaluation conservatrice, essentielle dans les environnements industriels critiques, où la défaillance d'un seul composant peut compromettre l'ensemble du système.

Dans la méthode MSLD, ce principe est étendu pour inclure les états intermédiaires. Même si aucun paramètre ne dépasse la limite stricte de panne, la présence d'un paramètre dans un état fortement dégradé impacte l'état global. Cela permet une détection précoce des risques, offrant ainsi aux équipes de maintenance la possibilité d'intervenir avant l'apparition d'une défaillance critique. Cette approche améliore la fiabilité globale du système et permet d'optimiser les stratégies de maintenance.

Les états individuels sont ensuite regroupés en catégories explicites telles que *optimal*, *stable*, *instable*, et *vulnérable*. Cette catégorisation simplifie l'analyse pour les experts de terrain, tout en fournissant une vue hiérarchisée de la santé du système. En visualisant ces états, les techniciens peuvent rapidement identifier les composants nécessitant une intervention, facilitant ainsi la prise de décision. La Figure 2.4c illustre comment ces états sont attribués en fonction des plages discrétisées.

2.2.4 Résultats

Après avoir introduit la méthode MSLD, cette section présente les résultats obtenus et discute de son efficacité. L'approche utilisée suit le processus décrit dans la section précédente (Section 2.2.3), combinant discrétisation des paramètres et catégorisation afin

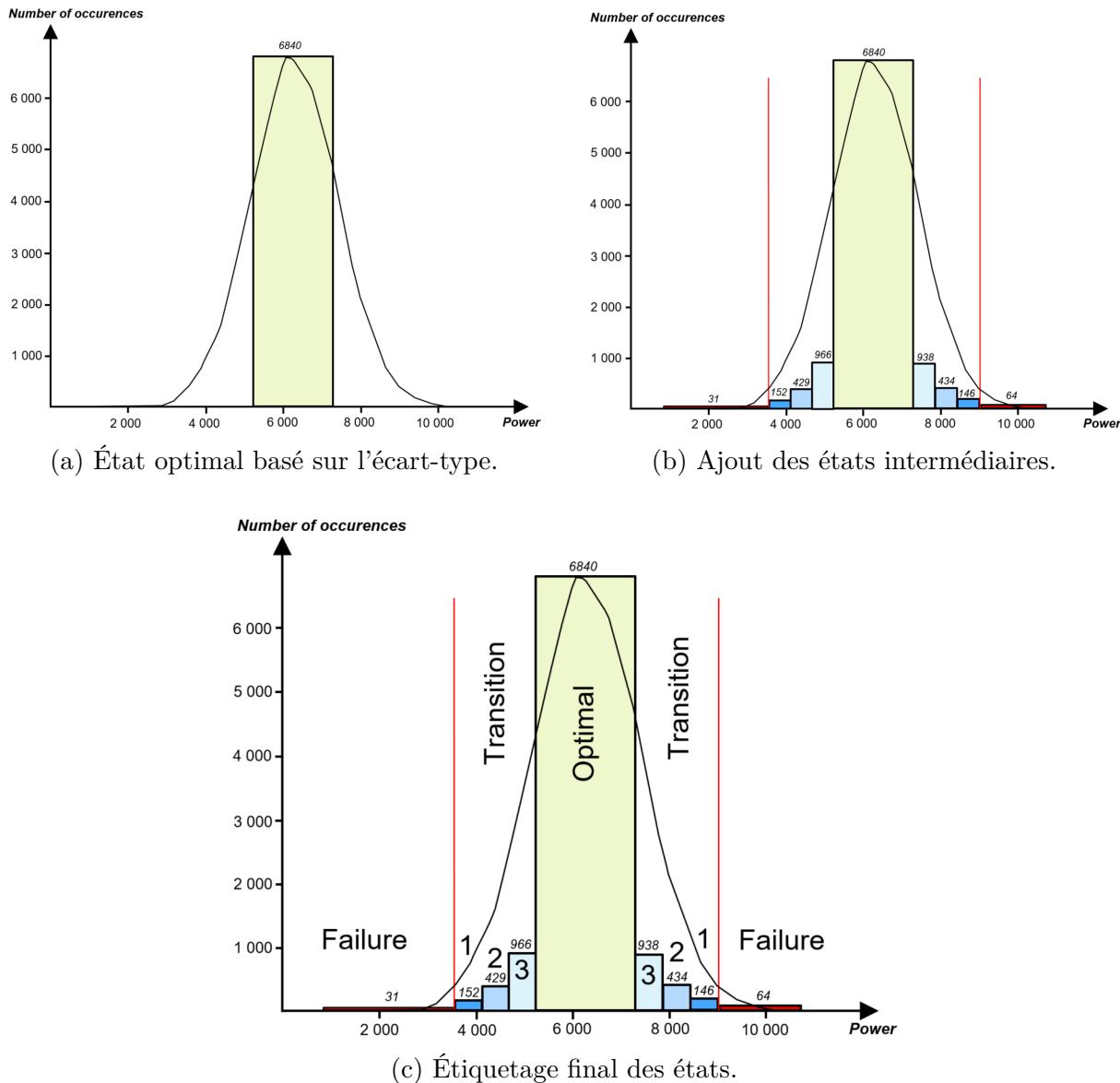


FIGURE 2.4 – Illustration des étapes de discréétisation et d'étiquetage dans MSLD.

de déterminer des états de santé. Dans un premier temps, cette approche est appliquée à un paramètre unique, la puissance, afin d'évaluer la pertinence de la discréétisation sur une variable spécifique. Ensuite, cette discréétisation est généralisée à l'ensemble des paramètres de la base de données afin de générer les états de santé du système à chaque entrée de la base de données.

L'ensemble de données utilisé est la base AI4I (Artificial Intelligence for Industry), qui contient 10 000 observations décrivant l'état d'un système industriel simulé. Elle regroupe

cinq variables d'entrée, dont la température, la vitesse de rotation, le couple et l'usure de l'outil. Un indicateur de panne global est fourni, résultant de cinq modes de pannes indépendants : usure excessive de l'outil, dissipation thermique insuffisante, défaillance énergétique, surcharge mécanique et pannes aléatoires.

Dans un premier temps, la méthode MSLD est appliquée au paramètre de puissance. Le Tableau 2.2 compare les résultats obtenus avec trois techniques de discréétisation : *Equal Width* (EW), *Equal Frequency* (EF) et MSLD. La répartition des observations dans les différentes catégories d'états varie considérablement selon la méthode employée.

TABLE 2.2 – Différentes distributions pour le paramètre "Puissance" avec différentes techniques de discréétisation.

| | Panne | Vulnérable | Instable | Stable | Optimal |
|------|-------|------------|----------|--------|---------|
| EW | 95 | 451 | 1415 | 3240 | 4799 |
| EF | 95 | 2453 | 2464 | 2453 | 2439 |
| MSLD | 95 | 298 | 863 | 1904 | 6840 |

La méthode EF a tendance à concentrer un grand nombre de valeurs aux extrémités, ce qui peut fausser l'interprétation des états de santé et déséquilibrer le diagnostic. À l'inverse, EW et MSLD conservent mieux la distribution naturelle des données, assurant une distinction plus progressive entre les états. Cependant, EW répartit uniformément les intervalles sans prendre en compte la répartition réelle des données, tandis que MSLD adapte dynamiquement ses seuils aux caractéristiques de la variable analysée.

Dans un second temps, la discréétisation MSLD est généralisée à l'ensemble des paramètres de la base AI4I afin de définir un état de santé global. Conformément à la méthodologie décrite en Section 2.2.3, l'état global d'un échantillon est déterminé par son paramètre le plus dégradé. Cette étape permet de catégoriser chaque observation selon cinq niveaux de santé, de l'état *optimal* à la *panne*.

Le Tableau 2.3 compare la répartition des états du système obtenue avec différentes méthodes de discréétisation. La méthode binaire sert de référence et illustre la perte d'information liée à l'absence d'états de transition. Elle classe en effet 96,5% des observations comme optimales et 3,5% en panne, sans offrir aucune nuance.

L'analyse des distributions des méthodes plus granulaires révèle une limite critique des approches standards (EW et EF) dans ce contexte multivarié. En appliquant le principe du "composant le plus faible" (où l'état global est le `max()` des états des paramètres), les deux méthodes aboutissent à un résultat statistiquement problématique, puisqu'aucune

des 10000 observations n'est classée comme "Optimale". Cela suggère que leurs seuils, trop larges ou mal positionnés, font que la moindre fluctuation normale sur un seul paramètre suffit à faire basculer l'ensemble du système dans un état de surveillance. Un tel indicateur, incapable de confirmer un état de fonctionnement nominal, aurait une utilité opérationnelle limitée.

Il convient de noter qu'en l'absence d'étiquettes de vérité terrain pour les états de santé intermédiaires, une comparaison quantitative basée sur des métriques d'erreur traditionnelles n'est pas directement applicable. L'évaluation se concentre donc sur la pertinence opérationnelle des distributions obtenues. Il s'agit notamment d'évaluer la capacité de chaque méthode à identifier un état de fonctionnement nominal, une condition nécessaire à la mise en place d'un système de surveillance crédible.

À l'inverse, la méthode MSLD démontre une capacité de discrimination supérieure. En définissant sa zone optimale sur des bases statistiques ($\mu \pm \sigma$), elle se montre plus robuste à cet effet d'agrégation. Elle parvient à identifier 998 observations comme étant dans un état "Optimal". Ce résultat est crucial car il prouve que MSLD est la seule des trois approches à pouvoir distinguer un fonctionnement véritablement sain des premiers signes de dégradation. En conclusion, l'analyse factuelle des distributions montre que,

TABLE 2.3 – Tableau des différentes distributions des états du système en fonction de la technique de discrétisation.

| | Panne | Vulnérable | Instable | Stable | Optimal |
|---------|-------|------------|----------|--------|---------|
| Binaire | 348 | 0 | 0 | 0 | 9652 |
| EW | 348 | 3851 | 5186 | 615 | 0 |
| EF | 348 | 7421 | 1928 | 303 | 0 |
| MSLD | 348 | 2074 | 3290 | 3290 | 998 |

dans le cadre de cette comparaison, MSLD fournit un indicateur de santé plus réaliste et interprétable. Sa capacité à préserver une classe "Optimale" non-nulle, contrairement aux approches EW et EF, évite une distorsion statistique qui rend ces dernières irréalistes pour une application industrielle et offre ainsi une condition nécessaire à une maintenance prévisionnelle efficace.

2.2.5 Discussion et limitations

La méthode MSLD proposée permet une approche plus nuancée et flexible dans la création d'états de santé, tout en offrant des avantages clairs en termes de granularité et d'interprétabilité. Cependant, elle n'est pas exempte de limitations.

Un avantage notable dans le cadre de cette étude est la disponibilité préalable des seuils de défaillance pour chaque paramètre. Ces seuils ont permis d'appliquer directement la méthode MSLD sans nécessiter d'estimations ou de processus de calibration supplémentaires. Cela a facilité l'identification des états critiques et amélioré la précision de la discrétisation. Cependant, dans d'autres cas réelles, ces seuils ne sont pas toujours connus ou peuvent être difficilement définissables. L'absence de ces seuils complique l'application directe de la méthode et impose alors le recours à des techniques d'estimation ou à l'expertise humaine pour les déterminer.

De plus, bien que MSLD améliore la granularité des états de santé, elle peut parfois complexifier les résultats, surtout lorsque le nombre d'états est trop élevé. Cette granularité accrue nécessite une attention particulière lors de l'interprétation pour éviter des diagnostics trop conservateurs ou une surcharge d'informations pour les opérateurs.

Une autre limite méthodologique réside dans le traitement indépendant de chaque paramètre avant l'agrégation finale. La méthode MSLD discrétise chaque variable séparément avant d'appliquer le principe du "composant le plus faible". Bien qu'efficace, cette approche ne capture pas explicitement les corrélations entre les paramètres. Une perspective d'amélioration serait d'apprendre des zones de fonctionnement directement dans l'espace multivarié. Des techniques de clustering non supervisé pourraient, par exemple, identifier des régions jointes correspondant à des états de santé spécifiques. L'avantage majeur serait de modéliser des interactions complexes (par exemple, une température élevée n'est critique que si elle est couplée à un couple important), menant à une définition des états potentiellement plus fidèle à la physique du système. Cette piste constitue une perspective de recherche pour de futurs travaux.

Enfin, la méthode repose également sur un prétraitement rigoureux des données, notamment la gestion des valeurs manquantes, afin de garantir des résultats fiables et exploitables. En l'absence de ces étapes de préparation, la qualité des états de santé générés pourrait être compromise.

Conclusion

Ce chapitre a exploré les enjeux, les méthodologies et les solutions proposées pour la création d'un état de santé dans un contexte industriel, en mettant en lumière l'importance de la discrétisation pour structurer des données complexes et irrégulières. L'état de santé, élément central des stratégies de maintenance prévisionnelle, permet non seulement de suivre la condition des équipements mais aussi d'anticiper les défaillances et d'optimiser les interventions.

Les approches traditionnelles, basées sur des modèles physiques, bien que robustes et explicables, présentent des limites face à la complexité croissante des systèmes modernes. À l'inverse, les méthodes basées sur les données exploitent la puissance des algorithmes d'apprentissage automatique pour détecter des schémas complexes, mais elles sont souvent confrontées à des défis liés à l'interprétabilité et à la qualité des données.

Pour répondre à ces problématiques, ce chapitre a introduit la méthode MSLD (Multivariable System Labeling through Discretization), développée spécifiquement dans cette étude.

MSLD offre une approche pour structurer les données multivariées et créer des états de santé plus nuancés et interprétables. En introduisant des états de transition entre les conditions optimales et les seuils critiques, cette méthode améliore la granularité des diagnostics tout en facilitant la détection précoce des dégradations.

Les résultats expérimentaux ont permis de comparer la pertinence des indicateurs de santé générés. L'analyse a en effet mis en évidence une limite importante des approches standards (EW et EF) dans notre contexte multivarié : l'application du principe du "composant le plus faible" conduit à la disparition de la classe "Optimale". Un indicateur incapable de confirmer un fonctionnement nominal s'avère difficilement exploitable pour le suivi d'un équipement. À l'inverse, MSLD a démontré sa capacité à préserver cette classe, fournissant ainsi un indicateur de santé plus discriminant et mieux aligné avec une application industrielle réaliste.

Néanmoins, la méthode MSLD n'est pas exempte de limitations. Sa dépendance aux seuils de défaillance, qui peuvent nécessiter l'intervention d'experts pour être définis avec précision, introduit un certain degré de subjectivité. De plus, l'augmentation de la granularité peut complexifier l'interprétation des résultats, surtout lorsque le nombre d'états devient trop élevé. Cependant, cette interaction avec les experts peut également être une opportunité pour affiner la compréhension du système et enrichir la pertinence des états.

2.2. Proposition d'une méthode de création d'état de santé basée sur la discréétisation

de santé définis.

Ces limitations ouvrent la voie à de futures recherches, notamment pour automatiser l'identification des seuils critiques.

Le chapitre suivant explorera l'application de la méthode MSLD et d'autres techniques de diagnostic sur des bases de données publiques. Cette analyse comparative permettra d'évaluer la généralisabilité des approches proposées et de valider leur efficacité dans des contextes variés.

MODÉLISATION D'ÉTAT DE SANTÉ ET DIAGNOSTIC PAR RÉSEAU BAYÉSIEN

Introduction

Le chapitre précédent a introduit la méthode MSLD, une approche permettant de transformer des mesures continues en états de santé discrets et interprétables. Ce chapitre constitue la première étape de validation de ces états : il s'agit de démontrer qu'un modèle probabiliste, ici un réseau bayésien statique, peut diagnostiquer de manière fiable l'état de santé d'un équipement à un instant donné. Cette validation est un prérequis indispensable avant d'aborder la prédiction de l'évolution temporelle de ces états. Dans un premier temps, plusieurs bases de données publiques sont analysées afin de sélectionner celles les plus pertinentes pour notre étude. Ces bases sont ensuite transformées pour les adapter aux contraintes du contexte industriel visé, notamment à travers l'ajout de données manquantes et la création de seuils d'alerte. Ces transformations permettent d'obtenir des jeux de données plus proches des conditions réelles d'acquisition et de diagnostic, tout en conservant leur accessibilité publique.

Une fois ces bases adaptées, la méthode MSLD est appliquée pour générer des états de santé discrets à partir des mesures continues. Ces états reflètent différents niveaux de dégradation, allant de l'état optimal à des situations d'alerte ou de panne. Le réseau bayésien est ensuite construit pour modéliser les relations entre ces états et les variables mesurées.

L'approche repose ici sur des réseaux bayésiens statiques, dont la structure est enrichie par l'ajout de noeuds intermédiaires de diagnostic afin de réduire la complexité du modèle. L'objectif de cette modélisation est de valider que le réseau est capable de reproduire la logique du diagnostic issue de la méthode MSLD, en utilisant les bonnes mesures au bon instant.

Les résultats sont évalués sur plusieurs bases transformées, à l'aide de métriques clas-

siques telles que la précision, le rappel et le F1-score macro. Les matrices de confusion sont utilisées pour analyser finement la qualité du diagnostic produit par les modèles et mettre en évidence les éventuelles erreurs.

Ce travail constitue une étape intermédiaire dans la mise en place d'un système complet de maintenance prévisionnelle. Il permet de s'assurer que le diagnostic instantané est fiable, avant de passer à la prédiction de l'évolution dans le temps.

3.1 Utilisation des bases de données publiques en maintenance prévisionnelle

3.1.1 Motivations et contraintes

La maintenance prévisionnelle repose sur l'analyse de données industrielles dans le but d'anticiper les défaillances des équipements et d'améliorer la fiabilité des systèmes. L'accès à des données de qualité est donc un enjeu central pour le développement, l'évaluation et la mise en œuvre des algorithmes d'apprentissage.

Comme l'illustre l'analyse comparative présentée dans les tableaux 3.1 et 3.2, les bases de données exploitées dans ce domaine sont, pour la plupart, issues de travaux académiques ou expérimentaux. Leur origine, qu'elle soit simulée ou issue d'essais réels, provient généralement d'expérimentations menées sur des scénarios peu complexes et dans des conditions fortement contrôlées. Quelques rares bases publiques sont construites à partir de données collectées en environnement industriel réel, mais elles sont en plus faible proportion.

Les bases simulées sont généralement utilisées lorsque l'acquisition de données réelles en quantité suffisante est trop coûteuse, trop longue ou difficile à mettre en œuvre sur le système étudié. C'est notamment le cas pour certains équipements critiques ou complexes, comme les turboréacteurs, pour lesquels il est délicat d'observer directement des défaillances réelles sur de longues périodes. Dans de telles situations, les données sont générées à l'aide de modèles physiques ou hybrides, permettant de simuler des trajectoires de dégradation dans des conditions maîtrisées (SAXENA et al., 2008). Si ces jeux de données offrent un bon cadre pour expérimenter et comparer des méthodes, ils ne capturent pas toujours l'ensemble des dynamiques et incertitudes propres aux environnements industriels réels.

Dans ce contexte, l'un des enjeux majeurs réside dans la capacité des méthodes dé-

veloppées à être transposées sur des systèmes réels à partir de données simulées. C'est pourquoi des efforts sont souvent menés pour enrichir les jeux de données synthétiques par des scénarios proches de la réalité, ou pour valider les modèles sur des cas industriels accessibles à plus petite échelle. Dans notre étude, cette problématique est abordée à travers la transformation de bases publiques pour y intégrer des contraintes réalistes.

À l'inverse, les bases issues du terrain permettent de confronter les algorithmes à des situations concrètes, plus représentatives du fonctionnement réel des systèmes. Elles peuvent révéler des comportements inattendus ou des corrélations difficiles à anticiper, ce qui en fait un support important pour valider la robustesse et la transférabilité des approches. Toutefois, leur accès reste limité, et leur traitement nécessite souvent des adaptations spécifiques en raison des conditions d'acquisition.

Dans de nombreux cas, les bases de données acquises en environnement réel restent dans le domaine privé pour des raisons de confidentialité industrielle. Les informations qu'elles contiennent sont souvent sensibles, car elles peuvent révéler des détails sur les performances, la conception ou les faiblesses des équipements utilisés. Cette contrainte freine leur diffusion au sein de la communauté scientifique et limite les possibilités de collaboration ou de validation croisée entre différents travaux. Notre cas d'étude s'inscrit précisément dans ce contexte. Les données utilisées ici proviennent d'un environnement industriel réel et présentent plusieurs contraintes spécifiques liées à la nature du système et aux modalités d'acquisition. Cette complexité impose un effort méthodologique supplémentaire, mais permet en retour de construire des modèles plus adaptés aux réalités de la maintenance sur le terrain.

La prochaine partie présente la comparaison de différentes bases de données publiques de maintenance prévisionnelle en s'attardant sur les critères de sélection des bases retenues, ainsi que sur les méthodes de prétraitement pour les exploiter pour la suite de l'étude.

3.1.2 Sélection de bases de données

La sélection d'une base de données publique appropriée est une étape cruciale pour le développement et l'évaluation des algorithmes de maintenance prévisionnelle. Les jeux de données publics offrent des ressources précieuses pour tester des modèles, mais ils présentent souvent des différences qui peuvent limiter leur représentativité par rapport aux environnements industriels réels. Afin d'identifier les jeux de données les plus adaptés pour cette étude, une analyse comparative a été réalisée sur plusieurs jeux de données

publics couramment utilisés en maintenance prévisionnelle.

Cette comparaison se concentre sur les caractéristiques clés des jeux de données, notamment le nombre de paramètres, le volume d'échantillons, la présence de données temporelles et la fréquence d'échantillonnage. Ces critères ont été retenus en raison de leur impact direct sur la qualité des modèles de prédiction et de leur importance dans les environnements industriels réels.

En complément, d'autres paramètres ont été pris en compte afin d'évaluer dans quelle mesure ces bases de données reflètent les contraintes typiques d'un cas d'usage industriel. L'objectif est d'identifier les jeux de données qui présentent des conditions similaires à celles du système étudié, notamment en ce qui concerne l'irrégularité des acquisitions, la présence de données manquantes ou la répartition des états de défaillance. Ces aspects sont essentiels pour garantir que les modèles développés soient robustes et généralisables à des contextes industriels exigeants.

Un premier critère fondamental est l'irrégularité des acquisitions, un défi courant en maintenance prévisionnelle. Cette caractéristique est particulièrement fréquente dans des domaines comme l'avionique ou l'automobile, où les contrôles sont effectués à des intervalles irréguliers et où les interruptions de mesure peuvent être causées par des contraintes opérationnelles (WANG et WANG, 2020). L'absence d'un flux de données continu peut compliquer la détection des tendances de dégradation et nécessiter l'adaptation des méthodes d'apprentissage automatique à des séries temporelles discontinues.

Un second critère essentiel est la présence des données manquantes, qui peut résulter de fréquences d'acquisition variables selon les paramètres mesurés ou d'erreurs dans le processus d'enregistrement (HELWIG ; PIGNANELLI et SCHÜTZE, 2015). Ces lacunes dans les données nécessitent des stratégies pour éviter qu'elles ne biaissent l'apprentissage des modèles prédictifs. Les approches de suppression, d'imputation ou d'apprentissage de représentation sont des solutions envisageables, mais elles dépendent fortement de la structure des données disponibles (REN et al., 2023). Il est donc essentiel de sélectionner des jeux de données reflétant ce type de contraintes pour évaluer la robustesse des algorithmes face à ces imperfections.

Enfin, la taille du jeu de données et le volume d'échantillons disponibles influencent directement la capacité des modèles d'apprentissage à généraliser leurs prédictions. Un faible nombre d'échantillons peut limiter l'efficacité des algorithmes en raison d'une diversité insuffisante des données d'entraînement. Cette contrainte est particulièrement critique pour les modèles nécessitant de grands volumes de données, comme les réseaux de neu-

3.1. Utilisation des bases de données publiques en maintenance prévisionnelle

rones. Des approches comme la génération de données synthétiques, notamment à l'aide de GANs, ont été explorées pour pallier ce problème (HAKAMI, 2020). Cependant, il reste préférable d'utiliser des bases contenant une diversité naturelle suffisante pour éviter l'introduction d'artéfacts liés à la synthétisation des données.

Le tableau 3.1 résume ces critères appliqués à une sélection de jeux de données publics, en mettant en évidence les différences en termes de taille et de structure temporelle.

TABLE 3.1 – Comparaison des jeux de données selon des critères de taille et de caractéristiques temporelles.

| Nom | Nombre de paramètres | Nombre d'échantillons | Horodatage (timestamp) | Durée d'échantillonnage | Fréquence d'échantillonnage | Échantillonnage irrégulier |
|---|----------------------|-----------------------|------------------------|-------------------------|-----------------------------|----------------------------|
| AI4I (Matzka, 2020) | 5 | 10,000 | . | . | . | X |
| Bearing (Lee et al., 2007) | 4-8 | 7,588 | X | 2 mois | 10 min | . |
| CBM NPP (Coraddu et al., 2016) | 18 | 11,934 | . | . | . | . |
| CMAPSS (Saxena et al., 2008) | 26 | 265,256 | X | . | . | . |
| GD (Birgelen et Niggemann, 2018) | 18/22 | 16,220 | X | 13 min | 50 ms | . |
| HS (Helwig ; Pignanelli et Schütze, 2015) | 43,680 | 2,205 | X | 37 heures | 0,01/0,1/1 s | X |
| Li-ion Battery (Saha et Goebel, 2007) | 12 | 2,167 | X | . | . | . |
| MetroPT-3 (Veloso et al., 2022) | 15 | 1,516,948 | X | 7 mois | 10 s | . |
| Milling (Agogino et Goebel, 2007) | 13 | 1,500,000 | X | . | 4 ms | X |
| PPD (von Birgelen et al., 2018) | 25 | 228,425 | X | . | . | . |
| PRONOSTIA (Nectoux et al., 2012) | 3 | 24,069 | X | 57 heures | 40 µs/0,1 s | X |
| UFD (Gyamfi et al., 2018) | 37/44/52 | 540 | . | . | . | . |
| AI4I-PMDI (Autran et al., 2024a) | 5 | 10,000 | X | 4 ans | jours-semaines | X |

TABLE 3.2 – Comparaison des jeux de données selon les critères de défaillances et d'origine des données.

| Nom | % de défaillances (nombre) | Nombre d'états | Données manquantes | Parc d'équipement | Origine des données | Licence |
|----------------|----------------------------|----------------|--------------------|-------------------|---------------------|-------------|
| AI4I | 3% (338) | 5 | . | . | Synthétique | CC-BY |
| Bearing | . | . | . | . | Réel | CC0 |
| CBM NPP | . | . | . | . | Simulation | CC0 |
| CMPASS | 0,3% (100) | . | . | X | Simulation | CC0 |
| GD | 0,3% (50) | 2 | . | . | Réel | CC-BY-NC-SA |
| HS | . | 11 | . | . | Réel | CC-BY |
| Li-ion Battery | . | . | . | X | Simulation | CC0 |
| MetroPT-3 | . | . | . | . | Réel | CC-BY |
| Milling | . | . | . | . | Réel | CC0 |
| PPD | . | . | . | X | Réel | CC-BY-SA |
| PRONOSTIA | . | . | . | X | Réel | CC-BY |
| UFD | . | 2/3/4 | . | . | Réel | CC-BY |
| AI4I-PMDI | 3% (338) | 5 | X | X | Synthétique | CC-BY |

Le tableau 3.1 met en évidence la diversité des jeux de données publics en termes de taille et de caractéristiques temporelles. Certains jeux, comme *MetroPT-3*, contiennent un très grand nombre d'échantillons avec une fréquence d'échantillonnage régulière, tandis que d'autres, comme *HS* ou *PRONOSTIA*, présentent un échantillonnage irrégulier. L'irrégularité des acquisitions est une contrainte fréquente dans les environnements industriels réels, où les mesures ne sont pas toujours effectuées à des intervalles fixes. Cependant, peu de jeux de données publics intègrent explicitement cette caractéristique. Seuls certains jeux, tels qu'*AI4I*, *HS* et *PRONOSTIA*, incluent des données avec une acquisition irrégulière.

Au-delà des aspects temporels et de la taille des jeux de données, il est également essentiel d'examiner les caractéristiques liées aux défaillances et à l'origine des données. Le tableau 3.2 fournit une analyse comparative axée sur ces critères. On y observe que la majorité des jeux de données présentent un faible pourcentage de défaillances, en raison de la proportion importante d'échantillons sains par rapport aux échantillons en panne. Une partie des jeux de données propose des états explicites pour chaque observation, facilitant ainsi l'interprétation des critères de défaillance. Par ailleurs, les bases analysées se distinguent par leurs origines : certaines proviennent de simulations, d'autres de données industrielles réelles, et un plus petit nombre sont des données synthétiques générées artificiellement. Toutefois, aucun de ces jeux de données n'intègre simultanément toutes les contraintes que l'on peut rencontrer dans un contexte industriel réel.

En analysant ces contraintes, deux jeux de données ont été retenus pour la suite de cette étude : *AI4I 2020* et *Genesis Demonstrator*. Leur sélection repose sur les résultats des tableaux 3.1 et 3.2, mais également sur des considérations spécifiques à la nature des mesures, la présence d'étiquettes de panne, et leur complémentarité vis-à-vis des contraintes industrielles étudiées.

Le jeu de données *AI4I 2020* a été retenu pour la maîtrise qu'il offre sur la génération et la structure des données. Il s'agit d'un jeu synthétique simulant le comportement d'un système industriel, avec un mécanisme explicite de génération des pannes reposant sur cinq modes de défaillance distincts. Cette structuration permet un contrôle total sur les paramètres impliqués dans chaque défaillance, ce qui facilite l'analyse et la validation des méthodes de diagnostic développées. Il contient 10 000 échantillons, dont 338 sont annotés comme représentant une défaillance, soit environ 3 % du jeu. De plus, la simplicité de ses paramètres et sa taille modérée en font un jeu adapté pour expérimenter les étapes initiales de la méthodologie. Cependant, ce jeu présente des limites importantes

pour simuler un environnement réaliste : il ne contient ni données manquantes, ni irrégularité d’acquisition, ni identification d’équipements, et les seuils de panne ne sont pas directement exploitables. Pour pallier ces limites, une version enrichie du jeu, nommée AI4I-PMDI, a été constituée. Elle introduit des données manquantes, des conditions d’acquisition irrégulières, des identifiants de machine, ainsi qu’une formalisation explicite des seuils de panne utilisés dans la suite de l’étude.

En complément, la base de données Genesis Demonstrator a été sélectionnée pour son origine réelle et la représentativité de ses conditions d’acquisition en environnement industriel. Ce jeu de données a été collecté sur un démonstrateur physique dans le cadre des projets européens OPAK et IMPROVE. Il contient 16 220 observations temporelles multivariées acquises toutes les 50 ms, ainsi que des étiquettes de panne manuellement annotées. Dans la version labellisée, 50 échantillons sont associés à une défaillance, soit environ 0,3 % du total. Ces anomalies correspondent à des événements concrets tels que le grippage ou le décrochage du chariot linéaire, introduits de manière contrôlée mais réaliste. Les pannes sont documentées et validées, ce qui en fait un support fiable pour les phases de diagnostic. Contrairement à AI4I, les signaux de capteurs présents dans Genesis résultent de perturbations physiques réelles, enregistrées lors d’un fonctionnement en boucle fermée avec un automate industriel. Elle permet donc d’évaluer la robustesse des modèles dans des conditions proches de celles du terrain et complète utilement AI4I pour tester la transférabilité des approches proposées.

Après avoir justifié le choix de ces jeux de données, la partie suivante présente les transformations apportées afin de les adapter à notre contexte industriel réel, en vue d’y appliquer la méthode MSLD pour la création des états de santé.

3.1.3 Prétraitement des bases de données

Transformation structurelle des bases de données

Comme l’analyse précédente l’a démontré, les jeux de données publics ne reflètent que partiellement les contraintes des environnements industriels. Afin de pouvoir évaluer les algorithmes dans des conditions plus réalistes, il est donc nécessaire de transformer les bases de données retenues.

Pour simuler de telles contraintes industrielles, plusieurs modifications ont été appliquées aux bases de données utilisées dans cette étude. Un premier ajustement concerne l’introduction de données manquantes. Dans les environnements industriels, tous les para-

mètres ne sont pas mesurés à chaque inspection : cela dépend du type de contrôle effectué. Afin de reproduire cette réalité, un paramètre indiquant le type de contrôle a été ajouté, chaque type étant associé à un sous-ensemble spécifique de variables observées. Cette organisation permet d'introduire des données manquantes de manière structurée, en accord avec les protocoles de maintenance différenciés observés en pratique.

Un second ajustement consiste à intégrer une logique de gestion de parc d'équipements. Dans un environnement industriel réel, les systèmes étudiés appartiennent souvent à des ensembles hétérogènes d'équipements similaires, mais non identiques. L'ajout d'un identifiant système permet de simuler cette variabilité inter-équipements et d'évaluer la robustesse des algorithmes face à cette diversité.

Ces transformations ont été pensées pour s'approcher des conditions de fonctionnement observées dans notre cas d'étude, tout en conservant la structure d'origine des jeux de données. La Figure 3.1 illustre les étapes de transformation appliquées.

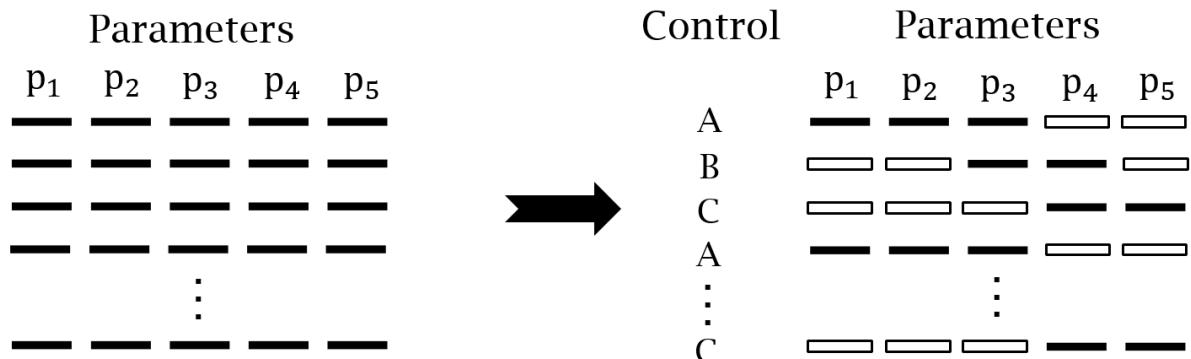


FIGURE 3.1 – Processus de transformation des bases de données vers des versions adaptées aux contraintes industrielles.

Les transformations structurelles étant réalisées, une étape complémentaire consiste à définir des seuils d'alerte sur les paramètres, afin de mieux relier les mesures aux états de santé.

Création de seuils d'alerte

Une autre limitation fréquente des bases de données concerne l'absence de seuils explicites permettant de relier directement les mesures à un état de santé. Dans notre contexte, l'état global d'un système est déterminé à partir de l'état de ses mesures individuelles : un paramètre critique suffit à considérer que l'équipement est en défaillance. La méthode MSLD repose sur ce principe en agrégeant les états des mesures discrétisées.

Cependant, cette logique n'est pas présente dans la majorité des jeux de données disponibles. Dans ceux-ci, l'état du système est souvent déterminé par la réussite ou l'échec d'une tâche globale, indépendamment des valeurs individuelles des paramètres. Ainsi, une même valeur mesurée peut correspondre tant à un fonctionnement normal qu'à une situation de panne, rendant impossible la définition stricte de seuils de défaillance.

Pour compenser cette absence, nous introduisons des seuils d'alerte sur les paramètres critiques. Ces seuils ne définissent pas une panne certaine, mais marquent une zone à partir de laquelle le risque de défaillance devient significatif. Ils sont construits à partir de l'historique des données, en identifiant les valeurs extrêmes atteintes lors des épisodes de défaillance, tout en restant en dehors des plages typiques du fonctionnement sain. L'objectif est d'engendrer un signal d'alerte suffisamment sensible, sans générer trop de faux positifs.

Concrètement, les seuils sont définis de manière à inclure toutes les valeurs observées en situation de défaillance, tout en maximisant la probabilité qu'un dépassement soit réellement associé à une panne. Ce sont donc des seuils d'alerte, et non de panne stricte.

Une fois ces seuils établis, il est possible d'associer à leur dépassement une probabilité de panne conditionnelle. Par exemple, si un paramètre dépasse le seuil dans 20 cas, dont 10 correspondent à des états de panne, la probabilité de défaillance associée à ce dépassement est de 50 %. Cette information probabiliste peut être exploitée dans les modèles de diagnostic pour affiner les décisions, notamment dans les réseaux bayésiens où ces notions d'incertitude sont intégrées nativement.

L'introduction de ces seuils d'alerte constitue une étape fondamentale pour rapprocher les bases de données utilisées du fonctionnement observé sur le terrain. Elle garantit la compatibilité avec les hypothèses de la méthode MSLD, et permet d'obtenir des états de santé cohérents avec la logique de notre cas d'étude.

Sur cette base transformée et enrichie, la méthode MSLD peut maintenant être appliquée pour générer les états de santé nécessaires aux tâches de diagnostic et de pronostic.

Application de la méthode MSLD pour la création des états de santé

La méthode MSLD (Multivariable System Labeling through Discretization) est appliquée aux jeux de données transformés afin d'établir une classification structurée des états de santé des équipements. Cette approche permet de segmenter les variables continues en catégories discrètes, facilitant l'interprétation des tendances de dégradation et l'identification des niveaux de risque.

À l'origine, la base de données AI4I fournit uniquement une indication binaire de panne (Panne ou OK), sans distinguer les degrés intermédiaires de dégradation. L'application de MSLD permet d'affiner cette information en introduisant plusieurs niveaux d'états de santé. Les paramètres tels que la température, la vitesse de rotation et le couple sont ainsi discrétisés selon cinq classes : optimal, stable, instable, alerte, et panne lorsque cela est applicable.

La discrétisation s'appuie sur les seuils définis lors de l'étape de prétraitement. Entre la classe optimale et celle d'alerte, deux classes de transition sont ajoutées : stable et instable. Elles permettent de capter la dégradation progressive sans attendre que la mesure atteigne le seuil d'alerte. Les seuils de panne, en revanche, ne sont définis que dans les cas où la base de données d'origine les fournit explicitement. C'est notamment le cas pour la base AI4I, où les règles de défaillance sont connues et intégrées dans le processus.

Pour les bases de données ne disposant pas de seuils de panne précis, comme la base Genesis Demonstrator, seule la classe d'alerte est définie comme seuil critique. Dans ce cas, aucune classe panne n'est ajoutée, car la relation directe entre mesure et panne n'est pas explicitement fournie. La méthode MSLD permet alors de structurer les états jusqu'à la classe d'alerte, fournissant une base robuste pour l'anticipation des dégradations. L'application de ces principes à la base Genesis Demonstrator permet d'obtenir une nouvelle version enrichie, nommée GD-PMDI, construite selon les mêmes logiques que AI4I-PMDI.

Dans les cas où la base d'origine indique une panne avérée, comme dans AI4I, l'état de santé est directement remplacé par la classe de panne correspondante. Cette règle garantit la cohérence avec les informations initiales et maintient la fiabilité du diagnostic.

L'adaptabilité de MSLD constitue un autre atout, permettant d'ajuster dynamiquement le nombre d'états en fonction des besoins de l'analyse. Cette flexibilité est essentielle pour adapter la méthode aux spécificités des systèmes industriels et garantir une représentation pertinente de l'état des équipements.

L'intégration de cette segmentation sur les jeux de données transformés fournit une base structurée pour l'analyse des performances des modèles de diagnostic. Les états ainsi définis sont ensuite exploités par les algorithmes de maintenance prévisionnelle pour affiner les prédictions et améliorer la prise de décision.

L'étape suivante consiste à sélectionner les modèles de diagnostic les plus adaptés pour exploiter ces informations, en tenant compte des contraintes et objectifs du système étudié.

3.2 Choix et justification du modèle de diagnostic

3.2.1 Critères de sélection des algorithmes

La maintenance prévisionnelle repose sur des algorithmes capables d’anticiper les défaillances des équipements et d’évaluer leur évolution au fil du temps. L’un des premiers objectifs est de prévoir l’évolution des paramètres critiques des équipements afin de détecter les tendances anormales susceptibles d’entraîner des défaillances. Lorsqu’un paramètre dépasse un seuil critique, le risque de panne augmente, ce qui permet aux équipes de maintenance d’anticiper les interventions nécessaires. Cependant, la simple prédiction de valeurs ne suffit pas toujours à caractériser des environnements complexes où les interactions entre paramètres influencent la fiabilité globale du système.

L’estimation des probabilités de panne constitue une approche complémentaire, en offrant une vision plus nuancée du risque. Contrairement aux seuils fixes, une modélisation probabiliste permet d’exprimer un degré de confiance associé à chaque prédiction et d’adapter les stratégies de maintenance en conséquence. Cette approche améliore la planification des interventions en tenant compte de l’incertitude et de la variabilité des systèmes.

Au-delà de l’anticipation des pannes, la prise en compte de la dimension temporelle est essentielle. Les algorithmes doivent être capables d’analyser l’évolution des équipements sur une période donnée, afin de mieux comprendre les dynamiques de dégradation et d’identifier les fenêtres critiques où le risque est maximal. Cette capacité permet d’optimiser la gestion des ressources en maintenance et d’éviter des interventions inutiles ou tardives.

Un diagnostic précis des pannes est également indispensable pour identifier les causes sous-jacentes des anomalies détectées. Une simple détection d’anomalies ne suffit pas à orienter efficacement les équipes de maintenance ; il est nécessaire d’attribuer les dysfonctionnements à des origines spécifiques afin de proposer des actions correctives ciblées. Cette identification des causes permet de réduire les temps d’arrêt, d’optimiser les réparations et d’améliorer la fiabilité globale des équipements.

L’efficacité de ces prédictions repose en grande partie sur l’interprétabilité des modèles utilisés. Dans les environnements industriels, les décisions doivent être justifiables et compréhensibles par les experts métier. Un modèle prédictif, aussi performant soit-il, sera difficilement adopté s’il ne permet pas aux utilisateurs de comprendre comment et pourquoi une décision a été prise. Une bonne interprétabilité facilite la validation des résul-

tats, l'identification des facteurs influençant les prédictions et l'ajustement des stratégies de maintenance en fonction des retours terrain.

Cette complémentarité entre modélisation statistique et expertise métier favorise une adoption plus large des outils de maintenance prévisionnelle. Un modèle transparent et aligné sur la compréhension des experts sera plus facilement intégré aux processus industriels, garantissant une meilleure prise en compte des recommandations algorithmiques dans la planification des interventions.

Les différentes approches de diagnostic seront comparées afin d'évaluer leur capacité à répondre aux critères de sélection définis précédemment.

3.2.2 Comparaison des approches de diagnostic

Les approches de diagnostic en maintenance prévisionnelle reposent principalement sur l'utilisation d'algorithmes de machine learning (ML) capables d'analyser les données collectées sur les systèmes industriels afin de prédire les défaillances potentielles et d'optimiser les interventions. Les méthodes classiques de ML sont largement adoptées dans ce domaine en raison de leur flexibilité et de leur capacité à traiter de vastes volumes de données. Cependant, chaque algorithme présente ses propres avantages et limites, notamment en termes d'interprétabilité, de gestion des incertitudes et de capacité à intégrer les connaissances d'experts.

Les algorithmes basés sur les arbres de décision, tels que les forêts aléatoires (Random Forest) (BREIMAN, 2001), sont fréquemment utilisés pour leur robustesse. Cette méthode consiste à agréger les prédictions d'un grand nombre d'arbres de décision entraînés sur des sous-ensembles aléatoires des données, réduisant ainsi le risque de surajustement. Ces modèles offrent une certaine transparence car il est possible de visualiser les règles de décision menant aux prédictions. Cependant, leur capacité à modéliser des relations complexes entre variables est limitée, surtout lorsque les données présentent des corrélations non linéaires ou des dépendances temporelles.

Les machines à vecteurs de support (SVM) (CORTES et VAPNIK, 1995) sont populaires pour les tâches de classification. Leur principe est de trouver l'hyperplan qui sépare les données en différentes classes tout en maximisant la marge, c'est-à-dire la distance entre cet hyperplan et les points les plus proches de chaque classe. Elles sont efficaces même dans les cas de jeux de données de petite taille. Cependant, leur interprétabilité est réduite, surtout lorsqu'elles utilisent des noyaux non linéaires, et leur performance dépend fortement du choix des hyperparamètres et du processus de validation croisée.

Les réseaux de neurones et leurs variantes profondes, telles que les réseaux de neurones convolutifs (CNN) (FUKUSHIMA, 1980) et les réseaux récurrents (RNN) (RUMELHART ; HINTON ; WILLIAMS et al., 1985), sont capables de modéliser des relations complexes et non linéaires entre les variables. Ils excellent dans la détection de schémas complexes et dans le traitement de données séquentielles. Cependant, ces modèles sont souvent qualifiés de "boîtes noires" en raison de leur faible interprétabilité. Il est difficile de comprendre précisément comment une prédition est générée, ce qui limite leur adoption dans les contextes industriels nécessitant des décisions explicables et justifiables.

Les méthodes de régression linéaire et logistique, bien que plus simples, sont encore largement utilisées pour leur interprétabilité directe. Ces modèles permettent de comprendre facilement les relations entre les variables d’entrée et les résultats. Cependant, leur capacité à capturer des relations complexes est limitée, ce qui peut affecter leur précision, en particulier dans les environnements industriels complexes où les phénomènes non linéaires sont fréquents.

Les approches séquentielles comme les modèles ARIMA (Autoregressive integrated moving average) (Box et al., 1970) et les LSTM (Long Short-Term Memory) (HOCHREITER et SCHMIDHUBER, 1997) sont particulièrement adaptés aux séries temporelles, un aspect central en maintenance prévisionnelle. Les LSTM, par exemple, peuvent modéliser la dépendance temporelle des données, ce qui permet de mieux anticiper l’évolution des paramètres critiques. Toutefois, ces modèles souffrent également d’un manque d’interprétabilité et nécessitent souvent un grand volume de données pour atteindre des performances optimales.

Les limites communes à ces approches résident dans leur difficulté à gérer les incertitudes et à intégrer directement les connaissances d’experts dans les processus de modélisation. La plupart des algorithmes de ML traitent les données de manière déterministe et nécessitent des volumes importants de données pour produire des résultats fiables. Ils sont également sensibles aux déséquilibres de classes et aux données manquantes, des défis courants dans les environnements industriels.

Le tableau 3.3 présente une comparaison des principales approches de machine learning utilisées en maintenance prévisionnelle selon des critères clés : la capacité à prédire les valeurs des paramètres, l’estimation des probabilités de panne, la prise en compte de la dimension temporelle, les capacités de diagnostic, l’intégration des connaissances d’experts et l’interprétabilité.

La colonne *Note* du tableau indique le nombre total de critères remplis par chaque

méthode. Cette note permet d'avoir une vue synthétique de la couverture fonctionnelle de chaque approche : plus la note est élevée, plus l'algorithme satisfait de critères jugés importants dans le cadre d'un diagnostic complet en maintenance prévisionnelle. Elle facilite ainsi l'identification des méthodes les plus polyvalentes et adaptées aux contraintes des environnements industriels complexes.

TABLE 3.3 – Comparaison entre différentes approches de machine learning pour la maintenance prévisionnelle.

| Algorithme | Prédiction paramètres | Probabilité panne | Prédiction temporelle | Diagnostic panne | Connaissances expertes | Interprétabilité | Référence | Note |
|---------------------|-----------------------|-------------------|-----------------------|------------------|------------------------|------------------|---|------|
| Elastic Net | X | | X | X | | X | (LANGONE ; CUZZOCREA et SKANTZOS, 2020) | 4 |
| Logistic Regression | X | | X | X | | X | (COSTELLO ; WEST et MCARTHUR, 2017) | 4 |
| Regression Tree | X | | X | X | | X | (ZHENG ; MALBASA et KEZUNOVIC, 2013) | 4 |
| ARIMA | X | | | | | X | (LIN et al., 2019) | 3 |
| GPR | X | | | | | X | (RICHARDSON ; OSBORNE et HOWEY, 2017) | 2 |
| BN | X | X | | X | X | X | (FERREIRO et al., 2012) | 5 |
| DBN | X | X | X | X | X | X | (LEWIS et GROTH, 2020) | 6 |
| Analyse de survie | | X | X | | | X | (CHEN et al., 2020) | 3 |
| Random Forest | X | | | X | | | (JIMENEZ-CORTADI et al., 2020) | 2 |
| SVM | X | | | X | | | (ORRÙ et al., 2020) | 2 |
| LSTM | X | | X | X | | | (YUAN ; WU et LIN, 2016) | 3 |
| VAE | X | | X | | | X | (ZHAO et al., 2019a) | 3 |
| GAN | X | | | X | | | (YOON ; JARRETT et SCHAAR, 2019) | 2 |
| Transformer | X | | X | | | | (LIM et al., 2021) | 2 |

ARIMA : Autoregressive Integrated Moving Average, BN : Bayesian Network, DBN : Dynamic Bayesian Network, GAN : Generative Adversarial Network, GPR : Gaussian Process Regression, LSTM : Long Short-Term Memory, SVM : Support Vector Machine, VAE : Variational AutoEncoder.

L'analyse comparative met en évidence que les méthodes classiques de machine learning présentent des forces variées, mais qu'aucune ne satisfait pleinement à l'ensemble des critères nécessaires pour un diagnostic optimal en maintenance prévisionnelle. Les réseaux bayésiens dynamiques, par exemple, se distinguent par leur capacité à gérer les incertitudes, à intégrer les connaissances d'experts et à prendre en compte la dimension temporelle, ce qui les rend particulièrement adaptés à des contextes industriels complexes.

Dans la section suivante, nous examinerons en détail les raisons qui justifient le choix des réseaux bayésiens comme approche privilégiée pour le diagnostic en maintenance prévisionnelle. Nous analyserons leurs capacités uniques, notamment en matière d'intégration des connaissances d'experts, d'interprétabilité et de gestion des incertitudes.

3.2.3 Choix des réseaux bayésiens

Les réseaux bayésiens offrent une approche adaptée aux défis de la maintenance prévisionnelle en raison de leur capacité à modéliser les relations probabilistes entre les paramètres d'un système. Ils permettent d'intégrer les connaissances des experts et de modéliser les incertitudes inhérentes aux environnements industriels. Grâce à leur structure graphique, ils assurent une interprétation claire des dépendances entre variables, ce qui les rend particulièrement intéressants pour le diagnostic et la prédiction des pannes. Comparés à d'autres méthodes comme les LSTM ou les modèles de forêts aléatoires, ils se distinguent par leur transparence et leur capacité à expliciter les relations causales, tandis que ces autres méthodes fonctionnent souvent comme des *boîtes noires*, rendant difficile l'interprétation des résultats.

Un aspect fondamental des réseaux bayésiens est leur capacité d'inférence bidirectionnelle, qui permet d'analyser les relations entre les variables du système dans les deux sens. D'une part, l'inférence causale permet de propager l'information vers l'avant : en connaissant l'état actuel de certaines variables, il devient possible d'estimer la probabilité d'occurrence de futurs événements, notamment des pannes éventuelles. Par exemple, si une augmentation de la température et une baisse de la pression sont observées, le réseau bayésien peut estimer la probabilité qu'une défaillance survienne dans les heures suivantes. Cette capacité est essentielle pour anticiper les dégradations et optimiser les stratégies de maintenance préventive.

D'autre part, l'inférence diagnostique (*rétropropagation*) permet de propager l'information vers l'arrière afin d'identifier les causes probables d'un événement observé. Si une panne est détectée, le réseau peut ajuster les probabilités des variables influençant di-

rectement ou indirectement cette défaillance pour remonter aux origines du problème. Cela permet, par exemple, de répondre à la question : "Sachant qu'une défaillance a été observée, quels paramètres anormaux en sont probablement la cause ?". Grâce à cette approche, les équipes de maintenance peuvent orienter plus efficacement leurs interventions en ciblant les facteurs contributifs plutôt que de se baser uniquement sur des inspections systématiques.

Cette double capacité d'inférence améliore la pertinence des réseaux bayésiens dans les environnements industriels, où il est crucial non seulement de prévoir les pannes mais aussi d'en comprendre les origines pour optimiser les actions correctives. Contrairement aux méthodes telles que les SVM ou les modèles ARIMA, qui ne fonctionnent que dans une seule direction en exploitant uniquement les tendances passées pour la prédiction, les réseaux bayésiens offrent une flexibilité accrue. Ils permettent une analyse complète des systèmes en intégrant à la fois les relations de causalité et les mécanismes sous-jacents aux défaillances, ce qui renforce leur efficacité pour la prise de décision en maintenance prévisionnelle.

Les réseaux bayésiens se distinguent également par leur gestion des incertitudes. Contrairement aux modèles déterministes, ils associent à chaque variable une probabilité de présence ou d'occurrence, ce qui permet d'estimer la fiabilité des décisions prises. Cette caractéristique est essentielle lorsque les données disponibles sont partielles, bruitées ou présentent des incohérences. L'intégration de cette incertitude dans les prédictions permet aux opérateurs de mieux évaluer les risques et de prendre des décisions en tenant compte des niveaux de confiance associés aux résultats du modèle. À l'inverse, des approches comme la régression logistique ou les arbres de décision offrent peu de mécanismes pour quantifier ces incertitudes, ce qui peut conduire à des décisions moins robustes dans des environnements industriels complexes.

L'interprétabilité des réseaux bayésiens et leur capacité à intégrer la connaissance métier constituent un autre avantage clé. Le modèle représente explicitement les relations de dépendance entre variables sous forme de graphes probabilistes, facilitant ainsi la compréhension des mécanismes qui influencent la santé de l'équipement. Cette transparence permet une collaboration directe avec les experts du domaine, qui peuvent non seulement valider la structure du modèle, mais aussi contribuer à sa construction. En effet, il est possible de définir les relations causales du graphe ou d'estimer les probabilités initiales en se basant sur leur savoir-faire, ce qui est particulièrement précieux lorsque les données d'apprentissage sont rares ou incomplètes. Cette fusion de la connaissance experte et

des données rend le modèle plus robuste et renforce la confiance des équipes de terrain. Par comparaison, des modèles comme les GANs ou les transformateurs, bien que puissants, manquent souvent de cette capacité d'explication et de co-construction.

Enfin, l'extension des réseaux bayésiens sous forme dynamique (*DBN*) permet d'intégrer la dimension temporelle dans les analyses. Cela rend possible le suivi de l'évolution des paramètres au fil du temps et l'identification de tendances qui pourraient conduire à des pannes futures. Grâce à cette approche, il est envisageable d'optimiser la planification des opérations de maintenance en anticipant les périodes critiques où le risque de défaillance est maximal. Les DBN se comparent favorablement aux modèles récurrents tels que LSTM, qui bien qu'efficaces pour la modélisation des séries temporelles, nécessitent des volumes de données importants pour l'entraînement et ne permettent pas une explicabilité aussi directe.

Le choix des réseaux bayésiens repose donc sur leur capacité à répondre aux exigences spécifiques de la maintenance prévisionnelle. Ils combinent une modélisation probabiliste rigoureuse, une interprétation claire des résultats, une gestion avancée des incertitudes et une inférence bidirectionnelle permettant d'analyser les symptômes et leurs causes de manière cohérente. Contrairement aux modèles d'apprentissage profond qui nécessitent de grandes quantités de données et offrent peu d'interprétabilité, les réseaux bayésiens permettent d'exploiter efficacement des jeux de données limités tout en restant compréhensibles par les experts métier. Ces caractéristiques en font une solution robuste et adaptée aux besoins des environnements industriels.

Dans la section suivante, la mise en œuvre du réseau bayésien est détaillée, en précisant les étapes de modélisation, d'apprentissage et d'inférence utilisées pour exploiter pleinement ses capacités.

3.3 Modélisation et mise en œuvre du réseau bayésien

3.3.1 Fondements des réseaux bayésiens

Théorie des graphes

Les réseaux bayésiens s’appuient sur la théorie des graphes pour représenter les relations probabilistes entre les variables d’un système. Ils reposent sur une structure de graphe orienté acyclique (*Directed Acyclic Graph*, DAG), où chaque nœud représente une variable aléatoire et chaque arc traduit une dépendance conditionnelle. Avant d’aborder leur application en maintenance prévisionnelle, il est utile de rappeler quelques notions fondamentales des graphes.

Un graphe est une structure mathématique définie comme un couple (V, E) , où :

- V est l’ensemble des sommets (ou nœuds) ;
- E est l’ensemble des arêtes (ou arcs) reliant ces sommets.

Les graphes se distinguent selon la nature de leurs relations. Dans un graphe non-orienté, les connexions entre les nœuds sont symétriques : si un nœud X_1 est relié à un nœud X_2 , alors la connexion est réciproque. Ces structures sont souvent utilisées pour modéliser des relations de corrélation, sans relation de causalité directe.

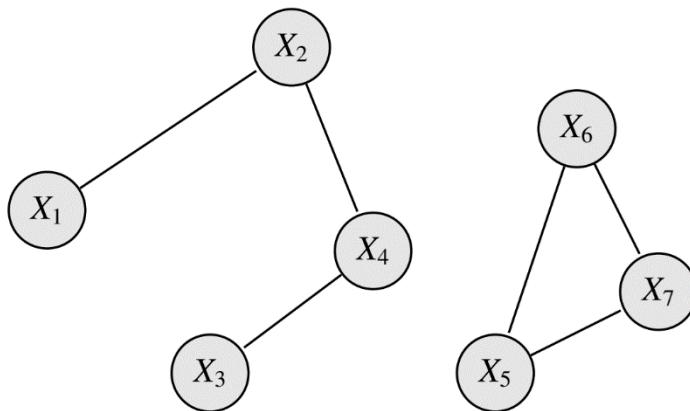


FIGURE 3.2 – Exemples de graphes non-orientés.

À l’inverse, un graphe orienté impose une direction aux relations entre les nœuds. Un arc dirigé de X_1 vers X_2 signifie que X_1 influence X_2 . Cette structure est particulièrement utile pour modéliser les relations causales entre variables.

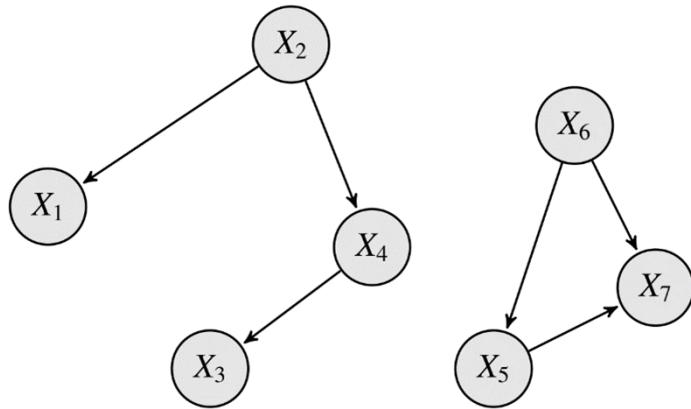


FIGURE 3.3 – Exemples de graphes orientés.

Dans un graphe orienté, une séquence de nœuds reliés par des arcs dirigés est appelée chemin. Lorsqu'un chemin revient à son point de départ, il forme un cycle. Un graphe orienté acyclique (DAG) est une structure où aucun cycle ne peut exister, garantissant ainsi une organisation hiérarchique des relations entre variables. Cette propriété est essentielle pour les réseaux bayésiens, car elle permet une factorisation efficace des distributions de probabilité et assure la cohérence des dépendances conditionnelles.

Les réseaux bayésiens exploitent ces propriétés pour modéliser les interactions entre différents paramètres d'un système. Ils permettent ainsi d'estimer les probabilités d'occurrence de divers états et d'identifier les causes probables d'anomalies. En combinant connaissances d'experts et données historiques, ils ajustent dynamiquement les probabilités conditionnelles, ce qui en fait un outil particulièrement pertinent pour la maintenance prévisionnelle.

Définition et structure du réseau bayésien

Un réseau bayésien est un modèle probabiliste structuré sous la forme d'un graphe orienté acyclique (DAG). Il représente les relations de dépendance conditionnelle entre variables aléatoires et permet d'inférer la probabilité d'événements futurs à partir des observations disponibles.

Formellement, un réseau bayésien est défini par un couple $B = (G, P)$, où :

- $G = (V, E)$ est un graphe orienté acyclique constitué d'un ensemble de nœuds V , représentant les variables aléatoires, et d'un ensemble d'arcs E , modélisant les dépendances conditionnelles ;

- P regroupe les distributions de probabilités conditionnelles associées à chaque variable.

Chaque variable X_i suit une distribution conditionnelle $P(X_i|Pa(X_i))$, où $Pa(X_i)$ désigne ses parents dans le graphe. Grâce à cette structure factorisée, la distribution conjointe des variables du réseau est donnée par :

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i|Pa(X_i))$$

Cette factorisation réduit la complexité du calcul des probabilités en exploitant l’indépendance conditionnelle entre variables non directement connectées. Elle permet ainsi de structurer efficacement la propagation de l’information et l’inférence probabiliste dans le réseau.

L’inférence dans un réseau bayésien repose sur le théorème de Bayes, qui permet de mettre à jour la probabilité d’un événement en fonction de nouvelles observations. Si une variable X dépend de Y , sa probabilité conditionnelle peut être exprimée sous la forme :

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

où $P(X|Y)$ est la vraisemblance, $P(Y)$ est la probabilité a priori de Y et $P(X)$ est une constante de normalisation. Ce théorème est central dans l’actualisation des croyances dans un réseau bayésien, car il permet d’inférer la distribution des variables cachées à partir des données observées.

Trois éléments fondamentaux définissent un réseau bayésien : les variables aléatoires, les relations de dépendance et les distributions de probabilités conditionnelles. Les variables, représentées par les nœuds du graphe, peuvent être discrètes ou continues en fonction du problème étudié. Dans un contexte industriel, elles correspondent souvent à des grandeurs mesurées telles que la température, la pression ou la vibration d’un équipement. Les relations de dépendance, représentées par les arcs dirigés, traduisent l’impact d’une variable sur une autre. Enfin, chaque nœud est associé à une table de probabilités conditionnelles décrivant la manière dont il est influencé par ses parents. Ces distributions peuvent être estimées à partir de données historiques ou définies manuellement par des experts lorsqu’aucune donnée n’est disponible.

Un réseau bayésien peut être construit de deux manières. Une première approche repose sur l’expertise métier pour définir la structure et les probabilités conditionnelles. Dans de nombreux systèmes industriels, les relations entre variables sont bien connues et

peuvent être intégrées directement dans le modèle. Par exemple, un expert en maintenance sait qu'une température moteur élevée combinée à une faible pression est un indicateur fort de dysfonctionnement. Même en l'absence de données, ces connaissances peuvent être formalisées sous forme de relations causales dans le réseau.

Une seconde approche consiste à apprendre la structure du réseau et ses paramètres à partir des données disponibles. Cette méthode est particulièrement utile lorsque les relations entre variables sont mal connues et nécessitent une analyse statistique pour être découvertes. Toutefois, les modèles appris automatiquement peuvent être difficiles à interpréter sans ajustements manuels. Dans la pratique, une approche hybride combinant expertise humaine et apprentissage automatique est souvent privilégiée. C'est cette démarche qui a été retenue dans notre étude, afin de capitaliser à la fois sur la connaissance métier disponible et sur les informations extraites des données.

Grâce à leur capacité à modéliser les relations probabilistes et à mettre à jour dynamiquement les probabilités, les réseaux bayésiens sont particulièrement adaptés à la maintenance prévisionnelle. Leur structure permet d'intégrer des observations et d'ajuster les prévisions en fonction des nouvelles données. L'étape suivante consiste à examiner les méthodes d'apprentissage de la structure et des paramètres afin d'optimiser ces modèles.

3.3.2 Apprentissage de la structure et des paramètres

Apprentissage de la structure

L'apprentissage de la structure d'un réseau bayésien vise à déterminer l'organisation optimale du graphe représentant les relations de dépendance entre les variables aléatoires. Cette étape est essentielle pour garantir une modélisation efficace et interprétable. Trois approches principales permettent de construire cette structure : les méthodes basées sur la recherche d'indépendances, celles utilisant un critère de score et les approches hybrides combinant ces deux stratégies (HECKERMAN ; GEIGER et CHICKERING, 1995 ; SPIRITES ; GLYMOU et SCHEINES, 1993).

Les méthodes fondées sur la recherche d'indépendances exploitent des tests statistiques pour identifier les relations conditionnelles entre les variables. Elles reposent sur le principe selon lequel si deux variables sont indépendantes sous certaines conditions, il n'est pas nécessaire d'ajouter un arc entre elles. L'algorithme PC (SPIRITES ; GLYMOU et SCHEINES, 1993), largement utilisé, applique le test du χ^2 (CHERNOFF et LEHMANN, 1954) pour

évaluer ces relations et simplifier progressivement la structure du graphe en supprimant les liens superflus. Bien que cette méthode offre une interprétation claire des dépendances identifiées, elle nécessite un volume de données important pour garantir des résultats fiables et peut être sensible au bruit, ce qui peut entraîner des erreurs dans la structure du réseau.

Une autre approche consiste à sélectionner la meilleure structure de graphe en maximisant un critère de score. Elle repose sur l’exploration d’un vaste espace de solutions, où chaque structure candidate est évaluée selon une fonction de score permettant de mesurer sa pertinence. Parmi les scores les plus couramment utilisés, on distingue plusieurs catégories :

- **Scores bayésiens** : Ces scores reposent sur l’estimation des probabilités conditionnelles en appliquant des principes bayésiens. Ils incluent BD (*Bayesian Dirichlet*) (HECKERMAN ; GEIGER et CHICKERING, 1995), K2 (COOPER et HERSOVITS, 1992), BDe (*Bayesian Dirichlet equivalent*) (HECKERMAN ; GEIGER et CHICKERING, 1995) et BDeu (*Bayesian Dirichlet equivalent uniform*) (BUNTINE, 1991).
- **Scores basés sur l’information** : Ces scores cherchent à équilibrer la qualité du modèle et sa complexité pour éviter le surajustement. On retrouve MDL (*Minimum Description Length*) (SUZUKI, 1993), AIC (*Akaike Information Criterion*) (AKAIKE, 1974) et BIC (*Bayesian Information Criterion*).
- **Scores basés sur la vraisemblance** : Ces méthodes évaluent la structure en maximisant la cohérence entre le modèle et les données. Cela inclut LL (*Log-Likelihood*) (BOUCKAERT, 1995) et NML (*Normalized Maximum Likelihood*) (ROOS et al., 2008).
- **Scores fondés sur l’information mutuelle** : Ces scores exploitent la dépendance entre variables pour guider l’apprentissage. Un exemple notable est MIT (*Mutual Information Tests*) (CAMPOS, 2006).

L’optimisation de ces scores implique d’explorer un nombre exponentiel de structures possibles, rendant une recherche exhaustive impraticable au-delà d’un certain nombre de variables. Pour surmonter cette difficulté, des méthodes heuristiques ont été développées afin d’explorer efficacement l’espace des solutions.

- L’algorithme K2 (COOPER et HERSOVITS, 1992) construit progressivement la structure du réseau en ajoutant des parents à chaque variable suivant un ordre prédéfini. Toutefois, ce choix d’ordre influence directement la structure obtenue et l’algorithme ne permet pas de supprimer des arcs une fois ajoutés, ce qui limite la

flexibilité du modèle.

- L'algorithme *Greedy Search* ou *Hill-Climbing* applique une optimisation gloutonne en modifiant le graphe par ajouts, suppressions ou inversions d'arcs. À chaque itération, la meilleure structure selon le score choisi est conservée. Bien que cette méthode permette une convergence rapide, elle peut être piégée dans des optima locaux.
- L'algorithme *Tabu Search* (GLOVER et LAGUNA, 1993) améliore le *Hill-Climbing* en autorisant temporairement des solutions sous-optimales pour échapper aux optima locaux. Il conserve un historique des structures explorées afin d'éviter de revenir inutilement sur des configurations précédemment rejetées.

Les approches hybrides combinent les tests d'indépendance et l'optimisation par score pour tirer parti des avantages de chaque méthode. Dans un premier temps, des tests statistiques éliminent les relations non pertinentes et réduisent l'espace de recherche. Ensuite, un algorithme d'optimisation affine la structure en maximisant un critère de score. L'algorithme BENEDICT (ACID et CAMPOS, 2001) suit cette logique en s'appuyant sur des tests d'indépendance pour initialiser la structure, avant d'appliquer une optimisation. Une autre stratégie hybride, proposée par (CAMPOS ; FERNÁNDEZ-LUNA et PUERTA, 2003), alterne entre ces deux types d'approches de manière itérative afin d'améliorer progressivement la qualité du modèle.

La Figure 3.4 illustre l'évolution chronologique des stratégies d'apprentissage de structure, en classant les algorithmes par grandes familles. On observe que les approches initiales, dans les années 1990, se concentraient principalement sur des méthodes pures : d'un côté les algorithmes basés sur les contraintes (en vert sur le schéma), comme le pionnier *PC*, et de l'autre les approches par score (en jaune), telles que *K2*. À partir des années 2000, le schéma met en évidence une tendance claire à l'hybridation (en bleu), où les deux stratégies sont combinées pour en mutualiser les forces. Des algorithmes comme *MMHC* (*Max-Min Hill-Climbing*) émergent comme des solutions robustes, en utilisant des tests d'indépendance pour restreindre l'espace de recherche avant d'appliquer une optimisation par score. Le diagramme montre également que la recherche s'est poursuivie au sein de chaque famille, avec des raffinements constants, illustrant une quête continue pour un meilleur compromis entre précision, robustesse et efficacité computationnelle.

L'apprentissage de la structure est une étape clé dans la construction d'un réseau bayésien. Il permet d'adapter la modélisation aux spécificités du domaine étudié tout en maintenant un équilibre entre complexité et interprétabilité. Une fois la structure définie,

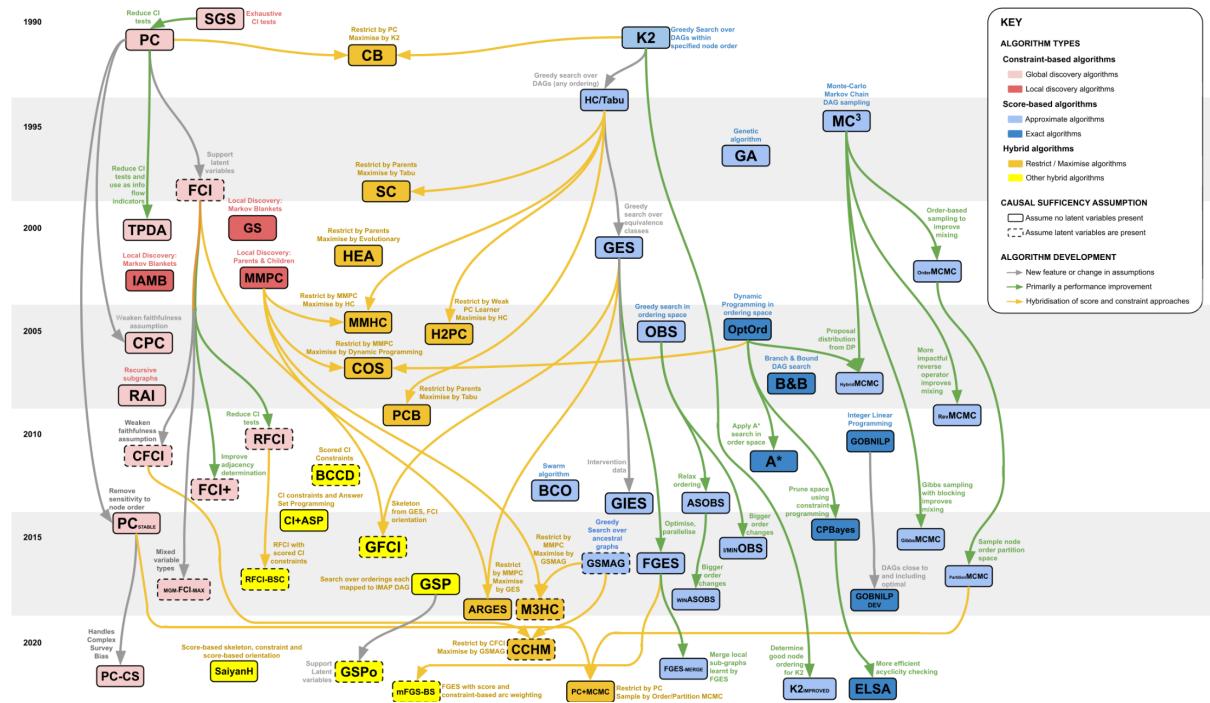


FIGURE 3.4 – Évolution des algorithmes d'apprentissage de la structure des réseaux bayésiens (KITSON et al., 2023).

l'étape suivante consiste à estimer les paramètres du modèle, c'est-à-dire les probabilités conditionnelles associées aux nœuds, afin d'obtenir une représentation quantitative fiable du système.

Apprentissage des paramètres

Une fois la structure du réseau bayésien déterminée, l'étape suivante est l'apprentissage des paramètres, qui consiste à estimer les distributions de probabilités conditionnelles $P(X_i|Pa(X_i))$ pour chaque variable X_i . Pour cette étude, l'algorithme ***Expectation-Maximization (EM)*** (DEMPSTER ; LAIRD et RUBIN, 1977) a été retenu. Ce choix est justifié par sa capacité à gérer efficacement les données manquantes, une contrainte centrale de notre cas d'étude où les mesures peuvent être irrégulières.

L'algorithme EM est une méthode itérative utilisée pour trouver les paramètres du maximum de vraisemblance d'un modèle probabiliste en présence de variables latentes ou de données manquantes. Soit X l'ensemble des variables observées et Z l'ensemble des variables latentes (les données manquantes), la log-vraisemblance des paramètres θ est définie par :

$$\ell(\theta) = \log P(X|\theta) = \log \sum_Z P(X, Z|\theta)$$

La marginalisation sur les variables latentes Z rend souvent ce calcul intraitable. L'algorithme EM contourne cette difficulté en alternant deux étapes jusqu'à convergence.

La première, l'étape E (Expectation), consiste à calculer l'espérance de la log-vraisemblance des données complètes, conditionnellement aux observations X et aux paramètres courants $\theta^{(t)}$. Cette espérance est notée $Q(\theta|\theta^{(t)})$:

$$Q(\theta|\theta^{(t)}) = E_{Z|X,\theta^{(t)}}[\log P(X, Z|\theta)] = \sum_Z P(Z|X, \theta^{(t)}) \log P(X, Z|\theta)$$

Cette étape revient à estimer la distribution des données manquantes en se basant sur les paramètres actuels.

La seconde, l'étape M (Maximization), met ensuite à jour les paramètres en maximisant la fonction Q calculée à l'étape E :

$$\theta^{(t+1)} = \arg \max_{\theta} Q(\theta|\theta^{(t)})$$

L'algorithme garantit une augmentation de la log-vraisemblance $\ell(\theta)$ à chaque itération jusqu'à atteindre un optimum local. Grâce à ce mécanisme, l'algorithme EM a permis d'ajuster dynamiquement les probabilités conditionnelles de notre modèle en prenant en compte l'incertitude liée aux observations absentes.

D'autres méthodes d'estimation ont été considérées. L'estimation par maximum de vraisemblance (MLE), qui estime les probabilités à partir des fréquences observées, est simple mais peu robuste face aux données manquantes, ce qui la rendait moins adaptée à notre contexte. De même, l'estimation bayésienne, bien que très performante pour intégrer des connaissances a priori, a été jugée moins directe que l'approche EM pour traiter la problématique spécifique des variables latentes dans notre jeu de données.

Grâce à l'application de l'algorithme EM, les relations entre les différentes variables de surveillance de l'équipement ont pu être quantifiées de manière fiable. L'estimation des probabilités conditionnelles à partir des données collectées, même incomplètes, a permis de finaliser un modèle capable d'évaluer les risques de défaillance et d'anticiper les anomalies.

À l'issue de cet apprentissage, le réseau bayésien est entièrement paramétré et modélise avec précision les dépendances au sein du système. Pour exploiter pleinement son potentiel dans un cadre opérationnel, il est cependant indispensable d'y intégrer un méca-

nisme synthétisant ces informations en un diagnostic global, qui sera l’objet de la section suivante.

Ajout des nœuds de diagnostic

Le diagnostic global est défini comme le pire état observé parmi les paramètres mesurés. Lors de l’apprentissage du réseau bayésien, seules les relations entre les paramètres sont apprises sans inclure l’état global de santé. Une fois la structure apprise, un nœud supplémentaire représentant l’état de santé global est ajouté. Ce nœud applique une règle d’agrégation prenant en compte le pire état des paramètres connectés.

Pour gérer la complexité inhérente à l’agrégation d’un grand nombre de variables, une architecture hiérarchique avec des nœuds de diagnostic intermédiaires est adoptée. Cette structure, illustrée conceptuellement par la Figure 3.5, permet de regrouper les paramètres en sous-systèmes logiques avant de calculer l’état de santé final.

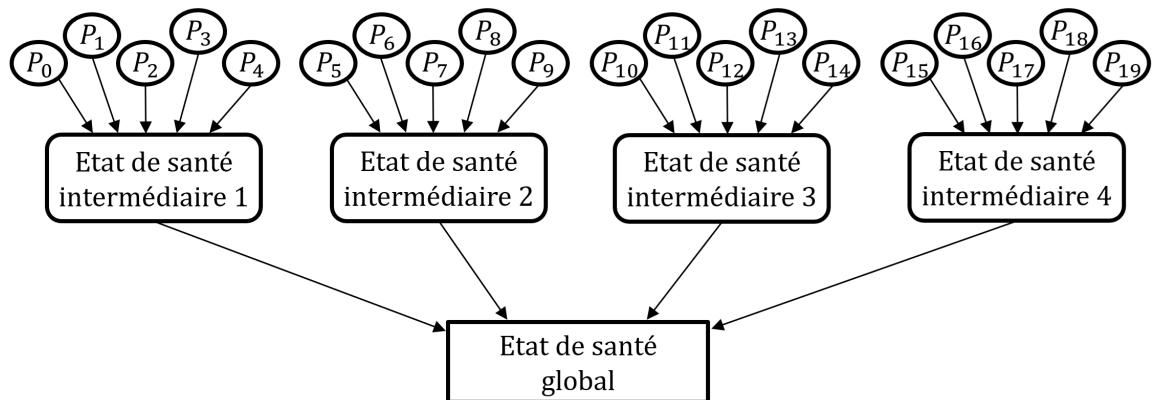


FIGURE 3.5 – Architecture du réseau bayésien de diagnostic avec nœuds intermédiaires pour la réduction de la complexité.

Cette approche est essentielle car la complexité du réseau sans nœuds intermédiaires, liée au nombre total de paramètres connectés directement au nœud diagnostic global, serait prohibitive. La taille de la table de probabilité conditionnelle (CPT) d’un nœud est donnée par :

$$\text{Taille_CPT} = c^n \times c$$

où n est le nombre de parents et c est le nombre de classes discrètes par paramètre. Dans notre cas, nous avons $y = 20$ paramètres et $c = 5$ classes. La taille du CPT du nœud

diagnostic global sans noeuds intermédiaires est donc :

$$\text{Taille_CPT_sans} = 5^{20} \times 5 = 4.77 \times 10^{14}$$

Cette complexité est inexplorable dans un contexte industriel.

Pour réduire cette complexité, des noeuds de diagnostic intermédiaires sont introduits. Chaque noeud intermédiaire regroupe un sous-ensemble de paramètres, applique localement la règle du pire état, et est ensuite relié au noeud diagnostic global.

Le nombre optimal de noeuds intermédiaires x est déterminé en minimisant la complexité totale suivante :

$$\text{Complexité}(x) = (c - 1) \times \left(y \times c^x + c^{\frac{y}{x}} \right)$$

avec $c = 5$, $y = 20$.

Pour déterminer x optimal, la formule est calculée pour différentes valeurs de x :

$$\begin{aligned} x = 1 : \quad & 4 \times (20 \times 5^1 + 5^{20}) = 4 \times (100 + 9.54 \times 10^{13}) = 3.82 \times 10^{14} \\ x = 2 : \quad & 4 \times (20 \times 5^2 + 5^{10}) = 4 \times (500 + 9765625) = 39082500 \\ x = 3 : \quad & 4 \times (20 \times 5^3 + 5^{6.66}) = 4 \times (2500 + 78125) = 322500 \\ x = 4 : \quad & 4 \times (20 \times 5^4 + 5^5) = 4 \times (12500 + 3125) = 62500 \\ x = 5 : \quad & 4 \times (20 \times 5^5 + 5^4) = 4 \times (62500 + 625) = 250500 \end{aligned}$$

L'évolution de la complexité en fonction de x est représentée sur la Figure 3.6.

Le minimum est atteint pour $x = 4$. Avec $x = 4$ noeuds intermédiaires, chaque noeud intermédiaire reçoit $y/x = 5$ parents. La taille des CPTs intermédiaires est :

$$\text{Taille_CPT_intermediaire} = 5^5 \times 5 = 15625$$

Il y a 4 noeuds intermédiaires, donc la complexité des CPTs intermédiaires est :

$$4 \times 15625 = 62500$$

Le noeud diagnostic global reçoit 4 parents (les noeuds intermédiaires), avec 5 classes

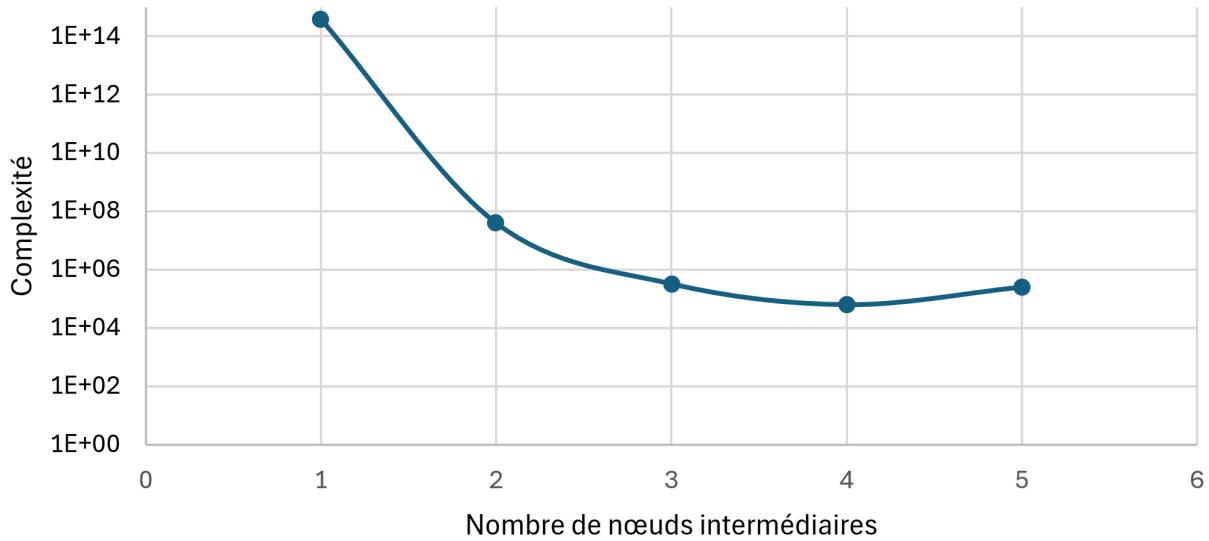


FIGURE 3.6 – Évolution de la complexité (échelle logarithmique) en fonction du nombre de noeuds intermédiaires.

chacun. La taille de sa CPT est :

$$\text{Taille_CPT_global} = 5^4 \times 5 = 3125$$

La complexité totale avec nœuds intermédiaires est donc :

$$\text{Complexité_totale_avec} = 62500 + 3125 = 65625$$

Le rapport de complexité entre le modèle sans nœuds intermédiaires et celui avec est :

$$\frac{4.77 \times 10^{14}}{65625} \approx 7.27 \times 10^9$$

L’introduction des nœuds intermédiaires réduit donc la complexité d’un facteur supérieur à 10^9 . Cette réduction drastique rend possible l’exploitation du modèle sans compromettre la logique du pire état appliquée pour le diagnostic global.

Avec cette structure optimisée, le modèle est prêt pour l’étape suivante, à savoir l’inférence probabiliste, permettant d’estimer l’état global à partir des mesures observées. Cette étape est détaillée dans la section suivante.

3.3.3 Inférence

L’inférence dans un réseau bayésien consiste à calculer la distribution de probabilité marginale d’une ou plusieurs variables en fonction des valeurs observées des autres variables du réseau. Cette opération permet d’estimer la probabilité d’un événement donné à partir des connaissances disponibles et joue un rôle fondamental en modélisation probabiliste et en apprentissage automatique.

L’inférence exacte repose sur le théorème de Bayes et permet d’obtenir des résultats précis. Cependant, cette approche devient rapidement coûteuse en temps de calcul et en mémoire lorsque la structure du réseau est complexe ou contient un grand nombre de variables. Dans de tels cas, la mise à jour des probabilités nécessite des calculs exponentiels, rendant l’inférence exacte impraticable.

Reprendons l’exemple du moteur électrique, où H représente l’humidité, T la température et F l’état du moteur. Si l’humidité est observée comme élevée ($H = \text{élevée}$), l’inférence permet d’estimer la probabilité que le moteur tombe en panne ($P(F = \text{panne} | H = \text{élevée})$). Cette probabilité peut être calculée par marginalisation en utilisant les probabilités conditionnelles définies dans le réseau bayésien :

$$P(F = \text{panne} | H = \text{élevée}) = P(F = \text{panne} | T = \text{élevée})P(T = \text{élevée} | H = \text{élevée}) + P(F = \text{panne} | T = \text{normale})P(T = \text{normale} | H = \text{élevée})$$

En supposant les valeurs suivantes :

$$P(F = \text{panne} | T = \text{élevée}) = 0.8, \quad P(T = \text{élevée} | H = \text{élevée}) = 0.7$$

$$P(F = \text{panne} | T = \text{normale}) = 0.1, \quad P(T = \text{normale} | H = \text{élevée}) = 0.3$$

Le calcul donne :

$$P(F = \text{panne} | H = \text{élevée}) = (0.8 \times 0.7) + (0.1 \times 0.3) = 0.56 + 0.03 = 0.59$$

Ainsi, lorsque l’humidité est élevée, la probabilité de panne du moteur est de 59 %. Ce résultat peut être utilisé pour ajuster les décisions de maintenance en anticipant une

intervention préventive.

Lorsque la structure du réseau est complexe et que le calcul exact devient trop coûteux, des méthodes d’inférence approximative sont utilisées. Les méthodes Monte-Carlo Markov Chain (MCMC) et la propagation de croyance sont parmi les plus courantes.

La méthode MCMC repose sur l’échantillonnage probabiliste en utilisant une chaîne de Markov pour générer des observations conformes à la distribution de probabilité recherchée. Cette approche est particulièrement efficace pour traiter des réseaux complexes, bien qu’elle nécessite un nombre suffisant d’itérations pour garantir une estimation fiable.

La propagation de croyance est une méthode déterministe qui propage des messages de probabilité à travers le réseau pour estimer les distributions marginales des variables. Elle est souvent employée dans les réseaux où les dépendances conditionnelles suivent une structure arborescente ou faiblement cyclique, permettant une convergence plus rapide des probabilités.

Le choix entre une approche exacte ou approximative dépend des contraintes de l’application, notamment en termes de temps de calcul et de précision requise. Si une grande exactitude est indispensable et que la structure du réseau le permet, l’inférence exacte est privilégiée. En revanche, pour des systèmes complexes où les calculs exacts deviennent inexploitables, les méthodes approximatives offrent un compromis efficace entre rapidité et précision.

Les réseaux bayésiens et leurs méthodes d’inférence sont largement utilisés dans des domaines tels que la médecine, la finance, l’ingénierie et la biologie. Ils permettent de modéliser des relations causales complexes et d’effectuer des prédictions probabilistes pour la prise de décision. Grâce à leur flexibilité, ces modèles s’adaptent à des contextes variés nécessitant l’exploitation d’informations partielles ou incertaines.

3.4 Résultats et analyses

3.4.1 Évaluation des performances

L'évaluation des performances du réseau bayésien repose sur une analyse approfondie des résultats obtenus sur l'ensemble de test. L'objectif est de comparer les prédictions du modèle avec les valeurs réelles en utilisant des métriques standards permettant de quantifier sa fiabilité. Parmi celles-ci, la précision, le rappel, le score F1 et la spécificité sont calculés afin de caractériser différents aspects de la performance du modèle.

L'évaluation commence par l'inférence sur les données de test, suivie de la comparaison des prédictions avec les classes réelles. Cette comparaison est synthétisée dans une matrice de confusion, qui offre une visualisation détaillée des erreurs commises par le modèle. Pour un problème de classification à C classes, cette matrice de taille $C \times C$ contient à la position (i, j) le nombre d'échantillons de la classe réelle i qui ont été classés comme appartenant à la classe j . Elle permet d'identifier les principales confusions et d'ajuster les décisions en conséquence.

Dans le cas d'une classification binaire entre un état nominal et une défaillance, la matrice de confusion se présente comme suit :

| | Prédiction positive | Prédiction négative |
|-----------------|---------------------|---------------------|
| Classe positive | Vrai Positif (TP) | Faux Négatif (FN) |
| Classe négative | Faux Positif (FP) | Vrai Négatif (TN) |

TABLE 3.4 – Matrice de confusion

À partir de cette matrice, les métriques suivantes sont calculées :

La précision mesure la proportion des prédictions positives qui sont correctes par rapport au nombre total de prédictions positives. Elle est particulièrement utile lorsque les faux positifs ont un coût élevé, par exemple en évitant des interventions de maintenance inutiles :

$$\text{Précision} = \frac{TP}{TP + FP} \quad (3.1)$$

Le rappel évalue la capacité du modèle à identifier les vrais cas de défaillance. Il est essentiel dans les situations où les faux négatifs doivent être minimisés, car une panne non détectée peut entraîner des coûts de maintenance corrective importants :

$$\text{Rappel} = \frac{TP}{TP + FN} \quad (3.2)$$

La spécificité mesure la capacité du modèle à éviter les fausses alertes en quantifiant la proportion de prédictions négatives correctes par rapport au total des vrais négatifs :

$$\text{Spécificité} = \frac{TN}{TN + FP} \quad (3.3)$$

Le score F1 est une mesure combinée de la précision et du rappel, offrant un indicateur équilibré lorsque ces deux métriques doivent être prises en compte conjointement :

$$\text{F1-score} = 2 \cdot \frac{\text{Précision} \cdot \text{Rappel}}{\text{Précision} + \text{Rappel}} \quad (3.4)$$

Dans un problème de classification multi-classes, le score F1 est calculé pour chaque classe, puis une moyenne macro est appliquée pour éviter qu'une classe majoritaire ne domine l'évaluation globale :

$$\text{F1-score}_{\text{macro}} = \frac{1}{C} \sum_{c=1}^C \text{F1-score}_c \quad (3.5)$$

L'interprétation des métriques dépend des objectifs spécifiques du système de maintenance prévisionnelle. Si l'objectif est de minimiser les cas de défaillance non détectés, le rappel est prioritaire, quitte à générer davantage de fausses alertes. À l'inverse, si l'objectif est de limiter les interventions inutiles, la précision est privilégiée pour éviter les fausses alertes. Dans notre étude, la priorité est donnée à la précision afin de limiter les interventions injustifiées et d'optimiser les ressources de maintenance. Toutefois, une attention particulière est également portée au rappel sur la classe critique des défaillances, afin d'éviter de manquer des pannes importantes.

Une autre approche permet d'évaluer la qualité de la structure générée par l'apprentissage du réseau bayésien, lorsque la structure idéale est connue. La distance de Hamming est une méthode couramment utilisée pour quantifier les différences entre la structure générée et une structure de référence optimale. Elle mesure le nombre de modifications nécessaires pour transformer la structure apprise en la structure idéale, en comptabilisant les ajouts, suppressions et inversions d'arcs. Si la distance de Hamming est faible, cela signifie que la structure générée est proche de la structure idéale. Cependant, cette méthode présente certaines limites, notamment son incapacité à prendre en compte les erreurs de classification ou les erreurs de mesure pouvant affecter les données d'entrée.

De plus, dans notre cas d'étude, la structure idéale du réseau n'est pas disponible, ce qui rend cette approche inapplicable pour l'évaluation de notre modèle.

En l'absence d'une structure de référence, l'évaluation repose donc exclusivement sur les performances prédictives du modèle, analysées à travers les métriques de classification. L'étude des résultats obtenus à partir des données de test permettra ainsi de juger la pertinence des relations apprises par le réseau bayésien et d'identifier d'éventuels axes d'amélioration.

3.4.2 Analyse des résultats

L'évaluation des performances du modèle de diagnostic constitue une étape intermédiaire essentielle avant son intégration dans un cadre dynamique. L'objectif est de vérifier que le réseau bayésien est capable de reproduire fidèlement les états de santé définis via la méthode MSLD, en utilisant les mesures disponibles à un instant donné. Il s'agit ici de valider la cohérence entre la structure du réseau, les données d'entrée et la logique d'agrégation appliquée à la construction des étiquettes. Cette validation est indispensable pour garantir que le modèle reflète correctement la définition de l'état de santé, et pour préparer son utilisation dans un réseau bayésien dynamique.

La Figure 3.7 illustre la structure du réseau bayésien appris sur la base de données *Genesis-PMDI*, après transformation. Les nœuds bleus représentent les variables physiques mesurées sur l'équipement, comme la vitesse, la position, le courant ou encore la force. Ces variables sont discrétisées à l'aide de la méthode MSLD, selon quatre niveaux de dégradation allant de l'état optimal à l'état d'alerte.

Les nœuds orange correspondent à des diagnostics intermédiaires, chacun lié à un sous-ensemble réduit de paramètres afin de limiter la complexité du modèle. Ils appliquent une logique d'agrégation locale, basée sur le pire état parmi les mesures qui leur sont connectées. Le nœud vert, en sortie, représente l'état de santé global de l'équipement. Il hérite de la même logique d'agrégation, appliquée cette fois aux diagnostics intermédiaires. Cette structuration reflète ainsi la définition même des états de santé selon MSLD, tout en facilitant l'explicabilité des inférences réalisées.

L'inférence a été réalisée sur deux jeux de données transformés : *AI4I-PMDI* et *Genesis-PMDI*. Le but est de vérifier si le modèle permet de reconstruire correctement l'état de santé, tel qu'il a été défini à partir des mesures.

Les performances obtenues sont parfaites, ce qui est cohérent avec la construction du modèle. Les états intermédiaires et l'état final ont été définis selon une logique identique

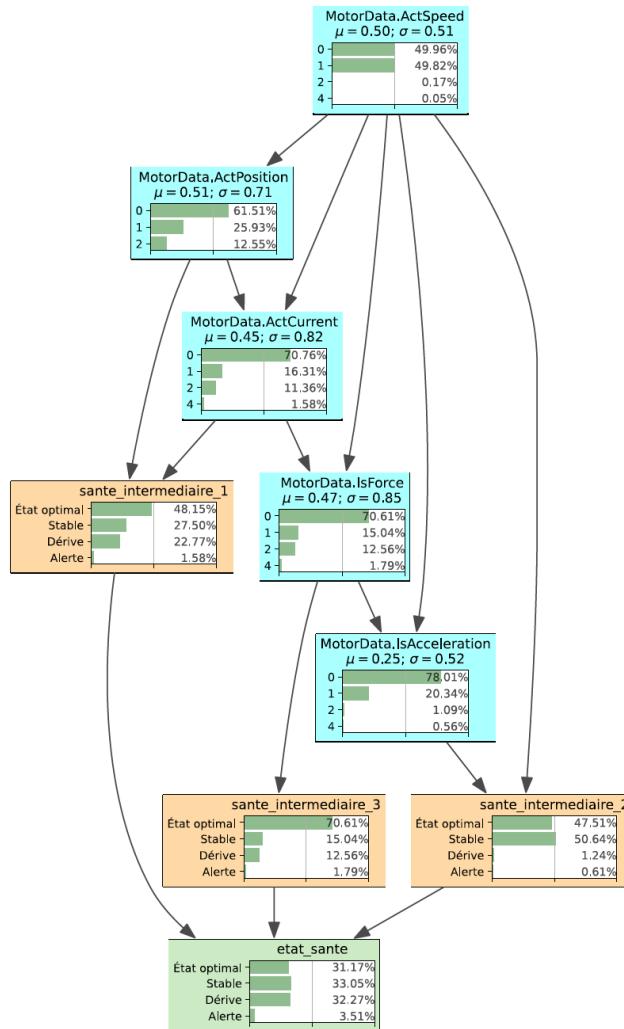


FIGURE 3.7 – Réseau bayésien appris sur la base Genesis-PMDI (sans évidence)

à celle utilisée pour générer les étiquettes, à savoir la prise en compte du pire état mesuré. Ces résultats confirment que le réseau appris est capable de reproduire exactement la structure de décision utilisée lors du prétraitement. Ils valident ainsi l’usage de ce réseau comme composant de diagnostic, en vue de son intégration dans une approche dynamique.

Pour évaluer la robustesse du modèle dans un contexte plus proche de la réalité industrielle, des données manquantes ont été introduites dans les bases transformées. Elles simulent l’absence de certaines mesures, en lien avec le type de contrôle réalisé à chaque instant.

Cette modification impacte directement le diagnostic. Si une mesure critique est absente, le réseau n’a plus l’information nécessaire pour identifier un état de dégradation,

TABLE 3.5 – Performances du diagnostic sur les jeux de données transformés (sans données manquantes)

| Base de données | Précision | Rappel | F1-score macro |
|-----------------|-----------|--------|----------------|
| AI4I-PMDI | 1.00 | 1.00 | 1.00 |
| Genesis-PMDI | 1.00 | 1.00 | 1.00 |

et peut donc sous-estimer l'état global.

TABLE 3.6 – Impact de l'introduction de données manquantes sur les performances du diagnostic

| Base de données | F1-score sans trous | F1-score avec trous | Écart |
|-----------------|---------------------|---------------------|-------|
| AI4I-PMDI | 1.00 | 0.88 | -0.12 |
| Genesis-PMDI | 1.00 | 0.87 | -0.13 |

On observe en moyenne une baisse de 12 % du F1-score macro. Ce résultat est attendu, car l'absence de certaines mesures masque parfois l'information critique nécessaire à un diagnostic précis. Ce comportement met en évidence une limite fondamentale des approches classiques, mais justifie aussi l'intérêt des réseaux bayésiens, qui sont capables de raisonner même en présence d'incertitude ou d'informations incomplètes.

L'objectif principal de cette phase était de valider que le réseau reproduit correctement les états de santé à partir des bonnes observations. Cette vérification étant satisfaisante, le modèle peut désormais être intégré dans une approche dynamique, où les observations futures seront anticipées à l'aide d'un modèle temporel.

Il convient de noter que l'ensemble de cette méthodologie a également été appliquée aux données propriétaires d'ArianeGroup. Cependant, pour des raisons de confidentialité, la nature de ces données et a fortiori les résultats obtenus ne peuvent être ni fournis ni commentés dans ce manuscrit.

Le chapitre suivant introduira les réseaux bayésiens dynamiques, dans lesquels les mesures à l'instant futur seront prédites à partir des états passés, permettant ainsi d'estimer l'état de santé prévisionnel.

Conclusion

Ce chapitre a posé les fondations du diagnostic probabiliste appliqué à des bases de données publiques à l'aide des réseaux bayésiens. Il a permis d'établir une méthodologie

complète allant de la sélection des données à la validation des performances d'un modèle de diagnostic fiable, robuste et interprétable.

La première partie a mis en lumière les limites des bases de données publiques existantes vis-à-vis des contraintes industrielles réelles. Pour y remédier, plusieurs transformations ont été appliquées afin de reproduire des conditions d'acquisition plus représentatives du terrain. Cela inclut l'introduction de données manquantes via des types de contrôles, l'organisation des équipements en parc, et la création de seuils d'alerte permettant de définir des états de santé même en l'absence de pannes explicitement annotées.

Une fois les données structurées, la méthode *MSLD* a été appliquée pour discréteriser les paramètres continus et générer des niveaux d'état de santé reflétant différentes phases de dégradation. Cette étape a été essentielle pour permettre l'apprentissage supervisé d'un réseau bayésien, tout en garantissant la compatibilité entre la logique de détection développée et les classes utilisées pour l'entraînement.

Le modèle de diagnostic a ensuite été construit autour d'un réseau bayésien structuré selon les dépendances apprises à partir des données. Des nœuds de diagnostic intermédiaires ont été introduits afin de maîtriser la complexité computationnelle liée à l'agrégation d'un grand nombre de paramètres. Cette structuration respecte la logique de la méthode *MSLD*, selon laquelle l'état de santé global est défini à partir du paramètre le plus dégradé.

Les résultats obtenus ont confirmé la validité de l'approche : les performances atteignent 100 % lorsque toutes les mesures nécessaires sont présentes. Ces résultats étaient attendus, car la structure du réseau est conçue pour refléter précisément la logique utilisée pour générer les labels. L'objectif de cette phase n'était donc pas de tester la capacité prédictive, mais de s'assurer que le réseau bayésien reproduit correctement le raisonnement de diagnostic à partir de mesures connues.

L'analyse a également montré que les performances se dégradent en présence de données manquantes, ce qui reflète les difficultés d'un diagnostic incomplet lorsque certaines mesures critiques ne sont pas disponibles. Ce constat renforce l'intérêt de considérer l'évolution temporelle des paramètres, car certaines mesures non observées à un instant donné peuvent être estimées à partir de leur dynamique passée.

Ce chapitre a ainsi validé la capacité des réseaux bayésiens statiques à diagnostiquer l'état d'un équipement lorsque les mesures sont disponibles. Ces constats motivent naturellement l'introduction des réseaux bayésiens dynamiques, qui seront au cœur du chapitre suivant. Nous y développerons une méthodologie complète pour intégrer la di-

mension temporelle et prédire l'évolution des états de santé à partir de l'évolution des paramètres mesurés.

MODÉLISATION TEMPORELLE DE LA DÉGRADATION DE SYSTÈMES INDUSTRIELS PAR RÉSEAUX BAYÉSIENS DYNAMIQUES

Introduction

Le chapitre précédent a établi et validé une méthode de diagnostic statique, démontrant qu'un réseau bayésien pouvait estimer de manière fiable l'état de santé d'un système à un instant t . Cette étape a confirmé la pertinence des états de santé générés par la méthode MSLD. Cependant, l'objectif final de la maintenance prévisionnelle est d'anticiper les défaillances en modélisant l'évolution temporelle de ces états. Pour répondre à cet objectif, ce chapitre étend l'approche statique en introduisant les Réseaux Bayésiens Dynamiques (RBD), des modèles capables de traiter des séquences temporelles et de prédire les transitions entre les états de santé successifs.

Ce chapitre est consacré à la conception, à l'application et à l'évaluation de modèles fondés sur les RBD. L'apport méthodologique principal ne réside pas dans une extension théorique des RBD, mais dans le développement d'une démarche complète et reproductible pour leur application à des données industrielles complexes, caractérisées par une acquisition ponctuelle, irrégulière et incomplète. Cette démarche s'articule autour de la mise en forme des données pour transformer des enregistrements épars en séquences temporelles exploitables, de l'adaptation de l'apprentissage par l'intégration de contraintes métier pour pallier la rareté des données, et de la quantification de la robustesse du modèle. Une analyse viendra notamment justifier la capacité du modèle à gérer les données manquantes, un avantage souvent affirmé mais rarement démontré dans ce contexte. Dans cette optique, le chapitre exposera les fondements théoriques des RBD et détaillera les

étapes de cette mise en œuvre.

Cela inclura une étude approfondie de la structuration temporelle des données, de l'apprentissage des transitions, de leur intégration au sein d'un modèle diagnostique global, ainsi que de la justification des choix de conception du modèle, particulièrement concernant la gestion de la complexité temporelle. Des jeux de données transformés, tels que *AI4I-PMDI* et *Genesis-PMDI*, servent de support expérimental pour valider la capacité du RBD à prédire les évolutions d'état de santé sur des horizons courts.

Une section suivante est dédiée à la méthodologie globale déployée tout au long de cette thèse (Section 4.3.1). Elle explicitera la démarche adoptée, depuis la définition de l'état de santé jusqu'à la modélisation dynamique, en soulignant les apports de chaque étape et en offrant une réflexion critique sur l'ensemble du processus.

Dans cette perspective, l'analyse critique de la compatibilité entre les approches dynamiques issues de la recherche académique et les contraintes spécifiques rencontrées dans les environnements industriels réels constituera un point central. À travers une comparaison approfondie entre jeux de données simulés et données de terrain, le chapitre mettra en évidence les limites des cas académiques et les défis posés par la complexité et l'imperfection des données industrielles.

Enfin, ce chapitre proposera une analyse des limites du travail réalisé et des perspectives d'amélioration (Section 4.3.3), ainsi que des recommandations méthodologiques et organisationnelles pour le déploiement de la maintenance prévisionnelle fondée sur les RBD en contexte opérationnel. Ce chapitre examinera également en profondeur le rôle central de l'expert métier, les exigences en matière de qualité de données, la nécessité d'une adaptation continue du modèle, ainsi que les perspectives d'évolution vers des architectures hybrides.

4.1 Construction d'un modèle dynamique à partir des réseaux bayésiens

4.1.1 Du diagnostic à la prédiction : principe des réseaux bayésiens dynamiques

Les RBD constituent une extension naturelle des RB statiques, intégrant l'évolution temporelle des variables dans la modélisation probabiliste. Alors que les RB classiques modélisent les relations de dépendance conditionnelle entre des variables aléatoires à un

instant fixe, les RBD permettent de capturer la dynamique des systèmes en représentant explicitement les liens entre les états successifs. Cette capacité est essentielle dans des contextes tels que la maintenance prévisionnelle, où l'objectif n'est pas uniquement de diagnostiquer l'état actuel du système, mais également d'anticiper son évolution à court ou moyen terme. Les RBD permettent ainsi de prédire les futurs états de santé à partir des observations et inférences précédentes, en tenant compte des processus de dégradation ou de transition observés dans le système.

Un RBD repose sur deux composantes principales : un réseau initial B_0 , qui modélise les dépendances entre variables à l'instant $t = 0$ ou à l'instant initial de la séquence, et un réseau de transition B_T , qui représente l'évolution entre deux instants successifs t et $t + 1$. Ce formalisme conduit à une représentation temporelle appelée *Two-Time-Slice Bayesian Network* (2TBN) (MURPHY, 2002). Dans cette représentation, chaque variable est indexée temporellement et peut être liée à ses homologues passées et présentes.

Hypothèses du modèle dynamique

La construction d'un RBD s'appuie sur plusieurs hypothèses fondamentales permettant de maîtriser la complexité du modèle tout en assurant sa capacité de généralisation.

L'hypothèse principale est celle de Markov d'ordre 1, stipulant que l'état futur d'un système dépend uniquement de son état actuel et non de l'historique complet. Cela se traduit mathématiquement par la propriété suivante :

$$P(X_{t+1}|X_t, X_{t-1}, \dots, X_0) = P(X_{t+1}|X_t)$$

Cette simplification est à la base de nombreux modèles dynamiques, parmi lesquels les chaînes de Markov (FELLER, 1991) et les filtres de Kalman (WELCH ; BISHOP et al., 1995). Elle est souvent une approximation raisonnable pour les systèmes physiques où l'état courant résume l'information pertinente du passé.

Une autre hypothèse forte est la stationnarité, également appelée homogénéité temporelle, des probabilités de transition. Plus formellement, pour un RBD où l'état X_t à chaque instant t est constitué des variables (X_1^t, \dots, X_N^t) , la stationnarité globale de la transition $X_t \rightarrow X_{t+1}$, qui assure la constance des probabilités p_{ij} définies précédemment, découle de l'invariance temporelle des mécanismes locaux. Ainsi, pour chaque variable X_m composant l'état, l'ensemble de ses parents $\text{Pa}(X_m^{k+1})$ est défini de manière identique à chaque instant k , et sa distribution de probabilité conditionnelle $P(X_m^{k+1}|\text{Pa}(X_m^{k+1}))$ est

également la même pour tout instant k . En conséquence de cette invariance des mécanismes locaux de transition, la probabilité de transition globale d'un état complet $X_t = i$ à un état complet $X_{t+1} = j$ ne dépend pas de l'instant t :

$$\forall t \geq 0, \forall i, j \in \mathcal{S}, \quad P(X_{t+1} = j | X_t = i) = p_{ij}$$

où p_{ij} est la probabilité de transition de l'état global i vers l'état global j , constante dans le temps. L'indice temporel discret k utilisé pour les distributions locales est ici analogue à t .

Enfin, une troisième hypothèse utilisée dans ce travail est que la structure intra-temporelle du RBD hérite directement de celle du RB statique défini pour le diagnostic et décrit au Chapitre 3. La partie intra-temporelle du RBD, qui concerne les relations entre variables au sein d'une même tranche de temps (*time-slice*) t ou $t + 1$, est donc identique à celle du modèle statique. Seule la composante temporelle, constituée des arcs inter-tranches reliant les variables de la tranche t à celles de la tranche $t + 1$, est ajoutée pour modéliser les dynamiques de transition.

Fonctionnement général du RBD

Le RBD est formalisé par une paire de graphes bayésiens (B_0, B_T) .

- B_0 représente la distribution de probabilité initiale des variables à l'instant $t = 0$, $P(X_0)$, définie par :

$$P(X_0) = \prod_{i=1}^n P(X_0^i | \text{Pa}(X_0^i))$$

où X_0^i est la variable i à l'instant initial et $\text{Pa}(X_0^i)$ l'ensemble de ses parents dans la structure du RB statique. Ces relations sont intra-tranche.

- B_T représente le modèle de transition, qui spécifie la distribution de probabilité des variables à l'instant $t + 1$ conditionnellement aux variables à l'instant t , $P(X_{t+1} | X_t)$, pour toute transition $t \rightarrow t + 1$:

$$P(X_{t+1} | X_t) = \prod_{i=1}^n P(X_{t+1}^i | \text{Pa}(X_{t+1}^i))$$

où les parents $\text{Pa}(X_{t+1}^i)$ d'une variable X_{t+1}^i peuvent être des variables de la tranche $t + 1$, ce qui correspond à des arcs intra-tranche, et/ou des variables de la tranche t , ce qui correspond à des arcs inter-tranches.

La distribution conjointe sur une séquence temporelle complète de longueur T_{max} , allant de l'instant 0 à T_{max} , est alors donnée par :

$$P(X_0, X_1, \dots, X_{T_{max}}) = P(X_0) \prod_{k=0}^{T_{max}-1} P(X_{k+1} | X_k)$$

Cette factorisation, qui découle de l'hypothèse de Markov d'ordre 1 et de la stationnarité, permet une gestion efficace de l'inférence temporelle.

Représentation 2TBN, déroulement temporel et illustration

La représentation 2TBN est une méthode compacte et modulaire pour décrire le RBD, se concentrant sur la modélisation de la transition entre deux tranches de temps adjacentes t et $t + 1$. Elle permet de définir la structure et les paramètres du modèle de transition B_T . Le 2TBN comprend deux types d'arcs :

- Les arcs intra-tranche, également appelés arcs intra-état, connectant les variables entre elles au sein de la tranche t (pour B_0) ou au sein de la tranche $t + 1$ (pour B_T). Ils reflètent la structure du RB statique.
- Les arcs inter-tranches, également appelés arcs inter-état, connectant des variables de la tranche t à celles de la tranche $t + 1$. Ils décrivent les relations de dépendance temporelle directe.

Pour illustrer concrètement cette structure 2TBN, considérons un système de maintenance prévisionnelle pour un moteur électrique, où l'on suit l'évolution de variables physiques au cours du temps :

- T_k : température du moteur à l'instant k .
- V_k : vitesse de rotation à l'instant k .
- S_k : état de santé de l'équipement à l'instant k .

La structure intra-tranche, qui concerne les relations au sein de chaque instant k , repose sur des relations physiques ou expertes connues. Par exemple, l'état de santé S_k peut dépendre de la température T_k et de la vitesse V_k au même instant : $S_k \leftarrow T_k, V_k$. Les transitions temporelles modélisent l'évolution de chaque variable d'un instant k à $k + 1$. Par exemple, la température T_{k+1} évolue en fonction de son état précédent T_k selon $P(T_{k+1} | T_k)$. De même, la vitesse V_{k+1} suit une transition $P(V_{k+1} | V_k)$. Quant à l'état de santé S_{k+1} , son évolution est modélisée par $P(S_{k+1} | S_k, T_{k+1}, V_{k+1})$, ce qui indique une dépendance à son état antérieur S_k ainsi qu'à la température et à la vitesse courantes à l'instant $k + 1$.

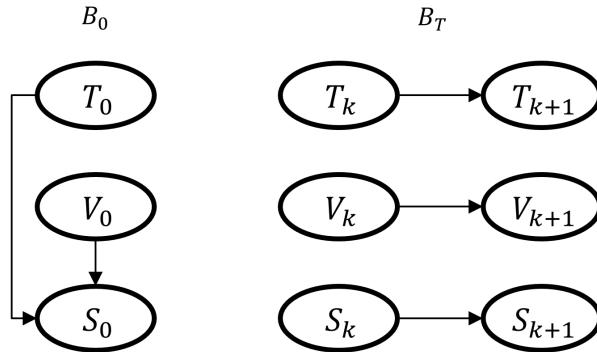


FIGURE 4.1 – Illustration du réseau de l'instant initial et de transition pour l'exemple du moteur électrique.

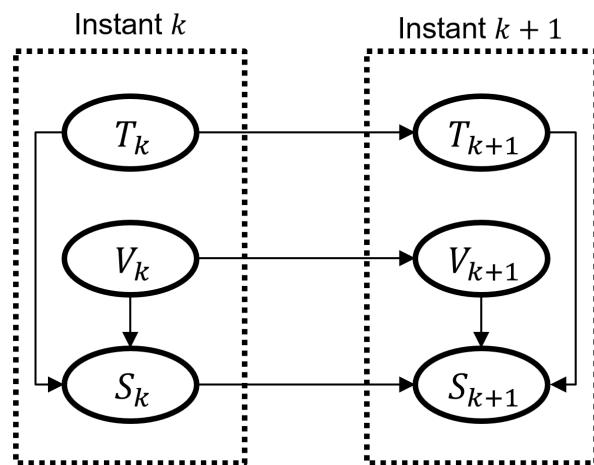


FIGURE 4.2 – Illustration du RBD pour l'exemple du moteur électrique.

Ce type de modèle, comme celui illustré par la Figure 4.2 pour le moteur électrique, permet d'anticiper les dégradations futures de l'équipement à partir de l'état courant et des observations, en tenant compte de l'évolution naturelle des paramètres. Il peut également être utilisé pour effectuer une inférence en ligne, le filtrage, permettant de mettre à jour l'état de santé et les distributions des autres variables au fur et à mesure que de nouvelles observations deviennent disponibles. Cet exemple illustre ainsi comment la modélisation par RBD offre une approche structurée et interprétable pour le pronostic en maintenance prévisionnelle.

Le modèle de transition 2TBN est ensuite utilisé pour représenter l'évolution sur plusieurs pas de temps grâce au principe du déroulement temporel. Le déroulement temporel, ou (*unrolling*), consiste à instancier une séquence de copies du modèle de transition B_T

pour chaque pas de temps, en commençant par la distribution initiale B_0 . La Figure 4.3 illustre ce processus.

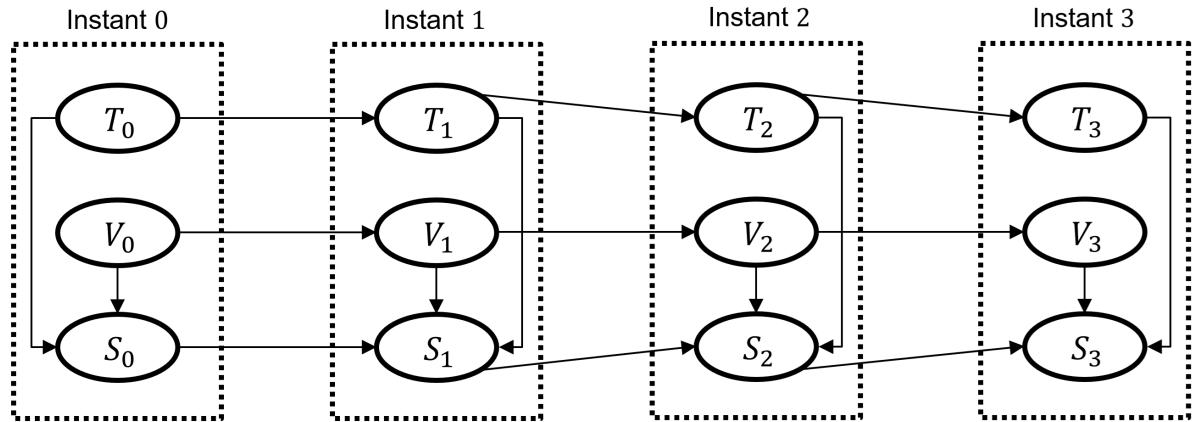


FIGURE 4.3 – Déroulement temporel d'un RBD (variables T, V, S) sur quatre instants ($t = 0$ à $t = 3$).

Chaque tranche temporelle $k + 1$ est générée à partir de la tranche k en utilisant le même modèle de transition B_T . Les relations entre tranches permettent de propager les croyances et d'effectuer des inférences au fil du temps. Ainsi, la distribution conjointe du RBD sur $T_{max} + 1$ instants, de X_0 à $X_{T_{max}}$, est obtenue par :

$$P(X_0, X_1, \dots, X_{T_{max}}) = \left(\prod_{j=1}^n P(X_j^0 \mid \text{Pa}(X_j^0)) \right) \prod_{k=0}^{T_{max}-1} \left(\prod_{i=1}^n P(X_i^{k+1} \mid \text{Pa}(X_i^{k+1})) \right)$$

Cette représentation permet d'implémenter des algorithmes d'inférence standards pour les RBD. Ceux-ci incluent plusieurs opérations :

- Le filtrage (*Filtering*) : consiste en l'estimation des variables non observées de l'état courant X_k , étant donné les observations disponibles jusqu'à l'instant k .
- Le lissage (*Smoothing*) : permet d'estimer les variables non observées d'un état passé X_j , où l'indice j est inférieur à k , à partir de l'ensemble des observations jusqu'à l'instant k .
- La prédiction (*Prediction*) : vise à estimer les variables d'un état futur X_j , où l'indice j est supérieur à k , sur la base des observations disponibles jusqu'à l'instant k .

Ces opérations dépendent du niveau d'observation des variables au sein de chaque état.

4.1.2 Méthodologie de création d'un RBD pour la maintenance prévisionnelle industrielle

La mise en œuvre d'un RBD pour la maintenance prévisionnelle industrielle suit une méthodologie rigoureuse, allant de la structuration temporelle des données à l'intégration des transitions dans un modèle global. Cette section décrit l'ensemble des étapes nécessaires à la construction d'un RBD applicable en contexte industriel, en s'appuyant sur les bases de données transformées comme indiqué dans les chapitres précédents et les concepts introduits ici. Cette méthodologie se veut une contribution pour les industriels souhaitant développer une approche de maintenance prévisionnelle qui tire parti des atouts des RBD, notamment leur capacité à modéliser l'évolution temporelle des équipements et à quantifier les incertitudes de manière interprétable. L'application de RBD aux bases de données publiques est détaillée dans la section suivante (Section 4.2).

Préparation des données temporelles

Un RBD repose sur des données séquentielles, structurées autour d'une progression dans le temps. Il est donc nécessaire de réorganiser les données initiales, disponibles sous forme de relevés individuels, en séries temporelles pour chaque équipement. Les jeux de données transformés, tels que *AI4I-PMDI* et *Genesis-PMDI* qui sont détaillés au Chapitre 3, contiennent pour chaque ligne un identifiant machine et un horodatage ou un ordre implicite qui permet de reconstituer la chronologie. Cela permet de regrouper les observations par équipement, puis de les trier selon l'ordre chronologique. Chaque série est ainsi une séquence temporelle décrivant l'évolution des mesures et de l'état de santé d'un système particulier.

Cette étape vise à établir une continuité entre les observations, en conservant uniquement les séquences présentant une progression temporelle cohérente et une longueur suffisante pour l'apprentissage. Le filtrage des séries s'opère donc selon des critères stricts, définis en amont : une longueur de séquence minimale est requise pour permettre un apprentissage significatif, et un seuil sur l'intervalle de temps maximal entre deux mesures consécutives est fixé pour écarter les séquences présentant des interruptions jugées trop longues. Toute série ne respectant pas ces seuils est automatiquement écartée afin d'éviter de biaiser la modélisation dynamique.

Génération des paires temporelles (X_t, X_{t+1})

Une fois les séries temporelles reconstituées pour chaque équipement, le passage à la modélisation dynamique se fait par la création de paires d'observations successives. Chaque paire associe une configuration de l'ensemble des variables du système, comprenant les paramètres mesurés et l'état de santé, à un instant t et la configuration observée à l'instant suivant $t + 1$ pour ce même équipement.

Il est ici fondamental de préciser que chaque instant, t ou $t + 1$, correspond à l'ensemble des mesures relevées lors d'une unique visite de maintenance. Par conséquent, une paire (X_t, X_{t+1}) ne représente pas l'évolution entre deux mesures rapprochées au sein de la même intervention, mais bien la transition du système dans l'intervalle de temps, potentiellement long et irrégulier, qui sépare deux opérations de maintenance consécutives.

Ces paires (X_t, X_{t+1}) sont générées par une fenêtre glissante le long de chaque série temporelle. Si une série pour un équipement a N observations, elle donnera lieu à $N - 1$ paires de transition. Pour un vecteur de variables à l'instant t noté $X_t = (X_1^t, X_2^t, \dots, X_n^t)$, et à l'instant $t + 1$ noté $X_{t+1} = (X_1^{t+1}, X_2^{t+1}, \dots, X_n^{t+1})$, chaque exemple d'apprentissage pour le modèle de transition est la concaténation de ces deux vecteurs. Cette structure est essentielle car elle permet au modèle d'apprendre comment l'état global du système à $t + 1$ dépend de l'état global à t . Elle permet de capturer l'évolution individuelle de chaque paramètre ainsi que les interactions temporelles entre les différentes variables du système. Les paires générées sont ensuite validées en s'assurant que l'intervalle de temps Δt entre deux observations successives t et $t + 1$ reste dans une plage considérée comme raisonnable et pertinente pour capturer une transition significative, afin d'assurer la pertinence des dynamiques modélisées.

Apprentissage des transitions temporelles

Le cœur de la modélisation dynamique réside dans l'apprentissage des dépendances temporelles entre les variables d'un instant à l'autre, c'est-à-dire l'apprentissage du modèle de transition B_T . Il s'agit d'estimer les distributions de probabilités conditionnelles (DPC) $P(X_i^{t+1} | \text{Pa}(X_i^{t+1}))$ pour chaque variable X_i^{t+1} du système à l'instant $t + 1$. Les parents $\text{Pa}(X_i^{t+1})$ d'une variable X_i^{t+1} peuvent être d'autres variables X_j^{t+1} de la même tranche de temps, ce qui constitue des dépendances intra-tranche, ou des variables X_k^t de la tranche de temps précédente, ce qui correspond à des dépendances inter-tranches.

L'apprentissage se fait typiquement en deux étapes :

1. Apprentissage de la structure : La structure du graphe inter-tranche, définissant les arcs reliant les nœuds de la tranche t à ceux de la tranche $t + 1$, est apprise afin d'identifier les dépendances temporelles les plus significatives. Des algorithmes d'apprentissage de structure, par exemple ceux basés sur des scores comme BDeu (BUNTINE, 1991) ou AIC (AKAIKE, 1974), ou sur des tests d'indépendance conditionnelle, sont appliqués au jeu de données de paires (X_t, X_{t+1}) .

Afin de garantir que le modèle appris soit non seulement statistiquement performant mais aussi physiquement cohérent, l'apprentissage est guidé par un ensemble de règles issues de la connaissance métier. Ces contraintes, établies en collaboration avec les experts industriels, permettent d'injecter une connaissance a priori dans le modèle :

- Interdire les arcs allant de $t + 1$ vers t , afin de respecter le principe de causalité temporelle.
- Limiter le nombre de parents pour chaque nœud, pour maîtriser la complexité du modèle.
- Forcer ou interdire certaines connexions spécifiques, en se basant sur la connaissance physique du système. Par exemple, un expert peut imposer qu'une variable X_i^{t+1} dépende au minimum de son état précédent X_i^t .

La structure intra-tranche à $t + 1$ est généralement supposée identique à celle de B_0 , le RB statique du Chapitre 3, conformément à l'hypothèse de stationnarité de la structure intra-tranche.

2. Apprentissage des paramètres : Une fois la structure du 2TBN définie, incluant la structure intra-tranche à $t + 1$ et les structures inter-tranches, les paramètres de transition, c'est-à-dire les DPC pour chaque nœud de la tranche $t + 1$, sont estimés à partir des données de paires (X_t, X_{t+1}) . Pour les variables discrètes, ce qui est le cas ici après application de la méthode MSLD, cela revient souvent à compter les occurrences des différentes configurations parent-enfant dans les données pour estimer les probabilités.

Le choix de la méthode d'estimation dépend alors du volume de données et de la nécessité d'intégrer des connaissances a priori. L'approche par maximum de vraisemblance (MLE) est simple et efficace lorsque les données sont abondantes, car elle estime les probabilités directement à partir des fréquences observées. Cependant, elle est sensible aux données rares et peut attribuer une probabilité nulle à une configuration jamais vue. À l'inverse, l'estimation bayésienne, en utilisant

un a priori de Dirichlet, est plus robuste avec des données limitées. Elle évite le problème des probabilités nulles en introduisant une "pseudo-observation" initiale et permet d'intégrer formellement une connaissance experte.

La discréétisation préalable par MSLD réduit la taille des espaces de probabilité, ce qui facilite l'estimation même avec un volume de données modéré et aide à prévenir le problème des DPC avec des effectifs nuls.

L'ensemble des relations, structure et paramètres, ainsi obtenus pour les transitions de t à $t + 1$ forme le bloc de transition B_T du RBD, représentant un modèle probabiliste des évolutions d'un pas de temps à l'autre.

Intégration au modèle statique de diagnostic

L'étape finale consiste à assembler les éléments dynamiques, à savoir le modèle de transition B_T , avec le modèle statique initial B_0 , qui a été validé au Chapitre 3. Le modèle B_0 décrit, à chaque instant, la relation entre les mesures observées et l'état de santé via un RB classique. Ce dernier est structuré autour de nœuds intermédiaires, qui sont des paramètres physiques discréétisés, et d'un nœud final représentant l'état de santé global de l'équipement.

Dans le RBD, la structure intra-tranche de B_0 est utilisée pour définir les dépendances au sein de la première tranche de temps. Pour les tranches suivantes, telles que $t + 1, t + 2, \dots$, la structure intra-tranche est également héritée de ce RB statique, et les dépendances inter-tranches sont définies par B_T . Ainsi, chaque tranche temporelle est un RB complet, dont les variables sont reliées aux tranches précédentes par les dépendances temporelles apprises. Les variables mesurées à $t + 1$ sont donc prédites, ou plus précisément leur distribution de probabilité est estimée, à partir de celles à t en utilisant B_T . L'état de santé à $t + 1$ est ensuite inféré à partir des distributions des mesures à $t + 1$, en utilisant la logique de diagnostic du RB statique. Cette logique est définie dans les relations intra-tranche à $t + 1$.

Le modèle global prend ainsi la forme d'un RBD, spécifié par la paire (B_0, B_T) , et peut être déroulé dans le temps pour l'inférence. Ce réseau permet, à partir d'un état initial observé et de nouvelles observations au fil du temps, de mettre à jour les croyances sur l'état du système, ce qui correspond au filtrage, et de simuler l'évolution future des paramètres et de l'état de santé d'un équipement, ce qui correspond à la prédition. Il constitue ainsi un outil potentiel performant pour le pronostic, en combinant une logique de diagnostic fiable avec une capacité à anticiper les dégradations à venir.

La construction du RBD ainsi réalisée repose sur un enchaînement cohérent d'étapes, depuis la structuration temporelle des données jusqu'à l'intégration dans un modèle séquentiel. Elle permet de transposer les capacités d'explication et de diagnostic des RB à un cadre temporel, indispensable pour répondre aux exigences de la maintenance prévisionnelle.

4.2 Application des RBD à des jeux de données industriels

4.2.1 Construction de jeux de données temporels à partir des données transformées

Alignement et gestion des séquences temporelles

L'élaboration d'un modèle dynamique requiert des données organisées en séquences temporelles, qui reflètent l'évolution d'un système d'un instant à l'autre. Or, les bases de données industrielles brutes, ainsi que les versions transformées *AI4I-PMDI* et *Genesis-PMDI* utilisées dans cette étude, se présentent initialement comme des collections d'enregistrements. Bien que ces enregistrements possèdent un identifiant d'équipement et un horodatage, ils ne constituent pas directement les paires $(t, t + 1)$ nécessaires au modèle.

Pour leur conférer une dimension temporelle exploitable par un RBD, une première transformation est appliquée :

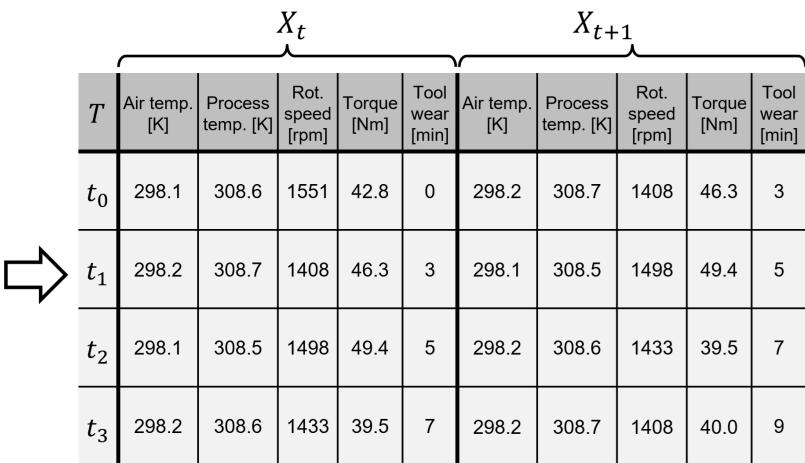
- Les données sont d'abord groupées par identifiant d'équipement.
- Ensuite, pour chaque équipement, les enregistrements correspondants sont triés chronologiquement d'après l'horodatage fourni. Pour *AI4I-PMDI*, cet horodatage avait été introduit afin de simuler des acquisitions irrégulières. Pour *Genesis-PMDI*, qui provient d'un système réel en conditions expérimentales, l'ordre d'acquisition est présumé chronologique.
- Cette suite ordonnée d'observations pour un même équipement constitue son historique complet.

L'hypothèse sous-jacente est que cet ordre reflète la séquence effective des états du système. Une attention particulière est accordée à la cohérence de ces historiques. Les séquences présentant des sauts temporels jugés excessifs ou des anomalies manifestes dans l'ordre des données sont ainsi écartées ou traitées avec précaution. Une fois cet alignement temporel réalisé pour chaque équipement, il devient alors possible de construire les paires d'états consécutifs.

Constitution des échantillons pour l'analyse temporelle

À partir des séries temporelles alignées pour chaque équipement, un nouveau jeu de données est élaboré, destiné spécifiquement à l'apprentissage du modèle de transition du

RBD, le 2TBN. Chaque ligne, ou échantillon, de ce nouveau jeu de données représente une transition observée entre deux instants successifs t et $t + 1$ pour un équipement donné. Concrètement, pour une série de mesures d'un équipement comportant N observations, $N - 1$ paires de transition sont générées. Chaque paire est formée par le vecteur des variables à l'instant t et celui des variables à l'instant $t + 1$. Ainsi, chaque variable du modèle initial, telles que la température, la pression, ou l'état de santé issu de la méthode MSLD, est dupliquée : une version pour l'instant t et une autre pour l'instant $t + 1$. Si X_j est la j -ième variable du système, un échantillon de transition, noté s par exemple, se présentera sous la forme : $(x_{1,s}^t, x_{2,s}^t, \dots, x_{n,s}^t, x_{1,s}^{t+1}, x_{2,s}^{t+1}, \dots, x_{n,s}^{t+1})$.



| T | X _t | | | | | | X _{t+1} | | | | | |
|-------|----------------|-------------------|------------------|-------------|-----------------|--|------------------|-------------------|------------------|-------------|-----------------|--|
| | Air temp. [K] | Process temp. [K] | Rot. speed [rpm] | Torque [Nm] | Tool wear [min] | | Air temp. [K] | Process temp. [K] | Rot. speed [rpm] | Torque [Nm] | Tool wear [min] | |
| t_0 | 298.1 | 308.6 | 1551 | 42.8 | 0 | | 298.2 | 308.7 | 1408 | 46.3 | 3 | |
| t_1 | 298.2 | 308.7 | 1408 | 46.3 | 3 | | 298.1 | 308.5 | 1498 | 49.4 | 5 | |
| t_2 | 298.1 | 308.5 | 1498 | 49.4 | 5 | | 298.2 | 308.6 | 1433 | 39.5 | 7 | |
| t_3 | 298.2 | 308.6 | 1433 | 39.5 | 7 | | 298.2 | 308.7 | 1408 | 40.0 | 9 | |

FIGURE 4.4 – Exemple de création d'une base de données temporelle avec un instant t et $t + 1$.

Une fois ces paires (X_t, X_{t+1}) construites, un prétraitement final est appliqué au besoin. Notamment, toutes les variables continues, c'est-à-dire celles qui n'auraient pas été discrétisées au préalable pour le modèle B_0 , sont converties en utilisant les mêmes intervalles et la même logique, par exemple avec les seuils issus de MSLD, pour les colonnes représentant les instants t et $t + 1$. Cette homogénéisation est importante pour que le RBD opère avec des variables catégorielles cohérentes à travers les tranches de temps et puisse ainsi réutiliser la logique de diagnostic du RB statique.

Au terme de cette procédure, deux jeux de données spécifiquement structurés pour l'apprentissage des transitions dynamiques sont obtenus. Pour *AI4I-PMDI*, 9980 transitions ont été générées à partir des 10000 observations initiales, après regroupement des séquences par machine. Pour *Genesis-PMDI*, 16219 transitions ont été construites de manière similaire. Ces jeux de données constituent la matière première pour l'apprentissage

et l'évaluation des modèles RBD abordés dans les sections suivantes.

4.2.2 Résultats expérimentaux et analyse des performances

Métriques d'évaluation et protocole

L'évaluation des réseaux bayésiens dynamiques pour la prédiction de l'état de santé à l'instant $t + 1$ s'appuie sur un ensemble de métriques classiques en classification. La précision, ou *accuracy*, mesure la proportion totale de prédictions correctes. Le rappel global, ou *recall*, quantifie la capacité à identifier correctement toutes les instances de chaque classe, y compris les minoritaires. Enfin, le score F1 macro, qui est la moyenne harmonique des deux précédentes, offre une mesure équilibrée de la performance. Une attention particulière est portée aux classes critiques, *Alerte* et *Panne*, pour évaluer la sensibilité du modèle à ces événements rares.

Le protocole expérimental commence par la séparation des jeux de données de transitions (X_t, X_{t+1}) en un ensemble d'apprentissage et un ensemble de test, avec une répartition respective de 70% et 30%. Pour garantir la représentativité des classes, une séparation stratifiée est appliquée à la base *AI4I-PMDI*, tandis qu'un découpage aléatoire simple est utilisé pour *Genesis-PMDI*. Le modèle RBD est ensuite entraîné sur les données d'apprentissage : sa structure intra-tranche hérite du RB statique validé au Chapitre 3, tandis que les dépendances inter-tranches et les paramètres de transition sont appris. Finalement, le modèle est évalué sur sa capacité à prédire l'état de santé futur S_{t+1} à partir des informations disponibles, ou évidences, à l'instant t , notées X_t , et potentiellement à $t + 1$, notées $X_{t+1} \setminus S_{t+1}$.

Prédiction des paramètres et états de santé à l'horizon $t + 1$

Les performances des modèles RBD pour la prédiction à l'horizon $t + 1$ sont résumées dans le Tableau 4.1.

TABLE 4.1 – Résultats du RBD à l'horizon $t + 1$ sur les deux bases transformées.

| Base de données | Accuracy | Rappel global | F1-score macro | Précision panne | Rappel panne | Précision alerte | Rappel alerte |
|-----------------|----------|---------------|----------------|-----------------|--------------|------------------|---------------|
| AI4I-PMDI | 0,71 | 0,56 | 0,55 | 0,00 | 0,00 | 0,60 | 0,71 |
| Genesis-PMDI | 0,83 | 0,70 | 0,71 | 0,00 | 0,00 | 0,42 | 0,34 |

Les résultats sur *Genesis-PMDI* révèlent des performances encourageantes, avec une exactitude de 83% et un F1-score de 71%, indiquant une bonne aptitude à modéliser les

transitions majoritaires. La performance sur *AI4I-PMDI* est plus modeste, affichant une exactitude de 71% et un F1-score de 55%. Cette différence est attribuable à la complexité accrue de ce jeu de données, qui simule des conditions industrielles plus réalistes.

Concernant la détection des classes critiques, le modèle s'avère incapable d'identifier la classe *Panne*, pour laquelle la précision et le rappel sont nuls dans les deux cas. Cette lacune est une conséquence directe de la faible fréquence des transitions vers un état de panne dans les données d'apprentissage. Elle s'explique également par le fait que le passage en panne ne dépend pas toujours directement de l'état précédent : une défaillance peut survenir brutalement, sans trajectoire clairement identifiable par le modèle. En revanche, le modèle montre une capacité notable à détecter les états d'*Alerte* sur *AI4I-PMDI*, avec un rappel de 71% et une précision de 60%, un résultat positif pour l'anticipation des dégradations.

Analyse de la robustesse aux données manquantes

L'un des avantages affirmés des réseaux bayésiens est leur capacité à gérer les données manquantes sans nécessiter d'imputation préalable. Cette propriété a été un facteur clé dans le choix de cette approche pour notre cas d'étude, où les données sont structurellement incomplètes. Théoriquement, le RBD gère une observation manquante en utilisant les lois de probabilités pour marginaliser la variable inconnue. Autrement dit, il considère tous les états possibles de la variable manquante, pondérés par leur probabilité conditionnelle calculée à partir des variables observées. La prédiction finale est donc une moyenne de toutes les prédictions possibles, ce qui la rend robuste à une absence partielle d'information.

Les performances obtenues sur la base *AI4I-PMDI*, qui contient des données manquantes introduites de manière contrôlée, constituent une démonstration indirecte de cette robustesse. Le modèle atteint une exactitude de 71% et un rappel de 71% sur la classe *Alerte*, des résultats qui n'auraient pas été possibles si le modèle avait été incapable de raisonner avec des informations incomplètes. La prédiction est donc correcte dans le sens où elle est probabilistiquement juste. Cependant, sa précision et sa certitude dépendent des informations disponibles. Si une variable manquante est un indicateur clé de l'évolution de l'état de santé, son absence augmentera l'incertitude de la prédiction, ce qui se traduira par une distribution de probabilité plus étalée sur les états futurs possibles. Le modèle reste fonctionnel, mais sa capacité à trancher de manière définitive est naturellement réduite.

Analyse qualitative et inférences temporelles

Au-delà des métriques quantitatives, le RBD offre la possibilité d'effectuer des analyses qualitatives sur l'évolution des équipements. Pour illustrer concrètement cette capacité d'inférence, la Figure 4.5 présente le résultat d'une prédiction sur le jeu de données *AI4I-PMDI*.

Dans cet exemple, un état initial, correspondant à la tranche de gauche à l'instant t , est fourni comme évidence : toutes les variables physiques sont observées, et la variable d'état de santé **etat_sante** est diagnostiquée en 'Instable', soit l'état 2. Le RBD propage cette information pour inférer les distributions de probabilité à l'instant $t + 1$. On observe que la variable d'état de santé futur **etat_sante_next** reste le plus probablement en 'Instable' avec une probabilité de 64%, mais le modèle quantifie également un risque de 7,9% de transition vers un état 'Alerte' et de 1,8% de passer en 'Panne'. Cette analyse fine, impossible avec de nombreux autres modèles, permet à un expert de prendre une décision éclairée en pesant la probabilité de chaque scénario.

Cette capacité à prédire un pas de temps peut être itérée pour simuler des trajectoires d'évolution. Par exemple, en initialisant un équipement dans un état d'Alerte, le modèle prédit que cet état a une forte probabilité de près de 60% de se maintenir à l'instant suivant. Ces prédictions de trajectoires illustrent la nature du comportement dynamique appris et peuvent révéler si le modèle a bien capturé des tendances réalistes, telles que la persistance des états dégradés.

Il faut cependant noter que l'incertitude des prédictions tend à augmenter avec l'allongement de l'horizon, car les erreurs d'estimation se propagent et s'accumulent à chaque pas de temps. Bien que l'évaluation se concentre ici sur l'horizon $t + 1$, cette capacité à dérouler le modèle est fondamentale pour des applications de pronostic à plus long terme, comme l'estimation de la Durée de Vie Utile Restante, ou DVUR.

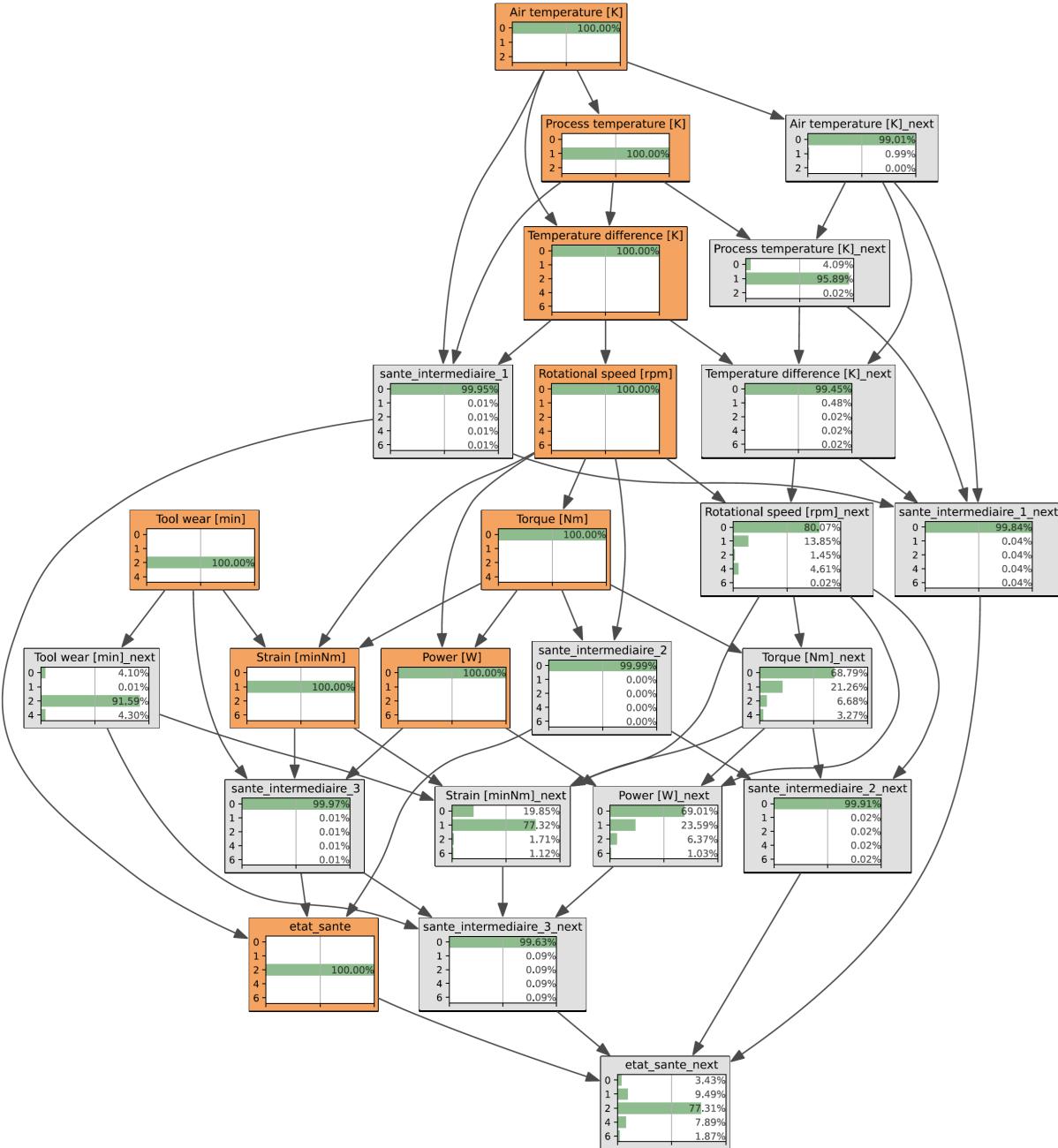


FIGURE 4.5 – Exemple d’inférence dans le RBD sur le jeu de données AI4I-PMDI. L’observation d’un état de ‘Instable’ (état 2) à l’instant t (variables à gauche) est utilisée comme évidence pour estimer la distribution de probabilité de chaque variable à l’instant $t+1$ (variables à droite, identifiées par le suffixe $_next$).

4.3 Discussion approfondie, analyse méthodologique et perspectives

Cette section propose une réflexion élargie sur les travaux présentés dans ce chapitre et, plus globalement, dans l'ensemble de cette thèse. Elle débute par une analyse rétrospective de la méthodologie générale adoptée, suivie d'une discussion sur l'adéquation des modèles RBD aux contextes industriels réels, en soulignant les divergences éventuelles avec les approches académiques plus traditionnelles. Enfin, elle identifie les limitations spécifiques du présent travail de recherche et esquisse des pistes pour des travaux futurs ainsi que des recommandations pour une mise en œuvre industrielle efficace.

4.3.1 Analyse rétrospective de la méthodologie de thèse

La méthodologie s'est articulée autour de plusieurs phases interdépendantes, chacune apportant une contribution spécifique à l'objectif final.

Étape 1 : Étude du contexte et positionnement de la maintenance prévisionnelle (Chapitre 1). Cette phase initiale a permis de poser les fondations théoriques et pratiques de l'étude. Elle a consisté en une revue approfondie de la littérature sur la maintenance industrielle, ses différentes stratégies (corrective, préventive, conditionnelle, prévisionnelle), et les enjeux économiques et opérationnels qui y sont associés. L'objectif était de comprendre l'évolution des pratiques, d'identifier les verrous scientifiques et technologiques persistants, et de justifier la pertinence d'explorer des approches fondées sur les modèles graphiques probabilistes, en particulier les réseaux bayésiens, pour répondre aux défis de la maintenance prévisionnelle. Le cas d'étude spécifique, avec ses contraintes propres (telles que la complexité des systèmes, l'acquisition irrégulière des données et la nécessité d'interprétabilité), a également été introduit et caractérisé, orientant ainsi les choix méthodologiques ultérieurs. La contribution principale de cette étape a été de définir clairement le périmètre de la recherche et les problématiques spécifiques à étudier.

Étape 2 : Élaboration d'un Indicateur d'État de Santé Interprétable et Opérationnel (Chapitre 2). Face à la réalité des données industrielles, où les états de santé des équipements ne sont pas toujours explicitement définis ou le sont de manière trop binaire (par exemple, sain ou défaillant), un effort méthodologique conséquent a été consacré à la création d'un indicateur de santé plus fin et plus signifiant. La méthode MSLD a été développée à cet effet, elle transforme les variables continues issues des capteurs en

un ensemble d'états discrets, en se fondant sur l'analyse statistique de la distribution de chaque paramètre et sur l'intégration de seuils de défaillance connus ou estimés. L'apport majeur de MSLD est de fournir une granularité accrue dans l'évaluation de la condition de l'équipement, facilitant ainsi la détection précoce des dégradations et offrant des niveaux d'alerte intermédiaires. Cet indicateur de santé multi-niveaux, sémantiquement riche, est devenu la variable cible pour les modèles de diagnostic et de pronostic développés par la suite.

Étape 3 : Développement d'un Modèle de Diagnostic Statique par Réseau Bayésien (Chapitre 3). Une fois l'état de santé défini, l'étape suivante a consisté à construire un modèle capable de diagnostiquer cet état à un instant t donné, en se basant sur les observations des paramètres physiques. Les RB statiques ont été choisis pour cette tâche car ils permettent de modéliser l'incertitude de manière probabiliste, de gérer nativement les données manquantes, d'intégrer des connaissances expertes dans la structure du modèle et de fournir des diagnostics interprétables grâce à la visualisation des relations entre variables. Un travail important de prétraitement et de transformation des données a été réalisé, notamment avec la création du jeu de données *AI4I-PMDI* à partir d'AI4I, pour simuler des conditions industrielles plus现实. La démarche a compris la justification du choix des RB face à d'autres algorithmes d'apprentissage automatique, l'apprentissage de la structure et des paramètres du RB statique, et une évaluation rigoureuse de ses performances en termes de classification des états de santé. Cette étape a permis de valider la pertinence des RB pour le diagnostic dans un contexte de données imparfaites.

Étape 4 : Extension à la Modélisation Temporelle par Réseau Bayésien Dynamique (Chapitre 4). Le présent chapitre a marqué le passage du diagnostic statique au pronostic à court terme, en étendant l'approche par RB aux RBD. L'objectif était de prédire l'évolution de l'état de santé d'un instant t à un instant $t+1$. Cela a nécessité la transformation des données en séquences temporelles et la construction d'un modèle 2-TBN. La structure intra-tranche du RBD a hérité de celle du RB statique, tandis que les dépendances inter-tranches, qui constituent le modèle de transition, ont été apprises à partir des paires d'observations successives. L'apport de cette étape réside dans la capacité à prendre en compte la dynamique du système de dégradation et à fournir une prédition probabiliste de l'état de santé futur, même avec des données bruitées ou incomplètes. La démarche a inclus la formalisation de la construction du RBD, son apprentissage sur les jeux de données *AI4I-PMDI* et *Genesis-PMDI*, et l'évaluation de ses performances prédictives à l'horizon $t+1$.

Il est important de souligner que cette démarche structurée a été validée au-delà du cadre académique. Elle a en effet été appliquée de manière identique sur le cas d'étude industriel réel de l'entreprise, confirmant sa pertinence opérationnelle, bien que les données et résultats spécifiques ne puissent être présentés ici pour des raisons de confidentialité.

Néanmoins, cette démarche globale, bien que cohérente, comporte des limites inhérentes qui méritent d'être soulignées. Celles-ci tiennent notamment aux hypothèses simplificatrices nécessaires à la modélisation, comme la stationnarité des processus ou la gestion de la temporalité, mais aussi à la forte dépendance de l'approche à la qualité des données initiales et à l'expertise métier. Ces différents points, qui ouvrent la voie à des perspectives d'amélioration, seront analysés en détail plus loin dans ce chapitre.

4.3.2 Adéquation des modèles RBD aux conditions industrielles : limites et enjeux

Les méthodes de maintenance prévisionnelle fondées sur l'apprentissage automatique, y compris les RBD, sont fréquemment développées et évaluées à partir d'études de cas académiques ou de jeux de données publics standardisés. Si ces approches constituent un socle important pour les développements théoriques et la comparaison objective des algorithmes, leur transposition à des environnements industriels réels soulève des défis considérables, liés aux divergences fondamentales entre ces deux contextes. Cette section analyse ces divergences et leurs conséquences pratiques pour la mise en œuvre de systèmes de diagnostic et de pronostic efficaces.

Limites des jeux de données académiques et simulés

La majorité des bases de données académiques couramment utilisées pour le développement et l'évaluation des modèles prédictifs sont issues soit de simulations, à l'instar de la base de données CMAPSS pour les turboréacteurs, soit d'expérimentations menées dans des conditions de laboratoire fortement contrôlées, comme les bancs d'essai de roulements PRONOSTIA.

Ces jeux de données académiques présentent plusieurs avantages indéniables pour la recherche. Ils offrent un contrôle précis des scénarios de dégradation, une annotation souvent claire et fiable des événements de défaillance ou des indicateurs de santé, une certaine homogénéité dans la qualité et la structure des données. De plus, ils facilitent la reproductibilité des expériences ainsi que la comparaison des performances entre différentes

approches algorithmiques. Toutefois, ces caractéristiques appréciables résultent aussi de simplifications importantes par rapport à la complexité du monde réel. Les dynamiques de dégradation y sont fréquemment modélisées de manière linéaire, monotone ou discrète, sans toujours prendre en compte les effets de synergie, les défaillances multiples, ou les interactions complexes entre sous-systèmes. Or, ce sont précisément ces phénomènes complexes qui constituent les difficultés majeures de la maintenance prévisionnelle en conditions réelles.

Les conditions opérationnelles telles que la charge, la vitesse, ou l'environnement sont souvent maintenues constantes ou varient selon des profils prédéfinis, et les sources d'aléas ou de bruit dans les mesures sont limitées ou modélisées de façon simplifiée. De plus, les capteurs y sont généralement supposés parfaitement fiables, les acquisitions de données régulières et complètes, et les défaillances distinctement marquées dans le temps, parfois même avec une indication précise du RUL. Si un tel niveau de suivi est envisageable sur un banc d'essai ou une ligne de production fixe, il contraste fortement avec le contexte de suivi d'un parc d'équipements, où les données sont par nature collectées de façon plus sporadique et incomplète.

Cette idéalisation, bien qu'utile pour isoler et étudier des aspects spécifiques des algorithmes, induit une vision souvent trop optimiste de la performance atteignable en maintenance prévisionnelle. Elle permet de valider des concepts et de comparer des modèles dans un cadre normalisé, mais ne garantit nullement que les modèles ainsi développés seront directement transposables, performants, ou même applicables dans un environnement de production industriel. Le passage du laboratoire à l'usine, ou du simulateur au système réel, exige de confronter les modèles à un contexte beaucoup plus incertain, caractérisé par des données incomplètes et hétérogènes, des comportements non linéaires et non stationnaires.

Dans le cas spécifique des réseaux bayésiens dynamiques, cette simplification des jeux de données académiques se manifeste notamment par une représentativité parfois limitée des transitions temporelles. Dans des données simulées ou de laboratoire, les séquences de dégradation peuvent apparaître lisses et progressives, et les changements d'état de santé suivre une évolution claire et prévisible. Or, dans un système industriel réel, cette linéarité est rarement la norme. Les transitions peuvent être abruptes, influencées par des facteurs externes non mesurés, ou interrompues et modifiées par des mécanismes de régulation, des actions de maintenance corrective non systématiquement enregistrées, ou des changements de configuration du système. C'est précisément face à ces dynamiques

complexes qu'un modèle probabiliste comme un RBD démontre sa valeur en agissant comme un outil d'aide à la décision et en fournissant à l'expert les éléments pour arbitrer entre le surcoût d'une maintenance préventive et le risque financier ou sécuritaire d'une panne non anticipée.

Défis inhérents aux données industrielles réelles

Lorsqu'on se tourne vers les données issues directement d'environnements industriels en exploitation, les caractéristiques rencontrées diffèrent radicalement de celles des jeux de données académiques.

Une première contrainte majeure est liée à la qualité et à la disponibilité des données. Celles-ci sont souvent collectées à des fréquences irrégulières. Cette irrégularité s'explique par des contraintes opérationnelles fortes. Alors que des systèmes fixes dans des usines récentes peuvent bénéficier d'un monitoring quasi-continu, le suivi d'un parc d'équipements mobiles ou dispersés, comme dans notre cas d'étude, impose des collectes de données intermittentes. L'acquisition est alors fortement dépendante de la disponibilité des systèmes pour la mesure, par exemple lors du retour d'un véhicule au dépôt. Ces fréquences sont alors typiquement dictées par des impératifs opérationnels, tels que la disponibilité des équipements ou les cycles de production, ou par des plans de maintenance, comme les inspections périodiques, plutôt que par les besoins de la modélisation. Les séries temporelles peuvent ainsi être fragmentées, avec des interruptions de collecte, des périodes d'échantillonnage variables, et une hétérogénéité significative entre les machines d'un même parc concernant les capteurs installés et les paramètres suivis. L'hétérogénéité des capteurs et des protocoles d'acquisition est une source fréquente de données manquantes structurelles : certains paramètres ne sont tout simplement jamais mesurés pour certains équipements ou le sont seulement dans des conditions particulières, par exemple lors d'arrêts machine spécifiques pour des contrôles non destructifs. Ces données manquantes sont difficiles à reconstituer sans une connaissance approfondie du système et de son contexte opérationnel, et les techniques d'imputation peuvent introduire des biais importants.

Une seconde difficulté concerne la qualité et la fiabilité des annotations, en particulier pour les états de défaillance. Dans de nombreux contextes industriels, les défaillances ne sont pas systématiquement précédées d'une identification claire dans les données. Les pannes sont souvent repérées *a posteriori*, à la suite d'interventions humaines, ou d'arrêts de production, sans qu'un lien direct et temporellement précis puisse toujours être établi avec les signaux des capteurs enregistrés avant l'événement. La définition même

d'une défaillance peut varier, désignant par exemple une panne franche, une dégradation fonctionnelle ou une perte de performance, et les seuils critiques associés aux paramètres mesurés peuvent être flous ou dépendre du contexte. Cela complique considérablement l'étiquetage des données, pourtant nécessaire à l'apprentissage supervisé des modèles de diagnostic et de pronostic. Face à ce défi, une bonne pratique consiste à mettre en place, en collaboration avec les experts métier, un protocole d'annotation rigoureux qui définit explicitement les différents états de santé et les critères objectifs pour leur identification.

La variabilité inter-équipement et intra-équipement représente également un enjeu de taille. Même au sein d'un parc d'équipements supposés identiques, chaque machine peut présenter des signatures de fonctionnement et des trajectoires de dégradation différentes en raison de son âge, de son historique d'utilisation, de son environnement de travail spécifique, ou des réglages et maintenances antérieures. De plus, un même équipement peut voir son comportement évoluer dans le temps, ce qui correspond à un phénomène de non-stationnarité, en raison de l'usure, de changements de configuration, ou de modifications des conditions opératoires. Pour gérer cette diversité, la méthodologie s'oriente vers des modèles locaux plutôt que globaux. L'approche consiste alors à développer des systèmes personnalisés pour chaque unité, ou à créer des mécanismes d'adaptation qui tiennent compte des régimes de fonctionnement spécifiques.

Enfin, l'échelle temporelle pertinente pour la détection et la prédiction de certaines dégradations peut s'étendre sur des périodes très longues, allant de plusieurs semaines ou mois jusqu'à plusieurs années pour des équipements robustes ou des dégradations lentes telles que la corrosion ou la fatigue. Cela implique de disposer d'historiques de données longs, continus et fiables, ce qui n'est pas toujours réalisable en pratique. Les interruptions dans la collecte, les changements de systèmes d'information, les remplacements de capteurs qui peuvent s'accompagner de calibrages différents, ou les modifications majeures des équipements peuvent altérer la continuité et la cohérence des séquences temporelles et compliquer l'exploitation des données passées pour l'apprentissage de modèles prédictifs à long terme. La mise en place d'une gouvernance des données rigoureuse, incluant la documentation de ces changements et l'application de techniques d'harmonisation, devient alors une étape préliminaire indispensable pour exploiter de tels historiques.

L'ensemble de ces contraintes spécifiques illustre les défis méthodologiques majeurs que posent les données industrielles réelles. Elles soulignent l'importance cruciale de développer des modèles et des approches capables de gérer explicitement l'incertitude, l'incomplétude, l'hétérogénéité et la variabilité, tant au niveau des données d'entrée que des

systèmes eux-mêmes. L'approche par réseau bayésien dynamique, au cœur de cette thèse, se présente comme une réponse prometteuse à une partie de ces problématiques, notamment par sa capacité à modéliser les relations de dépendance et les transitions d'état dans un cadre probabiliste robuste, tout en permettant l'intégration d'informations manquantes ou incertaines et de connaissances expertes. Cependant, comme cela sera discuté ultérieurement, des défis subsistent, notamment en ce qui concerne l'apprentissage à partir de données de défaillance extrêmement rares.

Réponse méthodologique aux contraintes industrielles de l'étude

Le cas d'étude industriel au cœur de cette thèse, introduit dans le premier chapitre, incarne parfaitement les défis décrits précédemment. Il s'agit d'une acquisition de données ponctuelle, irrégulière, et marquée par une incomplétude structurelle ainsi que par la rareté des défaillances avérées. La méthodologie fondée sur les réseaux bayésiens dynamiques a été spécifiquement conçue pour surmonter ces obstacles, qui rendent les approches de maintenance prévisionnelle traditionnelles souvent inopérantes.

Face au problème de l'acquisition ponctuelle et irrégulière, qui interdit un suivi continu des paramètres, le défi a été surmonté en structurant les données autour des visites de maintenance. Cette approche transforme une chronologie irrégulière en une séquence logique et régulière d'événements. Chaque instant t correspond ainsi à une visite de maintenance complète, et l'instant $t + 1$ à la visite suivante. On recrée de cette manière une forme de régularité séquentielle, même si la durée réelle entre deux visites reste variable et que le détail des dates au sein d'une même intervention est perdu. Le modèle de transition apprend alors à modéliser l'évolution d'un état à un autre entre deux inspections consécutives, en se concentrant sur le changement d'état plutôt que sur sa cinétique fine.

Comme justifié en détail au Chapitre 3, le choix des réseaux bayésiens a été motivé par leur capacité intrinsèque à gérer l'incertitude. Cette propriété s'est avérée importante face à nos données structurellement incomplètes, nous permettant de réaliser des inférences même en l'absence de certaines mesures, sans recourir à des techniques d'imputation potentiellement biaisées. Le réseau bayésien dynamique peut ainsi réaliser des inférences et fournir des prédictions probabilistes même en l'absence de certaines mesures. Cette capacité, fondamentale pour notre cas d'étude, permet d'exploiter l'ensemble des données collectées, aussi fragmentaires soient-elles, sans écarter d'observations précieuses.

Enfin, le défi majeur de la rareté des données de panne a orienté la stratégie de modélisation vers la détection précoce des états de dégradation intermédiaires. La méthode

MSLD a été développée dans ce but précis, c'est-à-dire créer un indicateur de santé à plusieurs niveaux définissant par exemple les états Optimal, Stable, Instable et Alerte. Les résultats expérimentaux montrent que si la prédiction de la classe Panne est difficile, le modèle parvient en revanche à anticiper la transition vers les états d'Alerte. Cette approche déplace l'objectif de la maintenance prévisionnelle de la prédiction de la défaillance finale, souvent inaccessible, vers l'identification des trajectoires de dégradation avancée, offrant ainsi aux experts un levier d'action concret et anticipé.

En conclusion, la démarche méthodologique de cette thèse ne s'est pas limitée à l'application d'un modèle standard. Elle a consisté à concevoir une solution complète, depuis la création d'un indicateur de santé granulaire via la méthode MSLD jusqu'à la modélisation de son évolution temporelle par un réseau bayésien dynamique. Ce dernier s'est ainsi révélé être un cadre de modélisation particulièrement pertinent, non pas malgré les imperfections des données, mais précisément grâce à sa capacité à les intégrer de manière cohérente et probabiliste.

4.3.3 Limitations spécifiques du travail de thèse et recommandations

Au-delà des défis généraux liés à l'application des modèles en contexte industriel, le travail de recherche mené dans cette thèse présente également des limites qui lui sont propres, lesquelles ouvrent la voie à des améliorations et extensions futures. Cette section détaille ces limitations intrinsèques et propose des recommandations pour une mise en œuvre industrielle plus robuste et efficace des RBD.

Analyse des limites intrinsèques aux modèles développés et à l'approche actuelle

Plusieurs aspects des modèles et de la méthodologie employés dans cette thèse méritent une analyse critique quant à leurs limitations.

- 1. Gestion de la Temporalité Irrégulière et des Δt Variables :** Le modèle RBD de type 2-TBN utilisé suppose que la nature des dépendances $P(X_{t+1}|X_t)$ ne change pas avec la durée exacte Δt séparant t et $t + 1$. En d'autres termes, une transition observée sur un intervalle de quelques heures est traitée de la même manière qu'une transition s'étalant sur plusieurs jours lors de l'apprentissage des DPC du modèle de transition. En réalité, la probabilité qu'un système transite d'un

état de santé à un autre peut être fortement influencée par le temps écoulé. Cette simplification peut conduire à une sous-estimation du risque si le Δt d'une nouvelle observation est court alors que la dégradation est rapide, ou à une surestimation si le Δt est long mais que le système demeure stable.

2. **Sensibilité à la Rareté des Données de Défaillance Avérée :** Les résultats expérimentaux, présentés à la Section 4.2.2, ont clairement montré la difficulté, voire l'incapacité, des modèles RBD appris à prédire la classe Panne. Ceci est une conséquence directe du très faible nombre d'exemples de transitions menant effectivement à cet état critique dans les jeux de données disponibles, situation fréquente dans les environnements industriels où la maintenance préventive est déjà en place pour éviter les pannes. Bien que la détection des états d'alerte précurseurs soit une contribution précieuse, la capacité limitée à anticiper la panne imminente reste une contrainte majeure pour un système de maintenance prévisionnelle complet. Une piste de recherche prometteuse pour pallier ce manque consiste à recourir à des jumeaux numériques. Ces modèles de simulation physique pourraient être utilisés pour générer des scénarios de dégradation jusqu'à la panne complète, enrichissant ainsi les jeux de données d'apprentissage sur ces événements critiques.
3. **Hypothèse de Stationnarité des Processus :** Le formalisme standard des RBD, et celui employé ici, repose sur l'hypothèse de stationnarité des probabilités de transition, le modèle B_T étant le même pour toutes les transitions $t \rightarrow t + 1$, et de la structure intra-tranche. Or, les systèmes industriels et leurs processus de dégradation peuvent évoluer avec le temps en raison de l'usure cumulative, de changements dans les conditions d'exploitation, de modifications des protocoles de maintenance, ou de remplacements de composants. Un modèle appris sur des données historiques peut ainsi perdre progressivement de sa pertinence et de sa précision si la dynamique réelle du système dévie par rapport à ce qui a été appris, illustrant le phénomène de dérive. Pour contrer ce phénomène, une approche adaptative est nécessaire, par exemple via un réapprentissage périodique du modèle sur les données récentes ou via des mécanismes détectant les changements de régime pour déclencher une mise à jour.
4. **Dépendance à la Qualité et aux Paramètres de la Discréétisation (MSLD) :** La performance globale du système de diagnostic (Chapitre 3) et de pronostic (Chapitre 4) est intrinsèquement liée à la pertinence des états de santé définis par la méthode MSLD, décrite au Chapitre 2. Bien que MSLD vise une discréétisation

adaptative et sémantiquement riche, le choix du nombre d'états de transition, la manière dont les seuils statistiques fondés sur μ et σ sont combinés avec les seuils de défaillance fournis par l'expert, et la méthode d'agrégation pour obtenir l'état global du système peuvent influencer la granularité et la sensibilité de la détection des dégradations. Une discréétisation sous-optimale pourrait masquer des dynamiques fines, créer des classes d'états difficiles à distinguer pour le modèle, ou rendre les transitions moins prévisibles.

5. Complexité en termes de Calcul pour les Systèmes à Très Grande Échelle :

L'apprentissage de la structure et des paramètres des réseaux bayésiens, et plus encore des RBD, ainsi que l'inférence exacte, peuvent exiger des ressources de calcul importantes pour des systèmes impliquant un grand nombre de variables interdépendantes ou nécessitant le suivi de très longues séquences temporelles. Bien que les techniques de discréétisation aident à gérer la complexité des DPC, le passage à l'échelle pour des applications industrielles de très grande envergure, telle qu'une usine entière avec des milliers de capteurs, reste un défi. Ce défi peut cependant être géré en pratique : le nombre de variables par modèle peut être ajusté pour respecter un budget de temps de calcul, et le suivi d'un parc se prête naturellement à une parallélisation massive, chaque équipement pouvant être traité indépendamment.

6. Validation en Conditions Opérationnelles Réelles et Intégration des Actions de Maintenance :

L'évaluation des modèles a été principalement réalisée sur des jeux de données académiques, bien que *AI4I-PMDI* ait été modifié pour simuler des aspects réels, et sur des données expérimentales *Genesis-PMDI* issues d'un environnement contrôlé. Une validation à grande échelle en environnement de production réel, impliquant une boucle de rétroaction avec les équipes de maintenance et une évaluation de l'impact effectif des prédictions sur les décisions et les coûts, n'a pas été menée dans le cadre de cette thèse. De plus, le modèle actuel ne prend pas explicitement en compte l'effet des actions de maintenance, qu'elles soient correctives, préventives planifiées, ou qu'elles résultent des alertes du système prédictif, sur l'état de santé et les dynamiques de transition futures. Une action de maintenance peut significativement altérer la trajectoire de dégradation, un aspect non modélisé ici. Pour des travaux futurs, une perspective d'amélioration majeure consisterait à intégrer ces interventions comme des variables explicites dans le RBD, dont l'activation modifierait dynamiquement les probabilités de transition afin de simuler l'effet d'une réparation ou d'un remplacement sur l'état de

santé du système.

Préconisations pour une mise en œuvre industrielle robuste

Le déploiement réussi d'un système de maintenance prévisionnelle fondé sur les RBD en contexte industriel requiert une approche méthodique et une attention particulière à plusieurs aspects clés, qui vont au-delà de la simple élaboration du modèle.

Avant toute chose, il est fondamental de s'assurer de la qualité et de la pertinence du modèle de diagnostic statique (RB) servant de fondation à chaque tranche temporelle du RBD. Ce réseau initial doit fidèlement refléter la structure causale, ou du moins les dépendances pertinentes, entre les paramètres mesurés et l'état de santé du système, tel que défini par exemple via la méthode MSLD. Sans un diagnostic fiable à l'instant t , toute tentative de prédiction de l'état à $t + 1$ sera compromise. Cela implique une validation rigoureuse de ce RB statique, si possible avec le concours d'experts du domaine.

La collecte et la préparation des données constituent des étapes critiques et souvent sous-estimées. Il est impératif de mettre en place des procédures de surveillance et d'acquisition de données qui soient non seulement rigoureuses, notamment en ce qui concerne les capteurs calibrés, l'horodatage précis et l'intégrité des données, mais aussi adaptées à la temporalité des phénomènes de dégradation étudiés. Les variables doivent être mesurées à des fréquences compatibles avec la vitesse d'évolution des dégradations que l'on souhaite anticiper. Il est également essentiel d'historiser toutes les interventions de maintenance, changements de configuration ou événements opérationnels inhabituels dans un format numérique structuré et facilement exploitable. Ces informations contextuelles sont en effet essentielles pour interpréter correctement les séquences de données et pour comprendre les ruptures potentielles dans les trajectoires de dégradation.

Lors de la phase de préparation des données pour le RBD, l'identification des séquences temporelles complètes et cohérentes par équipement, ainsi que la mise en correspondance des observations pour former les paires $(t, t + 1)$, doivent être effectuées avec une grande rigueur. La gestion des données manquantes doit être explicite : soit elles sont traitées par le formalisme probabiliste du RBD, lequel peut marginaliser les variables inconnues, soit des stratégies d'imputation prudentes et justifiées doivent être envisagées si leur proportion est trop importante.

Pendant la phase d'apprentissage du RBD, il est fortement recommandé d'introduire des contraintes sur la structure temporelle, c'est-à-dire les arcs inter-tranches, si certaines dépendances, ou absences de dépendance, sont connues a priori par les experts du système.

Cela permet de guider l'apprentissage, d'éviter la détection de relations statistiquement significatives mais causalement absurdes, et de réduire la complexité de l'espace de recherche des structures. De même, l'évaluation des performances du RBD ne doit pas se limiter aux scores globaux, tels que la précision ou le score F1. Une analyse qualitative et détaillée des cas de transitions mal anticipées, des fausses alertes, et des périodes où les prédictions du modèle sont les plus incertaines, ce qui peut être indiqué par une variance élevée dans les distributions de probabilité prédictives, est indispensable pour comprendre les faiblesses du modèle et identifier les leviers d'amélioration.

Sur le plan opérationnel, la mise en œuvre d'un RBD nécessite des outils de visualisation et d'interprétation adaptés aux utilisateurs finaux, comme les opérateurs de maintenance ou les experts des systèmes. Il est essentiel que ces utilisateurs puissent non seulement consulter les prédictions de l'état de santé futur, mais aussi comprendre les facteurs qui y contribuent, par exemple en identifiant quelles variables à t influencent le plus l'état à $t + 1$, et visualiser les probabilités associées à chaque état prédict. Un tableau de bord interactif, intégrant les mesures brutes, les états de santé diagnostiqués et prédicts, leur évolution temporelle, et potentiellement des indicateurs de confiance, favorise une meilleure appropriation de l'outil et une prise de décision plus éclairée.

Enfin, un aspect important pour la pérennité du système est la mise à jour périodique et l'adaptation continue du modèle. Les systèmes industriels et leurs modes de défaillance évoluent. Un modèle RBD, même initialement performant, peut voir sa précision décliner avec le temps. Il est donc indispensable d'anticiper, dès la phase de conception, une stratégie de réapprentissage ou de mise à jour du modèle. Cela peut prendre la forme de réapprentissages complets à intervalles réguliers, de mises à jour incrémentales des paramètres sur des fenêtres temporelles glissantes, ou de mécanismes plus sophistiqués de détection de dérive, qui déclenchent une adaptation lorsque les performances du modèle se dégradent ou que la distribution des données d'entrée change significativement.

En synthèse, l'analyse menée permet de formuler plusieurs recommandations opérationnelles claires pour le déploiement de cette méthodologie. L'approche par RBD s'est avérée pertinente pour un pronostic à court terme, correspondant à la prochaine visite de maintenance, mais son usage pour des horizons plus lointains doit être mené avec prudence en raison de l'accumulation de l'incertitude. Le modèle démontre une bonne tolérance aux données incomplètes, mais sa performance dépend des variables manquantes. Il est donc crucial d'assurer la disponibilité des capteurs les plus influents, identifiés lors de l'apprentissage de la structure, pour garantir une prédition fiable des états critiques.

Concernant le volume de données, bien qu'il soit difficile de définir un seuil universel, nos expériences suggèrent qu'un historique de plusieurs dizaines de transitions par équipement est un minimum requis. Enfin, les résultats montrent une difficulté à prédire les pannes franches mais une bonne capacité à anticiper les états dégradés. L'objectif opérationnel du modèle doit donc se concentrer sur la détection précoce des transitions vers des états comme l'état *Alerte*, qui fournit un signal d'action fiable pour la maintenance.

Collaboration essentielle avec l'expert métier

Dans tout projet de maintenance prévisionnelle, l'expert métier, qu'il soit ingénieur de maintenance, technicien spécialisé ou opérateur expérimenté, joue un rôle central et irremplaçable, ce qui est particulièrement manifeste pour le développement et le déploiement de systèmes fondés sur les réseaux bayésiens dynamiques. Loin d'être un simple utilisateur passif du modèle prédictif, il en est un acteur clé et un partenaire à part entière, impliqué à toutes les étapes du processus, de la conception initiale à l'interprétation des résultats et à la prise de décision opérationnelle.

Lors de la phase de construction du modèle, l'expert métier apporte une connaissance précieuse et souvent tacite sur le fonctionnement du système, ses composants critiques, les modes de défaillance connus, et les relations de cause à effet probables entre les paramètres mesurés et les signes de dégradation. Cette expertise est inestimable. Elle permet de valider la pertinence des variables sélectionnées pour la surveillance, de définir ou de confirmer les seuils critiques utilisés dans la méthode MSLD, et de valider la sémantique des états de santé intermédiaires comme Stable, Instable et Vulnérable. De plus, l'expert peut guider l'apprentissage de la structure du réseau bayésien en imposant ou en interdisant certaines liaisons entre variables afin d'assurer la cohérence du modèle. Par exemple, il saura qu'une augmentation de la charge sur un moteur peut influencer sa température, mais que l'influence inverse est moins directe. Cette collaboration permet d'ancrer le modèle dans la réalité physique et opérationnelle du système étudié, augmentant ainsi sa robustesse et sa crédibilité.

Dans la phase d'interprétation des résultats et de prise de décision, l'expert joue un rôle de filtre critique et de contextualisation. Les prédictions issues d'un RBD sont, par nature, probabilistes, et fournissent une distribution de probabilité sur les états de santé futurs possibles. L'expert est donc le plus à même de contextualiser ces prédictions en les reliant à l'historique de l'équipement, aux conditions opératoires et aux maintenances passées. Il peut ainsi juger de la pertinence et de l'urgence d'une alerte en tenant compte des consé-

quences potentielles d'une défaillance, que ce soit en termes de sécurité, de production ou de coûts. Finalement, c'est à lui que revient la responsabilité de prendre les décisions de maintenance adaptées : simple surveillance renforcée, planification d'une inspection approfondie, déclenchement d'une intervention préventive ou, dans les cas critiques, arrêt immédiat de l'équipement.

L'interface homme-machine (IHM) du système de maintenance prévisionnelle doit être conçue pour faciliter ce rôle : elle doit offrir une lecture claire et intuitive des résultats, comme les probabilités, les tendances ou les variables influentes, permettre à l'expert d'explorer l'information en détail, et potentiellement d'intégrer ses propres observations ou annotations. Enfin, le retour d'expérience de l'expert est le moteur d'une boucle d'amélioration continue, instaurant un véritable cercle vertueux. Cette approche, souvent qualifiée d'"expert-in-the-loop" et illustrée par la Figure 4.6, permet d'affiner le modèle avec le temps. L'expert valide ou contredit les prédictions en situation réelle et peut signaler des comportements nouveaux, des dérives de capteurs ou des cas où le modèle s'est trompé. Ces observations qualitatives et quantitatives sont alors collectées pour ajuster périodiquement les paramètres du modèle, comme les seuils de discréétisation ou les probabilités de transition. Ce processus itératif permet de trouver progressivement les paramètres optimaux et d'adapter le réseau aux évolutions du système, maintenant ainsi sa performance au fil du temps.

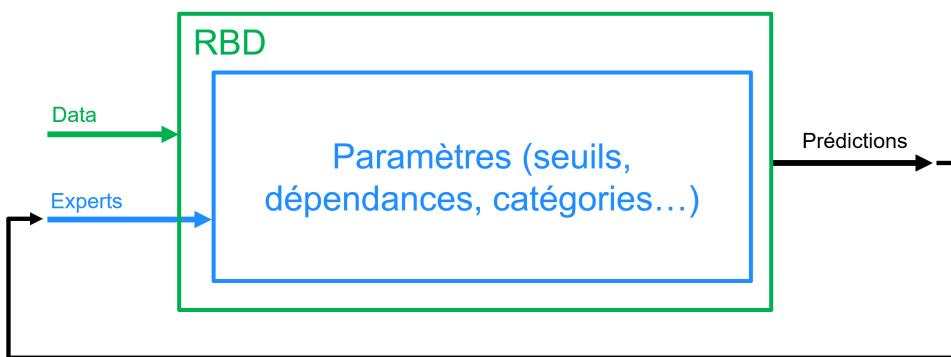


FIGURE 4.6 – Schéma du cercle vertueux de l'expert-in-the-loop. L'expert contribue à la définition des paramètres initiaux et utilise son retour sur les prédictions pour affiner le modèle de manière continue.

En somme, l'expert métier agit en tant que partenaire essentiel au système adaptatif. Son rôle dans la co-construction, la validation et l'évolution du modèle est déterminant pour le succès et la durabilité du système de maintenance prévisionnelle.

4.3.4 Perspectives d'évolution et pistes de recherche futures

Le travail présenté dans cette thèse, bien qu'apportant des contributions à la modélisation de la dégradation par RBD, ouvre la voie à de nombreuses perspectives d'amélioration et d'extension. Ces pistes visent à surmonter certaines des limitations identifiées et à enrichir les capacités des modèles pour des applications industrielles encore plus performantes et robustes.

Vers des modèles RBD plus adaptatifs, aux dynamiques temporelles plus sophistiquées et robustes aux déséquilibres

Une première direction de recherche concerne l'amélioration de la capacité des RBD à gérer la non-stationnarité et l'évolution des systèmes industriels. Comme discuté précédemment, l'hypothèse de stationnarité des paramètres du modèle de transition B_T peut être mise à l'épreuve dans des environnements dynamiques. Des mécanismes d'apprentissage adaptatif ou incrémental pourraient être intégrés pour permettre au RBD d'ajuster ses paramètres, voire sa structure, en continu ou périodiquement à mesure que de nouvelles données deviennent disponibles. Des techniques de détection de dérive pourraient signaler quand le modèle n'est plus en adéquation avec la réalité du système, déclenchant un double processus : d'une part, un réapprentissage automatique pour ajuster les paramètres, et d'autre part, une sollicitation des experts pour interpréter la cause de cette dérive et valider les adaptations du modèle. Cela pourrait impliquer l'utilisation de fenêtres temporelles glissantes pour l'estimation des probabilités, ou des approches bayésiennes séquentielles qui mettent à jour les distributions des paramètres de manière récursive.

Une autre piste importante est une meilleure prise en compte de la temporalité irrégulière des observations. Le modèle 2-TBN standard ne modélise pas explicitement la durée Δt entre les tranches de temps t et $t + 1$. Des extensions, telles que les réseaux bayésiens en temps continu (*Continuous Time Bayesian Networks*, CTBN) (FORRESTER et al., 2019 ; NODELMAN ; SHELTON et KOLLER, 2012) qui modélisent les taux de transition entre états, ou l'intégration de Δt comme une variable d'influence dans les DPC du modèle de transition, pourraient offrir des prédictions plus précises lorsque les intervalles d'acquisition sont très variables. Des modèles semi-Markoviens, où la durée passée dans un état influence la probabilité de transition, pourraient également être explorés.

La gestion de la rareté des données de défaillance avérée demeure un défi majeur. Des

techniques avancées de traitement des classes déséquilibrées pourraient être appliquées spécifiquement à l'apprentissage des transitions vers les états de panne. Cela inclut des méthodes de ré-échantillonnage, comme la méthode SMOTE pour générer des exemples synthétiques de la classe minoritaire ou le sous-échantillonnage de la classe majoritaire, ainsi que l'utilisation de fonctions de coût asymétriques lors de l'apprentissage qui pénalisent davantage les erreurs sur la classe Panne, ou encore des approches de détection d'anomalies où les états de panne sont considérés comme des événements rares et anormaux par rapport au fonctionnement nominal.

Enfin, si l'hypothèse de Markov d'ordre 1 s'avère trop restrictive pour certains systèmes complexes dont l'avenir dépend d'un historique plus long, des recherches pourraient porter sur des modèles RBD d'ordre supérieur, où X_{t+1} dépendrait par exemple de X_t et X_{t-1} . Cependant, cela doit être fait avec prudence pour éviter une explosion de la complexité, notamment en utilisant des techniques de sélection de caractéristiques pour identifier les éléments les plus pertinents de l'historique, ou des structures factorisées pour représenter les dépendances d'ordre supérieur.

Intégration de modèles hybrides et d'approches complémentaires

Bien que les réseaux bayésiens dynamiques offrent une modélisation causale et probabiliste robuste, ainsi qu'une bonne interprétabilité, ils peuvent être avantageusement combinés avec d'autres approches d'apprentissage automatique, notamment les techniques d'apprentissage profond, également connu sous le terme anglais *Deep Learning*, pour créer des modèles hybrides tirant parti des forces de chaque paradigme.

Les réseaux de neurones profonds, tels que les réseaux de neurones récurrents (RNN, LSTM, GRU) ou les transformers, excellent dans la capture de dépendances temporelles complexes. Par exemple, les LSTM utilisent des mécanismes de portes (*gates*) pour mémoriser ou oublier sélectivement des informations sur de longues séquences, ce qui leur permet de modéliser des dynamiques qui échappent aux modèles plus simples. Une piste prometteuse consiste à utiliser ces modèles pour :

- L'extraction de caractéristiques (ou *feature extraction* en anglais) : Un réseau de neurones pourrait prétraiter les données brutes pour en extraire des représentations latentes ou des indicateurs de santé synthétiques plus informatifs, qui seraient ensuite utilisés comme variables d'entrée discrétisées dans le RBD.
- La modélisation des transitions : Les probabilités de transition $P(X_i^{t+1}|\text{Pa}(X_i^{t+1}))$ au sein du RBD pourraient elles-mêmes être estimées par un modèle neuronal,

permettant de capturer des relations plus complexes que ne le permettent les tables de probabilités conditionnelles classiques, surtout si certaines variables parentes sont continues.

- La prédiction directe de certaines variables : Des modèles neuronaux pourraient être employés pour prédire l'évolution future de certains paramètres physiques clés. Ces prédictions, accompagnées de leur incertitude estimée, seraient ensuite intégrées comme données d'observation dans le RBD pour inférer l'état de santé global.

Cette hybridation permettrait de bénéficier de la puissance de modélisation des réseaux neuronaux pour les aspects complexes et de type "boîte noire" du système, tout en conservant la structure explicative, la gestion de l'incertitude et la capacité d'intégration de connaissances a priori des réseaux bayésiens pour le raisonnement de plus haut niveau.

Une autre exemple de recommandation serait la modélisation explicite de l'impact des actions de maintenance. Le RBD pourrait être étendu pour inclure des nœuds représentant les différents types d'actions de maintenance possibles, telles qu'une réparation mineure, un remplacement de composant ou une révision majeure. Ces nœuds d'action influencerait alors les probabilités de transition de l'état de santé, permettant de simuler l'effet d'une intervention, par exemple un retour à un état Optimal ou Stable avec une certaine probabilité après une maintenance. Cela rendrait les prédictions plus réalistes en tenant compte du cycle de vie complet de l'équipement, y compris les interventions humaines, et ouvrirait la voie à l'optimisation des politiques de maintenance, notamment pour déterminer quand intervenir et quel type d'intervention choisir.

L'amélioration de la robustesse et de la sensibilité de la méthode de discréétisation MSLD est également une perspective. Des études plus approfondies sur l'impact du nombre d'états, la définition des seuils, et la méthode d'agrégation pourraient être menées. Des techniques d'optimisation pourraient être explorées pour ajuster dynamiquement ces paramètres en fonction des caractéristiques des données ou des objectifs de performance du système de maintenance.

Enfin, le développement d'outils d'aide à la décision plus avancés et de protocoles de validation sur le terrain rigoureux est essentiel pour traduire les avancées de la recherche en bénéfices industriels concrets. Cela inclut la conception d'IHM qui non seulement présentent les prédictions du RBD de manière claire et exploitable, mais permettent également aux experts métier d'interagir avec le modèle, de tester des scénarios hypothétiques ou de type "si/alors", et de fournir un retour d'information structuré.

En conclusion, l'adoption de modèles hybrides, l'amélioration de l'adaptabilité temporelle, la prise en compte explicite des interventions, et une validation poussée sur le terrain représentent des voies prometteuses pour renforcer les capacités des systèmes de maintenance prévisionnelle fondés sur les RBD, les rendant ainsi encore plus pertinents et efficaces face aux défis complexes des environnements industriels modernes.

Conclusion

Ce chapitre a introduit, développé et évalué une approche de modélisation dynamique par réseaux bayésiens dans le cadre de la maintenance prévisionnelle. En étendant la structure statique d'un réseau bayésien à une représentation temporelle cohérente, le RBD permet non seulement de diagnostiquer un état présent mais également d'anticiper son évolution, typiquement à l'horizon d'un pas de temps.

Une méthodologie complète a été proposée et justifiée, depuis la définition et la création d'états de santé interprétables grâce à la méthode MSLD, la transformation des jeux de données pour refléter les contraintes industrielles, et la construction d'un modèle de diagnostic statique robuste, jusqu'à l'élaboration d'un modèle 2-TBN capable de capturer les transitions d'état.

Les résultats expérimentaux, obtenus sur les jeux de données simulés *AI4I-PMDI* et expérimentaux *Genesis-PMDI*, démontrent la pertinence de l'approche, notamment pour la détection précoce des situations d'alerte, bien que la prédiction des pannes rares demeure un défi majeur.

La discussion approfondie a mis en lumière la démarche méthodologique globale de la thèse, ses apports en termes d'interprétabilité et de gestion de l'incertitude, ainsi qu'une réflexion critique sur les limites des approches académiques face aux réalités du terrain et sur les limitations spécifiques du travail présenté. Les défis liés à la rareté des classes critiques, à la gestion de la temporalité irrégulière, à l'hypothèse de stationnarité et aux données manquantes ont été soulignés, ouvrant la voie à des perspectives d'amélioration significatives. Parmi celles-ci, l'apprentissage adaptatif, l'exploration de modèles temporels plus sophistiqués, les approches hybrides, la prise en compte de l'impact des opérations de maintenance, et une implication encore plus étroite des experts métier apparaissent comme des pistes fructueuses.

En définitive, le RBD, tel que développé et appliqué dans cette thèse, se distingue par sa capacité à concilier robustesse face à des données imparfaites, explicabilité des

mécanismes modélisés, et adaptation temporelle. Il constitue une brique essentielle vers un système de pronostic fiable, évolutif et opérationnel, apte à répondre aux contraintes de données imparfaites, de variabilité structurelle, et de justification des décisions critiques dans des environnements industriels complexes.

CONCLUSION GÉNÉRALE

La maintenance prévisionnelle est essentielle pour optimiser les opérations industrielles, mais son déploiement dans des contextes de données complexes reste un défi majeur. Cette thèse a proposé une méthodologie complète pour la PdM, spécifiquement adaptée aux systèmes industriels où l'acquisition de données est irrégulière, partielle, et où les historiques de défaillance sont limités. L'objectif central était de développer une approche robuste et interprétable, de la création d'indicateurs d'état de santé jusqu'au pronostic à court terme, en utilisant les réseaux bayésiens.

Synthèse des Contributions

La réponse à la problématique de la PdM en contexte de données difficiles s'est articulée autour de contributions méthodologiques clés. Initialement, une analyse du contexte industriel a mis en évidence les limites des stratégies de maintenance traditionnelles et les défis spécifiques liés à l'acquisition de données non continues, justifiant le besoin d'approches innovantes.

Pour pallier l'absence d'indicateurs de santé détaillés, la méthode MSLD a été développée. Cette contribution majeure permet de créer, par discréétisation non supervisée, des états de santé HI multi-niveaux et sémantiquement riches comme Optimal, Stable, Prudent, Vulnérable et Défaillance, à partir de données brutes, même acquises irrégulièrement. MSLD améliore la granularité du diagnostic et facilite la détection précoce des dégradations, en définissant l'état global par le principe du maillon faible.

La phase de diagnostic a été ensuite abordée avec des RB statiques. Ce choix a été motivé par leur capacité à gérer l'incertitude et les données manquantes, tout en offrant une structure interprétable. Des bases de données publiques, AI4I et Genesis Demonstrator, ont été transformées en AI4I-PMDI et Genesis-PMDI pour mieux simuler les contraintes industrielles. La construction des RB, incluant des noeuds intermédiaires pour gérer la complexité, a validé leur aptitude à reproduire fidèlement la logique de diagnostic issue de MSLD.

Enfin, l'approche a été étendue au pronostic à court terme via les RBD. À partir

de données séquentielles et de paires d'observations X_t, X_{t+1} , les modèles RBD ont été construits, leur structure intra-tranche héritant des RB statiques et leurs dépendances inter-tranches étant apprises. Les résultats expérimentaux ont confirmé la capacité des RBD à prédire l'évolution de l'état de santé à l'horizon $t + 1$, notamment pour anticiper les états d'alerte, cruciaux pour une maintenance proactive.

Le message innovant central de cette thèse réside dans la proposition d'un cadre méthodologique intégré, MSLD vers RB puis RBD. Ce cadre est spécifiquement conçu et validé pour des environnements où les données sont imparfaites, offrant une voie pragmatique pour valoriser des jeux de données difficiles et améliorer les stratégies de maintenance.

Périmètre et Limites de l'Étude

Cette recherche s'est concentrée sur les systèmes industriels où l'acquisition de données de capteurs est ponctuelle, typiquement lors d'interventions de maintenance, générant des données irrégulières, potentiellement incomplètes, avec peu d'instances de défaillance. L'accent a été mis sur les modèles probabilistes interprétables.

Plusieurs limites sont à reconnaître. La méthode MSLD, bien qu'adaptative, peut nécessiter une expertise métier pour définir certains seuils de défaillance lorsque ceux-ci ne sont pas explicitement connus. Les modèles RBD utilisés s'appuient sur des hypothèses simplificatrices telles que la dépendance Markovienne d'ordre 1 et la stationnarité des transitions. Ces approximations peuvent ne pas toujours refléter la totalité de la complexité des processus de dégradation réels, notamment sur de longues périodes ou face à des régimes variables. Le formalisme 2TBN standard ne modélise pas non plus explicitement l'influence d'un Δt variable entre observations. La prédiction des pannes franches est restée un défi majeur, principalement en raison de leur extrême rareté dans les jeux de données d'apprentissage, limitant l'apprentissage supervisé pour ces événements. Bien que la structuration avec noeuds intermédiaires aide à gérer la complexité des CPT, l'application à des systèmes de très grande échelle pourrait poser des défis computationnels. Enfin, la validation a principalement reposé sur des jeux de données transformés et expérimentaux. Une validation industrielle à grande échelle, avec intégration complète et modélisation de l'impact des actions de maintenance, reste une étape future nécessaire.

Perspectives de Recherche et Recommandations

Les contributions et limites de cette thèse ouvrent plusieurs perspectives de recherche et permettent de formuler des recommandations pour l'application industrielle.

L'amélioration des modèles RBD constitue une première voie. L'intégration de mécanismes d'apprentissage adaptatif pour gérer la non-stationnarité et la dérive des modèles est une piste. Une meilleure prise en compte de l'irrégularité des Δt via des formalismes comme les CTBN ou l'inclusion de Δt comme variable d'influence est souhaitable. Pour la prédiction des pannes rares, des techniques avancées de gestion des classes déséquilibrées ou de détection d'anomalies méritent d'être explorées.

L'exploration de modèles hybrides est une deuxième direction prometteuse. L'apprentissage profond pourrait être utilisé pour l'extraction de caractéristiques à partir de signaux bruts ou pour modéliser des dynamiques complexes, ces éléments alimentant ensuite les RBD pour le raisonnement et la gestion de l'incertitude.

Une troisième perspective importante est l'intégration explicite de l'impact des actions de maintenance dans les RBD. Modéliser comment les interventions affectent les transitions d'état de santé rendrait les prédictions plus réalistes et permettrait d'optimiser les politiques de maintenance.

La méthode MSLD pourrait également être affinée, notamment par des recherches sur l'optimisation du nombre d'états ou l'automatisation partielle de la définition des seuils critiques.

Pour le déploiement industriel, plusieurs recommandations émergent. La qualité du modèle de diagnostic statique initial est primordiale. Des procédures rigoureuses de collecte et de préparation des données, incluant les informations contextuelles et de maintenance, sont indispensables. L'intégration de connaissances expertes dans la définition de la structure des RB et RBD peut grandement améliorer leur pertinence. Le développement d'HMI claires et interactives est crucial pour l'adoption par les utilisateurs et l'amélioration continue des modèles. Enfin, une stratégie de mise à jour et d'adaptation continue des modèles est nécessaire pour garantir leur performance à long terme face à l'évolution des systèmes.

En conclusion, cette thèse a cherché à fournir des bases solides pour la PdM en contexte de données difficiles, en privilégiant une approche probabiliste et interprétable. Les perspectives de recherche et les recommandations suggèrent que des avancées futures permettront de développer des systèmes de PdM encore plus performants et adaptés aux exi-

gences complexes de l'industrie. La collaboration étroite avec les experts métier demeurera un facteur clé de succès pour traduire ces avancées en bénéfices industriels tangibles.

TABLE DES FIGURES

| | | |
|-----|--|-----|
| 1 | Évolution du nombre d'articles scientifiques annuels sur le thème de la maintenance prévisionnelle dans la base de données d'Elsevier. | 2 |
| 1.1 | Comparaison des stratégies de maintenance et de leurs temporalités d'intervention (RAN et al., 2019). | 12 |
| 1.2 | Les étapes clés du processus de maintenance prévisionnelle. | 16 |
| 1.3 | Illustration des contraintes d'acquisition des données dans le cas d'étude : acquisition ponctuelle et irrégulière, données manquantes et rareté des pannes. | 28 |
| 2.1 | Cycle opérationnel de l'état de santé : de la donnée brute à la décision de maintenance. | 34 |
| 2.2 | Processus de répartition des étiquettes par discréétisation. | 45 |
| 2.3 | Processus d'étiquetage par discréétisation dans la méthode MSLD. | 46 |
| 2.4 | Illustration des étapes de discréétisation et d'étiquetage dans MSLD. | 48 |
| 3.1 | Processus de transformation des bases de données vers des versions adaptées aux contraintes industrielles. | 63 |
| 3.2 | Exemples de graphes non-orientés. | 74 |
| 3.3 | Exemples de graphes orientés. | 75 |
| 3.4 | Évolution des algorithmes d'apprentissage de la structure des réseaux bayésiens (KITSON et al., 2023). | 80 |
| 3.5 | Architecture du réseau bayésien de diagnostic avec nœuds intermédiaires pour la réduction de la complexité. | 82 |
| 3.6 | Évolution de la complexité (échelle logarithmique) en fonction du nombre de nœuds intermédiaires. | 84 |
| 3.7 | Réseau bayésien appris sur la base Genesis-PMDI (sans évidence) | 90 |
| 4.1 | Illustration du réseau de l'instant initial et de transition pour l'exemple du moteur électrique. | 100 |
| 4.2 | Illustration du RBD pour l'exemple du moteur électrique. | 100 |

| | | |
|-----|---|-----|
| 4.3 | Déroulement temporel d'un RBD (variables T, V, S) sur quatre instants ($t = 0$ à $t = 3$). | 101 |
| 4.4 | Exemple de création d'une base de données temporelle avec un instant t et $t + 1$ | 108 |
| 4.5 | Exemple d'inférence dans le RBD sur le jeu de données AI4I-PMDI. L'observation d'un état de 'Instable' (état 2) à l'instant t (variables à gauche) est utilisée comme évidence pour estimer la distribution de probabilité de chaque variable à l'instant $t + 1$ (variables à droite, identifiées par le suffixe <code>_next</code>). | 112 |
| 4.6 | Schéma du cercle vertueux de l'expert-in-the-loop. L'expert contribue à la définition des paramètres initiaux et utilise son retour sur les prédictions pour affiner le modèle de manière continue. | 126 |

LISTE DES TABLEAUX

| | | |
|-----|--|-----|
| 1.1 | Comparaison des types de maintenance : avantages et inconvénients | 19 |
| 2.1 | Synthèse des limites des approches existantes pour la création d'états de santé. | 42 |
| 2.2 | Différentes distributions pour le paramètre "Puissance" avec différentes techniques de discréétisation. | 49 |
| 2.3 | Tableau des différentes distributions des états du système en fonction de la technique de discréétisation. | 50 |
| 3.1 | Comparaison des jeux de données selon des critères de taille et de caractéristiques temporelles. | 60 |
| 3.2 | Comparaison des jeux de données selon les critères de défaillances et d'origine des données. | 60 |
| 3.3 | Comparaison entre différentes approches de machine learning pour la maintenance prévisionnelle. | 70 |
| 3.4 | Matrice de confusion | 87 |
| 3.5 | Performances du diagnostic sur les jeux de données transformés (sans données manquantes) | 91 |
| 3.6 | Impact de l'introduction de données manquantes sur les performances du diagnostic | 91 |
| 4.1 | Résultats du RBD à l'horizon $t + 1$ sur les deux bases transformées. | 109 |

BIBLIOGRAPHIE

- AAMODT, Agnar et PLAZA, Enric, « Case-Based Reasoning : Foundational Issues, Methodological Variations, and System Approaches », in : *AI Commun.* 7.1 (1994), 39–59.
- ACHOUCH, Mounia et al., « On Predictive Maintenance in Industry 4.0 : Overview, Models, and Challenges », in : *Applied Sciences* 12.16 (2022).
- ACID, S. et CAMPOS, L. de, « A hybrid methodology for learning belief networks : Benedict », in : *International Journal of Approximate Reasoning* (2001), 235–262.
- AFNOR, *Fascicule de Documentation NF X60-000 - Maintenance industrielle*, Afnor EDITIONS, 2016.
- AGOGINO, A. et GOEBEL, K., *Milling Data Set*, rapp. tech., <https://www.nasa.gov/intelligent-systems-division/discovery-and-systems-health/pcoe/pcoe-data-set-repository/>, NASA Ames Research Center, 2007.
- AGRAHARI, Rupesh et al., « Applications of Bayesian network models in predicting types of hematological malignancies », in : *Scientific Reports* (2018), p. 1-12.
- AGRAWAL, Rakesh et SRIKANT, Ramakrishnan, « Fast Algorithms for Mining Association Rules in Large Databases », in : *Proceedings of the 20th International Conference on Very Large Data Bases*, VLDB '94, San Francisco, CA, USA : Morgan Kaufmann Publishers Inc., 1994, 487–499, ISBN : 1558601538.
- AIVALIOTIS, P. ; GEORGOULIAS, K. ; ARKOULI, Z. et MAKRIS, S., « Methodology for enabling Digital Twin using advanced physics-based modelling in predictive maintenance », in : *Procedia CIRP* (2019).
- AKAIKE, H., « A new look at the statistical model identification », in : *IEEE transactions on automatic* (1974), p. 716-723.
- AKERBP, *Valhal Oil Platform Dataset*, <https://openindustrialdata.com/data/>, Accessed : 2024-06-03, 2022.
- ALFEO, Antonio L. ; CIMINO, Mario G.C.A. et VAGLINI, Gigliola, « Degradation stage classification via interpretable feature learning », in : *Journal of Manufacturing Systems* (2021).
- ALLAH BUKHSH, Zaharah ; SAEED, Aaqib ; STIPANOVIC, Irina et DOREE, Andre G., « Predictive maintenance using tree-based classification techniques : A case of railway

-
- switches », in : *Transportation Research Part C : Emerging Technologies* 101 (2019), p. 35-54.
- AMIHAI, Ido et al., « An Industrial Case Study Using Vibration Data and Machine Learning to Predict Asset Health », in : *2018 IEEE 20th Conference on Business Informatics (CBI)*, t. 01, 2018, p. 178-185.
- AMIN, Md. Tanjin ; KHAN, Faisal et IMTIAZ, Syed, « Fault detection and pathway analysis using a dynamic Bayesian network », in : *Chemical Engineering Science* (2018), p. 777-790.
- AMRUTHNATH, Nagdev et GUPTA, Tarun, « A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance », in : 2018.
- AN, Dawn ; KIM, Nam Ho et CHOI, Joo Ho, « Practical options for selecting data-driven or physics-based prognostics algorithms with reviews », in : *Reliability Engineering & System Safety* (2015).
- ARENA, S. ; FLORIAN, E. ; ZENNARO, I. ; ORRÙ, P.F. et SGARBOSSA, F., « A novel decision support system for managing predictive maintenance strategies based on machine learning approaches », in : *Safety Science* 146 (2022), p. 105529.
- ARIAS CHAO, Manuel ; KULKARNI, Chetan ; GOEBEL, Kai et FINK, Olga, « Fusing physics-based and deep learning models for prognostics », in : *Reliability Engineering & System Safety* 217 (2022), p. 107961.
- ASGARPOUR, Masoud et SØRENSEN, John Dalsgaard, « Bayesian Based Diagnostic Model for Condition Based Maintenance of Offshore Wind Farms », in : *Energies* (2018).
- ASHRAFI, Maryam ; DAVOUDPOUR, Hamid et KHODAKARAMI, Vahid, « Risk assessment of wind turbines : Transition from pure mechanistic paradigm to modern complexity paradigm », in : *Renewable and Sustainable Energy Reviews* (2015), p. 347-355.
- ATAMURADOV, Vepa ; MEDJAHER, Kamal ; DERSIN, Pierre ; LAMOUREUX, Benjamin et ZERHOUNI, Noureddine, « Prognostics and health management for maintenance practitioners-Review, implementation and tools evaluation. », in : *International Journal of Prognostics and Health Management* 8.3 (2017), p. 1-31.
- AUTRAN, Jean-Victor ; KUHN, Véronique ; DIGUET, Jean-Philippe ; DUBOIS, Matthias et BUCHE, Cédric, « AI4I-PMDI : Predictive maintenance datasets with complex industrial settings' irregularities », in : *Knowledge-Based and Intelligent Information & Engineering Systems : Proceedings of the 28th International Conference KES-2024, Seville, Spain, 11-13 September 2023*, t. 246, Procedia Computer Science, Elsevier, 2024a, p. 1201-1209.

-
- « Discretization Strategies for Improved Health State Labeling in Multivariable Predictive Maintenance Systems », in : *Proceedings of the 13th International Conference on Data Science, Technology and Applications, DATA 2024, Dijon, France, July 9-11, 2024*, SCITEPRESS, 2024b, p. 434-441.
- AYVAZ, Serkan et ALPAY, Koray, « Predictive maintenance system for production lines in manufacturing : A machine learning approach using IoT data in real-time », in : *Expert Systems with Applications* 173 (2021), p. 114598.
- BAPTISTA, Márcia et al., « Forecasting fault events for predictive maintenance using data-driven techniques and ARMA modeling », in : *Computers & Industrial Engineering* 115 (2018), p. 41-53.
- BARALDI, P. ; MANGILI, F. et ZIO, E., « A Kalman Filter-Based Ensemble Approach With Application to Turbine Creep Prognostics », in : *IEEE Transactions on Reliability* (2012).
- BERNARDO, José M. et SMITH, Adrian F. M., *Bayesian Theory*, John Wiley et Sons, 1994.
- BIRGELEN, Alexander von et NIGGEMANN, Oliver, « Anomaly Detection and Localization for Cyber-Physical Production Systems with Self-Organizing Maps », in : *IMPROVE - Innovative Modelling Approaches for Production Systems to Raise Validatable Efficiency : Intelligent Methods for the Factory of the Future*, Berlin, Heidelberg : Springer Berlin Heidelberg, 2018, p. 55-71.
- BOKRANTZ, J. ; SKOOGH, A. ; YLIPÄÄ, T. et STAHLRE, J., « Smart maintenance : A research agenda for industrial maintenance management », in : *International Journal of Production Economics* 191 (2017), p. 154-169.
- BONDY, John Adrian et MURTY, Uppaluri Siva Ramachandra, *Graph theory*, Springer Publishing Company, Incorporated, 2008.
- BOUCKAERT, R. R., « Bayesian Belief Networks : From Construction to Inference », in : *PhD thesis, University of Utrecht* (1995).
- BOUSDEKIS, Alexandros ; LEPENIOTI, Katerina ; APOSTOLOU, Dimitris et MENTZAS, Gregoris, « A Review of Data-Driven Decision-Making Methods for Industry 4.0 Maintenance Applications », in : *Electronics* 10.7 (2021).
- BOX, George EP ; JENKINS, Gwilym M ; REINSEL, Gregory C et LJUNG, Greta M, *Time series analysis : forecasting and control*, Holden-Day, 1970.
- BREIMAN, Leo, « Random Forests », in : *Machine Learning* 45.1 (2001), p. 5-32.

-
- BROCAL, Francisco ; GONZÁLEZ, Cristina ; KOMLJENOVIC, Dragan ; KATINA, Polinpapilho F et SEBASTIÁN, Miguel A, « Emerging risk management in industry 4.0 : an approach to improve organizational and human performance in the complex systems », in : *complexity* 2019.1 (2019), p. 2089763.
- BUDACH, Lukas et al., « The effects of data quality on machine learning performance », in : *arXiv preprint arXiv :2207.14529* (2022).
- BUNTINE, W. L., « Theory refinement on Bayesian networks », in : *Proc. UAI'91* (1991), 52–60.
- BUTLER, K.L., « An expert system based framework for an incipient failure detection and predictive maintenance system », in : *Proceedings of International Conference on Intelligent System Application to Power Systems*, 1996, p. 321-326.
- BUTTS, C.T., « Social network analysis : A methodological introduction », in : *Asian Journal of Social Psychology* 11.1 (2008), p. 13-41.
- CAI, Baoping ; HUANG, Lei et XIE, Min, « Bayesian Networks in Fault Diagnosis », in : *IEEE Transactions on Industrial Informatics* (2017).
- CAI, Baoping ; LIU, Hanlin et XIE, Min, « A real-time fault diagnosis methodology of complex systems using object-oriented Bayesian networks », in : *Mechanical Systems and Signal Processing* (2016), p. 31-44.
- CAI, Baoping et al., « Application of Bayesian Networks in Reliability Evaluation », in : *IEEE Transactions on Industrial Informatics* (2018), p. 2146-2157.
- CALABRESE, Francesca ; REGATTIERI, Alberto ; BORTOLINI, Marco ; GAMBERI, Mauro et PILATI, Francesco, « Predictive Maintenance : A Novel Framework for a Data-Driven, Semi-Supervised, and Partially Online Prognostic Health Management Application in Industries », in : *Applied Sciences* 11.8 (2021).
- CAMPOS, L. M. de, « A scoring function for learning Bayesian networks based on mutual information and conditional independence tests », in : *Journal of Machine Learning Research* (2006), 2149–2187.
- CAMPOS, L. M. De ; FERNÁNDEZ-LUNA, J. M. et PUERTA, J. M., « An iterated local search algorithm for learning bayesian networks with restarts based on conditional independence tests », in : *International Journal of Intelligent Systems* (2003), 221–235.
- CANIZO, Mikel et al., « Real-time Predictive Maintenance for Wind Turbines Using Big Data Frameworks », in : *Journal of Manufacturing Science and Engineering*, 2017.

-
- CANNARILE, Francesco ; BARALDI, Piero et ZIO, Enrico, « An evidential similarity-based regression method for the prediction of equipment remaining useful life in presence of incomplete degradation trajectories », in : *Fuzzy Sets and Systems* (2019).
- CARBERY, Caoimhe M. ; WOODS, Roger et MARSHALL, Adele H., « A Bayesian network based learning system for modelling faults in large-scale manufacturing », in : *2018 IEEE International Conference on Industrial Technology (ICIT)*, 2018, p. 1357-1362.
- CARVALHO, Thyago Peres et al., « A systematic literature review of machine learning methods applied to predictive maintenance », in : *Computers & Industrial Engineering* (2019).
- CATLETT, J., « On changing continuous attributes into ordered discrete attributes », in : *Machine Learning — EWSL-91*, sous la dir. d'Yves KODRATOFF, Berlin, Heidelberg : Springer Berlin Heidelberg, 1991, p. 164-178.
- CHEN, Chong et al., « Predictive maintenance using cox proportional hazard deep learning », in : *Advanced Engineering Informatics* 44 (2020), p. 101054.
- CHEN, Chuang ; LU, Ningyun ; JIANG, Bin et WANG, Cunsong, « A Risk-Averse Remaining Useful Life Estimation for Predictive Maintenance », in : *IEEE/CAA Journal of Automatica Sinica* 8.2 (2021), p. 412-422.
- CHEN, J. ; ROBERTS, C. et WESTON, P., « Fault detection and diagnosis for railway track circuits using neuro-fuzzy systems », in : *Control Engineering Practice* 16.5 (2008), p. 585-596.
- CHEN, Ricky T. Q. ; RUBANOVA, Yulia ; BETTENCOURT, Jesse et DUVENAUD, David, « Neural Ordinary Differential Equations », in : *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18, Montréal, Canada : Curran Associates Inc., 2018, 6572–6583.
- CHERNOFF, H. et LEHMANN, E., « The use of maximum likelihood estimates in χ^2 tests for goodness of fit », in : *The Annals of Mathematical Statistics* (1954), 579–586.
- CHIACHÍO, Juan ; CHIACHÍO, Manuel ; PRESCOTT, Darren et ANDREWS, John, « A knowledge-based prognostics framework for railway track geometry degradation », in : *Reliability Engineering & System Safety* 181 (2019), p. 127-141.
- CHOCKALINGAM, Sabarathinam ; PIETERS, Wolter ; TEIXEIRA, André et GELDER, Pieter van, « Bayesian Network Models in Cyber Security : A Systematic Review », in : *Secure IT Systems* (2017).
- CLIMENTE-ALARCON, Vicente ; NAIR, Devi ; SUNDARIA, Ravi ; ANTONINO-DAVIU, Jose A. et ARKKIO, Antero, « Combined Model for Simulating the Effect of Transients on a

-
- Damaged Rotor Cage », in : *IEEE Transactions on Industry Applications* 53.4 (2017), p. 3528-3537.
- COMPARE, Michele ; BARALDI, Piero et ZIO, Enrico, « Challenges to IoT-enabled predictive maintenance for industry 4.0 », in : *IEEE Internet of Things Journal* 7.5 (2019), p. 4585-4597.
- COMPARE, Michele et ZIO, Enrico, « Predictive Maintenance by Risk Sensitive Particle Filtering », in : *IEEE Transactions on Reliability* 63.1 (2014), p. 134-143.
- CONSTANTINOU, Anthony Costa ; FENTON, Norman ; MARSH, William et RADLINSKI, Lukasz, « From complex questionnaire and interviewing data to intelligent Bayesian network models for medical decision support », in : *Artificial Intelligence in Medicine* (2016), p. 75-93.
- COOPER, G. et HERSOVITS, E., « A Bayesian method for the induction of probabilistic networks from data », in : *Machine Learning* (1992), 309-347.
- CORADDU, Andrea et al., « Machine learning approaches for improving condition-based maintenance of naval propulsion plants », in : *Proceedings of the Institution of Mechanical Engineers, Part M : Journal of Engineering for the Maritime Environment* 230.1 (2016), p. 136-153.
- CORTES, Corinna et VAPNIK, Vladimir, « Support-vector networks », in : *Machine Learning* 20 (1995), p. 273-297.
- COSTELLO, Jason J. A. ; WEST, Graeme M. et MCARTHUR, Stephen D. J., « Machine Learning Model for Event-Based Prognostics in Gas Circulator Condition Monitoring », in : *IEEE Transactions on Reliability* 66.4 (2017), p. 1048-1057.
- COULIBALY, Lassana ; KAMSU-FOGUEM, Bernard et TANGARA, Fana, « Rule-based machine learning for knowledge discovering in weather data », in : *Future Generation Computer Systems* 108 (2020), p. 861-878.
- CUI, Lingli ; WANG, Xin ; WANG, Huaqing et MA, Jianfeng, « Research on Remaining Useful Life Prediction of Rolling Element Bearings Based on Time-Varying Kalman Filter », in : *IEEE Transactions on Instrumentation and Measurement* 69.6 (2020), p. 2858-2867.
- CUMMINS, Logan et al., « Explainable predictive maintenance : A survey of current methods, challenges and opportunities », in : *IEEE access* (2024).
- DANGUT, Maren David ; SKAF, Zakwan et JENNIONS, Ian K., « An integrated machine learning model for aircraft components rare failure prognostics with log-based dataset. », in : *Isa Transactions* (2020).

-
- « Handling Imbalanced Data for Aircraft Predictive Maintenance Using the BACHE Algorithm », in : *Appl. Soft Comput.* 123 (2022).
- DAS, Swagatam ; DATTA, Shounak et CHAUDHURI, Bidyut B., « Handling data irregularities in classification : Foundations, trends, and future challenges », in : *Pattern Recognition* 81 (2018), p. 674-693.
- DAVARI, Narjes et al., « A Survey on Data-Driven Predictive Maintenance for the Railway Industry », in : *Sensors* (2021).
- DAVIS, Mark HA, « Piecewise-deterministic Markov processes : A general class of non-diffusion stochastic models », in : *Journal of the Royal Statistical Society : Series B (Methodological)* 46.3 (1984), p. 353-376.
- DEAN, T. et KANAZAWA, K., « A model for reasoning about persistence and causation », in : *Comput. Intell.* (1989), 142–150.
- DEMPSTER, A. P. ; LAIRD, N. M. et RUBIN, D. B., « Maximum Likelihood from Incomplete Data via the EM Algorithm », in : *Journal of the Royal Statistical Society. Series B (Methodological)* 39.1 (1977), p. 1-38.
- DOUGHERTY, James ; KOHAVI, Ron et SAHAMI, Mehran, « Supervised and Unsupervised Discretization of Continuous Features », in : *Machine Learning Proceedings 1995*, sous la dir. d'Armand PRIEDITIS et Stuart RUSSELL, San Francisco (CA) : Morgan Kaufmann, 1995, p. 194-202.
- DOYEN, Laurent et GAUDOIN, Olivier, « Classes of imperfect repair models based on reduction of failure intensity or virtual age », in : *Reliability Engineering & System Safety* 84.1 (2004), p. 45-56.
- DRURY, Brett ; VALVERDE-REBAZA, Jorge ; MOURA, Maria-Fernanda et ANDRADE LOPES, Alneu de, « A survey of the applications of Bayesian networks in agriculture », in : *Engineering Applications of Artificial Intelligence* (2017), 29–42.
- EMMANUEL, Tlamelo et al., « A survey on missing data in machine learning », in : *Journal of Big data* 8 (2021), p. 1-37.
- EXCELLENCE, Prognostics Center of, *Prognostics Data Repository*, sous la dir. de NASA, URL : <https://www.nasa.gov/intelligent-systems-division/discovery-and-systems-health/pcoe/pcoe-data-set-repository/> (visité le 2025).
- FAUSING OLESEN, Jonas et SHAKER, Hamid Reza, « Predictive Maintenance for Pump Systems and Thermal Power Plants : State-of-the-Art Review, Trends and Challenges », in : *Sensors* 20.8 (2020).

-
- FAYYAD, Usama M et IRANI, Keki B, « Multi-interval discretization of continuous-valued attributes for classification learning », in : *Ijcai*, t. 93, 1993, p. 1022-1029.
- FELLER, William, *An introduction to probability theory and its applications, Volume 2*, t. 2, John Wiley & Sons, 1991.
- FERREIRO, Susana ; ARNAIZ, Aitor ; SIERRA, Basilio et IRIGOIEN, Itziar, « Application of Bayesian networks in prognostics for a new Integrated Vehicle Health Management concept », in : *Expert Systems with Applications* 39.7 (2012), p. 6402-6418.
- FIGUEROA, Rosa L. ; ZENG-TREITLER, Qing ; KANDULA, Sasikiran et NGO, Long H., « Predicting sample size required for classification performance », in : *BMC Medical Informatics and Decision Making* 12 (2012).
- FIRESTONE, Simon M. et al., « Understanding the associations between on-farm biosecurity practice and equine influenza infection during the 2007 outbreak in Australia », in : *Preventive Veterinary Medicine* (2013), p. 28-36.
- FLORES, M. Julia ; NICHOLSON, Ann E. ; BRUNSKILL, Andrew ; KORB, Kevin B. et MASCARO, Steven, « Incorporating expert knowledge when learning Bayesian network structure : A medical case study », in : *Artificial Intelligence in Medicine* (2011), p. 181-204.
- FLORIAN, Eleonora ; SGARBOSSA, Fabio et ZENNARO, Ilenia, « Machine learning-based predictive maintenance : A cost-oriented model for implementation », in : *International Journal of Production Economics* 236 (2021), p. 108114.
- FORRESTER, Tyler ; HARRIS, Mark ; SENECA, Jacob et SHEPPARD, John, « Continuous Time Bayesian Networks in prognosis and health management of centrifugal pumps », in : *Proceedings of the Annual Conference of the PHM Society*, t. 11, 1, 2019.
- FOSSIER, Simon et ROBIC, Pierre-Olivier, « Maintenance of complex systems — From preventive to predictive », in : *2017 12th International Conference on Live Maintenance (ICOLIM)*, 2017, p. 1-6.
- FOX, Harriet et al., « A review of predictive and prescriptive offshore wind farm operation and maintenance », in : *Energies* 15.2 (2022), p. 504.
- FREEDMAN, David ; PISANI, Robert et PURVES, Roger, « Statistics (international student edition) », in : *Pisani, R. Purves, 4th edn. WW Norton & Company, New York* (2007).
- FREEMAN, Linton C., « Centrality in social networks conceptual clarification », in : *Social Networks* (1978), p. 215-239.

-
- FREEMAN, Linton C ; ROEDER, Douglas et MULHOLLAND, Robert R, « Centrality in social networks : II. Experimental results », in : *Social networks* 2.2 (1979), p. 119-141.
- F.R.S., Karl Pearson, « LIII. On lines and planes of closest fit to systems of points in space », in : *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2.11 (1901), p. 559-572.
- FU, Wenhan ; CHIEN, Chen-Fu et TANG, Lizhen, « Bayesian network for integrated circuit testing probe card fault diagnosis and troubleshooting to empower Industry 3.5 smart production and an empirical study », in : *Journal of Intelligent Manufacturing* (2022), 785-798.
- FUKUSHIMA, Kunihiko, « Neocognitron : A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position », in : *Biological cybernetics* 36.4 (1980), p. 193-202.
- GAAG, Linda C. van der ; BOLT, Janneke ; LOEFFEN, Willie et ELBERS, Armin, « Modelling Patterns of Evidence in Bayesian Networks : A Case-Study in Classical Swine Fever », in : *Lecture Notes in Computer Science* (2010), p. 675-684.
- GANDHI, Niketa ; ARMSTRONG, Leisa J. et PETKAR, Owaiz, « Predicting Rice Crop Yield using Bayesian Networks », in : *2016 International Conference on Advances in Computing, Communications and Informatics* (2016), p. 795-799.
- GARCÍA, Salvador ; LUENGO, Julián ; SÁEZ, José Antonio ; LÓPEZ, Victoria et HERRERA, Francisco, « A Survey of Discretization Techniques : Taxonomy and Empirical Analysis in Supervised Learning », in : *IEEE Transactions on Knowledge and Data Engineering* 25.4 (2013), p. 734-750.
- GASHI, Milot et THALMANN, Stefan, « Taking Complexity into Account : A Structured Literature Review on Multi-component Systems in the Context of Predictive Maintenance », in : *Information Systems*, sous la dir. de Marinos THEMISTOCLEOUS et Maria PAPADAKI, Cham : Springer International Publishing, 2020, p. 31-44, ISBN : 978-3-030-44322-1.
- GAY, Antonin et al., « A Short Review on the Integration of Expert Knowledge in Prognostics for PHM in Industrial Applications », in : *2021 5th International Conference on System Reliability and Safety (ICSR)*, 2021, p. 286-292.
- GHONEIM, Sherif ; RASHED, Amr et ELKALASHY, Nagy, « Fault Detection Algorithms for Achieving Service Continuity in Photovoltaic Farms », in : *Intelligent Automation & Soft Computing* 29 (2021).

GIARRATANO, Joseph C. et RILEY, Gary, *Expert systems : principles and programming*, sous la dir. de Boston : PWS Pub. Co., PWS Pub. Co., 1998, URL : <http://worldcat.org>.

GINKEL, Joost R. van, « Handling Missing Data in Principal Component Analysis Using Multiple Imputation », in : *Essays on Contemporary Psychometrics*, sous la dir. de L. Andries van der ARK ; Wilco H. M. EMONS et Rob R. MEIJER, Springer International Publishing, 2023, p. 141-161.

GLOVER, F. et LAGUNA, M., « Tabu search », in : *Modern Heuristic Techniques for Combinatorial Problems* (1993), 70–150.

GOHEL, Hardik A. ; UPADHYAY, Himanshu ; LAGOS, Leonel ; COOPER, Kevin et SANZETENE, Andrew, « Predictive maintenance architecture development for nuclear infrastructure using machine learning », in : *Nuclear Engineering and Technology* 52.7 (2020), p. 1436-1442.

GOYAL, Deepam et PABLA, BS, « The vibration monitoring methods and signal processing techniques for structural health monitoring : a review », in : *Archives of Computational Methods in Engineering* 23 (2016), p. 585-594.

GRABOT, Bernard, « Rule mining in maintenance : Analysing large knowledge bases », in : *Computers & Industrial Engineering* 139 (2020), p. 105501.

GROUP, EDP, *Inside a Boiler—Slagging Prediction*, <https://www.edp.com/en/innovation/open-data/data>, Accessed : 2024-06-03, 2020.

GUYON, Isabelle et ELISSEEFF, André, « An Introduction of Variable and Feature Selection », in : *J. Machine Learning Research Special Issue on Variable and Feature Selection* 3 (2003), p. 1157 -1182.

GYAMFI, Kojo Sarfo ; BRUSEY, James ; HUNT, Andrew et GAURA, Elena, « Linear dimensionality reduction for classification via a sequential Bayes error minimisation with an application to flow meter diagnostics », in : *Expert Systems with Applications* 91 (2018), p. 252-262.

HAKAMI, Ali, « Strategies for overcoming data scarcity, imbalance, and feature selection challenges in machine learning models for predictive maintenance », in : *Scientific Reports* (2020).

HAN, Kai ; WANG, Yunhe ; ZHANG, Chao ; LI, Chao et XU, Chao, « Autoencoder inspired unsupervised feature selection », in : *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, IEEE, 2018, p. 2941-2945.

-
- HANACHI, Houman ; LIU, Jie ; BANERJEE, Avisekh ; CHEN, Ying et KOUL, Ashok, « A Physics-Based Modeling Approach for Performance Monitoring in Gas Turbine Engines », in : *IEEE Transactions on Reliability* 64.1 (2015), p. 197-205.
- HAO, Liu ; JINSONG, Yu ; PING, Zhang et XINGSHAN, Li, « A review on fault prognostics in integrated health management », in : 2009.
- HARARY, Frank, *Graph theory (on Demand Printing of 02787)*, CRC Press, 2018.
- HE, Rui ; TIAN, Zhigang et ZUO, Ming J., « A semi-supervised GAN method for RUL prediction using failure and suspension histories », in : *Mechanical Systems and Signal Processing* 168 (2022), p. 108657.
- HECKERMAN, D. ; GEIGER, D. et CHICKERING, D. M., « Learning Bayesian networks : The combination of knowledge and statistical data », in : *Machine Learning* (1995), 197–243.
- HELWIG, Nikolai ; PIGNANELLI, Eliseo et SCHÜTZE, Andreas, « Condition monitoring of a complex hydraulic system using multivariate statistics », in : *2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*, 2015, p. 210-215.
- HENG, Aiwna ; ZHANG, Sheng ; TAN, Andy C.C. et MATHEW, Joseph, « Rotating machinery prognostics : State of the art, challenges and opportunities », in : *Mechanical Systems and Signal Processing* 23.3 (2009), p. 724-739.
- HMELO-SILVER, Cindy E et AZEVEDO, Roger, « Understanding complex systems : Some core challenges », in : *The Journal of the learning sciences* 15.1 (2006), p. 53-61.
- HOCHREITER, Sepp et SCHMIDHUBER, Jürgen, « Long Short-Term Memory », in : *Neural Comput.* 9.8 (1997), 1735–1780.
- HOLZINGER, Andreas, « Interactive machine learning for health informatics : when do we need the human-in-the-loop ? », in : *Brain Informatics* (2016).
- HONG, Jie ; MIAO, Xuewen ; HAN, Lei et MA, Yanhong, « Prognostics model for predicting aero-engine bearing Grade-life », in : *Proceedings of ASME Turbo Expo 2009* (2009).
- HU, Chao ; YOUN, Byeng D. et WANG, Pingfeng, « Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life », in : *2011 IEEE Conference on Prognostics and Health Management*, 2011, p. 1-10.
- HU, Chao ; YOUN, Byeng D. ; WANG, Pingfeng et TAEK YOON, Joung, « Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life », in : *Reliability Engineering & System Safety* 103 (2012), p. 120-135.

-
- HU, Yang ; BARALDI, Piero ; DI MAIO, Francesco et ZIO, Enrico, « A particle filtering and kernel smoothing-based approach for new design component prognostics », in : *Reliability Engineering and System Safety* (2015).
- HU, Yang ; MIAO, Xuewen ; SI, Yong ; PAN, Ershun et ZIO, Enrico, « Prognostics and health management : A review from the perspectives of design, development and decision », in : *Reliability Engineering & System Safety* 217 (2022), p. 108063.
- HU, Yaogang et al., « A probability evaluation method of early deterioration condition for the critical components of wind turbine generator systems », in : *Mechanical Systems and Signal Processing* 76-77 (2016), p. 729-741.
- HUANG, Lida ; CAI, Guoray ; YUAN, Hongyong et CHEN, Jianguo, « A hybrid approach for identifying the structure of a Bayesian network », in : *Expert Systems With Applications* (2019), p. 308-320.
- HUANG, Zeyi ; XU, Zhengguo ; KE, Xiaojie ; WANG, Wenhai et SUN, Youxian, « Remaining useful life prediction for an adaptive skew-Wiener process model », in : *Mechanical Systems and Signal Processing* 87 (2017), p. 294-306.
- HUSSAIN, Akhtar ; LEE, Seung-Jae ; CHOI, Myeon-Song et BRIKCI, Fouad, « An Expert System for Acoustic Diagnosis of Power Circuit Breakers and On-Load Tap Changers », in : *Expert Syst. Appl.* 42.24 (2016), 9426-9433.
- HUSSAIN, Sajid et GABBAR, Hossam A., « Vibration Analysis and Time Series Prediction for Wind Turbine Gearbox Prognostics », in : *International Journal of Prognostics and Health Management* (2020).
- HUYNH, Khac Tuan, « An adaptive predictive maintenance model for repairable deteriorating systems using inverse Gaussian degradation process », in : *Reliab. Eng. Syst. Saf.* (2021).
- HUYNH, Khac Tuan ; BARROS, Anne et BERENGUER, Christophe, « Multi-Level Decision-Making for The Predictive Maintenance of k -Out-of-n :F Deteriorating Systems », in : *IEEE Transactions on Reliability* (2015).
- IDE, Tsuyoshi, « Formalizing Expert Knowledge Through Machine Learning », in : 2016, p. 157-175, ISBN : 978-1-4939-3592-5.
- JAHANI, Salman ; ZHOU, Shiyu et VEERAMANI, Dharmaraj, « Stochastic prognostics under multiple time-varying environmental factors », in : *Reliability Engineering & System Safety* 215 (2021), p. 107877.
- JANG, J.-S.R., « ANFIS : adaptive-network-based fuzzy inference system », in : *IEEE Transactions on Systems, Man, and Cybernetics* 23.3 (1993), p. 665-685.

-
- JARDINE, A. K. S. et al., « Maintenance, replacement, and reliability : Theory and applications », in : *Springer Science & Business Media* (2006).
- JARDINE, Andrew K.S. ; LIN, Daming et BANJEVIC, Dragan, « A review on machinery diagnostics and prognostics implementing condition-based maintenance », in : *Mechanical Systems and Signal Processing recent* (2005).
- J.C, Verhagen Wim et W.M., De Boer Lennaert, « Predictive maintenance for aircraft components using proportional hazard models », in : *Journal of Industrial Information Integration* 12 (2018), p. 23-30.
- JIANG, Yu ; ZHU, Hua ; DING, Cong et PFEIFFER, Olivia, « A novel ensemble fuzzy model for degradation prognostics of rolling element bearings », in : *Journal of Intelligent & Fuzzy Systems* 37 (2019), p. 1-7.
- JIMENEZ, Juan José Montero ; SCHWARTZ, Sébastien ; VINGERHOEDS, Rob A. ; GRABOT, Bernard et SALAÜN, Michel, « Towards multi-model approaches to predictive maintenance : A systematic literature survey on diagnostics and prognostics », in : *Journal of Manufacturing Systems* (2020).
- JIMENEZ-CORTADI, Alberto ; IRIGOEN, Itziar ; BOTO, Fernando ; SIERRA, Basilio et RODRIGUEZ, German, « Predictive Maintenance on the Machining Process and Machine Tool », in : *Applied Sciences* 10.1 (2020).
- JORDON, James et al., « Synthetic Data—what, why and how ? », in : *arXiv preprint arXiv :2205.03257* (2022).
- JOVICIC, Eda ; PRIMORAC, Daria ; CUPIC, Marko et JOVIC, Alan, « Publicly Available Datasets for Predictive Maintenance in the Energy Sector : A Review », in : *IEEE Access* 11 (2023a), p. 73505-73520.
- « Publicly Available Datasets for Predictive Maintenance in the Energy Sector : A Review », in : *IEEE Access* 11 (2023b), p. 73505-73520.
- KALATHAS, Ilias et PAPOUTSIDAKIS, Michail, « Predictive Maintenance Using Machine Learning and Data Mining : A Pioneer Method Implemented to Greek Railways », in : *Designs* 5.1 (2021).
- KANAWADAY, Ameeth et SANE, Aditya, « Machine learning for predictive maintenance of industrial machines using IoT sensor data », in : *international conference on software engineering*, 2017.
- KARUPPUSAMY, P., « Machine Learning Approach to Predictive Maintenance in Manufacturing Industry - A Comparative Study », in : *Journal of Social and Clinical Psychology* (2020).

-
- KERBER, Randy, « ChiMerge : discretization of numeric attributes », in : *Proceedings of the Tenth National Conference on Artificial Intelligence*, AAAI'92, San Jose, California : AAAI Press, 1992, 123–128.
- KHAN, Samir et YAIRI, Takehisa, « A review on the application of deep learning in system health management », in : *Mechanical Systems and Signal Processing* (2018).
- KHAN, Z. J., « Predictive maintenance and Internet of Things », in : *Proc. Int. Conf. Comput., Electron. Electr. Eng.* 2021, 1–5.
- KITSON, Neville Kenneth ; CONSTANTINO, Anthony C. ; GUO, Zhigao ; LIU, Yang et CHOBTHAM, Kiattikun, « A survey of Bayesian Network structure learning », in : *Artificial Intelligence Review* (2023).
- KORDESTANI, Mojtaba ; SAIF, Mehrdad ; ORCHARD, Marcos E. ; RAZAVI-FAR, Roozbeh et KHORASANI, Khashayar, « Failure Prognosis and Applications—A Survey of Recent Literature », in : *IEEE Transactions on Reliability* 70.2 (2021), p. 728-748.
- KRAUS, Mathias et FEUERRIEGEL, Stefan, « Forecasting remaining useful life : Interpretable deep learning approach via variational Bayesian inferences », in : *Decision Support Systems* 125 (2019), p. 113100.
- KURREWAR, Harshad ; BEKAR, Ebru Turanouglu ; SKOOGH, Anders et NYQVIST, Per, « A Machine Learning Based Health Indicator Construction in Implementing Predictive Maintenance : A Real World Industrial Application from Manufacturing », in : *Advances in Production Management Systems. Artificial Intelligence for Sustainable and Resilient Production Systems*, sous la dir. d'Alexandre DOLGUI ; Alain BERNARD ; David LEMOINE ; Gregor von CIEMINSKI et David ROMERO, Cham : Springer International Publishing, 2021, p. 599-608, ISBN : 978-3-030-85906-0.
- LADYMAN, James ; LAMBERT, James et WIESNER, Karoline, « What is a complex system ? », in : *European Journal for Philosophy of Science* 3 (2013), p. 33-67.
- LAKSHMINARAYAN, Kamakshi ; HARP, Steven A ; GOLDMAN, Robert P ; SAMAD, Tariq et al., « Imputation of Missing Data Using Machine Learning Techniques. », in : *KDD*, t. 96, 1996, p. 140-145.
- LANGONE, Rocco ; CUZZOCREA, Alfredo et SKANTZOS, Nikolaos, « Interpretable Anomaly Prediction : Predicting anomalous behavior in industry 4.0 settings via regularized logistic regression tools », in : *Data & Knowledge Engineering* 130 (2020), p. 101850.

-
- LE SON, Khanh ; FOULADIRAD, Mitra ; BARROS, Anne ; LEVRAT, Eric et IUNG, Benoît, « Remaining useful life estimation based on stochastic deterioration models : A comparative study », in : *Reliability Engineering & System Safety* 112 (2013), p. 165-175.
- LECUN, Yann ; BENGIO, Y. et HINTON, Geoffrey, « Deep Learning », in : *Nature* 521 (2015), p. 436-44.
- LEE, Elisa T. et Go, Oscar T., « Survival Analysis in Public Health Research », in : *Annual Review of Public Health* 18.1 (1997), PMID : 9143714, p. 105-134, eprint : 10.1146/annurev.publhealth.18.1.105.
- LEE, J. ; QIU, H. ; YU, G. ; LIN, J. et SERVICES, Rexnord Technical, *Bearing Data Set*, rapp. tech., <https://www.nasa.gov/intelligent-systems-division/discovery-and-systems-health/pcoe/pcoe-data-set-repository/>, NASA Ames Research Center, 2007.
- LEE, Jay ; NI, Jun ; DJURDJANOVIC, Dragan ; QIU, Hai et LIAO, Haitao, « Intelligent prognostics tools and e-maintenance », in : *Computers in Industry* 57.6 (2006), p. 476-489.
- LEE, Jay et al., « Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications », in : *Mechanical Systems and Signal Processing* 42.1 (2014), p. 314-334.
- LEE, Jay et al., « Intelligent Maintenance Systems and Predictive Manufacturing », in : *Journal of Manufacturing Science and Engineering* (2020).
- LEE, Wo Jae et al., « Predictive Maintenance of Machine Tool Systems Using Artificial Intelligence Techniques Applied to Machine Condition Data », in : *26th CIRP Life Cycle Engineering (LCE)*, t. 65, Elsevier BV, 2017, p. 32-37.
- LEI, Yaguo et al., « Machinery health prognostics : A systematic review from data acquisition to RUL prediction », in : *Mechanical Systems and Signal Processing* (2018b).
- « Machinery health prognostics : A systematic review from data acquisition to RUL prediction », in : *Mechanical Systems and Signal Processing* 104 (2018a), p. 799-834.
- LEI, Yaguo et al., « Applications of machine learning to machine fault diagnosis : A review and roadmap », in : *Mechanical Systems and Signal Processing* 138 (2020), p. 106587.
- LEUKEL, Joerg ; GONZÁLEZ, Julian et RIEKERT, Martin, « Adoption of machine learning technology for failure prediction in industrial maintenance : A systematic review », in : *Journal of Manufacturing Systems* (2021).
- LEWIS, Austin D. et GROTH, Katrina M., « A Dynamic Bayesian Network Structure for Joint Diagnostics and Prognostics of Complex Engineering Systems », in : *Algorithms* (2020).

-
- LI, Hongfei et al., « Improving rail network velocity : A machine learning approach to predictive maintenance », in : *Transportation Research Part C-emerging Technologies* 45 (2014), p. 17-26.
- LI, Jianfeng ; XIAO, Mingqing ; LIANG, Yajun ; TANG, Xilang et LI, Chao, « Three-dimensional Simulation and Prediction of Solenoid Valve Failure Mechanism Based on Finite Element Model », in : *IOP Conference Series : Earth and Environmental Science* 108.2 (2018a), p. 022035.
- LI, Shiyang et al., « Enhancing the Locality and Breaking the Memory Bottleneck of Transformer on Time Series Forecasting », in : *Advances in Neural Information Processing Systems*, sous la dir. de H. WALLACH et al., t. 32, Curran Associates, Inc., 2019.
- LI, Weihua et al., « A perspective survey on deep transfer learning for fault diagnosis in industrial scenarios : Theories, applications and challenges », in : *Mechanical Systems and Signal Processing* 167 (2022), p. 108487.
- LI, Xiali ; DENG, Songting ; WANG, Song ; Lv, Zhengyu et WU, Licheng, « Review of Small Data Learning Methods », in : *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*, t. 02, 2018b, p. 106-109.
- LIAO, Linxia et KÖTTIG, Felix, « Review of Hybrid Prognostics Approaches for Remaining Useful Life Prediction of Engineered Systems, and an Application to Battery Life Prediction », in : *IEEE Transactions on Reliability* (2014).
- « A hybrid framework combining data-driven and model-based methods for system remaining useful life prediction », in : *Applied Soft Computing* 44 (2016), p. 191-199.
- LIM, Bryan ; ARIK, Sercan O. ; LOEFF, Nicolas et PFISTER, Tomas, « Temporal Fusion Transformers for interpretable multi-horizon time series forecasting », in : *International Journal of Forecasting* 37.4 (2021), p. 1748-1764.
- LIN, Chin-Yi ; HSIEH, Yu-Ming ; CHENG, Fan-Tien ; HUANG, Hsien-Cheng et ADNAN, Muhammad, « Time Series Prediction Algorithm for Intelligent Predictive Maintenance », in : *IEEE Robotics and Automation Letters* 4.3 (2019), p. 2807-2814.
- LIN, Lin ; LUO, Bin et ZHONG, ShiSheng, « Multi-objective decision-making model based on CBM for an aircraft fleet with reliability constraint », in : *International Journal of Production Research* 56.14 (2018), p. 4831-4848.
- LIN, Qin ; ADEPU, Sridha ; VERWER, Sicco et MATHUR, Aditya, « TABOR : A Graphical Model-based Approach for Anomaly Detection in Industrial Control Systems »,

-
- in : *Proceedings of the 2018 on Asia Conference on Computer and Communications Security* (2018), p. 525-536.
- LIU, Datong ; PANG, Jingyue ; ZHOU, Jianbao ; PENG, Yu et PECHT, Michael, « Prognostics for state of health estimation of lithium-ion batteries based on combination Gaussian process functional regression », in : *Microelectronics Reliability* 53.6 (2013), p. 832-839.
- LIU, Ningjing et al., « Automated food safety early warning system in the dairy supply chain using machine learning », in : *Food Control* 136 (2022), p. 108872.
- LIU, Ruonan ; YANG, Boyuan ; ZIO, Enrico et CHEN, Xuefeng, « Artificial intelligence for fault diagnosis of rotating machinery : A review », in : *Mechanical Systems and Signal Processing* (2018).
- LLOYD, S., « Least squares quantization in PCM », in : *IEEE Transactions on Information Theory* 28.2 (1982), p. 129-137.
- LUNDBERG, Scott M et LEE, Su-In, « A Unified Approach to Interpreting Model Predictions », in : *Advances in Neural Information Processing Systems*, sous la dir. d'I. GUYON et al., t. 30, Curran Associates, Inc., 2017.
- LUO, Bo ; WANG, Haoting ; LIU, Hongqi ; LI, Bin et PENG, Fangyu, « Early Fault Detection of Machine Tools Based on Deep Learning and Dynamic Identification », in : *IEEE Transactions on Industrial Electronics* (2019).
- LUO, Weichao ; HU, Tianliang ; YE, Yingxin ; ZHANG, Chengrui et WEI, Yongli, « A hybrid predictive maintenance approach for CNC machine tool driven by Digital Twin », in : *Robotics and Computer-Integrated Manufacturing* 65 (2020), p. 101974.
- MALBASA, Vuk ; ZHENG, Ce ; CHEN, Po-Chen ; POPOVIC, Tomo et KEZUNOVIC, Mladen, « Voltage Stability Prediction Using Active Machine Learning », in : *IEEE Transactions on Smart Grid* (2017).
- MARCOT, Bruce G. ; STEVENTON, J. Douglas ; SUTHERLAND, Glenn D. et McCANN, Robert K., « Guidelines for developing and updating Bayesian belief networks applied to ecological modeling and conservation », in : *Canadian Journal of Forest Research* (2006), p. 3063-3074.
- MARTIN PECH, Jaroslav Vrchota et BEDNÁŘ, Jiří, « Predictive Maintenance and Intelligent Sensors in Smart Factory : Review », in : *Sensors* (2021).
- MATHEW, Josey ; LUO, Ming et PANG, Chee Khiang, « Regression kernel for prognostics with support vector machines », in : *2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, 2017, p. 1-5.

-
- MATHEW, Vimala ; TOBY, Tom ; SINGH, Vikram ; RAO, B. Maheswara et KUMAR, M. Goutham, « Prediction of Remaining Useful Lifetime (RUL) of turbofan engine using machine learning », in : 2017.
- MATZKA, Stephan, « Explainable Artificial Intelligence for Predictive Maintenance Applications », in : *2020 Third International Conference on Artificial Intelligence for Industries (AI4I)*, 2020, p. 69-74.
- MCLACHLAN, Scott ; DUBE, Kudakwashe ; HITMAN, Graham A et FENTON, Norman E., « Bayesian networks in healthcare : Distribution by medical condition », in : *Artificial Intelligence In Medicine* (2020).
- MILLER, Kyle et DUBRAWSKI, Artur, « System-Level Predictive Maintenance : Review of Research Literature and Gap Analysis. », in : *arXiv : Artificial Intelligence* (2020).
- MOBLEY, R. Keith, *An Introduction to Predictive Maintenance 2nd edition*, sous la dir. de BUTTERWORTH-HEINEMANN, Elsevier, 2002, ISBN : 978-0-7506-7531-4.
- MOŁĘDA, Marek ; MAŁYSIAK-MROZEK, Bożena ; DING, Weiping ; SUNDERAM, Vaidy et MROZEK, Dariusz, « From corrective to predictive maintenance—A review of maintenance approaches for the power industry », in : *Sensors 23.13* (2023), p. 5970.
- MOSALLAM, A. ; MEDJAHER, K. ; et ZERHOUNI, N., « Data-driven prognostics and health management : A review of recent trends and future directions », in : *Annual Reviews in Control 45* (2016), p. 200-217.
- MULDERS, Michel et HAARMAN, Mark, *Predictive Maintenance 4.0 Beyond the hype : PdM 4.0 delivers results September 2018*, 2018, URL : <https://www.pwc.nl/en/insights-and-publications/services-and-industries/industrial-manufacturing/predictive-maintenance-4-0.html>.
- MURPHY, Kevin P, « Dynamic bayesian networks : Representation, inference and learning », thèse de doct., University of California, Berkeley, 2002.
- NAGPAL, Chirag ; YADLOWSKY, Steve ; ROSTAMZADEH, Negar et HELLER, Katherine, « Deep Cox Mixtures for Survival Regression », in : *Proceedings of Machine Learning Research* (2021).
- NARAYANAN, A. et MOORE, M., « Quantum-inspired genetic algorithms », in : *Proceedings of IEEE International Conference on Evolutionary Computation*, 1996, p. 61-66.
- NASH, David et HANNAH, Murray, « Using Monte-Carlo simulations and Bayesian Networks to quantify and demonstrate the impact of fertiliser best management practices », in : *Environmental Modelling & Software* (2011), p. 1079-1088.

-
- NEAGU, Daniel ; SOLEIMANI, Morteza et CAMPEAN, Felician, « Diagnostics and prognostics for complex systems : A review of methods and challenges », in : *Quality and Reliability Engineering International* (2021).
- NECTOUX, Patrick et al., « PRONOSTIA : An experimental platform for bearings accelerated degradation tests », in : *IEEE International Conference on Prognostics and Health Management, PHM'12*. Denver, United States, 2012, p. 1-8.
- NEERUKATTI, Rajesh Kumar ; LIU, Kuang C. ; KOVVALI, Narayan et CHATTOPADHYAY, Aditi, « Fatigue Life Prediction Using Hybrid Prognosis for Structural Health Monitoring », in : *Journal of Aerospace Information Systems* 11.4 (2014), p. 211-232, eprint : 10.2514/1.I010094.
- NEWMAN, Mark EJ, « The structure and function of complex networks », in : *SIAM review* 45.2 (2003), p. 167-256.
- NGUYEN, Khanh T.P. et MEDJAHER, Kamal, « A new dynamic predictive maintenance framework using deep learning for failure prognostics », in : *Reliability Engineering & System Safety* (2019).
- NODELMAN, Uri ; SHELTON, Christian R et KOLLER, Daphne, « Continuous time Bayesian networks », in : *arXiv preprint arXiv :1301.0591* (2012).
- NUNES, P. ; SANTOS, J. et ROCHA, E., « Challenges in predictive maintenance – A review », in : *CIRP Journal of Manufacturing Science and Technology* 40 (2023), p. 53-67.
- NYULÁSZI, L. et al., « Fault Detection and Isolation of an Aircraft Turbojet Engine Using a Multi-Sensor Network and Multiple Model Approach », in : *Acta Polytechnica Hungarica* 15 (2018), p. 189-209.
- OCHELLA, Sunday ; SHAFIEE, Mahmood et DINMOHAMMADI, Fateme, « Artificial intelligence in prognostics and health management of engineering systems », in : *Engineering Applications of Artificial Intelligence* 108 (2022), p. 104552.
- ORRÙ, Pier Francesco et al., « Machine Learning Approach Using MLP and SVM Algorithms for the Fault Prediction of a Centrifugal Pump in the Oil and Gas Industry », in : *Sustainability* 12.11 (2020), p. 4776.
- OUADAH, Abdelfettah ; ZEMMOUCHI-GHOMARI, Leila et SALHI, Nedjma, « Selecting an appropriate supervised machine learning algorithm for predictive maintenance », in : 119 (2022), p. 4277-4301.

-
- PANGUN, Park ; PIERGIUSEPPE, Di Marco ; HYEJEON, Shin et JUNSEONG, Bang, « Fault Detection and Diagnosis Using Combined Autoencoder and Long Short-Term Memory Network », in : *Sensors (Basel)* 19 (2019), p. 4612.
- PAOLANTI, Marina et al., « Machine Learning approach for Predictive Maintenance in Industry 4.0 », in : *IEEE asme international conference on mechatronic and embedded systems and applications*, 2018.
- PARIS, P. et ERDOGAN, F., « A Critical Analysis of Crack Propagation Laws », in : *Journal of Basic Engineering* 85.4 (1963), p. 528-533.
- PEARL, J., « Bayesian Networks : A Model of Self-Activated Memory for Evidential Reasoning », in : *UCLA Computer Science Department* (1985).
- PEARL, Judea, *Probabilistic Reasoning in Intelligent Systems : Networks of Plausible Inference*, San Francisco, CA, USA : Morgan Kaufmann Publishers Inc., 1988, ISBN : 1558604790.
- *Causality*, 2^e éd., Cambridge University Press, 2009.
- PEARSON, Karl, « LIII. On lines and planes of closest fit to systems of points in space », in : *The London, Edinburgh, and Dublin philosophical magazine and journal of science* 2.11 (1901), p. 559-572.
- PECCHIA, Antonio ; SHARMA, Aashish ; KALBARTZYK, Zbigniew ; COTRONEO, Domenico et IYER, Ravishankar K., « Identifying Compromised Users in Shared Computing Infrastructures : A Data-Driven Bayesian Network Approach », in : *2011 30th IEEE International Symposium on Reliable Distributed Systems* (2011), p. 127-163.
- PENG, Cheng et al., « A Remaining Useful Life Prognosis of Turbofan Engine Using Temporal and Spatial Feature Fusion », in : *Sensors* 21.2 (2021).
- PERTSELAKIS, Minas ; LAMPATHAKI, Fenareti et PETRALI, Pierluigi, « Predictive Maintenance in a Digital Factory Shop-floor : Data Mining on Historical and Operational Data coming from Manufacturers' Information Systems », in : (2019).
- PESTANA, D., *Diesel engine faults features dataset (3500-default)*, <https://data.mendeley.com/datasets/k22zxz29kr/1>, Accessed : 2024-06-03, 2020.
- POOLSAPPASIT, Nayot ; DEWRI, Rinku et RAY, Indrajit, « Dynamic Security Risk Management Using Bayesian Attack Graphs », in : *IEEE Transactions on Dependable and Secure Computing* (2012), p. 61-74.
- POTES RUIZ, Paula ; KAMSU FOGUEM, Bernard et GRABOT, Bernard, « Generating knowledge in maintenance from Experience Feedback », in : *Knowledge-Based Systems* 68 (2014), p. 4-20.

Predictive Maintenance Implementation Guide Data Sets,

<https://gallery.azure.ai/Experiment/Predictive-Maintenance-Implementation-Guide-Data-Sets-1>, Accessed : 2024-06-03, 2016.

- PRYTZ, Rune ; NOWACZYK, Sławomir ; ROGNVALDSSON, Thorsteinn et BYTTNER, Stefan, « Predicting the need for vehicle compressor repairs using maintenance records and logged vehicle data », in : *Engineering applications of artificial intelligence* (2015).
- PUGNO, Nicola ; CIAVARELLA, M ; CORNETTI, Pietro et CARPINTERI, Alberto, « A generalized Paris' law for fatigue crack growth », in : *Journal of the Mechanics and Physics of Solids* 54.7 (2006), p. 1333-1349.
- QIAN, Yuning ; YAN, Ruqiang et GAO, Robert X., « A multi-time scale approach to remaining useful life prediction in rolling bearing », in : *Mechanical Systems and Signal Processing* 83 (2017), p. 549-567.
- QIN, Taichun ; ZENG, Shengkui et GUO, Jianbin, « Robust prognostics for state of health estimation of lithium-ion batteries based on an improved PSO-SVR model », in : *Microelectronics Reliability* 55.9 (2015), Proceedings of the 26th European Symposium on Reliability of Electron Devices, Failure Physics and Analysis, p. 1280-1284.
- RAJ, Venkat ; BALASUBRAMANIAN, C. et KALIAPPAN, M., « Development of Big Data Predictive Analytics Model for Disease Prediction using Machine learning Technique », in : *Journal of Medical Systems* 43 (2019).
- RAN, Yongyi ; ZHOU, Xin ; LIN†, Pengfeng ; WEN, Yonggang et DENG, Rui long, « A Survey of Predictive Maintenance : Systems and Purposes and Approaches », in : *arXiv : Signal Processing* (2019).
- RASTOGI, Vrinda ; SRIVASTAVA, Sahima ; MISHRA, Manasi et THUKRAL, Rachit, « Predictive Maintenance for SME in Industry 4.0 », in : *2020 Global Smart Industry Conference (GloSIC)*, IEEE, 2020.
- RAZGON, Margarita et MOUSAVI, Alireza, « Relaxed Rule-Based Learning for Automated Predictive Maintenance : Proof of Concept », in : *Algorithms* 13.9 (2020), p. 219.
- REN, Lijuan ; WANG, Tao ; SEKHARI SEKLOULI, Aicha ; ZHANG, Haiqing et BOURAS, Abdelaziz, « A review on missing values for main challenges and methods », in : *Information Systems* 119 (2023), p. 102268.
- RESEARCH, Verified Market, *Predictive Maintenance Market Size And Forecast*, URL : <https://www.verifiedmarketresearch.com/product/predictive-maintenance-market/>.

-
- RESEARCH, Verified Market, *Predictive Maintenance Market Size And Forecast*, 2022, URL : \ \https : // www . verifiedmarketresearch . com / product / predictive - maintenance - market /.
- REYNOLDS, Douglas A et al., « Gaussian mixture models. », in : *Encyclopedia of biometrics* 741.659-663 (2009), p. 3.
- RIBEIRO, Marco Túlio; SINGH, Sameer et GUESTRIN, Carlos, « "Why Should I Trust You?" : Explaining the Predictions of Any Classifier », in : *CoRR* abs/1602.04938 (2016), arXiv : 1602.04938.
- RICHARDSON, Robert R. ; OSBORNE, Michael A. et HOWEY, David A., « Gaussian process regression for forecasting battery state of health », in : *Journal of Power Sources* 357 (2017), p. 209-219.
- RINALDI, Giovanni ; PILLAI, Ajit C ; THIES, Philipp R et JOHANNING, Lars, « Multi-objective optimization of the operation and maintenance assets of an offshore wind farm using genetic algorithms », in : *Wind Engineering* 44.4 (2020), p. 390-409.
- RISSANEN, Jorma, « Stochastic complexity and modeling », in : *The annals of statistics* (1986), p. 1080-1100.
- ROOS, T. ; SILANDER, T. ; KONTKANEN, P. et MYLLYMAKI, P., « Bayesian network structure learning using factorized NML universal models », in : *Proc. ITA '08* (2008).
- RUBANOVA, Yulia ; CHEN, Ricky T. Q. et DUVENAUD, David K, « Latent Ordinary Differential Equations for Irregularly-Sampled Time Series », in : *Advances in Neural Information Processing Systems*, sous la dir. de H. WALLACH et al., t. 32, Curran Associates, Inc., 2019.
- RUIZ-SARMIENTO, Jose-Raul et al., « A Predictive Model for the Maintenance of Industrial Machinery in the Context of Industry 4.0 », in : *Engineering Applications of Artificial Intelligence* (2019).
- RUMELHART, David E ; HINTON, Geoffrey E ; WILLIAMS, Ronald J et al., *Learning internal representations by error propagation*, 1985.
- SAHA, B. et GOEBEL, K., *Battery Data Set*, Accessed : 2024-06-03, 2007, URL : <https://www.nasa.gov/intelligent-systems-division/discovery-and-systems-health/pcoe/pcoe-data-set-repository/>.
- SAHAL, Radhya ; BRESLIN, John G. et ALI, Muhammad Intizar, « Big data and stream processing platforms for Industry 4.0 requirements mapping for a predictive maintenance use case », in : *Journal of Manufacturing Systems* (2020).

-
- SAHAMI, Mehran ; DUMAIS, Susan ; HECKERMAN, David et HORVITZ, Eric, « A Bayesian Approach to Filtering Junk E-Mail », in : *AAAI'98* (1998).
- SAJID, Anam ; ABBAS, Haider et SALEEM, Kashif, « Cloud-assisted IoT-based SCADA systems security : A review of the state of the art and future challenges », in : *Ieee Access* 4 (2016), p. 1375-1384.
- SAMATAS, Gerasimos G. ; MOUMGIAKMAS, Seraphim S. et PAPAKOSTAS, George A., « Predictive Maintenance Bridging Artificial Intelligence and IoT », in : (2021), p. 0413-0419.
- SAMET, Ahmed et al., « KSPMI : A Knowledge-based System for Predictive Maintenance in Industry 4.0 », in : *Robotics and Computer-Integrated Manufacturing* (2022).
- SAN MIGUEL, Maxi et al., « Challenges in complex systems science », in : *The European Physical Journal Special Topics* 214 (2012), p. 245-271.
- SAXENA, Abhinav ; CELAYA, Jose ; SAHA, Bhaskar ; SAHA, Sankalita et GOEBEL, Kai, « Evaluating algorithm performance metrics tailored for prognostics », in : *2009 IEEE Aerospace conference*, 2009, p. 1-13.
- SAXENA, Abhinav ; GOEBEL, Kai ; SIMON, Don et EKLUND, Neil, « Damage propagation modeling for aircraft engine run-to-failure simulation », in : *2008 International Conference on Prognostics and Health Management*, 2008, p. 1-9.
- SCANLON, Patricia ; KAVANAGH, Darren F. et BOLAND, Francis M., « Residual Life Prediction of Rotating Machines Using Acoustic Noise Signals », in : *IEEE Transactions on Instrumentation and Measurement* 62.1 (2013), p. 95-108.
- SCHENKELBERG, Kai ; SEIDENBERG, Ulrich et ANSARI, Fazel, « Analyzing the impact of maintenance on profitability using dynamic bayesian networks », in : *Procedia CIRP* 88 (2020), 13th CIRP Conference on Intelligent Computation in Manufacturing Engineering, 17-19 July 2019, Gulf of Naples, Italy, p. 42-47.
- SEIXAS, Flávio Luiz ; ZADROZNY, Bianca ; LAKS, Jerson ; CONCI, Aura et SAADE, Débora Christina Muchaluat, « A Bayesian network decision model for supporting the diagnosis of dementia, Alzheimer's disease and mild cognitive impairment », in : *Computers in Biology and Medicine* (2014), p. 140-158.
- SELCUK, Sule, « Predictive maintenance, its implementation and latest trends : » in : *Proceedings of the Institution of Mechanical Engineers, Part B : Journal of Engineering Manufacture* (2017).

-
- SERRADILLA, Oscar ; ZUGASTI, Ekhi et ZURUTUZA, Urko, « Deep learning models for predictive maintenance : a survey, comparison, challenges and prospect », in : *Applied Intelligence* (2020).
- SI, Xiao-Sheng ; WANG, Wenbin ; HU, Chang-Hua et ZHOU, Dong-Hua, « Remaining useful life estimation – A review on the statistical data driven approaches », in : *European Journal of Operational Research* 213.1 (2011), p. 1-14.
- SIEMENS, *The True Cost of Downtime 2024*, rapp. tech., Siemens, 2024, URL : https://assets.new.siemens.com/siemens/assets/api/uuid:1b43afb5-2d07-47f7-9eb7-893fe7d0bc59/TCOD-2024_original.pdf.
- SIKORSKA, J.Z. ; HODKIEWICZ, M. et MA, L., « Prognostic modelling options for remaining useful life estimation by industry », in : *Mechanical Systems and Signal Processing* 25.5 (2011), p. 1803-1836.
- SILVESTRIN, Luis ; HOOGENDOORN, Mark et KOOLE, Ger, « A Comparative Study of State-of-the-Art Machine Learning Algorithms for Predictive Maintenance », in : *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2019, p. 760-767.
- SIMÕES, António ; VIEGAS, José Manuel ; FARINHA, José Torres et FONSECA, Inácio, « The state of the art of hidden markov models for predictive maintenance of diesel engines », in : *Quality and Reliability Engineering International* 33.8 (2017), p. 2765-2779.
- SONG, Ya ; SHI, Guo ; CHEN, Leyi ; HUANG, Xinpei et XIA, Tangbin, « Remaining useful life prediction of turbofan engine using hybrid model based on autoencoder and bidirectional long short-term memory », in : *Journal of Shanghai Jiaotong University (Science)* 23.1 (2018), p. 85-94.
- SOONTRONCHAI, W., *IIOT Data of Wind Turbine*, <https://www.kaggle.com/datasets/wasuratme96/iiot-data-of-wind-turbine>, Accessed : 2024-06-03, 2019.
- SOUALHI, Abdenour ; MEDJAHER, Kamal et ZERHOUNI, Noureddine, « Bearing Health Monitoring Based on Hilbert–Huang Transform, Support Vector Machine, and Regression », in : *IEEE Transactions on Instrumentation and Measurement* (2015a).
- « Bearing Health Monitoring Based on Hilbert–Huang Transform, Support Vector Machine, and Regression », in : *IEEE Transactions on Instrumentation and Measurement* 64.1 (2015b), p. 52-62.
- SPIEGELHALTER, D. et LAURITZEN, S., « Sequential updating of conditional probabilities on directed acyclic structures », in : *Networks* (1990), 579–605.

-
- SPIRTES, P. ; GLYMOU, C. et SCHEINES, R., « Causation, Prediction, and Search », in : *Springer-Verlag* (1993).
- SUSTO, Gian Antonio ; BEGHI, Alessandro et DE LUCA, Cristina, « A Predictive Maintenance System for Epitaxy Processes Based on Filtering and Prediction Techniques », in : *IEEE Transactions on Semiconductor Manufacturing* 25.4 (2012), p. 638-649.
- SUSTO, Gian Antonio ; SCHIRRU, Andrea ; PAMPURI, Simone ; MCLOONE, Sean et BEGHI, Alessandro, « Machine Learning for Predictive Maintenance : A Multiple Classifier Approach », in : *IEEE Transactions on Industrial Informatics* (2015).
- SUZUKI, J., « A construction of Bayesian networks from databases based on an MDL principle », in : *Proc. UAI'93* (1993), 266–273.
- SWART, Petrus Daniël et VLOK, Pieter-Jan, « Failure Statistics : Budgeting Preventative Maintenance Activities Using Forecasted Work Orders », in : (2015), p. 321-337.
- TAN, Cher Ming et RAGHAVAN, Nagarajan, « Imperfect predictive maintenance model for multi-state systems with multiple failure modes and element failure dependency », in : *2010 Prognostics and System Health Management Conference*, 2010, p. 1-12.
- THOMAS, Douglas et WEISS, Brian, « Maintenance Costs and Advanced Maintenance Techniques in Manufacturing Machinery : Survey and Analysis », in : *International Journal of Prognostics and Health Management* (2021).
- TIAN, Jinpeng et XIONG, Rui, « Battery State of health estimation with fewer labelled data : a semi-supervised approach », in : *2023 IEEE 6th International Electrical and Energy Conference (CIEEC)*, 2023, p. 2086-2091.
- TOLSTIKOV, Andrei, « Machine Learning Approaches for Failure Type Detection and Predictive Maintenance », mém. de mast., University of Darmstadt, 2015.
- UCAR, Aysegul ; KARAKOSE, Mehmet et KIRIMCA, Necim, « Artificial intelligence for predictive maintenance applications : key components, trustworthiness, and future trends », in : *Applied Sciences* 14.2 (2024), p. 898.
- USCHOLD, Mike et GRUNINGER, Michael, « Ontologies : principles, methods and applications », in : *The Knowledge Engineering Review* 11.2 (1996), 93–136.
- VADIRAJA, Pramod et CHATTHA, Muhammad Ali, « A Survey on Knowledge integration techniques with Artificial Neural Networks for seq-2-seq/time series models », in : *ArXiv* abs/2008.05972 (2020).
- VARGAS, Ricardo Emanuel Vaz et al., « A realistic and public dataset with rare undesirable real events in oil wells », in : *Journal of Petroleum Science and Engineering* 181 (2019), p. 106223.

-
- VASWANI, Ashish et al., « Attention is All you Need », in : *Advances in Neural Information Processing Systems*, sous la dir. d'I. GUYON et al., t. 30, Curran Associates, Inc., 2017.
- VELOSO, Bruno ; GAMA, João ; RIBEIRO, Rita et PEREIRA, Pedro, « The MetroPT dataset for predictive maintenance », in : *Scientific Data* (2022).
- VINGERHOEDS, R.A. ; JANSSENS, P. ; NETTEN, B.D. et AZNAR FERNÁNDEZ-MONTESINOS, M., « Enhancing off-line and on-line condition monitoring and fault diagnosis », in : *Control Engineering Practice* 3.11 (1995), p. 1515-1528.
- VON BIRGELEN, Alexander ; BURATTI, Davide ; MAGER, Jens et NIGGEMANN, Oliver, « Self-Organizing Maps for Anomaly Localization and Predictive Maintenance in Cyber-Physical Production Systems », in : *Procedia CIRP* 72 (2018), p. 480-485.
- WANG, Biao ; LEI, Yaguo ; LI, Naipeng et LI, Ningbo, « A Hybrid Prognostics Approach for Estimating Remaining Useful Life of Rolling Element Bearings », in : *IEEE Transactions on Reliability* (2020).
- WANG, Pingfeng ; YOUN, Byeng D. et HU, Chao, « A generic probabilistic framework for structural health prognostics and uncertainty management », in : *Mechanical Systems and Signal Processing* 28 (2012a), p. 622-637.
- « A generic probabilistic framework for structural health prognostics and uncertainty management », in : *Mechanical Systems and Signal Processing* 28 (2012b), Interdisciplinary and Integration Aspects in Structural Health Monitoring, p. 622-637.
- WANG, Qin ; TAAL, Cees et FINK, Olga, « Integrating Expert Knowledge With Domain Adaptation for Unsupervised Fault Diagnosis », in : *IEEE Transactions on Instrumentation and Measurement* 71 (2022), p. 1-12.
- WANG, Xi et WANG, Chen, « Time series data cleaning with regular and irregular time intervals », in : *arXiv preprint arXiv :2004.08284* (2020).
- WANG, Yali ; LIMMER, Steffen ; OLHOFER, Markus ; EMMERICH, Michael T.M. et BÄCK, Thomas, « Vehicle Fleet Maintenance Scheduling Optimization by Multi-objective Evolutionary Algorithms », in : *2019 IEEE Congress on Evolutionary Computation (CEC)*, 2019, p. 442-449.
- WASSERMAN, Stanley et FAUST, Katherine, *Social Network Analysis : Methods and Applications*, Structural Analysis in the Social Sciences, Cambridge University Press, 1994.
- WELCH, Greg ; BISHOP, Gary et al., « An introduction to the Kalman filter », in : (1995).

-
- WILENSKY, Uri et RESNICK, Mitchel, « Thinking in levels : A dynamic systems approach to making sense of the world », in : *Journal of Science Education and technology* 8 (1999), p. 3-19.
- WUEST, Thorsten ; WEIMER, Daniel ; IRGENS, Christopher et THOBEN, Klaus-Dieter, « Machine learning in manufacturing : advantages, challenges, and applications », in : *Production and Manufacturing Research : An Open Access Journal* 4.1 (2016), p. 23-45.
- XIAO, Fu ; ZHAO, Yang ; WEN, Jin et WANG, Shengwei, « Bayesian network based FDD strategy for variable air volume terminals », in : *Automation in Construction* (2013), p. 106-118.
- XU, Dong et al., « Residual fatigue life prediction of ball bearings based on Paris law and RMS », in : *Chinese Journal of Mechanical Engineering* 25 (2012), 320–327.
- XU, Jiuping et XU, Lei, « Health management based on fusion prognostics for avionics systems », in : *Journal of Systems Engineering and Electronics* 22.3 (2011), p. 428-436.
- XU, Peng ; WEI, Guo ; SONG, Kai et CHEN, Yinsheng, « High-accuracy health prediction of sensor systems using improved relevant vector-machine ensemble regression », in : *Knowledge Based Systems* (2021).
- YAM, Richard C. M. ; TSE, Peter W. ; LI, L. et TU, P. L., « Intelligent Predictive Decision Support System (IPDSS) for Condition-Based Maintenance », in : *The International Journal of Advanced Manufacturing Technology* (2001).
- YOON, Jinsung ; JARRETT, Daniel et SCHAAR, Mihaela van der, « Time-series Generative Adversarial Networks », in : *Advances in Neural Information Processing Systems*, sous la dir. de H. WALLACH et al., t. 32, Curran Associates, Inc., 2019.
- YUAN, Mei ; WU, Yuting et LIN, Li, « Fault diagnosis and remaining useful life estimation of aero engine using LSTM neural network », in : *2016 IEEE International Conference on Aircraft Utility Systems (AUS)*, 2016, p. 135-140.
- ZADEH, L.A., « The role of fuzzy logic in the management of uncertainty in expert systems », in : *Fuzzy Sets and Systems* 11.1 (1983), p. 199-227.
- ZENISEK, Jan ; HOLZINGER, Florian et AFFENZELLER, Michael, « Machine learning based concept drift detection for predictive maintenance », in : *Computers & Industrial Engineering* (2019).
- ZHAI, Simon ; GEHRING, Benedikt et REINHART, Gunther, « Enabling predictive maintenance integrated production scheduling by operation-specific health prognostics with generative deep learning », in : *Journal of Manufacturing Systems* (2021).

-
- ZHANG, Peng et al., « Marine Systems and Equipment Prognostics and Health Management : A Systematic Review from Health Condition Monitoring to Maintenance Strategy », in : *Machines* (2022).
- ZHANG, Weiting ; YANG, Dong et WANG, Hongchao, « Data-Driven Methods for Predictive Maintenance of Industrial Equipment : A Survey », in : *IEEE Systems Journal* (2019).
- ZHAO, Jinlou et YANG, Liqian, « A bi-objective model for vessel emergency maintenance under a condition-based maintenance strategy », in : *SIMULATION* 94.7 (2018), p. 609-624.
- ZHAO, Minghang ; TANG, Baoping et TAN, Qian, « Bearing remaining useful life estimation based on time-frequency representation and supervised dimensionality reduction », in : *Measurement* 86 (2016), p. 41-55.
- ZHAO, Qingyu ; ADELI, Ehsan ; HONNORAT, Nicolas ; LENG, Tuo et POHL, Kilian M., « Variational AutoEncoder for Regression : Application to Brain Aging Analysis », in : *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019 : 22nd International Conference*, Shenzhen, China : Springer-Verlag, 2019a, 823–831, ISBN : 978-3-030-32244-1.
- ZHAO, Rui et al., « Deep learning and its applications to machine health monitoring », in : *Mechanical Systems and Signal Processing* 115 (2019b), p. 213-237.
- ZHAO, Xingang ; WANG, Xinyan et GOLAY, Michael W., « Bayesian Network-Based Fault Diagnostic System for Nuclear Power Plant Assets », in : *Nuclear Technology* (2023), p. 401-418.
- ZHENG, Ce ; MALBASA, Vuk et KEZUNOVIC, Mladen, « Regression tree for stability margin prediction using synchrophasor measurements », in : *IEEE Transactions on Power Systems* 28.2 (2013), p. 1978-1987.
- ZHENG, Shuai ; FARAHAT, Ahmed et GUPTA, Chetan, « Generative Adversarial Networks for Failure Prediction », in : *Machine Learning and Knowledge Discovery in Databases*, sous la dir. d'Ulf BREFELD et al., Cham : Springer International Publishing, 2020, p. 621-637, ISBN : 978-3-030-46133-1.
- ZHOU, Anmei ; YU, Dejie et ZHANG, Wenyi, « A research on intelligent fault diagnosis of wind turbines based on ontology and FMECA », in : *Advanced Engineering Informatics* 29.1 (2015), p. 115-125.

ZHOU, Haoxuan et al., « Construction of health indicators for condition monitoring of rotating machinery : A review of the research », in : *Expert Systems with Applications* 203 (2022), p. 117297.

ZONTA, T. ; DA COSTA, C. A. et DA ROSA RIGHI, R., « Predictive maintenance in the Industry 4.0 : A systematic literature review », in : *Computers & Industrial Engineering* 150 (2020), p. 106889.

ÇINAR, Zeki Murat et al., « Machine Learning in Predictive Maintenance towards Sustainable Smart Manufacturing in Industry 4.0 », in : *Sustainability* (2020).

Titre : DADDY : Détection par Anticipation De DYsfonctionnement dans des systèmes embarqués

Mot clés : Maintenance prévisionnelle, Réseaux Bayésiens Dynamiques, Systèmes Industriels Complexes, État de Santé, Diagnostic, Pronostic

Résumé : Cette thèse propose une méthodologie complète pour la maintenance prévisionnelle, adaptée aux systèmes industriels à données imparfaites. Fondée sur les réseaux bayésiens, cette approche offre un cadre de travail structuré et interprétable pour le diagnostic et la prévision.

Premièrement, une méthode originale est développée pour convertir les données brutes des capteurs en plusieurs niveaux de santé interprétables, comme "stable" ou "alerte". Deuxièmement, ces niveaux de santé permettent de construire un premier modèle de diagnostic qui utilise les réseaux bayésiens pour fonctionner même avec des informations

manquantes. Enfin, ce modèle de diagnostic instantané sert de fondation à un second modèle, dynamique, qui prévoit l'évolution future de l'état de santé.

La méthodologie est validée sur des jeux de données modifiés pour simuler les conditions industrielles réelles. Les résultats démontrent sa robustesse : le modèle de diagnostic conserve un score F1 de 87% malgré des données manquantes, tandis que le modèle prédictif anticipe 71% des états d'alerte, cruciaux pour éviter les pannes. Ce travail fournit une méthode concrète pour valoriser des données difficiles et déployer une maintenance prévisionnelle efficace.

Title: DADDY: Anticipatory Detection of Dysfunction in Embedded Systems

Keywords: Predictive Maintenance, Dynamic Bayesian Networks, Complex Industrial Systems, Health State, Diagnosis, Prognostics

Abstract: This thesis proposes a complete methodology for predictive maintenance, tailored to industrial systems with imperfect data. Based on Bayesian networks, this approach offers a structured and interpretable framework for diagnosis and forecasting.

First, an original method is developed to convert raw sensor data into several interpretable health levels, such as "stable" or "alert". Second, these health levels are used to build an initial diagnostic model that uses Bayesian networks to operate even with missing information. Finally, this snapshot diag-

nostic model serves as the foundation for a second, dynamic model, which predicts the future evolution of the health state.

The methodology is validated on datasets modified to simulate real-world industrial conditions. The results demonstrate its robustness: the diagnostic model maintains an F1-score of 87% despite missing data, while the predictive model anticipates 71% of the alert states that are crucial for preventing failures. This work provides a practical method to leverage difficult data and deploy an effective predictive maintenance system.