SPIM

Thèse de Doctorat

école doctorale **sciences pour l'ingénieur et microtechniques**

UFC UNIVERSITÉ DE FRANCHE-COMTÉ

Estimation du RUL par des approches basées sur l'expérience : de la donnée vers la connaissance

RACHA KHELIF

SPIM

Thèse de Doctorat

UFC

école doctorale **sciences pour l'ingénieur et microtechniques**

UNIVERSITÉ DE FRANCHE-COMTÉ

N° | X | X | X |

THÈSE présentée par

RACHA KHELIF

pour obtenir le

Grade de Docteur de
l'Université de Franche-Comté

Spécialité : **Automatique**

Estimation du RUL par des approches basées sur l'expérience : de la donnée vers la connaissance

Soutenue publiquement le 14 décembre 2015 devant le Jury composé de :

FARHAT FNAIECH	Rapporteur	Professeur, Université de Tunis
MICHEL COMBACAU	Rapporteur	Professeur, Université de Toulouse III Paul Sabatier
HASSANE ALLA	Examinateur	Professeur, Université de Grenoble Alpes
NADA MATTA	Examinateur	Maître de conférences-HDR, Université de Technologies de Troyes
NOUERDDINE ZERHOUNI	Directeur de thèse	Professeur, ENSMM
BRIGITTE CHEBEL-MORELLO	Co-encadrant	Maître de conférences, Université de Franche-Comté
SIMON MALINOWSKI	Co-encadrant	Maître de conférences, Université de Rennes 1

REMERCIEMENTS

Je dois toute ma gratitude à ceux et celles qui ont contribué à la réalisation de ce travail de thèse et qui ont transformé mon parcours de doctorante en une expérience exceptionnelle.

Tout d'abord, je tiens à exprimer ma sincère reconnaissance à Monsieur Noureddine

ZERHOUNI, professeur à l'ENSMM pour d'avoir dirigé ce travail. Je le remercie également pour ses conseils, ses critiques constructives et la confiance qu'il a placée en moi.

Je remercie chaleureusement Monsieur Simon MALINOWSKI, maître de conférences à l'université de Rennes 1 pour son encadrement et sa compétence.

J'exprime toute ma gratitude à Madame Brigitte CHEBEL-MORELLO, maître de conférences à l'université de Franche-Comté pour la confiance qu'elle m'a accordée, sa compétence, son implication et toutes les heures qu'elle a consacrées à l'encadrement de cette thèse. Je tiens à lui exprimer ma sincère gratitude pour sa disponibilité et son respect des délais serrés de relecture. J'ai aussi apprécié ses qualités humaines et sa sincère amitié.

Je remercie mesdames et messieurs les membres du jury de l'intérêt qu'ils ont bien voulu porter à mon travail. J'adresse toute ma reconnaissance aux :

- Professeur Farhat FNAICH et Professeur Michel COMBACAU d'avoir accepté de rapporter ce travail et d'y avoir apporté des précieuses remarques.
- Professeur Hassane ALLA et Docteur Nada MATTA d'avoir accepté d'examiner ce travail de thèse.

Je remercie le FEDER (Fond Européen de Développement Economique et Régional) qui a financé cette thèse dans le cadre d'un projet industriel nommé ALTIDE (Aide à La Traçabilité Intelligente Des Equipements)

Je tiens également à remercier tout le personnel de l'ENSMM, et du département pour l'ambiance chaleureuse qui régnait au département, et pour leur aide qui m'a été précieuse durant la réalisation de ces travaux. Je remercie en particulier mes collègues de l'équipe PHM.

Enfin, je remercie mes parents, mes frères et mon oncle Abdelkrim pour leur soutien affectif et pour leur présence tout au long de mon parcours. J'adresse un remerciement particulier à mon mari qui m'a encouragé et soutenu.

Un grand merci à Nesrine pour sa présence et son aide précieuse.

SOMMAIRE

TABLE DES FIGURES

LISTE DES TABLES

LISTE DES DÉFINITIONS

I

CONTEXTE ET PROBLÉMATIQUES

1

INTRODUCTION

1.1/ CONTEXTE

Avec l'augmentation de la complexité des systèmes industriels, on voit apparaitre de nombreux incidents à l'origine de défaillances causant des dégâts considérables sur les biens, l'environnement et les personnes.
Le 13 Aout 2003, la défaillance d'une turbine à la centrale hydroélectrique "Sayano-Shushenskay" en Russie a causé la mort de 75 personnes. Pourtant l'équipement défectueux avait vibré pendant un temps considérable sans qu'il n'y ait aucune intervention. Suite à l'incident, le réseau électrique local a été endommagé conduisant à une panne générale de courant (c.f. Figure 1.1-a).
Le 06 Juillet 2013, 47 personnes ont été tuées, à cause d'un déraillement d'un train transportant du pétrole au Lac-Mégantic en Québec. Ce déraillement a provoqué une série de violentes explosions. l'incident était dû à plusieurs facteurs , parmi lesquels des problèmes mécaniques de la locomotive de tête. En plus des 47 personnes mortes, une grande partie du centre-ville a été détruite. Le déraillement a causé la destruction de 30 bâtiments et 36 bâtiments ont été démolis par principe de protection contre la contamination. Le centre ville ainsi que la rivière et le lac adjacents ont été contaminés (les wagons de train ont déversé environ 6 millions de litres de pétrole brut, qui s'est enflammé) et près de 2000 personnes ont dû être évacuées (c.f. Figure 1.1-b).

(a). L'accident de la centrale hydroélectrique "Sayano-Shushenskay" a eu lieu [1],

(b). Vue d'hélicoptère de Lac-Mégantic le jour du déraillement [2].

FIGURE 1.1 – Conséquences de défaillances imprévues.

Force est de constater que la fiabilité des équipements a un impact sur la sécurité des biens et des personnes et peut engendrer, lorsque la maintenance est négligée, des incidents et des coût prohibitifs. Les conséquences d'une maintenance négligée ou mal

faite sont très lourdes et occasionnent des catastrophes qui non seulement ont des coûts élevés (dûs à l'arrêt de production, le remplacement des biens, etc.) mais qui causent en plus la perte des personnes. Ces catastrophes peuvent également avoir des conséquences désastreuses sur l'environnement en contaminant la région touchée par le drame. Les autorités sont obligées de lancer des opérations d'évacuation et de nettoyage afin de rétablir au mieux l'environnement, sans toutefois pouvoir, dans certain cas, éliminer la contamination de la région.

Par conséquent, les entreprises doivent disposer d'équipements fiables, bien-entretenus par un système de maintenance performant et bien organisé. Une bonne maintenance permet de prolonger la durée de vie des équipements tout en participant à une meilleure performance globale. Les entreprises cherchent des stratégies d'amélioration de leur système de maintenance. En effet, ces stratégies ont évolué de la maintenance corrective (exécutée après détection de panne) et de la maintenance préventive systématique (exécutée à des intervalles de temps préalables) vers des maintenances intelligentes comme la maintenance basée sur l'état (CBM), ou maintenance prédictive.

La CBM peut être faite durant la surveillance de fonctionnement du bien ou des paramètres significatif de ce fonctionnement pour estimer ou identifier l'état courant et décider des interventions.
La maintenance prédictive, durant le suivi continu du bien, projette dans le futur l'état courant pour prévoir les actions de maintenance à développer avant l'arrivée de la panne.

Nous nous intéressons à la maintenance prédictive et plus particulièrement au PHM (Prognostics and Health Management) qui est un processus de management de l'état de santé d'un équipement et du pronostic de son RUL.

1.2/ CADRE DE TRAVAIL

Nos travaux de thèse se placent dans le cadre du projet ALTIDE (Aide à La Traçabilité Intelligente Des Equipements) qui a comme objectifs de :

1. Réserver l'impact environnemental en augmentant la durée de vie des équipements afin de prolonger leur exploitation, en assurant la sécurité des biens et des personnes en contact avec ces équipements industriels.
2. Garantir la fiabilité des installations par une meilleure gestion de la maintenance, ce qui augmentera la disponibilité des équipements.
3. Développer une démarche durable en assurant le suivi de l'état de santé des équipements pendant leur phase d'exploitation et la traçabilité d'information tout au long du cycle de vie.
4. Assurer la continuité des services offerts par les équipements en anticipant les défaillances grâce aux méthodes de pronostic soutenant les méthodes de suivi de l'état de santé des composants.

Ce projet a permis le déploiement d'un système de traçabilité intelligente rendant disponible à tout moment les données, informations et connaissances relatives à l'équipement, grâce entre autre à une plateforme de e-maintenance distribuée sécurisée.

Dans ce contexte, cette thèse s'est intéressé au quatrième point c'est-à-dire assurer le suivi de l'état de santé d'un équipement et son pronostic de défaillance.

1.2.1/ PROBLÉMATIQUE

Pour atteindre ces objectifs, notre démarche s'est intéressée à identifier l'état de santé courant d'un nouveau composant/équipement et à estimer sa durée de vie résiduelle avant défaillance. Nous avons été amenés à faire le pronostic du temps restant avant défaillance en prenant appui sur un historique d'expériences de suivi d'état de santé couvrant tout le cycle de vie d'un composant critique.

Pour ce faire, nous avons développé des méthodes guidées par les données et plus particulièrement basées sur l'expérience.
Pour représenter et manipuler l'expérience nous avons pris appui sur le paradigme du raisonnement à partir de cas.

Étudier la défaillance et son évolution s'imposent. Pour ce faire, nous avons à disposition des données relatives à des expériences qui seront à la base de la représentation de cette dégradation. Ainsi, les problématiques suivantes se posent :

— Représentation de l'expérience.
— Conception de la phase de remémoration avec l'élaboration de mesures de similarités adaptées au problème.
— Estimation de l'état courant de défaillance et du RUL.

1.3/ DÉMARCHE SUIVIE

Pour solutionner ces problématiques, nous allons explorer plusieurs types de représentation de la défaillance.

Une première piste vise à élaborer des courbes de tendance. Les caractéristiques de cette première méthode sont les suivantes :

1. Les expériences sont modélisées par des courbes dépendant du temps grâce à différentes méthodes.
2. Le suivi en ligne est fait par la modélisation des données capteurs en courbe de tendance et l'évaluation de l'état courant sur cette courbe.
3. Le RUL est estimé à partir de ces courbes.

Une deuxième piste est conduite en exploitant directement les données capteurs, sans les modéliser par des courbes. Le temps restant avant défaillance sera associé à ces données capteurs. Ces données, qui sont des séries temporelles multidimensionnelles, seront traitées par des méthodes supervisées comme les SVR.

La résolution de ces problématiques de recherche est en lien direct avec le choix de la représentation de la défaillance.

Dans un premier lieu, nous allons développer une approche basée sur les instances (IBL) qui nécessitera de :

- **Définir une représentation temporelle de l'expérience par une série temporelle monodimensionnelle** (courbe de tendance) :
  Deux formalisations seront testées : (i) une première qui s'intéressera à l'état de santé du composant. Seule une partie des données capteurs en lien avec le

bon fonctionnement du composant et son état défaillant sera prise en compte. Un indicateur de santé sera construit sous forme de séries temporelles mono-dimensionnelles. (ii) une deuxième qui agrègera toutes les données pour former ainsi des trajectoires de dégradation.

- **Concevoir des mesures de similarité adaptées aux séries mono-dimensionnelles** :
  Deux mesures de similarité pondérées basées sur la distance euclidienne favorisant les données les plus proches de la défaillance (important pour le pronostic) seront testées. La première évoluera sur des fenêtres fixes parallèles alors que la deuxième sera projetée temporellement.
- **Identifier l'état de santé d'un nouveau composant suivi en ligne** :
  Lors du suivi en ligne d'un nouveau composant, ses données seront modélisées. Sur la courbe de tendance (indicateur de santé ou courbe de dégradation), l'état courant sera identifié grâce aux mesures de similarités projetées et le RUL en est déduit.

L'approche IBL développée évoluera ensuite vers une approche de RàpC intégrant de la connaissance, ce qui nous amène à :

- **Extraire de la connaissance à partir des données capteurs** :
  La connaissance extraite est de deux types : temporelle qui complètera la modélisation des indicateurs de santé et fréquentielle qui enrichira la présentation des instances.

Ensuite, nous nous sommes intéressés à une autre approche d'estimation du RUL en travaillant directement sur les données capteurs et en modifiant l' étape de remémoration. L'expérience sera représentée par des séries temporelles mono-dimensionnelles ou multidimensionnelles. Nous nous intéresserons particulièrement à :

- **L'extraction des caractéristiques** :
  Des caractéristiques telles que les coefficients de régression seront extraites à partir d'expériences. Ces caractéristiques déterminées sur une fenêtre donnée seront modélisées par une régression à vecteurs de support afin d'estimer le RUL.

Ces différentes pistes seront détaillées dans les trois chapitres de la partie "Contributions".

1.4/ ORGANISATION DU MÉMOIRE

En plus de l'introduction générale, ce mémoire est articulé en cinq chapitres.

Le chapitre 2 introduit les notions de PHM en insistant sur le pronostic dont on présentera un état de l'art qui décrit les différentes approches, les verrous et hypothèses qui lui sont associés. Enfin, le chapitre positionnera nos travaux par rapport à la littérature du pronostic.

Le chapitre 3 concerne la réalisation d'une approche de pronostic à partir d'instances. Un état de l'art sur l'utilisation de ces approches pour l'estimation du RUL est d'abord présenté. Ensuite, on détaillera chaque étape de l'approche, c'est-à-dire la formalisation des instances, la phase de remémoration et l'estimation du RUL. La formalisation de l'instance est faite de deux manières : une supervisée qui n'utilise que 20% des données capteurs et génère des indicateurs qui sont liés à l'état de santé du compo-

sant et une deuxième non-supervisée qui exploite pleinement l'ensemble des données mais qui génère des signaux qui ne sont pas directement liés à l'état de santé. Ces signaux sont appelés "trajectoires de dégradation". L'objectif de la phase de remémoration est de définir une mesure de similarité adaptée aux séries temporelles qui représentent les instances. La mesure de similarité sert à remémorer les expériences similaires qui seront utilisées afin d'estimer le RUL. L'approche est enfin appliquée sur deux types de composants critiques qui seront utilisés tout au long de ce travail : les turboréacteurs et les batteries. Les résultats obtenus seront évalués et comparés à la fin du chapitre.

Le chapitre 4 constituera un pas vers l'évolution de l'approche à base d'instances vers une approche de raisonnement à partir de cas (RàPC) typique. Cela sera fait par l'injection de connaissance. Ce chapitre définira d'abord les notions de raisonnement à partir de cas, ensuite un état de l'art sur l'utilisation de RàPC pour l'analyse et prédiction des séries temporelles sera présenté. Puis le chapitre décrira la formalisation des cas par l'injection de connaissance. L'ensemble des données sert à extraire des règles de connaissance temporelle qui complètent la modélisation des indicateurs de santé. Un deuxième type de connaissance, cette fois fréquentielle est utilisé afin d'enrichir la formalisation des cas et affiner la phase de remémoration. Enfin les effets de l'injection de cette connaissance seront étudiés en appliquant l'approche sur le même type de données utilisé précédemment.

Le chapitre 5 explorera une autre démarche qui tout comme les approches précédentes ne nécessite pas la définition d'un seuil de défaillance. Le RUL sera directement lié aux expériences en utilisant la régression à vecteurs de support (SVR). Les expériences seront soit mono-dimensionnelles (indicateurs de santé), soit multidimensionnelles (historique de dégradation). À partir de cette nouvelle représentation de l'expérience, des caractéristiques seront extraites et directement liées au RUL. Le chapitre présentera d'abord un état de l'art sur la régression à vecteurs de support et décrira ensuite l'approche proposée, c'est-à-dire l'extraction des caractéristiques, la phase d'apprentissage et l'estimation du RUL. La représentation multidimensionnelle de l'expérience exige une sélection de variable enveloppe qui sera présentée juste après la description de l'approche. Enfin, la méthode sera appliquée sur les mêmes jeux de données.

Enfin, le chapitre 6 concluera le travail de recherche développé dans cette thèse et discute des perspectives et des travaux futurs envisagés.

2

PRONOSTIC DE DÉFAILLANCE

2.1/ INTRODUCTION

La sûreté des biens et des personnes est un des enjeux majeurs en entreprise. Afin de maintenir les équipements en condition opérationnelle et d'assurer leur fiabilité tout en garantissant leur haute performance, les stratégies de maintenance ont évolué du correctif « fix-it-when-it-breaks» vers le prévisionnel «predict-prevent », et plus spécifiquement vers le **Pronostic et Health Management** (PHM). Nos travaux de thèse s'inscrivent pleinement dans la lignée du PHM qui est une discipline qui a débuté il y a une dizaine d'années et qui est en pleine effervescence.

L'objectif de ce chapitre est de définir ce qu'est le PHM, les notions qui lui sont associées et l'architecture permettant de mettre en place la stratégie associée à ce concept. La section 2.2 décrit les différents modules constituant cette architecture et permettra ainsi de positionner nos travaux de recherche dans ces modules. Les deux premiers, acquisition et traitement de données seront réalisés à l'aide d'outils et de méthodes existants. Par contre, trois modules susciteront notre intérêt : l'évolution de l'état de santé et le diagnostic dont on dressera un rapide inventaire à la section 2.3 et le module de pronostic qui fera l'objet d'une étude approfondie à la section 2.4. Il existe un certain nombre de travaux dans le domaine proposant différentes classifications. Nous en privilégierons une, celle qui fait consensus dans la littérature, pour établir un état de l'art sur les travaux relatifs au pronostic. Les problèmes de pronostic intéressent aussi bien le domaine académique que le domaine industriel et sont résolus à partir d'hypothèses de travail. Nous recenserons ces hypothèses en section 2.5 . Nous recenseros ensuite les défis que la communauté scientifique doit relever dans ce domaine. Enfin, nous positionnerons nos travaux par rapport à ces hypothèses de travail et ces défis scientifiques.

2.2/ PROGNOSTICS AND HEALTH MANAGEMENT (PHM)

2.2.1/ ARCHITECTURE DE PHM

Le PHM est une discipline en pleine effervescence qui s'intéresse à étudier les mécanismes de défaillance de systèmes réels dans le but de mieux gérer l'utilisation d'informations sur les conditions d'exploitation des équipements. Le PHM aborde les principales tâches liées à la détection de la dégradation, le diagnostic de pannes et la prévision et gestion pro-active des défaillances [Zio, 2012].

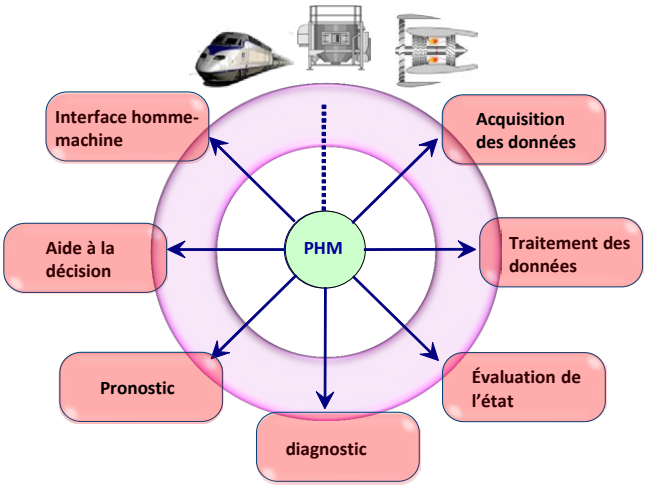


FIGURE 2.1 – Les modules de PHM.

Le processus de PHM est composé de sept modules qui sont illustrés à la figure 2.1

1. **L'acquisition des données :** consiste à mesurer des grandeurs physiques telles que la tension, le courant, la température, etc. à l'aide de capteurs, logiciels et observations humaines. Ces données sont obtenues grâce à un système d'acquisition qui collecte, pré-traite les données à envoyer aux autres modules et à stocker dans une base de données fiable et sûre.
2. **Le traitement des signaux/données :** analyse et interprète les signaux afin d'extraire des informations caractérisant le comportement du système, soit dans le domaine temporel et/ou fréquentiel.
3. **L'évaluation de l'état courant :** sera obtenue à partir de ces caractéristiques et permettra à l'aide de comportement nominal de détecter les différentes anomalies possibles.
4. **Le diagnostic :** correspond à la localisation et l'identification des causes des anomalies ou défaillances.
5. **Le pronostic :** s'appuie sur l'état actuel du système et le résultat de la détection et/ou du diagnostic pour prédire la durée de vie avant défaillance.
6. **L'aide à la décision :** concernant les stratégies de maintenance à mettre en œuvre pour maintenir en bon état le système. Ce module prend appui sur toutes les informations obtenues (état actuel du système, RUL, connaissance du contexte, etc.).
7. **L'interface homme-machine :** offre un moyen de présenter et stocker sous différentes formes les informations utiles.

2.2.2/ NOS TRAVAUX DE RECHERCHE

Nos travaux de recherche s'intègrent dans cette architecture. Nous serons amenés à traiter des données de surveillance issues de capteurs, que l'on considère comme des observations de l'état de santé du composant critique surveillé. Les données recueillies brutes ne sont pas directement exploitables et nécessitent d'être traitées auparavant.

1. Nous appliquerons des méthodes de traitement de signal, afin d'extraire à partir des signaux bruts des caractéristiques pertinentes pour le pronostic. Ces caractéristiques (*features*) peuvent être temporelles, fréquentielles, quand on a des signaux stationnaires, ou spatio-temporelles comme par exemple les transformées en ondelettes dans le cas de non-stationnarité. Dans nos travaux, nous utiliserons des méthodes temporelles et fréquentielles pour l'extraction de caractéristiques.
2. Nous développerons deux types de méthodes de pronostic. La première se base sur des caractéristiques extraites des signaux que l'on modélisera par des indicateurs qui donneront la tendance d'évolution de l'état de santé de l'élément surveillé. La deuxième se base directement sur les données capteurs.
3. Nous exploiterons les données fréquentielles et temporelles dans nos méthodes de pronostic.

Les modules de collecte de données et d'extraction de caractéristiques par des méthodes de traitement de signal seront utilisés dans nos travaux mais ne présenteront pas d'apport scientifique. Par contre, nous nous intéresserons à l'évaluation de l'état de santé qui peut être faite grâce aux trois modules suivants (évaluation de l'état courant, diagnostic et pronostic) et nous nous intéressons plus spécifiquement au module de pronostic dans lequel porteront nos apports. Avant de décrire nos travaux, nous présenterons un rapide état de l'art sur les méthodes d'évaluation de l'état de santé et sur le diagnostic de défaillance. Nous dédierons ensuite une section plus conséquente aux méthodes de pronostic de défaillance, notre domaine de recherche.

2.3/ ÉVALUATION DE L'ÉTAT DE SANTÉ ET DIAGNOSTIC DE DÉFAILLANCE

2.3.1/ ÉVALUATION DE L'ÉTAT DE SANTÉ

FIGURE 2.2 – Évaluation de l'état de santé.

Les pannes inattendues doivent être évitées car elles entraînent des arrêts intempestifs de la production ce qui occasionne des pertes financières non négligeables. Pour pallier à cette éventualité, nous devons étudier le mécanisme de défaillance qui hors cas de défaillance catalytique implique une évolution de la dégradation. Cette évolution s'observe par l'apparition de symptômes spécifiques à des états de santé dégradés de gravité de plus en plus importante. Afin d'enrayer ce processus de dégradation il est nécessaire d'évaluer avec précision l'état actuel de la dégradation.

L'évaluation de l'état de santé d'un système est appréhendé dans la littérature scientifique, comme un problème de classification de données capteurs. Les différents états correspondent chacun à un mode de fonctionnement de l'équipement (comme mode Normal, dégradé de type 1, dégradé de type 2 jusqu'au mode défaillant). Dans ce cadre-là, différents algorithmes de classification ont été appliqués. Nous décrivons ici les principales méthodes de l'état de l'art.

Les modèles de Markov cachés (Hidden Markov Models : HMM) sont des méthodes probabilistes connues dans les applications de surveillance grâce à leur capacité à identifier les états cachés du système. [Camci et al., 2006] proposent d'utiliser des modèles de Markov cachés hiérarchiques (hierarchical HMM : HHMM) pour estimer l'état de santé actuel des outils de forage utilisés dans le domaine du pétrole et des forêts des machine CNC [1] . [Dong et al., 2006] ont inclus la densité relative à la durée de l'état dans leurs travaux en se basant sur les modèles semi-markovien cachés (Hidden Semi Markov Models : HSMM). [Miao et al., 2007] proposent une modélisation HMM adaptative pour mettre à jour le modèle de manière à améliorer les performances de reconnaissance au cours du temps .

[Kim et al., 2012, Selak et al., 2014], ont exploité les propriétés des SVM (Support vector machines) pour estimer l'état de santé d'un équipement. [Wang et al., 2014] ont utilisé les SVM pour réaliser une classification en multi-catégories de l'usure des outils. [Liu et al., 2013] combinent les SVM avec les ondelettes afin de construire une nouvelle fonction noyau à base d'ondelettes de Morlet. L'approche développée a été utilisée pour classer les états de santé des roulements. [Zhang et al., 2015] ont utilisé les SVM pour la classification des états de machines rotatives. Les algorithmes de colonies de fourmis ont été intégrés pour l'optimisation des paramètres SVM et la sélection de variables.

Les réseaux de neurones ont aussi été utilisés pour déterminer l'état de santé courant. [Yu, 2011] a utilisé les cartes auto-adaptatives afin d'évaluer l'état de santé des roulements. Deux états ont été identifiés ; l'état normal et l'état défaillant. Dans [Moosavi et al., 2015], les auteurs ont utilisé un réseau de neurones afin de classer les différents niveaux de court-circuit. [Yeo et al., 2003] ont utilisé un système d'inférence flou basé sur les réseaux de neurones adaptatifs.

2.3.2/ DIAGNOSTIC DES DÉFAILLANCES

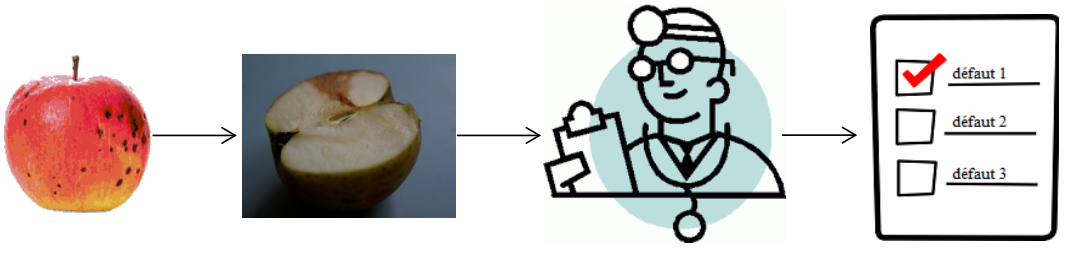


FIGURE 2.3 – Diagnostic de défaillance.

La norme européenne NF EN 13306 définit le diagnostic comme des actions visant à détecter les pannes, les localiser et identifier leurs causes. Les algorithmes

1. CNC : Une machine-outil à commande numérique, Computer Numerical Control en anglais.

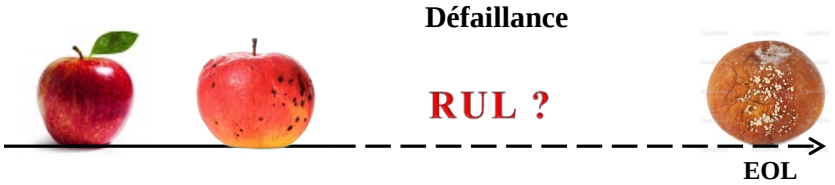
de diagnostic doivent être capables de détecter la performance du système, ses niveaux de dégradation, et ses défaillances (pannes). Cette détection est faite en fonction de changement de propriétés physiques par l'observation de phénomènes détectables [Vachtsevanos et al., 2006]. [Venkatasubramanian et al., 2003c, Venkatasubramanian et al., 2003a, Venkatasubramanian et al., 2003b] classent les techniques de diagnostic en trois catégories : les méthodes basées sur des modèles des procédés physiques, qu'ils soient quantitatifs (à base d'observateurs d'état d'estimation paramétrique de redondance analytique) ou qualitatifs (basées sur l'analyse fonctionnelle du système ou sur des modèles causaux) et les méthodes basées sur les données et plus particulièrement sur un historique de pannes. Bien que les approches guidées par des modèles puissent être plus efficaces que les approches sans modèles, il n'est pas toujours possible de modéliser un système complexe ([Jardine et al., 2006]). Dans de telles situations, l'approche guidée par les données est choisie.

[Subrahmanya et al., 2013] ont utilisé des réseaux de neurones dynamiques afin de diagnostiquer des systèmes non linéaires dynamiques avec des mesures partielles de l'état. [Mahmoud et al., 2013] ont développé une méthode basée sur un filtre de Kalman pour détecter et isoler les défaillances dans des systèmes électrohydrauliques. Dans [Wei et al., 2009], une banque de filtre de Kalman est utilisée afin de détecter et isoler les défaillances qui se produisent dans des moteurs d'avions. [Ricci et al., 2011] ont développé une méthode de diagnostic basée sur la décomposition en modes empiriques et la transformée de Hilbert. Les fonctions intrinsèques utilisées comme entrées au spectre sont automatiquement sélectionnées en créant des indices de mérites. [Georgoulas et al., 2015] ont proposé une approche de détection de pannes basée sur une représentation symbolique d'historique de vibration ou chaque symbole est lié à l'occurrence d'un type de panne. Dans [Hurdle et al., 2009, Bartlett et al., 2009], des techniques d'arbres de défaillances sont utilisées pour des problèmes de diagnostic.

Le diagnostic vise à identifier les problèmes une fois survenus, contrairement au pronostic qui a l'objectif de les anticiper. Les prochaines sections sont entièrement dédiées au pronostic, le cœur de notre travail.

2.4/ PRONOSTIC DE DÉFAILLANCE

2.4.1/ DÉFINITION



RUL: Date limite de consommation

FIGURE 2.4 – Pronostic de défaillance.

[Lee et al., 2014] aborde le pronostic en tentant de répondre aux 4 questions suivantes : (i) Comment la machine fonctionne-t-elle ? (ii) Quand tomberait- elle en panne ?

(iii) Quelle sera la faute initiale provoquant la panne ? (iv) Pourquoi la panne se produit ?

D'après la norme ISO 13381-1 [ISO, 2004], le pronostic de défaillances correspond à l'estimation de la durée de fonctionnement avant défaillance et du risque d'existence ou d'apparition ultérieure d'un ou de plusieurs modes de défaillance. Cette durée de fonctionnement avant défaillance est communément appelée RUL (Remaining Useful Life).

[Jardine et al., 2006] décomposent le pronostic en deux phases principales ; la première prédit le temps restant avant qu'une défaillance survienne en fonction de l'état actuel de la machine et de son profil de fonctionnement et le second prévoit la possibilité qu'une machine fonctionne sans faute ou échec jusqu'à une date ultérieure. Cependant, les travaux sur le pronostic couvrent principalement la première phase.

La norme ISO 13381-1 [ISO, 2004] d'autre part définit trois niveaux de pronostics :

1. Pronostic de mode de défaillance existant. (Niveau 1)
2. Pronostic de future mode de défaillance. (Niveau 2)
3. Pronostic de poste action. (Niveau 3)

— **Le pronostic de niveau 1** fournit une estimation de la durée de vie utile restante des composants / systèmes en se basant sur les progressions de chaque mode de défaillance diagnostiqué.
— **Le pronostic de niveau 2** évalue les effets possibles du mode de défaillance identifié sur les autres modes et modélise la progression possible de chaque dégradation potentielle afin d'estimer le pire des cas pour les composants / systèmes affectés. Par conséquent, les modèles décrivant les interactions des modes de défaillance doivent être développés à ce niveau.
— **Le pronostic de niveau 3** évalue l'impact des actions de maintenance sur ces modèles mentionnés ci-dessus [Sikorska et al., 2011].

Notre travail concerne le niveau 1 du pronostic, dont l'objectif est d'estimer correctement la durée de fonctionnement avant défaillance également appelée durée de vie résiduelle (le RUL).

2.4.2/ CLASSIFICATION DES APPROCHES DE PRONOSTIC

Beaucoup de travaux en pronostic existent et ont fait l'objet de maintes classifications. Il en existe au moins 8 et selon les critères considérés, les classes obtenues diffèrent d'un auteur à un autre. La première classification assez fréquemment utilisée, sous forme pyramidale a été proposée par [Lebold et al., 2001] où l'on distingue trois approches principales : le pronostic basé sur les modèles physiques, le pronostic piloté par les données, et le pronostic basé sur l'expérience. L'approche basée sur l'expérience aussi appelée approche basée sur la fiabilité utilise les données recueillies à partir du retour d'expérience pour estimer les paramètres des lois de fiabilité. Ces approches tendent à être des approches guidées par les données, ce qui explique pourquoi aujourd'hui le terme «les approches basées sur la fiabilité» est en train de disparaître.

Cette classification a évolué (comme l'a défini [Heng et al., 2009]) vers deux types d'approches, que l'on retrouve invariablement dans les différentes classifications : les approches guidées par les données et celles basées sur les modèles physiques. A ces approches [Peng et al., 2010] ajoutent les approches basées sur la connaissance et les approches hybrides (entre données et physique), tandis que [Byington et al., 2002,

Heng et al., 2009] ajoutent à ces deux classes les approches basées sur l'expérience.

Par contre la norme ISO 13381-1 :2004 [ISO, 2004] a proposé une classification différente en considérant d'autres critères : modélisation mathématique de la dégradation, exploitation de la connaissance et risque de détérioration basée sur l'espérance de vie :

— La norme considère les approches basées sur des modèles physiques comme une sous-catégorie de la classe des modèles mathématiques puisque les lois physiques sont modélisées mathématiquement.
— L'approche est dite basée sur la connaissance quand la connaissance experte, sous forme de règles de décision floues ou non, est exploitée dans la méthode de pronostic.
— L'approche est considérée comme une approche basée sur l'espérance de vie quand la durée de vie utile est déterminée par rapport au risque de détérioration attendue dans des conditions d'exploitation connues.

[Sikorska et al., 2011] dans sa classification remplacent la classe des approches guidée par les données par deux classes : les réseaux de neurones (RN) et les modèles d'espérance de vie (tels que les approches stochastiques, statistiques, bayésiennes, à base de modèles Markov cachés . . .). La classe RN correspond à la description de la classe des approches guidées par les données. Cependant, le critère de classification se base sur la construction du modèle prédictif et non pas sur le type d'élément surveillé. Dans ces conditions, la catégorie des approches guidées par les données peut être composée de ces deux types d'approches. Le tableau 2.1 présente un récapitulatif des différentes classifications.

TABLE 2.1 – Classification des méthodes de pronostic.

Auteur année	Classification	Sous-classes
[ISO, 2004]	Mathématique des modèles d'utilisation de vie.	Comportementale (physique de défaillance). statistique, probabiliste,RNA.
	Modèles d'utilisation de vie.	Les modèles fiabiliste, Les modèles de détérioration.
	Les modèles basés sur la connaissance.	Symptôme à base de règles / modèle de défauts, modèle de l'arbre des causes, RàPC.
[Jardine et al., 2006]	Approches statistiques, IA, pronostic basé sur les modèles physiques.	
[Vachtsevanos et al., 2006]	pronostic basé sur les modèles physiques. Les modèles guidés par les données. Les modèles basés sur l'expérience.	
[Heng et al., 2009]	Les modèles guidés par les données	modèles de projection simples (Lissage exponentiel), modèles autorégressifs, RNA, HMM, le filtrage particulaire, Techniques bayésiennes.
	pronostic basé sur les modèles physiques.	
[Peng et al., 2010]	Les modèles basés sur la connaissance	Les systèmes experts, Logique floue.
	Les modèles guidés par les données.	RNA,Techniques bayésiennes, HMM, taux de risque et le taux de risque proportionnel.
	Pronostic basé sur les modèles physiques.	
	Modèles hybrids.	
	Les modèles basés sur la connaissance.	Les systèmes experts, logique floue.
[Sikorska et al., 2011]	Modèles d'utilisation de vie.	Les modèles stochastiques, les modèles statistiques.
	RNA.	
	Pronostic basé sur les modèles physiques.	
[Lee et al., 2014]	Pronostic basé sur les modèles physiques. Les modèles guidés par les données. Modèles hybrides.	

2.4.3/ CLASSIFICATION DES APPROCHES DE PRONOSTIC SUIVI

Afin de proposer une classification la plus en adéquation avec nos travaux et ceux de la littérature scientifique, nous avons fait une cartographie des différents éléments composant les classifications étudiées au tableau 2.1 . Cette cartographoe est représentée à la figure 2.5.

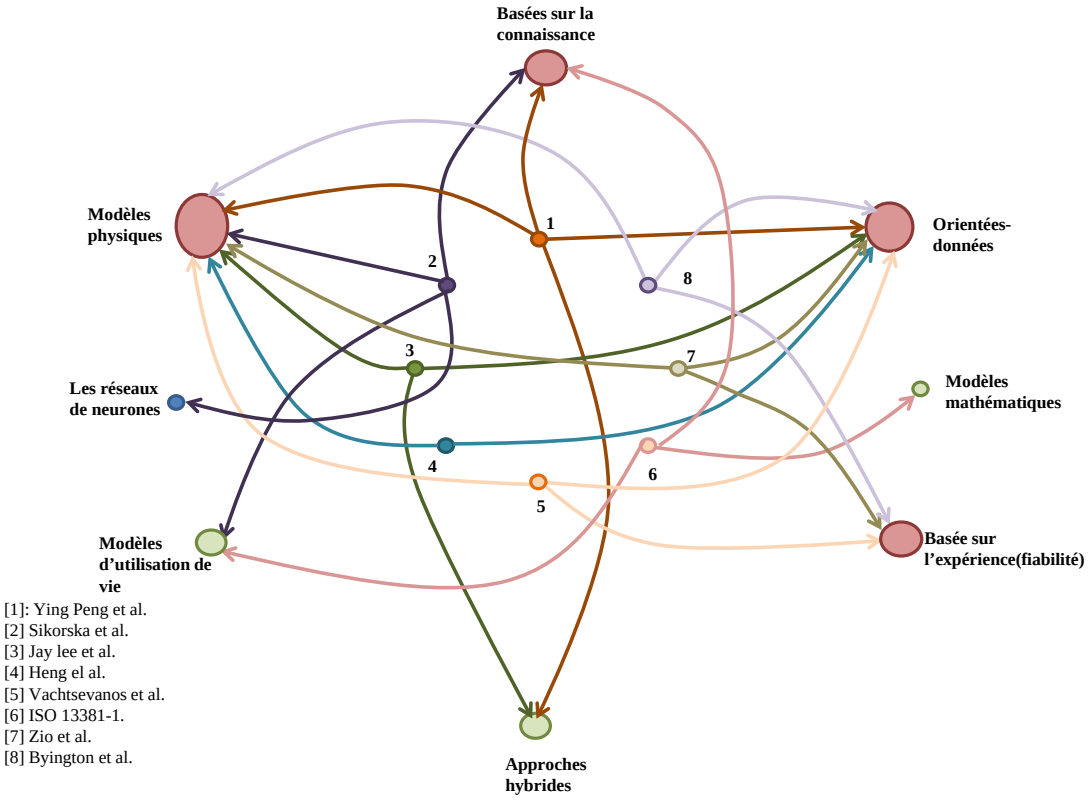


FIGURE 2.5 – Cartographie des classes de pronostic utilisées.

Les cercles représentant les classes sont d'autant plus grands que les classes sont citées. Nous retiendrons de cette cartographie les classes les plus représentées : modèles physiques, guidées par les données et basées sur la connaissance. Elles se différencient par rapport au type d'information sur lequel les modèles sont construits :

- À partir d'équation formalisant la loi physique de dégradation du composant.
- À partir de données capteurs sur le composant tout au long de son cycle de vie.
- À partir de la connaissance experte sur le comportement du composant et sa fatigue.

Les classes basées sur les modèles physiques et guidées par les données sont largement reconnues. En effet, la plupart des auteurs qui travaillent dans le domaine de pronostic aujourd'hui se mettent dans l'une de ces deux catégories [Li et al., 2000, Oppenheimer et al., 2002, Cadini et al., 2009, Myötyri et al., 2006, Peng et al., 2012, Fink et al., 2014, Maio et al., 2012, Xue et al., 2008]. Par contre, la catégorie des approches basées sur la connaissance soulève certains doutes. Les problèmes liés à la considération de cette catégorie en tant que classe sont :

- Peu de travaux déclarés.
- Les travaux déclarés traitent habituellement à la fois du diagnostic et du pronostic. La connaissance est utilisée pour traiter le problème de diagnostic. En

se basant sur les résultats du diagnostic, le pronostic est réalisé en utilisant d'autres approches généralement guidée par les données ([Biagetti et al., 2004, Choi et al., 1995, Butler, 1996]). Les exemples où ces approches ont été utilisées comme un outil principal pour la prédiction du RUL sont rares [Sikorska et al., 2011].

— Lors de l'utilisation de la connaissance pour la prédiction du RUL, cette connaissance est dérivée et modélisée à partir des données puisque on n'a pas réellement de connaissance experte sur la dégradation, ce qui nous permet de les associer à des approches guidées par les données. En effet ces approches ne sont pas assez matures pour être considérées comme une catégorie à part entière, il est plus approprié de les considérer comme une sous-catégorie des approches guidées par les données.

La classification la plus consensuelle dans la communauté PHM est de considérer les trois approches suivantes : les méthodes guidées par les données, les méthodes basées sur des modèles physiques et les méthodes hybrides, comme le montre la figure 2.6. Toutefois, on définira dans l'approche orientée données, 3 types de démarches : une statisticienne, une liée aux outils d'intelligence artificielle et plus spécifiquement à l'apprentissage automatique et une orientée connaissance. Cette dernière pourrait se fondre dans la démarche IA vu le nombre de travaux restreints dans le domaine, mais que l'on dissociera, car nous proposons dans nos travaux de développer une partie connaissance, pour aider à la résolution du problème de pronostic.
L'état de l'art des approches de pronostic (présenté dans la prochaine section) suivra cette classification.

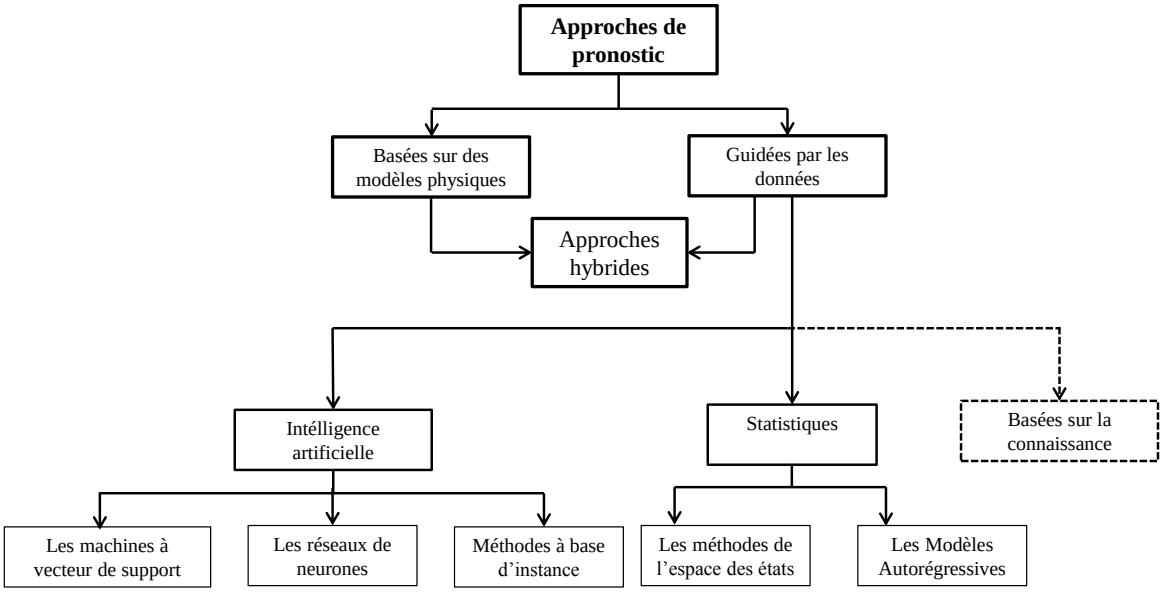


FIGURE 2.6 – Classification des approches de pronostic suivi.

2.4.3.1/ LES APPROCHES BASÉES SUR LES MODÈLES PHYSIQUES

Les approches basées sur des modèles physiques ou basées sur la physique de défaillance construisent des modèles analytiques qui sont directement liés aux processus physiques influençant la santé des composants. Ces approches nécessitent des

connaissances sur les lois physiques (mécanique, chimie, électricité, hydraulique etc). Le

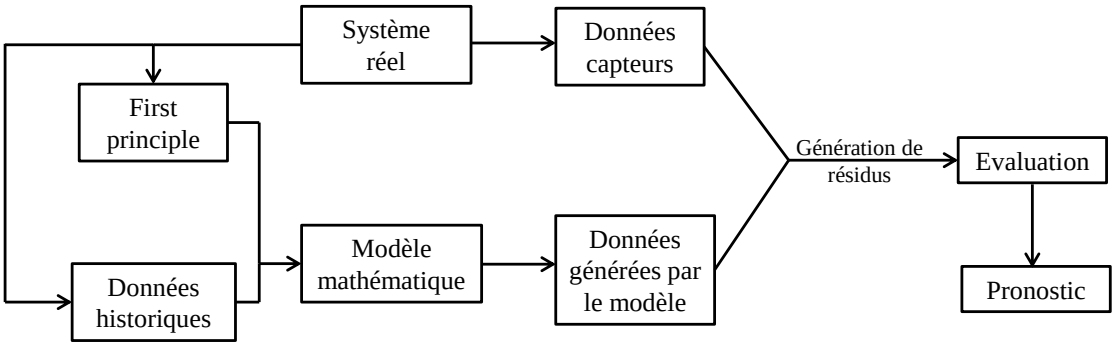


FIGURE 2.7 – Organigramme des approches basées sur des modèles physiques [Peng et al., 2010].

modèle physique peut être décrit par des systèmes dynamiques tels que les équations non-linéaires, les équations différentielles, la représentation d'état, etc.

Le principe du pronostic basé sur un modèle physique est résumé à la figure 2.7. Le comportement du système est représenté par un modèle analytique. Ensuite, un test de cohérence est effectué en comparant les données capteurs issues du système réel et les sorties du modèle analytique. Les résidus générés sont évalués. En présence d'un dysfonctionnement, les résidus dépassent un seuil de détection de défauts [Peng et al., 2010].

L'estimation du temps restant avant défaillance est basée sur la projection de comportement du système et de sa dégradation dans le futur. Les modèles physiques traitent par exemple des problèmes d'usures de matériau, de croissance de fissure, de cassure par fatigue et corrosion.

[Li et al., 2000] ont utilisé la loi de Paris [2] afin de prévoir le taux de croissance de défaut sur un roulement en développant un modèle de propagation de défaut déterministe qui lie le taux de croissance de défaut à la taille de la zone de défaut instantané et les constantes du matériels. D'autre part, [Li et al., 2005] ont modélisé la propagation des fissures de pignon droit en utilisant la loi de Paris et l'analyse par éléments finis (FEA) [3]. Le modèle FEA a été intégré pour calculer les champs de contraintes et de déformation en fonction de la charge des dents de pignon, la géométrie et les propriétés des matériaux. [Oppenheimer et al., 2002] ont modélisé la croissance de fissure d'un arbre du rotor en utilisant la loi de Forman [4] des fractures mécaniques linéaires.

2.4.3.2/ LES APPROCHES GUIDÉES PAR LES DONNÉES

Les approches guidées par les données cherchent à extraire à partir d'un historique de données de surveillance des modèles d'évolution du fonctionnement du système surveillé allant jusqu'à sa dégradation.

2. La loi de Paris est le modèle de croissance de fissures le plus populaire en mécanique de la rupture. Elle relie la gamme de facteurs d'intensité de contraintes à la fissure sous-critique
3. Une méthode numérique utilisée pour résoudre les équations différentielles
4. La loi de Forman est une variation de la loi de Paris qui prend en compte l'effet de différents paramètres comme la contrainte moyenne.

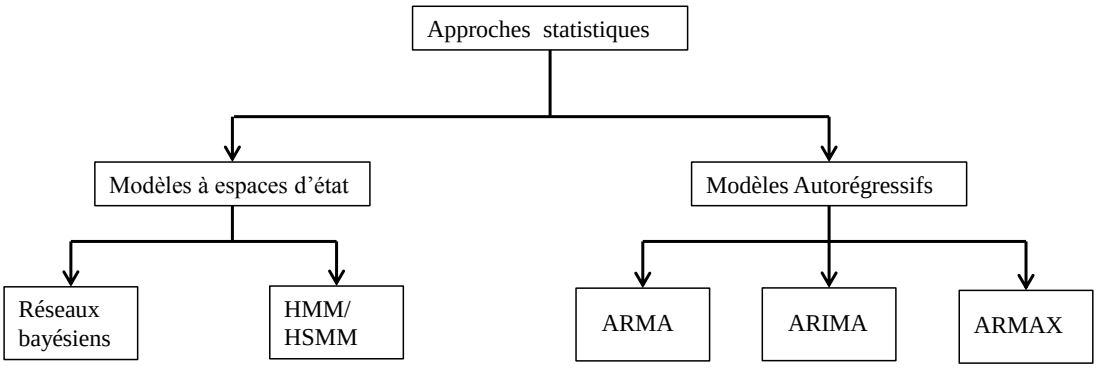


FIGURE 2.8 – La classification des approches statistiques.

Ces approches comportent deux phases. Une première hors ligne est dédiée à la compréhension et l'apprentissage du comportement de la dégradation. Puis, une deuxième phase en ligne estime l'état de santé courant du système et prédit sa durée de fonctionnement avant défaillance. Elles peuvent être classées en deux catégories : les approches statistiques et issues de l'intelligence artificielle.

1. **Les statistiques multidimensionnelles :** Nous passons en revue les méthodes les plus largement utilisées en pronostic, qui sont également illustrées à la figure 2.8.

   a.Les Modèles Auto-régressifs : Les modèles ARMA (modèles auto-régressifs et moyenne mobile), ARIMA (modèles moyenne mobile auto-régressif intégré) et ARMAX sont les méthodes de référence dans la modélisation des séries temporelles [Lewis et al., 1994, Tsay, 2000, Box et al., 2011]. Les modèles ARMA et ARMAX ne sont utilisés que pour des données stationnaires. Ils sont améliorés par l'application d'une opération d'intégration dans le modèle ARIMA. Les modèles ARMA ont été utilisés dans [Galati et al., 2006] afin d'ajuster les données de vibration des roulements des boites de transmission des hélicoptères, dans [Sinha, 2002] pour prédire les tendances des turbines à flux, et dans [Yan et al., 2004] afin d'estimer la durée de vie utile restante avant la défaillance d'un système de mouvement de porte d'ascenseur. [Wu et al., 2007] ont utilisé les modèles ARIMA pour prédire les caractéristiques vibratoires des machines rotatives.

   Les techniques ARMA ont également été intégrées dans le logiciel de pronostic et fusion de données « WatchdogagentTM » développé par le centre NFS, le centre de la maintenance intelligente des systèmes comme décrit par [Lee et al., 2006].

   Les modèles régressifs ne nécessitent pas un historique de dégradation et doivent être évalués de manière récursive jusqu'à atteindre un certain seuil, ce qui engendre une accumulation d'erreur systématique et détériore la performance du prédicteur.

   b. Les modèles à espace d'état

   i.Les modèles de Markov cachés : ou HMM ont été appliqués en pronostic [Rabiner et al., 1986, Baruah et al., 2005, Zhang et al., 2005] mais ne sont pas adaptés à représenter une structure temporelle. Ils ont été remplacés par les HSMM qui comptent une composante temporelle dans

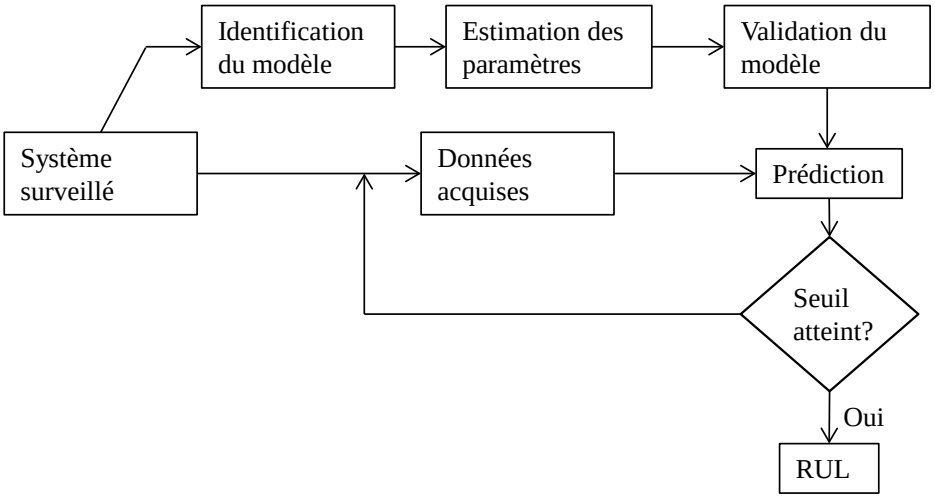


FIGURE 2.9 – Organigramme des approches auto-régressives pour la prédiction du RUL.

la structure de HMM [Wu et al., 2009, He et al., 2006, Dong et al., 2007, Bechhoefer et al., 2006]. [Wu et al., 2009], ont développé une approche basée sur les HSMM afin de prédire le RUL des actionneurs entraînés (motorisés) par un moteur asynchrone. [Liu et al., 2012], ont proposé de combiner les HSMM avec une méthode Monte Carlo séquentielle. Ces outils calculent les probabilités de transition entre les états de santé et leur durée alors que la méthode Monte Carlo définit la relation probabiliste entre les états de santé et les observations de l'équipement surveillé. Les HSMM ont aussi été utilisés pour le pronostic des défaillances d'hélicoptères [He et al., 2006]. [Dong et al., 2006] ont proposé un modèle HSMM pour le diagnostic et pronostic de UH-60 Blackhawk (un hélicoptère utilitaire moyen de l'armée américaine). Le HSMM formé peut être utilisé pour diagnostiquer et estimer l'état courant de santé de l'équipement, et sera à la base de l'estimation du RUL à l'aide d'une équation récursive composée de la durée dans les états et les paramètres du modèle estimé. [Bechhoefer et al., 2006] ont étudié l'applicabilité des HSMM pour prédire le RUL des arbres à cames utilisés dans les hélicoptères.

ii.Les réseaux bayésiens : sont des modèles graphiques directs, ils sont issus de la fusion de la théorie des graphes et des probabilités [5]. Le modèle RB a été étendu à un modèle réseau bayésien dynamique appliqué au pronostic grâce à la modélisation des changements de système au fil du temps. Il est en mesure de surveiller, mettre à jour et prédire les états futurs du système étudié.

[Ferreiro et al., 2012] ont montré l'utilité des réseaux bayésiens pour pronostiquer l'usure des freins des avions. Dans [Zaidan et al., 2015], une approche bayésienne hiérarchique a été proposée afin de construire un modèle probabiliste d'estimation du RUL. [Gebraeel et al., 2005], d'autre part, ont appliqué une approche bayésienne dans le but de mettre à

5. Un réseau bayésien est un graphe acyclique qui est constitué de nœuds, représentants les variables aléatoires, et d'arcs représentants la relation de probabilité de transition entre ces variables. Chaque nœud contient une distribution de probabilité conditionnelle qui décrit la relation entre la variable et ses parents.

jour la distribution des paramètres stochastiques du modèle exponentiel décrivant le processus de dégradation des paliers d'essai.

[Dong et al., 2008] ont étudié une méthode de pronostic basée sur les RBD afin de prédire la durée de vie utile restante des trépans d'une machine de forage vertical. Le RUL est considéré comme un état caché et est déduit des données invisibles en utilisant le modèle de prédiction ainsi que des règles d'inférence. [Hu et al., 2011], ont utilisé les RBD pour modéliser la propagation des défaillances dans un système complexe en tenant compte des interactions et des dépendances entre les sous-systèmes. [Sheppard et al., 2005] les ont utilisés pour modéliser l'incertitude des instruments tels que les systèmes d'augmentation de stabilité des hélicoptères.

2. L'intelligence artificielle ou apprentissage automatique : Nous recensons les trois types de méthodes les plus utilisées en pronostic : les réseaux de neurones artificiels (RNA), les méthodes de régression à vecteurs de support (SVR), et les méthodes à base d'instances (IBL).

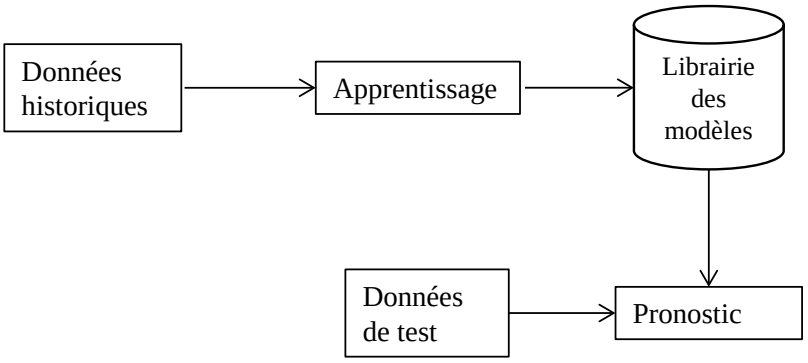


FIGURE 2.10 – Approches d'apprentissage automatique.

a.Les Réseaux de neurones : la plupart des approches d'apprentissage automatique développées en pronostic sont les réseaux de neurones artificiels et leurs variantes. En fait, les RNA sont connus pour leur capacité à modéliser des systèmes complexes, multidimensionnels et non-linéaires sans la nécessité d'une compréhension physique du comportement du système. Un RNA est une fonction d'approximation non linéaire à multiples entrées et sorties et peut être utilisé à différentes fins parmi lesquelles le pronostic. Selon les entrées du modèle RNA, le dernier a la souplesse nécessaire pour exécuter différentes tâches, comme la prédiction du temps restant avant la défaillance par exemple.

[Yam et al., 2001] ont utilisé les réseaux de neurones récursifs (RNR) et ont proposé un type de réseaux dynamiques, afin de prédire l'évolution de la défaillance dans un train planétaire. [Wang et al., 2001] ont développé un modèle combinant les ondelettes et les réseaux de neurones récursifs (récurrent Wavelet Neural Network (RWNN)) afin de pronostiquer la propagation de la fissure d'un roulement. Les réseaux neurones-ondelettes appartiennent à une nouvelle classe de réseaux de neurones (Wavelet Neural Networks) qui possède des caractéristiques d'identification et de classification

tout en stockant des informations temporelles. [Yu et al., 2006] ont proposé un réseau de neurones pour prédire le comportement d'un processus d'alésage au cours de son cycle de vie complet en utilisant les réseaux de neurones récursifs. Le travail prédit des signature qui caractérisent le processus en supposant qu'un historique de ces derniers est disponible.

b.Les machines à vecteurs de support : les machines à vecteurs de support (SVM) ont été développées par Vapnick et ses collègues [Vapnik et al., 1996]. Les SVM sont à la base d'un système d'apprentissage qui représente un espace de caractéristiques d'entrée dans un espace de plus grande dimension. Cela est appelé l'astuce de noyaux (kernel trick) et donne aux SVM la capacité à traiter des données non-linéaire. Ces méthodes ont une bonne généralisation et ont deux types d'application : la classification à vecteurs de support, utilisée pour résoudre le problème de diagnostic, et la régression à vecteurs de support, utilisée pour résoudre le problème de pronostic [Benkedjouh et al., 2013, Caesarendra et al., 2011, Widodo et al., 2011, Soualhi et al., 2015]

c.Méthodes à base d'instances : les approches à base d'instances prennent appui sur un ensemble de données, sur lequel des techniques d'apprentissage sont appliquées. Elles se basent sur le principe : "les expériences acquises lors de la résolution d'un problème peuvent être utilisées pour résoudre des cas similaires". Un algorithme souvent utilisé pour ce type d'approches est les k plus proches voisins. Une instance est généralement représentée par un vecteur dans un espace euclidien de dimension n et l'objectif est de trouver les instances similaires. L'avantage de ces méthodes est l'apprentissage incrémental. Quand un problème est résolu avec succès, l'expérience est conservée afin de résoudre de nouveaux problèmes similaires.

Les méthodes à base d'instances font partie du raisonnement à partir de cas (RàPC) qui est un paradigme de raisonnement par analogie. Les nouveaux problèmes sont résolus en se basant sur la connaissance spécifique acquise lors de la résolution d'anciens problèmes. Le RàPC s'appuie sur la philosophie : "les cas similaires produisent des résultats similaires" [Aamodt et al., 1994]

Il existe un certain nombre de travaux en IBL pour la résolution de problème de pronostic [Zio et al., 2010, Ramasso et al., 2013, Mosallam et al., 2013, Wang et al., 2008, Xue et al., 2008]. Nous ferons dans le chapitre 3, une étude approfondie de ces méthodes

2.4.3.3/ APPROCHES HYBRIDES

Les approches physiques sont généralement accompagnées d'un module en ligne qui permet d'estimer et mettre à jour les paramètres du modèle à partir des données de surveillance acquises. Ces approches combinant les modèles physiques, hors ligne et les modèles guidées par les données, en ligne sont dites hybrides. Par exemple, l'estimateur bayésien récursif est une approche importante pour intégrer des modèles du système avec les données de capteurs et activer les modèles dits hybrides. Les estimateurs bayésiens récursifs ont deux procédures communes à chaque itération : prédire et actualiser. Deux variantes importantes de l'estimateur bayésien sont le filtre de Kalman et le filtre particulaire. Le filtre de Kalman est approprié pour les modèles espace-

état linéaires avec un bruit gaussien tandis que le filtre particulaire peut être utilisé pour les modèles non linéaires avec du bruit non gaussien [An et al., 2013, Saha et al., 2011, Rigamonti et al., 2015, Peel, 2008, Lall et al., 2010].

TABLE 2.2 – Avantages et inconvénients des méthodes de pronostic.

Approches	avantages	Inconvénients
Les modèles physique	- Fournir les estimations les plus précises pour toutes les options de modélisation. - La sortie est facile à comprendre.	- Une connaissance détaillée et complète du comportement du système est exigée. - Il est difficile de construire les modèles.
Les réseaux bayésiens	- Basés sur une base théorique bien construite. - Facile de prédire davantage les états.	- Un historique de transition d'état et de données fatales est nécessaire. Ne peuvent pas modéliser les défauts précédemment imprévus.
Les modèles de markov Cachés	- Révèlent les changements des états cachés.	- Les hypothèses des HMMs ne sont pas pratiques dans le monde réel. Les HSMMs détendent les hypothèses, mais complique le modèle.
Les modèles auto-régressifs	- Pas besoin d'historique de défaillance. - Efficace du point de vue du calcul. - Aucune compréhension détaillée des mécanismes de défaillance n'est nécessaire.	- Les modèles ARMAs de base supposent la stationnarité du processus et de bruit. - Sensible au bruit et aux conditions initiales.
Les réseaux de neurones artificiels	- Modélise les systèmes analytiquement difficiles. - Capable de mettre en œuvre une reconnaissance des formes en ligne, précise et rapide.	- Le ré-entraînement du modèle est nécessaire si les conditions de fonctionnement changent. Une démarche opaque, aucune explication n'est donnée.
La régression à vecteurs de support	- Une bonne généralité - Capacité de traiter les données non-linéaires	- Le choix du noyau [Burges, 1998].
Les approches à base d'instances	- Apprentissage incrémental. - Entraînement rapide - Pas de perte d'information	- La performance dépend de la similarité entre les données.

Le tableau 2.2 fait un récapitulatif des avantages et des inconvénients des méthodes de pronostic que l'on vient de décrire.
Les approches basées sur les modèles physiques fournissent des estimations précises, mais sur un type de faute unique. Ces modèles posent problème quand apparaissent

d'autres fautes. Ces approches sont spécifiques à l'application et ne peuvent pas être généralisées. Quant aux systèmes complexes, la construction du modèle est très difficile et souvent couteuse. C'est pourquoi ces méthodes sont plus appropriées pour des applications où la précision l'emporte sur le coût (comme les application du domaine militaire). Par contre, les approches guidées par les données offrent le meilleur compromis entre coût, précision et complexité.
Les réseaux de neurones donnent des prédictions précises avec la capacité de modéliser les systèmes analytiquement difficiles. Par contre il sont des boites noires et sont basés sur une phase d'apprentissage qui est à répéter avec l'utilisation de nouvelles données.
Les méthodes de l'espace d'états révèlent les changements d'états et sont construit à partir d'une base théorique solide mais ils utilisent des hypothèses qui ne sont pas facilement satisfaites dans les application réelles.
Les modèles auto-régressifs n'exigent pas un historique de défaillance mais ils produisent une accumulation d'erreurs dûes aux prédictions faites sur des prédictions précédentes.
La régression à vecteurs de support est connue pour sa bonne généralité et sa capacité à traiter des données non-linéaires mais le choix du noyau doit être bien établi.
Les approches à base d'instances ont l'avantage d'un apprentissage incrémental et une phase d'entraînement rapide par contre la performance dépend de degré de similarité entre les instances utilisées.
Toutefois toutes ces méthodes sont mises en place à partir d'hypothèses de travail que nous allons aborder au prochain paragraphe.

2.5/ HYPOTHÈSES PÉRIMÈTRES ET DÉFIS

2.5.1/ LES HYPOTHÈSES DE TRAVAIL

En accord avec les travaux de [Uckun et al., 2008] sur le vieillissement de systèmes électromécaniques, nous recensons 4 hypothèses de travail :

- Le vieillissement d'un système dépend de son utilisation (par exemple du nombre de kilométrage d'une voiture), du temps passé (les batteries s'usent à cause de ses composants chimiques même si l'on ne s'en sert pas) et des conditions environnementales (des équipements se détériorent à cause de leur exposition au froid).
- Le vieillissement des composants est considéré comme un processus monotone qui peut se manifester dans la composition physique et chimique de ceux-ci.
- Les signes de vieillissement (directs ou indirects) doivent être observables et détectables avant que la défaillance ne se manifeste (pas de panne soudaine sans signe avant coureur, pas de panne catalytique).
- Les signes du vieillissement doivent être corrélés avec les modèles de vieillissement des composants et leur vie utile restante avant défaillance.

Ces hypothèses sont les hypothèses fondamentales dans la majorité des travaux dans le domaine. Pour que l'on puisse développer des méthodes de pronostic sur un système, ces hypothèses de travail doivent être remplies. Elles sont à la base des notions définissant si un système peut être pronostiqué, donc pronosticable, ou non.

2.5.2/ LES NOTIONS DE PRONOSTICABILITÉ ET DIAGNOSTICABILITÉ

Différents travaux dans le domaine des systèmes à événements discrets se sont intéressés aux notions de diagnosticabilité et pronosticabilité d'un système et ont proposé une formalisation de ces notions dans le contexte des langages formels.

Selon [Kumar et al., 2010], les termes pronosticabilité et prévisibilité sont interchangeables. La prévisibilité d'un langage est une condition plus forte que sa diagnosticabilité, [Sampath et al., 1995, Genc et al., 2006].

[Sampath et al., 1995] ont été les premiers à définir la diagnosticabilité sur la base de la condition à pouvoir détecter l'apparition d'une défaillance dans un délai fini. Selon [Rintanen et al., 2007], un système est diagnosticable si chaque séquence infinie d'événements comportant une défaillance se distingue d'une séquence infinie d'événements sans défaillance. Les auteurs ont introduit une notion de délai qui quantifie le nombre d'événements observés avant défaillance. [Pencolé, 2004] a adopté la définition de [Sampath et al., 1995] alors que [Jiang et al., 2004] ont ajouté à la définition de [Sampath et al., 1995] la notion de pré-diagnosticabilité comme une précondition pour la diagnosticabilité. Un système est pré-diagnosticable par rapport à des spécifications si chaque trace d'état défectueux possède un indicateur comme préfixe, où un indicateur est une trace d'état finie pour laquelle toutes les extensions infinies sont défectueuses.

La pronosticabilité est le dual de la diagnosticabilité. Cependant au lieu de considérer la capacité à détecter les défaillances dans un délai fini (hypothèse 3), la pronosticabilité s'intéresse à la prédiction de l'occurrence de défaillances à venir à condition que cette défaillance ne soit pas catalytique (hypothèse 3). [Jéron et al., 2007] définissent intuitivement la pronosticabilité comme la capacité à prédire l'inévitabilité strictement avant son apparition. [Genc et al., 2006] par contre formulent la prévisibilité (pronosticabilité) en se basant sur la condition que chaque trace de défaillance doit avoir un préfixe de non-défaillance. Ainsi, toute trace observable a la propriété de déterminer l'inévitabilité d'une défaillance dans un nombre d'étapes uniformément borné. Cela signifie qu'il est possible de déduire les futures occurrences de l'événement en prenant appui sur le nombre d'évènements des chaines qui ne contiennent pas l'événement à prédire. [Kumar et al., 2010] définissent un système pronosticable comme un système ayant des chaines de caractères pour lesquelles une défaillance peut se produire à la prochaine étape et se distingue des chaines qui ne garantissent pas l'occurrence de la défaillance à venir. [Cao, 1989] examine le problème de la prévisibilité comme un type spécial de projection entre deux langages (ensemble de trajectoires). Un système G est pronosticable à partir d'un autre système H si l'information contenue dans le langage L(G) peut être obtenue par les informations contenues dans L(H).

Ces définitions de pronosticabilité traduisent les hypothèses du paragraphe 2.5. Dans la pratique, nous devons vérifier que les données observées sur le système surveillé reflètent le vieillissement et sont corrélées à la dégradation, que la défaillance est observable. Si ces conditions ne sont pas satisfaites le système n'est pas pronosticable.

A la figure 2.11, un algorithme de vérification de la pronosticabilité d'un système est illustré. La nature des données est considérée comme une condition forte pour la pronosticabilité d'un système. Si les données ne reflètent pas le vieillissement et ne sont pas corrélées à la dégradation, le système ne peut pas être pronosticable. Si cette condition est satisfaite, il faut vérifier l'observabilité de la défaillance. La défaillance est prévisible seulement si elle est observable et s'il est possible de déduire son occurrence à partir

des chaines ou modèles de défaillance.

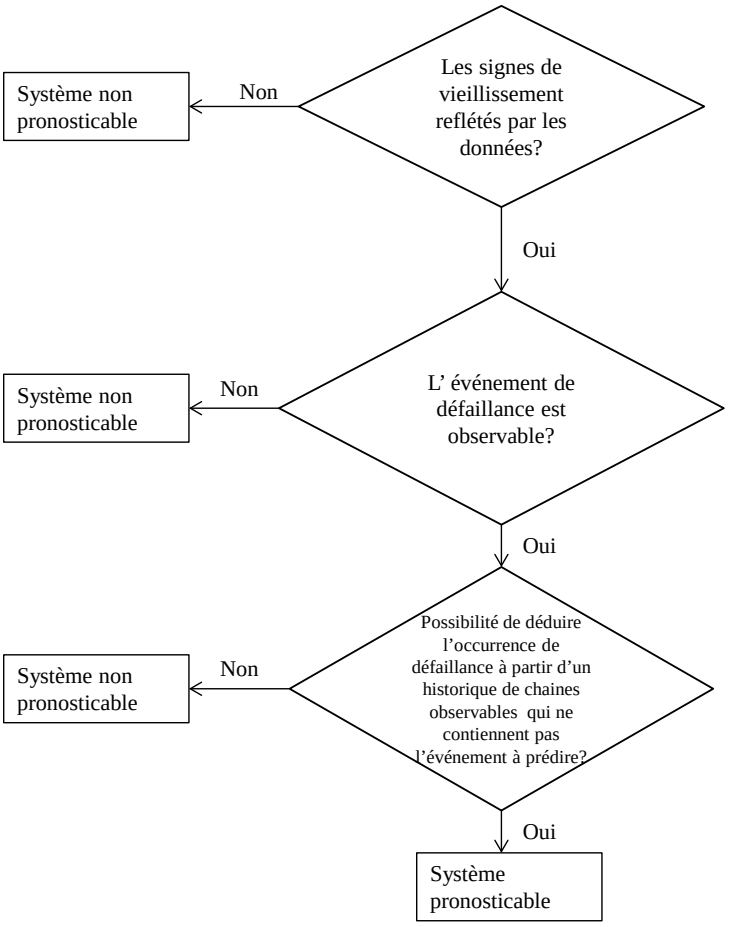


FIGURE 2.11 – Vérification de la pronosticabilité d'un système.

2.5.2.1/ LA NATURE DES DONNÉES

Pour vérifier la réalisation des hypothèses de travail, on doit avoir à disposition des données collectées sur le système à pronostiquer. Les données doivent être informatives. Les approches de diagnostic et pronostic construisent des modèles de dégradation pour évaluer et prévoir la défaillance. Par conséquent, la nature des données collectées à partir du système détermine si celui-ci est pronosticable ou non.

Ces données doivent respecter un minimum de propriétés, comme être cohérentes (elles ne contredisent pas), fiables (les capteurs ne donnent des données justes), disponibles (les capteurs sont fonctionnels), suffisantes (on a à disposition toutes les données requises).

On peut avoir accès à des données bruitées (dû à des problèmes de collecte de données) incertaines. Cela nécessitera un pré-traitement lors de la construction de modèles de pronostic. Cependant, il existe des algorithmes qui effectuent le pronostic en telles conditions en utilisant des techniques de gestion d'incertitude où l'historique de donnée est habituellement représenté par des distributions de probabilité plutôt que des valeurs

déterministes.

2.5.3/ LES DÉFIS AUJOURD'HUI EN PRONOSTIC

Bien que de nombreux algorithmes de pronostic aient été développés, il reste encore beaucoup de travail à réaliser. Plusieurs aspects doivent être étudiés et améliorés avant que l'on puisse appliquer avec efficacité le pronostic dans le domaine de l'industrie. Parmi les obstacles à l'application du pronostic, on liste les suivants :

- **La complexité du système** : Les systèmes réels sont difficiles à modéliser en raison de leur complexité. Ils comportent des composants en interactions entre eux, interactions difficilement observables. La plupart des travaux dans la modélisation guidée par les données sont faits sur le composant le plus critique du système, ce qui permet d'obtenir un RUL à surveiller pour éviter tout arrêt du système.
- **Les données de dégradation** : De nombreux modèles de pronostic, en particulier ceux guidés par les données nécessitent un historique de données allant jusqu'à la défaillance. Cependant, en pratique, les communautés industrielles ne permettent que très rarement à leurs équipements de fonctionner jusqu'à la défaillance. De plus, atteindre la défaillance peut prendre des années. Pour éviter de détériorer les équipements, des bancs d'étude sur la fatigue ou la dégradation de composants critiques sont mis en place. Ils simulent d'une manière accélérée les conditions de dégradation. Mais, on peut se poser la question d'adéquation entre le modèle de défaillance accéléré et le modèle de dégradation réel.
- **Actions de maintenance négligées** : Souvent, les algorithmes développés assument une propagation de défaillance irréversible. La plupart des approches reportées dans la littérature considèrent les effets de la maintenance comme du bruit. Pour une approche de pronostic efficace, il est nécessaire de surveiller les changements conséquents de la fiabilité de l'équipement.
- **Conditions d'exploitation non-connues** : Alors que beaucoup de modèles de pronostic sont développés pour des conditions d'utilisation constantes, les machines, dans la plupart des cas pratiques, sont utilisées dans un environnement variable comme pas exemple la vitesse, la charge, la température, etc.
- **La relation non-linéaire entre les indices de surveillance et l'état de santé actuel de composant** : Un bon nombre de techniques de pronostic utilisent les données de surveillance pour représenter l'état de santé de l'équipement surveillé. L'idée est d'utiliser un modèle de courbe de dégradation se construisant pas à pas. Pour ces techniques un seuil connu comme le seuil de défaillance doit être défini afin d'arrêter la prédiction et en déduire le RUL. La précision de la prédiction dépend fortement de l'hypothèse que la défaillance se produit au moment où la courbe dépasse le seuil prédéfini.
- **La dégradation n'est pas un processus monotone** : Le processus irréversible de dégradation d'un système n'implique pas nécessairement une progression monotone des caractéristiques de ce dernier. Par exemple, le comportement de vibration d'un roulement n'a pas toujours une croissance monotone [Wang, 2010].
- **Les difficultés de validation** : Contrairement aux autres domaines de prédiction où la validation peut être faite plus ou moins rapidement (ex : les prévisions météorologiques et ceux de bourse), les résultats du pronostic ne peuvent être vérifiés qu'une fois le vrai comportement du système connu, ce qui peut prendre des années. Un autre défi concerne l'évaluation des performances des algo-

rithmes de pronostic qui provient de l'absence d'indicateurs de performance communs. Les points forts et faibles des algorithmes ne sont pas absolus, ils varient selon les critères de performance utilisés.

2.5.4/ POSITIONNEMENT DE NOS TRAVAUX

Nos travaux, sont touchés par certains défis mentionnés à la section 2.5.3. Les systèmes complexes sont considérés comme des ensembles de composants critiques en interaction entre eux. Nous ne les appréhendons pas dans une démarche systémique, mais nous les représentons par le composant le plus critique. Celui-ci peut être un sous-système composé de plusieurs modules. Les données de dégradation d'autre part, ont été obtenues par des tests de dégradation accélérés en faisant l'hypothèse que le processus de dégradation accéléré est équivalent à celui réel. Quant à la dégradation, on suppose qu'elle est irréversible. On ne tient pas compte des actions de maintenance qui ont pu être effectuées. Nous n'avons pas spécifiquement étudié le problème de condition d'exploitation constantes ou variables, mais nous avons appliqué l'approche proposée sur un type de données recueillies sous différents profils d'utilisation. Finalement, on fait l'hypothèse que même si la dégradation n'est pas un processus monotone, des tendances monotones peuvent être extraites de ce dernier en faisant de l'extraction de caractéristiques.

Nous travaillons sur des données relatives à des expériences complètes de suivi d'état de fonctionnement d'un composant critique tout au long de son cycle de vie. On appelle une expérience un fait observé ou des épreuves destinées à vérifier une hypothèse ou à étudier des phénomènes [6]. Dans notre cas nous observons l'état de santé d'un composant afin d'obtenir des expériences de défaillance du composant qui ont fait l'objet de notre étude. L'expérience sera définie par une série temporelle qui peut être mono ou multidimensionnelle.

Cet historique de cas de défaillance sera exploité pour créer des modèles, qui seront réutilisés lors de la surveillance d'un nouveau composant de la même famille afin de prédire son comportement, son état courant et son temps restant avant défaillance. Un des paradigmes adaptés à ce type de résolution de problème est le raisonnement à partir de cas, que l'on introduira au chapitre 3 et développera au chapitre 4.

Nous nous sommes intéressés dans ce travail à quatre problématiques scientifiques différentes :

1. *L'extraction des caractéristiques et la construction d'indicateur de santé* : les données recueillies sont brutes et souvent bruitées ce qui les rend peu utilisables par les algorithmes de pronostic. Elles doivent donc être traitées afin d'extraire des informations pertinentes, appelées caractéristiques, qui sont liées à l'état de santé du composant et à la dégradation. Afin de mieux visualiser ces deux derniers éléments, les caractéristiques extraites des données sont réduites ou agrégées construisant des indicateurs de santé. La valeur de ces indicateurs quantifient l'état de santé et leur variation dans le temps traduit le comportement de la dégradation. Le verrou scientifique associé à cette problématique concerne la construction d'indicateurs de santé représentant le comportement non linéaire des composants. Ces indicateurs sont les représentants des expériences passées qui seront exploitées

6. http ://www.cnrtl.fr/definition/expérience

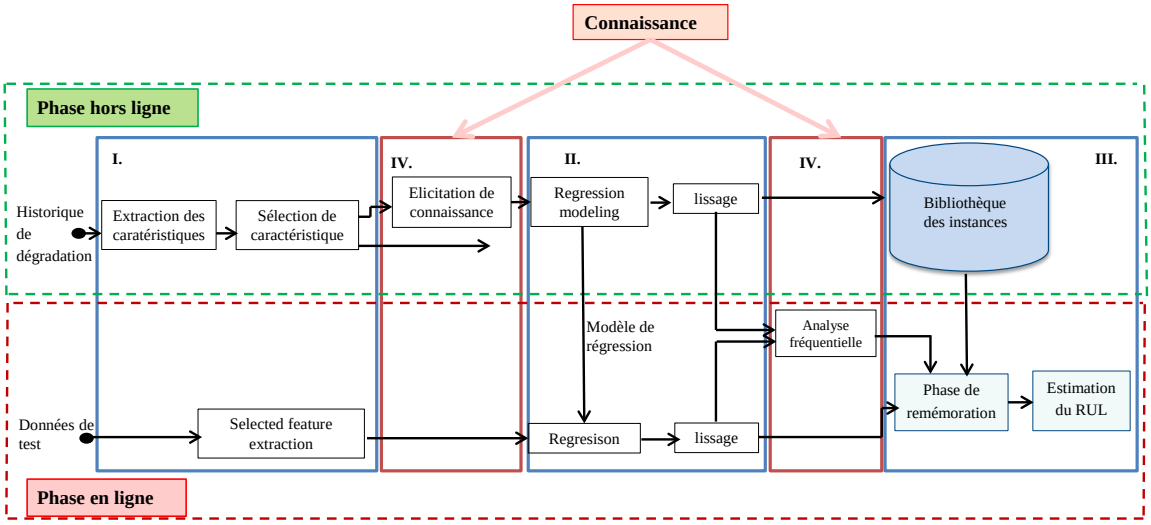


FIGURE 2.12 – Le schéma général de l'approche globale proposée.

pour la prédiction de la durée de vie résiduelle du composant. La solution envisagée est de modéliser la dégradation , ce qui nous amène au verrou suivant.

2. *La modélisation de la dégradation* : les indicateurs de santé sont issus d'un modèle de dégradation qui représente les différents états du composant. Une mauvaise modélisation conduit à de mauvaises prédictions ce qui explique l'obligation de bien choisir les outils de modélisation. Les approches orientées-données ne nécessitent pas une compréhension explicite des mécanismes de dégradation. Elles permettent de définir leur progression à partir des données. On considère trois outils de modélisation, dont deux supervisés. Le premier basé sur la régression linéaire tient compte de l'état de fonctionnement du composant et génère un indicateur de santé. Le deuxième outil supervisé est basé sur la régression à vecteurs de support et lie directement les caractéristiques à la durée de vie résiduelle. Le troisième outil est non supervisé. Il définit une tendance d'évolution en agrégeant les données capteurs en une série temporelle mono-dimensionnelle grâce à un modèle UKR (Unsupervised Kernel Regression).

3. *L'évaluation de l'état de santé courant et l'estimation de la durée de vie résiduelle avant défaillance* : l'estimation de l'état courant du composant est une information importante pour prédire avec précision le RUL. Nous avons identifier et localiser l'état de la dégradation grâce à des mesures de similarité effectuées sur des fenêtres glissantes. Le RUL est prédit en réutilisant les expériences mémorisées. Ce choix nous a permis de nous affranchir de la définition d'un seuil de défaillance, point problématique lors de l'estimation du RUL.

4. *L'extraction de la connaissance sous-jacente aux données et son utilisation pour la résolution de problème de pronostic* : Afin d'exploiter toutes les données à disposition, nous avons extrait de la connaissance des données capteurs, que l'on a utilisé pour améliorer la méthode de pronostic.

Nous proposons une approche de pronostic basée sur le paradigme de raisonnement à partir de cas. Suivant les différentes applications, et les différentes modélisations développées, nous retrouvons les 4 étapes schematisées à la figure 2.12. Le pronostic du RUL est fait en recherchant des similarités entre le comportement d'un nouvel

équipement et un historique de comportement connu sur son cycle de vie complet. Ces approches sont composées de deux étapes : une première hors-ligne construisant à partir de l'historique une librairie d'expérience contenant les solutions et une deuxième étape en-ligne appliquée aux données surveillées sur un nouvel équipement dont on construira une expérience qui sera comparée à la librairie des expériences. Deux types de connaissances sont associés à une expérience. Une temporelle qui complète la modélisation de l'expérience et une fréquentielle qui affine la phase de remémoration par l'ajout de l'aspect fréquentiel à la mesure de similarité. Afin d'améliorer les résultats, l'étape de remémoration est modifiée en extrayant des caractéristiques à partir des expériences et lier le RUL directement à ces caractéristiques. L'application de ses méthodes nécessite d'autre hypothèses en plus à celles définies par [Uckun et al., 2008] :

- Les données de surveillance sont disponibles et fiables. Les capteurs sont supposés fonctionner correctement et sans défaillance.
- L'état de santé de l'équipement est reflété par les données de surveillance.
- Des tendances monotones de dégradation sont supposées être disponibles, soit, directement en forme de données capteur, soit par des méthodes d'extraction de caractéristiques.
- L'approche est appliquée sur des composants de la même famille, c'est-à-dire des composants ayant le même comportement de dégradation. L'expérience de défaillance d'un certain composant (ex : un turboréacteur) peut seulement être réutilisée sur le même type de composants.

2.6/ CONCLUSION

La mise en place d'un système de maintenance efficace augmente la disponibilité des équipements ainsi que leur sécurité tout en diminuant les coûts attenants à la maintenance. Par conséquent, l'accent était mis sur le développement d'une stratégie de maintenance efficace appelée PHM dont on a présenté un état de l'art dans ce chapitre. Parmi les modules de PHM, on s'est intéressé plus particulièrement au pronostic. La littérature scientifique compte huit différentes classifications que l'on a cartographiées. On a retenu les classes les plus pertinentes qui sont au nombre de trois, à savoir : la classe guidée par les données, la classe des modèles physiques et la classe hybride. Les modèles physiques sont plus appropriés pour des applications à coût-justifié où la précision l'emporte sur la plupart des autres facteurs et les modèles physiques restent cohérents à travers les systèmes tandis que les approches guidées par les données offrent le meilleur compromis entre le coût, la précision et la complexité. La classe hybride est une combinaison de modèles physiques et approche guidées par les données. Nous avons établi un état de l'art sur les différentes approches qui constituent la classe guidée par les données. Ces approches sont : (i) les réseaux de neurones qui sont des boites noires et qui nécessite la répétition de l'apprentissage avec toutes nouvelles données. (ii) Les méthodes à espace d'état qui révèlent les changements d'état mais qui utilisent des hypothèses sont difficilement satisfaites dans les applications réelles. (iii) Les modèles auto-régressifs qui ne nécessitent pas un historique de dégradation mais qui produisent une accumulation d'erreur dûe aux prédictions faites sur des prédictions précédentes. (iv) Les méthodes à base d'instances qui ont une phase d'apprentissage rapide et incrémentale mais qui dépendent du degré de similarité entre les instances. (v) La régression à vecteurs de support qui a une bonne généralité et peut traiter des données non-linéaires mais qui doit être bien paramétrée.

Les algorithmes de pronostic décrits dans ce chapitre sont basés sur des hypothèses à vérifier, qui peuvent être représentées par la pronosticabilité d'un système. L'hypothèse majeure étant que les pannes ne peuvent pas être catalytiques. Cela se traduit pour les systèmes à évènement discret qu'il est possible de déduire les futures occurrences de la défaillance en prenant appui sur l'historique d'évènements de la chaine qui ne contient pas l'évènement à prédire.

Le pronostic aujourd'hui est confronté à plusieurs défis parmi lesquels : la complexité des systèmes, la non-linéarité des données et le seuil de défaillance. Ces verrous sont traités par la réalisation d'une approche de pronostic des composants critiques orientée données. Le système complexe sera présenté, par conséquent par son composant le plus critique.
Afin d'éviter la définition d'un seuil de défaillance, nous allons développer une approche basée sur l'expérience où l'expérience est définie par une série temporelle qui peut être mono ou multidimensionnelle et afin de pouvoir traiter la non-linéarité des données nous combinons l'approche à base d'expérience avec des méthodes connues pour leur capacité à traiter telles données comme les SVR et la régression non-supervisée des noyaux (UKR).

Dans le prochain chapitre, nous commencerons par mettre en place une approche basée sur l'expérience qui est l'apprentissage à partir d'instances. L'approche construit une bibliothèque d'instances formalisées comme trajectoires liées à la dégradation. En ligne, la nouvelle instance de test est comparée à la bibliothèque. Celles les plus similaires sont récupérées et utilisées pour estimer directement le RUL sans la nécessité de fixer un seuil.

II

CONTRIBUTION

3

PRONOSTIC ORIENTÉ EXPÉRIENCE FORMALISÉ PAR LES DONNÉES

3.1/ INTRODUCTION

À partir d'un historique d'expériences relatives à l'état de santé de composants critiques de même famille, nous allons appliquer une démarche de raisonnement à partir de cas (RàPC). On s'intéresse particulièrement à deux types de techniques de RàPC : le raisonnement à partir d'instances et le raisonnement à partir de cas. La première technique, qui n'utilise que les données, est une méthode classique en apprentissage automatique. Elle a fait l'objet de nombreux travaux dans le domaine du PHM, qui sont passés en revue en section 3.2. La deuxième technique se différencie tout d'abord par l'ajout de la connaissance et se poursuit par le développement des autres phases de raisonnement à partir de cas.

L'objectif principal de ce chapitre est de développer une méthode de raisonnement à partir d'instances en exploitant les données capteurs relatives à différentes expériences et en les convertissant en instances qui sont, soit des indicateurs reflétant le comportement de l'état de santé, soit des trajectoires de dégradation. La démarche que nous adopterons est composée de trois phases : la formalisation des expériences en instances, la remémoration, et l'estimation de l'état courant de santé et prédiction du RUL.

Pour la formalisation, on propose deux types de représentations pour transformer les données capteurs en une série temporelle mono-dimensionnelle donnant une tendance d'évolution de l'état du composant.

La première prendra appui sur l'état du composant et ne tiendra compte que des données pendant sa bonne marche et sa fin de vie. Cette méthode sera dite supervisée et générera un indicateur évoluant dans le temps qui reflètera ainsi l'état de santé du composant.

La deuxième représentation dite non supervisée permettra d'obtenir des trajectoires de dégradation par l'agrégation des données capteurs, sans tenir compte de l'état du composant. Ces trajectoires sont obtenues par l'exploitation complète des données capteurs sans toutefois être directement liées à l'état de santé.

Puis, en ce qui concerne la phase de remémoration, notre objectif est de définir une mesure de similarité adaptée aux séries temporelles traitées. On considère trois mesures : une distance euclidienne, une mesure de similarité entre blocs et une mesure de simi-

larité pondérée avec une projection temporelle entre blocs. Cette projection est réalisée grâce à des fenêtres glissantes qui permettent d'identifier en ligne l'état courant du composant.

La troisième phase estimera le RUL à partir de l'état courant du composant.

La suite de ce chapitre est organisé comme suit. La section 3.2 présentera un état de l'art des approches à base d'instances. La section 3.3 présentera les deux approches de formalisation de l'expérience que l'on propose d'utiliser : une supervisée et une non-supervisée. Ensuite, les sections 3.4, 3.5 décrivent la méthode de remémoration des expériences similaires et l'estimation de l'état de santé avec la prédiction du RUL. La section 3.6, présentera les différentes métriques d'évaluation des méthodes de pronostic. Ces différentes phases seront appliquées sur deux types de composants critiques (turboréacteurs et batteries). Les résultats obtenus seront présentées et comparés à la section 3.7.

3.2/ ÉTAT DE L'ART SUR LES APPROCHES À BASE D'INSTANCES

Habituellement, les approches de pronostic basées sur l'IBL (apprentissage à partir des instances) se composent de trois étapes illustrées à la figure 3.1 : (i) la formalisation de l'expérience en instance, (ii) la phase de remémoration qui définit un score de similarité entre les expériences (instances) , (iii) et la prédiction du RUL.

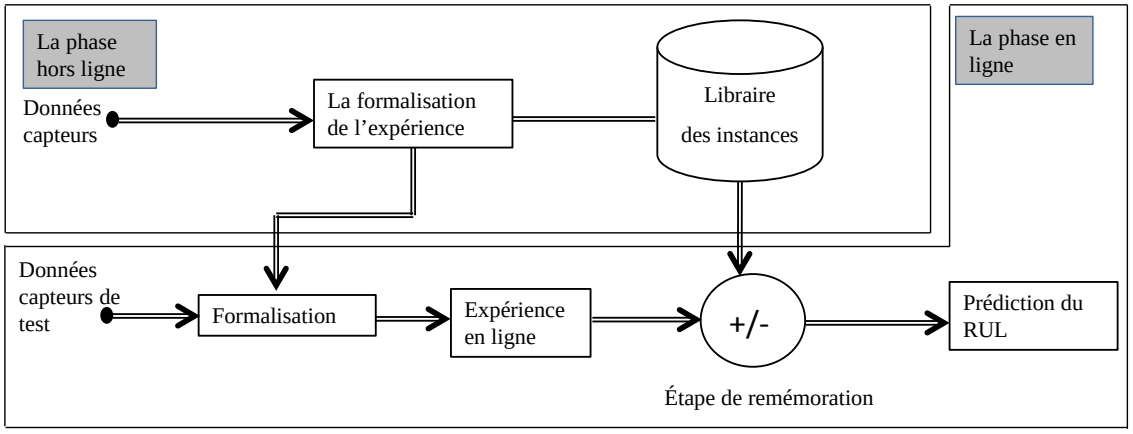


FIGURE 3.1 – Les approches basées sur la connaissance.

La formalisation de l'expérience doit capturer le comportement de dégradation du composant. En effet, l'expérience est un cas d'historique de défaillance. Elle peut être sauvegardée sous forme d'une instance présentée par des données capteurs où l'expérience est conservée dans sa présentation brute ou par un signal monotone représentant une courbe de tendance.

[Xue et al., 2008] ont représenté les instances par des vecteurs d'attributs de dimension N qui sont constitués par les données capteurs. [Zio et al., 2010] les ont formalisées sous forme d'observations multidimensionnelles qui représentent l'évolution du processus concerné. [Ramasso et al., 2013] ont représenté les instances comme des observations de défaillance auxquelles des masses de croyances ont été associées. Ces masses représentent l'incertitude et l'imprécision de l'état courant.

[Mosallam et al., 2013] d'autre part, ont réduit l'espace d'entrée en un signal mono-dimensionnel qui représente une instance. [Wang et al., 2008] les ont représentées par un indicateur de santé créé à partir d'observations filtrées (observations de début et fin de vie).
Dans cette approche, nous proposerons une représentation mono-dimensionnelle des instances. Pour cela, on présentera à la section 3.3 un état de l'art des travaux sur la construction des courbes de tendance à partir des données capteurs.

Pour la remémoration des expériences [Mosallam et al., 2013] ont utilisé la distance euclidienne pour comparer en ligne, la nouvelle trajectoire avec les instances de la librairie. Ce nouveau signal mono-dimensionnel a été construit en estimant les données capteurs grâce à un filtre bayésien. Une fois obtenu, il a été attribué à la classe appropriée en utilisant la méthode de K plus proches voisins. Ensuite, la distance euclidienne a été utilisée afin de remémorer l'instance la plus similaire à partir de la classe.
[Wang et al., 2008] ont aussi basé la phase de remémoration sur une distance euclidienne avec un paramètre de décalage.
[Zio et al., 2010] ont proposé une mesure de similarité point par point floue. Le score de cette mesure est basé sur une fonction d'appartenance floue qui établit un lien entre la différence entre les éléments des trajectoires et les valeurs d'appartenance. Les poids sont inversement proportionnels aux scores de la similarité. Le RUL est obtenu comme une moyenne pondérée des RUL des cas similaires.
[Ramasso et al., 2013] ont développé une mesure de similarité basée sur la distance euclidienne et les fenêtres glissantes. Seules les dernières fenêtres des observations constituant l'expérience sont glissées sur les données des instances de la librairie. L'instance ayant la distance minimale a été récupérée et utilisée afin de prédire le RUL et l'état de santé.

Dans la plupart des approches à base d'instances utilisées pour le pronostic, l'information contenue dans les données de surveillance mises à jour n'est pas entièrement utilisée. À la phase de remémoration, le score de similarité est soit défini en utilisant un vecteur d'attributs qui caractérise l'instance [Xue et al., 2008], soit en utilisant un vecteur de données qui caractérise les dernières observations [Zio et al., 2010, Ramasso et al., 2013]. Pour y remédier, [Mosallam et al., 2013] et [Wang et al., 2008] ont considéré tout l'historique d'observations avec une distance euclidienne point par point entre les trajectoires du test et d'apprentissage (instances). Ces derniers ont pleinement utilisé les données de surveillance mises à jour. Cependant, toutes les observations avaient la même influence sur le score de la similarité alors qu'il est connu que les observations tardives sont plus critiques vis à vis de la défaillance et doivent donc être considérées comme plus importantes (poids). [Wang, 2010] a amélioré son approche en privilégient les derniers blocs (les données de surveillance les plus récentes) des trajectoires d'apprentissage et en considérant la trajectoire en ligne (test) comme un seul bloc.

Dans ce travail, on propose des mesures de similarité par blocs. Les trajectoires d'apprentissage et les trajectoires test sont divisées en bloc. La division des trajectoires test offre plus de liberté pour manipuler la similarité entre les séries temporelles. Ces mesures de similarités considèrent l'historique complet d'observation tout en donnant plus de poids aux cycles tardifs et en identifiant l'état courant du composant.

3.3/ FORMALISATION DE L'INSTANCE

Il existe deux types de travaux représentant les données capteur en une série temporelle unidimensionnelle : (i) le premier type se fait par des méthodes non-supervisées de réduction des données en des séries temporelles uni-variées qu'on nommera trajectoires de dégradation, (ii) le deuxième type se fait par des méthodes supervisées tenant compte de l'état de santé du composant. Par conséquent, les trajectoires obtenues sont nommées par [Wang et al., 2008] indicateurs de santé. Un état de santé (voir définition 1) prend la valeur 1 quand le composant est en bonne santé et la valeur 0 quand il est en panne [1].

[Mosallam et al., 2013], ont utilisé l'analyse en composantes principales afin de réduire l'espace dimensionnel élevé des données en un espace à une dimension. [Benkedjouh et al., 2013] ont appliqué ISOMAP, qui est une généralisation non-linéaire de l'algorithme d'échelle multidimensionnelle (MDS : Multi-Dimensional Scaling). MDS permet à partir des distances euclidiennes entre points, de déterminer un système de coordonnées réduit qui préserve le distance. Le signal généré par ces deux travaux est nommé abusivement à notre sens, un indicateur de santé. Bien que les deux techniques soient des outils de réduction de dimension bien établis, le signal, considéré comme un indicateur de santé dans la littérature, pourrait refléter l'état de santé comme ne pas le refléter. Il peut ne pas représenter l'état de santé et ainsi produire des indicateurs de santé biaisés. En effet, le signal fournit seulement une représentation dimensionnelle réduite et compacte de l'espace des caractéristiques. En plus, certains des éléments sélectionnés au cours du processus de réduction peuvent ne pas être appropriés pour caractériser l'état de santé du composant. En outre, les mêmes étapes de réduction de données doivent être faites en ligne, ce qui rend ces approches moins adaptées pour des applications des systèmes réels.

Dans [Lee et al., 2006], la régression logistique, une méthode qui construit un modèle prédictif permettant d'estimer les variables de sortie à partir de variables cibles qualitatives, a été utilisée pour convertir les données multidimensionnelles en un indicateur de santé avec l'hypothèse qu'un équipement en bonne santé produit un sortie égale à 1, alors que le même équipement dans un état défectueux donne une sortie égale à 0. Comme indiqué par [Wang et al., 2008] , la régression logistique déforme les modèles originaux de dégradation du système. La courbe logistique est plate lorsque les valeurs approchent 0 et 1. Par conséquent, les indicateurs de santé produits par la régression logistique sont moins sensibles en début et fin de vie qu'au milieu (on n'arrive pas à faire la distinction entre les trajectoires comme les courbes sont plates au début et fin de vie), ce qui peut conduire à une plus grande erreur de prédiction lors de l'extrapolation de la courbe de l'indicateur de santé. Pour conserver les motifs originaux dans le signal, [Wang et al., 2008] ont proposé un modèle de régression linéaire utilisé comme une évaluation de la performance du système.
Dans cette approche, on s'intéresse aux deux types de travaux. Comme [Wang et al., 2008], on construit un indicateur de santé par un modèle de régression linéaire. En deuxième lieu, on fusionne les données capteurs par la régression non supervisée et on obtient aussi comme [Mosallam et al., 2013] une trajectoire de dégradation.

1. à cause de résidus et de lissage, les trajectoires obtenues ne sont pas strictement entre 0 et 1

Définition 1 : État de santé

Un état de santé est quantifie par une valeur entre 0 et 1 exprimant la capacité du composant à satisfaire aux exigences qui lui sont imposées.
Un bon état de santé est exprimé par une valeur proche de 1 et un mauvais état est exprimé par une valeur proche de 0.

3.3.1/ DES DONNÉES AUX INSTANCES

Soit $\mathcal{T}$ un ensemble d'apprentissage composé de $|\mathcal{T}|$ séries temporelles multivariées $T_1, \ldots, T_{|\mathcal{T}|}$. Chaque série temporelle de cette ensemble représente les données capteurs des équipements surveillés. La longueur de T_i est dénotée par $l(T_i)$. Avec cette notation, une série temporelle T_i appartenant à $\mathcal{T}$ peut être écrite comme $T_i = t_1^i, t_2^i, \ldots t_{l(T_i)}^i$. Puisque ces séries temporelles sont multivariées, $t_j^i \in \mathbb{R}^d, d \geq 2, 1 \leq i \leq |\mathcal{T}|, 1 \leq j \leq l(T_i)$.

Dans cette section, nous cherchons à transformer les $|\mathcal{T}|$ séries temporelles constituant l'ensemble d'apprentissage en $|\mathcal{T}|$ instances, $I_1, \ldots, I_{|\mathcal{T}|}$ à l'aide d'une fonction de régression $I_i = f_r(T_i), \forall i \in [1, |\mathcal{T}|]$.

Définition 2 : Instance

Une instance I est une série temporelle monodimensionnelle représentant l'historique de défaillance du composant T.

$$I = \{D_1, D_2, \ldots D_{l(T_i)}\},$$

où $D_j \in \mathbb{R}$ est un descripteur de l'instance obtenu grâce à la fonction de régression, $D_j = f_r(t_j), 1 \leq j \leq l(T)$.

Les données capteurs sont donc fusionnées afin de visualiser l'expérience et évaluer la dégradation du composant. On propose deux types de formalisations.
La première formalisation construit des indicateurs de santé à partir de l'historique de dégradation à l'aide de la régression linéaire, c'est-à-dire les descripteurs dans ce cas sont les valeurs de l'indicateur de santé . Les indicateurs obtenus ont l'avantage d'être directement liés à l'état de santé et permettent donc d'évaluer la condition de l'équipement. Par contre, seules les données de début et fin de vie sont utilisées afin d'entraîner le modèle de la régression. Par conséquent la phase d'apprentissage de cette modélisation ne tient compte que de 20% des données capteurs.
Dans un soucis d'amélioration des performances de cette méthode, et afin d'exploiter toutes les données capteurs, on propose une deuxième formalisation non-supervisée qui construit des trajectoires de dégradation à l'aide de la méthode de régression non-supervisée à base de noyau (UKR). Les descripteurs dans ce cas sont les valeurs de la trajectoire de dégradation. Toutefois, le signal obtenu n'est pas directement lié à l'état de santé mais il représente fidèlement et d'une manière compacte les données capteurs. Les deux types de formalisation sont décrites ci-après.

3.3.1.1/ INDICATEUR DE SANTÉ

Les indicateurs de santé sont obtenus grâce à la régression linéaire qui est une approche utilisée pour prédire une variable cible à partir de variables prédictives.

Soit $T_i = t^i_1, t^i_2, \ldots, t^i_{l(T_i)}$, une série temporelle multivariée où $t^i_j = t^i_{j,1}, t^i_{j,2}, \ldots, t^i_{j,N}$, N étant le nombre de capteurs considérés. On cherche à obtenir le vecteur $Y_i = Y^i_1, Y^i_2, \ldots, Y^i_{l(T_i)}$ qui constitue l'indicateur de santé (cf. définition 3).

Définition 3 : Indicateur de santé

Un indicateur de santé $Y_i = f_r(T_i)$ est défini comme comme une série temporelle uni-variée représentant l'évolution de l'état de santé du composant

$$Y^i_j = \beta_0 + \beta_1 t^i_{j,1} + \cdots + \beta_n t^i_{j,N} + \epsilon, \tag{3.1}$$

où $t^i_j = \{t^i_{j,1}, t^i_{j,2}, \ldots, t^i_{j,N}\}$ est le vecteur d'observation (données capteurs), et le coefficient β_p quantifie l'association entre la variable $t^i_{j,p}$ et la réponse Y^i_j. Le vecteur $B = \{\beta_p\}$ est le vecteur des paramètres du modèle de régression. ϵ est le vecteur aléatoire représentant les résidus.

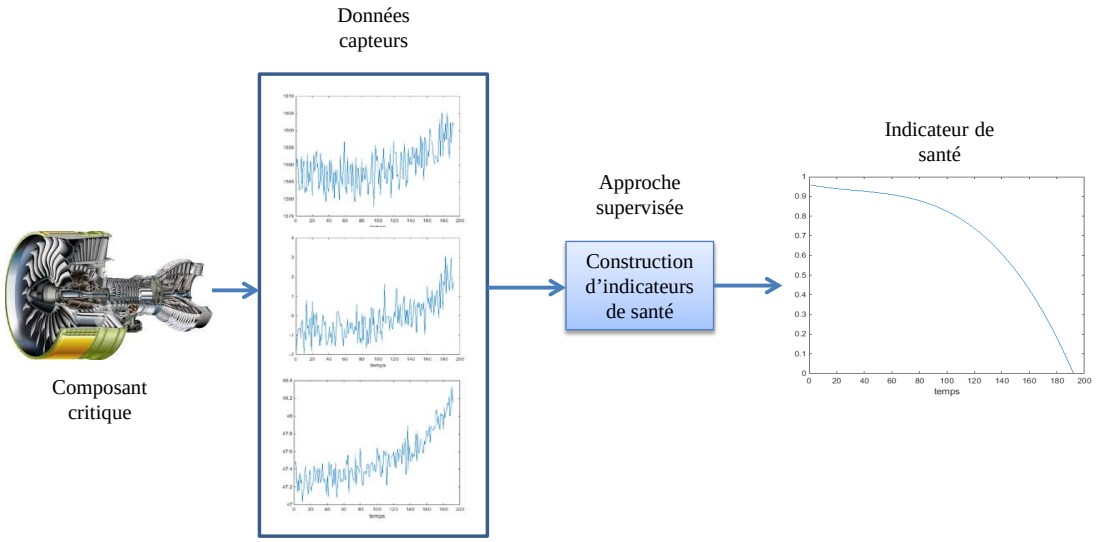


FIGURE 3.2 – Fusionner les données capteurs en un indicateur de santé.

Soit **B** l'ensemble de tous les vecteurs possible B. L'objectif est de trouver un vecteur qui minimise la somme des carrés résidus. Ce vecteur est obtenu par la méthode des moindres carrés.

Pour notre application, on suppose que le composant est dans un bon état au début de vie, ce qui signifie que son indicateur de santé vaut 1 et développe ensuite une faute qui conduit à la défaillance totale (indicateur de santé vaut 0). La combinaison des données capteurs et leurs valeurs d'indicateur de santé de chaque composant appartenant à l'ensemble d'apprentissage est utilisée hors ligne afin d'entraîner le modèle de régression et d'obtenir ses paramètres. Le modèle est ensuite utilisé hors ligne pour extrapoler les valeurs intermédiaires de l'indicateur de santé et en ligne pour construire l'indicateur de santé du nouveau composant.

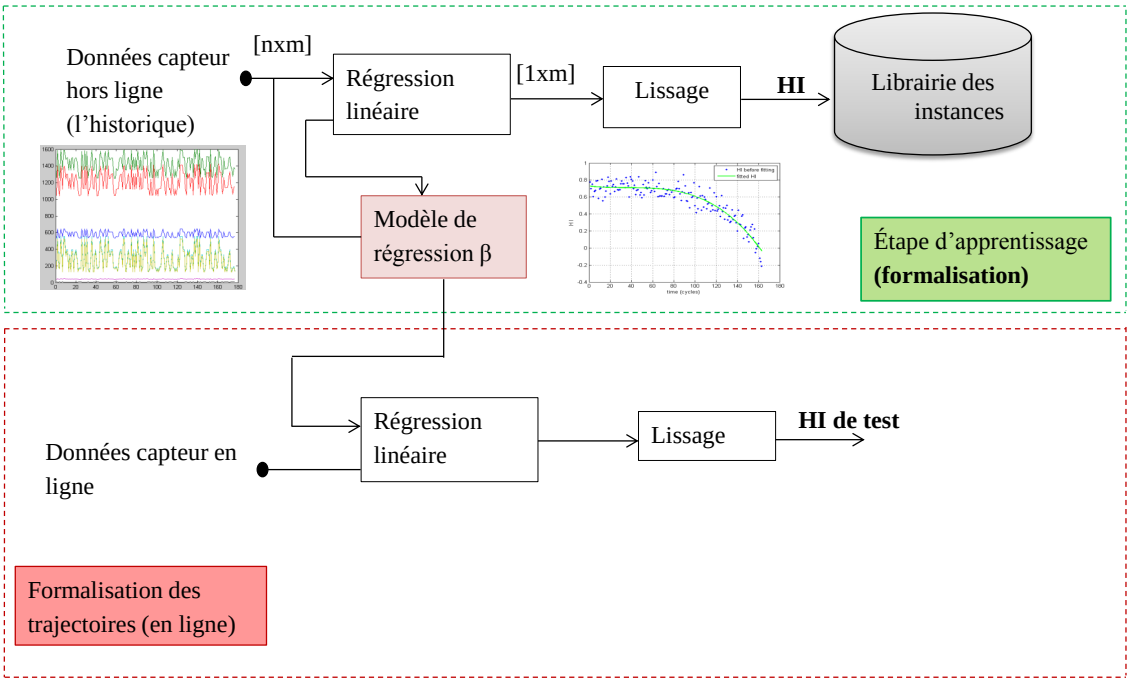


FIGURE 3.3 – Formalisation de l'expérience en utilisant la régression linéaire.

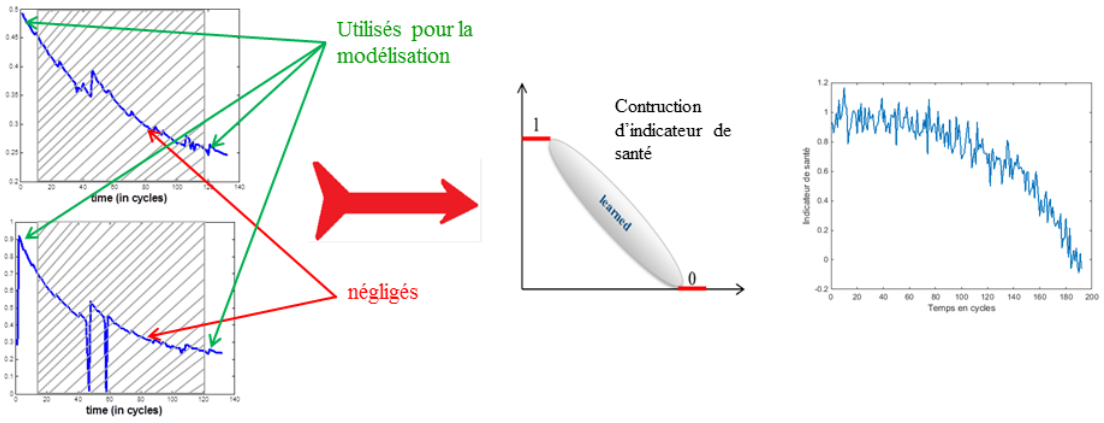


FIGURE 3.4 – Construction de l'indicateur de santé.

À partir de la figure 3.4, on constate que seulement 20% des données sont utilisées pour entrainer le modèle. Afin d'y remédier, on propose une approche non-supervisée qui sera détaillée par la suite.

3.3.1.2/ TRAJECTOIRE DE DÉGRADATION

Définition 4 : Trajectoire de dégradation

Une trajectoire de dégradation $TD = f_r(T_i)$ est définie comme une série temporelle uni-variée qui est une représentation fidèle et compacte des données capteurs multidimensionnelles liées au processus de la dégradation.

Soit $T_i = t_1^i, t_2^i, \ldots, t_{l(T_i)}^i$, une série temporelle multivariée où $t_j^i = t_{j,1}^i, t_{j,2}^i, \ldots, t_{j,N}^i$, N étant le nombre de capteursconsidérés. On cherche à obtenir le vecteur :
$TD_i = TD_1^i, TD_2^i, \ldots, TD_{l(T_i)}^i$ qui constitue la trajectoire de dégradation (cf. définition 4).

La régression non supervisée à base de noyau est une approche récente qui est utilisée pour obtenir une représentation dimensionnelle latente fidèle $TD_i = TD_1^i, TD_2^i, \ldots, TD_{l(T_i)}^i$ de l'ensemble des données observées (données capteurs dans notre cas)
$t_j^i = t_{j,1}^i, t_{j,2}^i, \ldots, t_{j,N}^i$.

La méthode a été proposée par Meinecke et Klanke comme une formulation non supervisée de l'estimateur «Nadaraya-Waston». L'idée est de généraliser l'estimateur au cas non supervisé d'apprentissage de fonction (function learning) [Meinicke et al., 2005].

Dans le cas supervisé, l'estimateur réalise une généralisation continue de la relation fonctionnelle entre les variables aléatoires X et Y, comme décrit dans l'équation

$$t_j^d = f(TD_j^d) = \sum_{i=1}^{N} t_{j,i}^d \frac{K_H(x - TD_{j,i}^d)}{\sum_m K_H(x - TD_{j,m}^d)}, \tag{3.2}$$

où K_H est la densité de noyau.

Comme indiqué par [Memisevic et al., 2003], la différence entre la régression supervisée et non supervisée réside dans l'utilisation des variables d'entrée. Dans le cas supervisé, cela devient un problème d'estimation d'une relation fonctionnelle entre les variables d'entrée et celles de sortie. Dans le cas non supervisé, l'espace des variables d'entrée est considéré comme manquant et doit être estimé ainsi que la relation fonctionnelle en trouvant l'ensemble des échantillons de sortie qui donne l'erreur de reconstruction minimale.

Afin de tirer la contrepartie non supervisée de l'estimateur, les auteurs dans [Meinicke et al., 2005], utilisent la même forme fonctionnelle de l'estimateur de régression kernel «Nadaraya-Waston», mais traitent les données d'entrée en tant que paramètres manquants. Cet ensemble de paramètres $TD_i = \{TD_i^j\}$ sert de représentation latente d'une dimension inférieure de l'ensemble de données d'origine $T_i = \{T_i^j\}$. La fonction devient :

$$\begin{cases} b_i(TD_j^i; TD_i) = \frac{K_H(x - TD_{j,i}^d)}{\sum_m K_H(x - TD_{j,m}^d)} \\ t_j^d = f(TD_j^i; TD_i) = \sum_i^N t_{j,i}^d b_i(TD_j^i; TD_i) = T_i b(TD_j^i; TD_i), \end{cases} \tag{3.3}$$

où $b_i(TD_j^i; TD_i)$ contient la fonction de base latente basé noyau et $f(TD_j^i; TD_i)$ est la fonction de noyau. L'objectif de la fonction non supervisée qui a été définie par [Meinicke et al., 2005] est de trouver une réalisation convenable de la cartographie entre le domaine latent et les données d'origine ainsi que la représentation latente associée. Ceci peut être réduit au problème de trouver la bonne densité latente du mélange (mixture) p. Avec ce modèle de densité latente, la fonction UKR peut être complètement spécifiée sans d'autres paramètres. Après avoir défini le modèle UKR, la phase d'apprentissage d'UKR consiste à minimiser l'erreur de reconstruction R, qui est l'erreur entre les données observées et celles reconstruites à partir du vecteur de variables latentes TD_i.

$$R(X) = \frac{1}{N} \sum_{i=1}^{N} \|TD_j^i - f(TD_j^i; TD_i)\|^2 \tag{3.4}$$

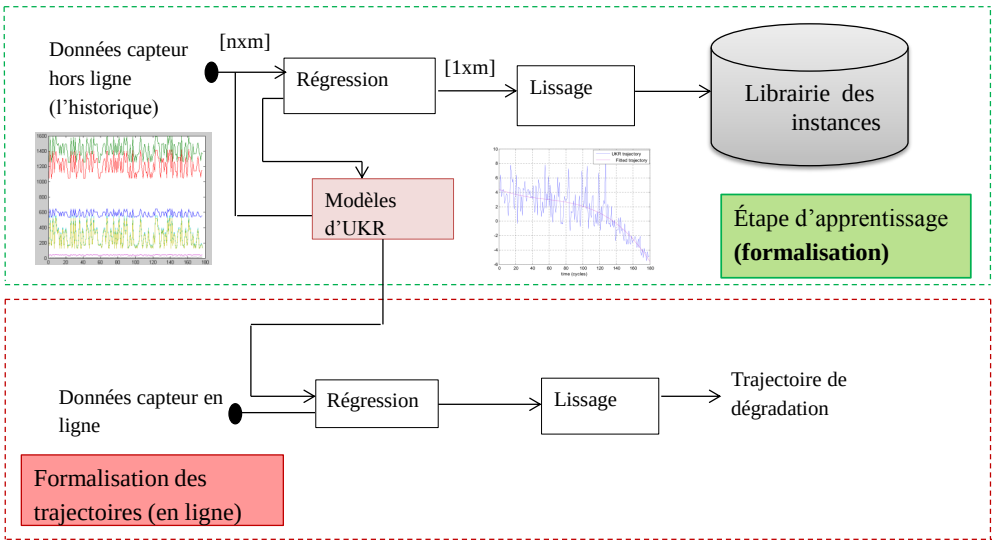


FIGURE 3.5 – Formalisation de l'expérience en utilisant UKR.

L'approche non-supervisée n'exige pas de réglage des valeurs d'apprentissage pour l'obtention du signal mono-dimensionnel. Par conséquent, les données donc sont entièrement utilisées pour obtenir les modèles UKR. Ces modèles sont appliqués sur les données de test pour obtenir la trajectoire de dégradation pour l'élément de test, comme le montrent les figures 3.5 et 3.6.

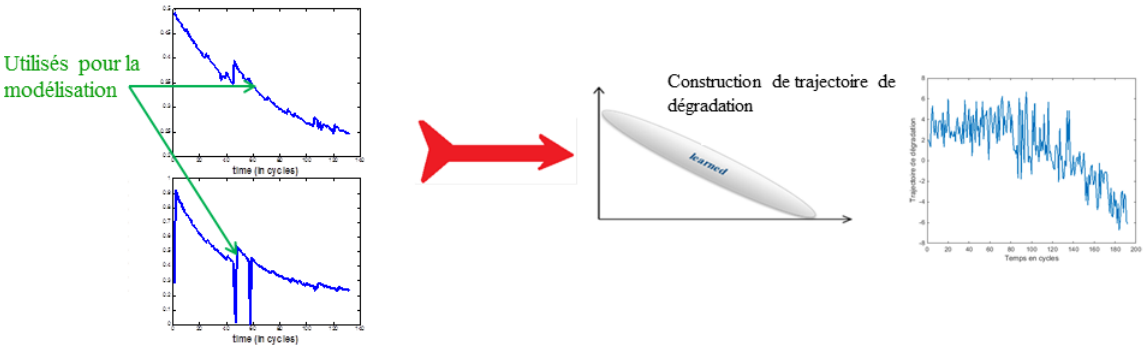


FIGURE 3.6 – Construction d'une trajectoire de dégradation.

Comme on peut constater d'après la figure 3.7, à partir de chaque moteur appartenant a l'ensemble d'apprentissage, un modèle est appris et sauvegardé dans la bibliothèque des modèles. Cette bibliothèque est ensuite utilisée pour formaliser les instances d'apprentissage et de test.

Pour une instance d'apprentissage, le modèle d'UKR correspondant est connu et directement utilisé pour construire la trajectoire de dégradation. Alors que pour l'élément de test, le modèle correspondant n'est pas connu mais supposé être l'un des modèles présents dans la bibliothèque. Afin d'identifier le bon modèle, tous les modèles de la bibliothèque sont d'abord utilisés (figure 3.7). À l'étape de la remémoration, seulement les modèles appropriés sont maintenus.

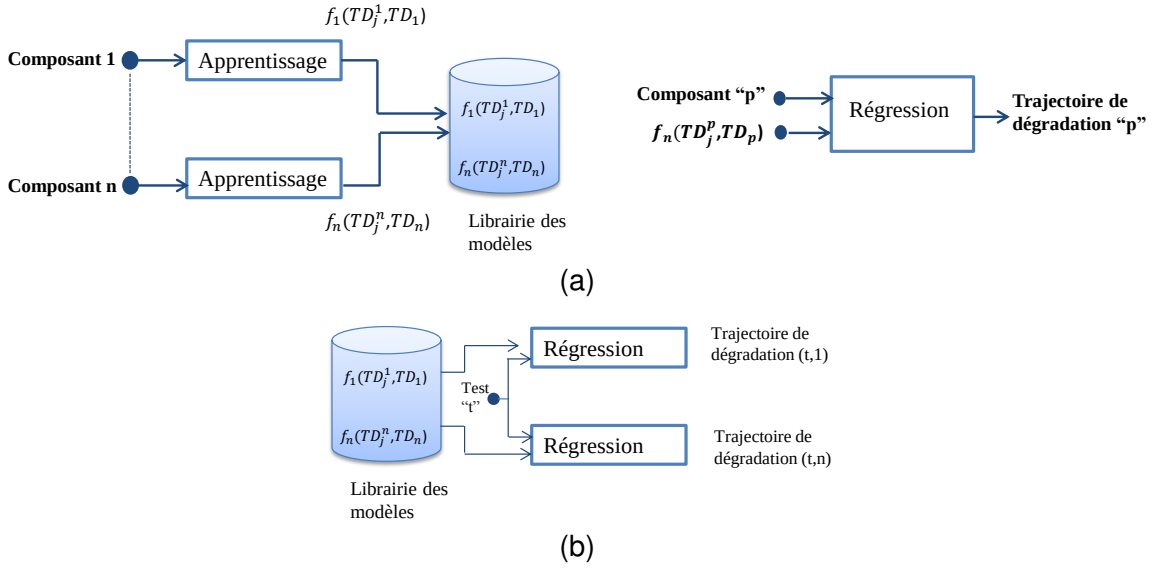


FIGURE 3.7 – Construction de la trajectoire de dégradation. (a) Pour un élément d'apprentissage "p" la trajectoire de dégradation est issue de son propre modèle UKR et (b) pour un élément en ligne tous les modèles de la librairies sont réutilisés pour produire n trajectoires.

3.3.2/ LISSAGE

Les données récoltées à partir des systèmes industriels sont bruitées, ce qui se traduit par des signaux fluctuants à la sortie de la régression. Afin d'enlever ce bruit, ces signaux sont traités et lissés. On a considéré deux types de lissage. Un ajustement de la courbe par un polynôme et la décomposition en modes empiriques. Pour l'ajustement par un polynôme, la fonction est donnée a priori et le dégrée de polynôme est à déterminer. Par contre, en utilisant la décomposition en modes empiriques la base de fonctions est construite à partir des propriétés du signal.

3.3.2.1/ AJUSTEMENT DE COURBE PAR UN POLYNÔME

Ajuster les données en utilisant une fonction polynomiale de la forme

$$y(x, W) = w_0 + w_1 x + w_2 x^2 + ... + w_M x^M = \sum_{j=0}^{M} w_j x^j, \tag{3.5}$$

où M est l'ordre du polynôme, et x^j désigne x élevé à la puissance j. Les coefficients de polynôme $w_0, \ldots, w_M$ sont notés par le vecteur W.

Les valeurs des coefficients sont déterminées en ajustant le polynôme aux données d'apprentissage. Cela peut être fait en minimisant une fonction d'erreur qui mesure l'inadéquation entre la fonction $y(x, W)$, pour une valeur donnée de W, et les points de données d'ensemble d'apprentissage. Un choix simple de la fonction d'erreur, qui est largement utilisée, est donné par la somme des carrées des erreurs entre chaque point x_n

et sa prédiction $y(x_n, W)$ par le modèle polynomial, de sorte qu'on minimise

$$E(w) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n, W) - x_n\}^2. \tag{3.6}$$

On résout le problème d'ajustement de courbe en choisissant la valeur de W pour lequel $E(W)$ et aussi faible que possible. Étant donné que la fonction d'erreur est une fonction quadratique des coefficients W, ses dérivés par rapport aux coefficients seront linéaire dans les éléments de W, et de sorte que la minimisation de la fonction d'erreur a une solution unique, désignée par w^*. Le polynôme résultant est donné par la fonction $y(x, w^*)$.

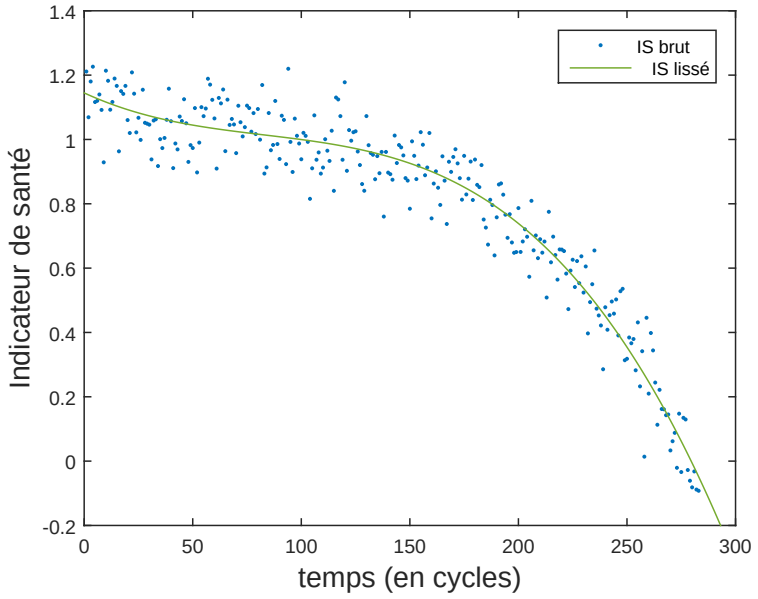


FIGURE 3.8 – L'ajustement de la courbe à un polynôme de 3ème degré.

3.3.2.2/ LA DÉCOMPOSITION EN MODES EMPIRIQUES

Les signaux résultant de la régression comme on le voit à la figure 3.8 sont très oscillant en raison de la nature bruitée des données capteurs. Ces signaux sont lissés grâce à la méthode de décomposition en modes empiriques, où un résidu permet de filtrer le bruit de la courbe et fournir ainsi un indicateur de santé.

L'EMD est un algorithme qui décompose les signaux en fonctions intrinsèques successives. Il a d'abord été proposé par [Huang et al., 1998].

Étant donné un signal x(t), l'algorithme effectif de EMD peut être résumé par cinq étapes, [Huang et al., 1998, Rilling et al., 2003] :

1. Identifier tous les extrêmes de $x(t)$
2. Interpoler entre les différents minimas (respectivement maximas), pour former une enveloppe $e_{min(t)}$ (respectivement $e_{max(t)}$).
3. Calculer la moyenne $m(t) = \frac{e_{min}(t)+e_{max}(t)}{2}$.
4. Extraire le détail $d(t) = x(t) - m(t)$.

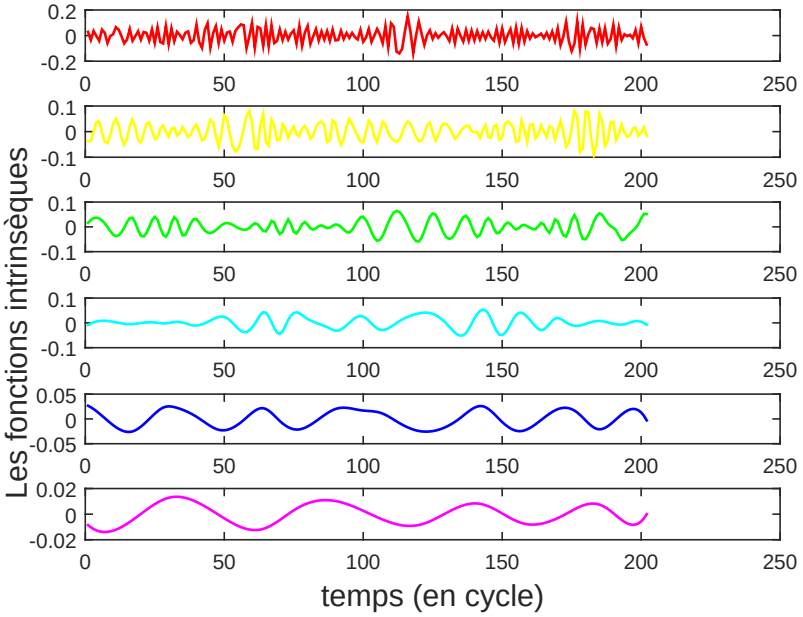


FIGURE 3.9 – Les fonctions intrinsèques obtenues avec la décomposition en modes empiriques.

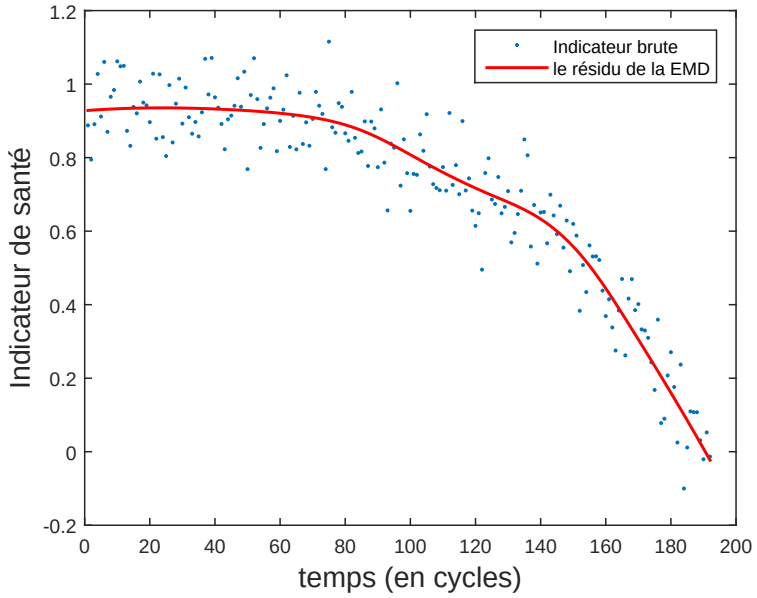


FIGURE 3.10 – Le résidu de la décomposition en modes empiriques.

5. Itérer les opérations 1 à 4 sur le résidu $m(t)$ jusqu'à ce que ce dernier soit constant ou monotone.

Pour la formalisation de l'expérience, deux approches ont été proposées : une qui construit des indicateurs de santé en utilisant la régression linéaire et une deuxième qui donne des trajectoires de dégradation à l'aide de la régression non-supervisée.
À cause de la nature bruitée des données capteurs, les trajectoires obtenues contiennent des fluctuations qui sont enlevés, soit en ajustant les trajectoires à des polynômes où le degré et le polynôme sont déjà prédéfinis ou par la décomposition du signal en modes

empiriques où la base de fonctions est construite à partir des propriétés du signal. L'ajustement par un polynôme est rapide par rapport à la décomposition en modes empiriques qui est une méthode itérative mais qui est aussi plus générale comme elle n'impose pas le type de fonction à utiliser.

Les expériences obtenues, à savoir les séries temporelles, sont sauvegardées sans la bibliothèque des instances et seront utilisées par la suite dans la phase de remémoration.

3.4/ REMÉMORATION DE L'INSTANCE

Dans les approches basées sur l'expérience, la phase de remémoration est d'une grande importance et dépend de la mesure de similarité ou de distance entre la nouvelle instance et les instances de la bibliothèques. Si les mesures sont mal-appropriées, la remémoration mènera à des prédictions erronées.

Les instances étant des séries temporelles, nous nous intéressons aux mesures de similarité entre séries temporelles.
Des travaux ont été faits dans le domaine et l'un des moyens les plus simples pour calculer la distance entre deux séries temporelles est de calculer la distance entre tous les points temporels, comme décrit dans la définition 5 [Aghabozorgi et al., 2015].

Définition 5 : Distance entre séries temporelles

Soit $T_i = \{t_1^i, t_2^i, \ldots, t_{l(T_i)}^i\}$ une série temporelle de longueur $l(T_i)$. Si la distance entre deux séries temporelles est définie sur tous les points, cette $dist(T_i, T_k)$ est la somme des distances entre les points individuels

$$dist(T_i, T_k) = \sum_{j=1}^{l(T_i} dist(t_j^i, t_j^k)$$

[Aghabozorgi et al., 2015] a recensé les mesures de similarité entre séries temporelles. Certaines mesures de similarité sont spécifiques à la représentation de la série temporelle telle que la "MINDIST"qui est spécifique à la représentation SAX (Symbolic Aggregate approXimation, approximation agrégée symbolique) des séries temporelles [Lin et al., 2007], d'autres peuvent être utilisées même sur les séries temporelles brutes. Parmi les mesures de similarité les plus connues et utilisées pour la comparaison des séries temporelles, on mentionne : (i) la distance de Hausdroff qui mesure l'éloignement de deux sous-ensembles d'un espace métrique, elle est adaptée pour estimer la similarité spatiale entre deux séries temporelles, (ii) distance basée sur les modèles de Markov Cachés, où chaque série temporelle est modélisée par un HMM . La distance entre deux trajectoires est ensuite définie comme la valeur absolue de la somme des probabilité de vraisemblance de trajectoires générées par leurs propres modèles moins la probabilité de vraisemblance de trajectoires générées par d'autres modèles, (iii) la mesure de déformation temporelle dynamique qui est une méthode qui calcule un appariement optimal entre deux séries temporelles avec certaines restrictions. Les séries sont déformées non-linéairement dans la dimension temporelle afin de déterminer une mesure de similarité indépendante de certaines variations non-linéaires, (iv) et la distance euclidienne. Les formules de ces différentes distance peuvent être trouvées dans [Zhang et al., 2006].

[Aghabozorgi et al., 2015, Keogh, 2006] ont identifié 3 objectifs au calcul de similarité :

i.Trouver des séries temporelles similaires dans le temps : l'objectif est de grou-

per les séries temporelles qui varient d'une manière similaire à chaque instant de temps. Ce genre de similarité est généralement basé sur la distance euclidienne [Ratanamahatana et al., 2005, Bagnall et al., 2005]. Cette similarité est coûteuse sur des séries temporelles brutes d'une très grande longueur. C'est pourquoi il y a un grand nombre de travaux sur des méthodes qui représentent d'une manière compacte les séries temporelles. Un aperçu général sur ces travaux a été fait par [Keogh et al., 2003].

ii. **Trouver des séries temporelles similaires dans la forme :** l'objectif est de partitionner les séries temporelles disposant d'une forme commune ensemble. Comme résultat, les méthodes élastiques [Agrawal et al., 1993, Aref et al., 2004] telles que la déformation temporelle dynamique [Chu et al., 2002] qui est utilisée pour calculer la dis-similarité entre les séries temporelles. Un exemple de cette approche est de grouper les prix des actions liées à différentes sociétés qui ont un motif commun dans leur stock indépendamment de son apparition dans la série temporelle [Bagnall et al., 2005].

iii. **Similarité dans le changement :** Dans ces approches, habituellement des méthodes de modélisation telles que les chaines de Markov cachées et les processus d'ARMA sont utilisées afin de modéliser la série temporelle [Smyth et al., 1997, Kalpakis et al., 2001, Xiong et al., 2002]. La similarité est ensuite mesurée à partir des paramètres du modèle ajusté. Les séries temporelles sont donc groupées en se basant sur la similarité des paramètres ajustés. Cette approche est appropriée pour des séries d'une très grande longueur et ne peut pas être appliquée sur des séries temporelles d'une longueur modérée ou courte [Wang et al., 2006].

Définition 6 : Instances similaires

Pour une instance *"problème"* I_p, l'ensemble $M_p = \{I_{p,1}, \ldots, I_{p,k}\}$ est dit être l'ensemble de meilleurs appariements, si toutes les instances y appartenant sont les k instances les plus similaires à l'instance I_p. C'est-à-dire, leur distance figure parmi les distances minimales ou leur score de similarité figure parmi les scores les plus élevés. Autrement dit : $\forall I_{i,p} \in M_p, d(I_p, I_{i,p}) \in \{distancesminimales\}$ ou $\forall I_{i,p} \in M_p, score(I_p, I_{i,p}) \in \{scoremaximaux\}$

Soit $\mathbf{I} = \{I_1, I_2, \ldots, I_{|\mathcal{T}|}\}$ la base des expériences et E_p est le cas problème, on cherche à remémorer à partir de la base, les expériences les plus similaires dans le temps au cas problème (cf. définitions 5, 6).

Pour quantifier cette similarité, on se base sur la distance euclidienne afin de définir des mesures adaptées au pronostic.

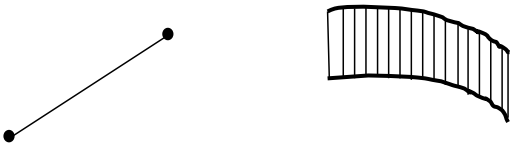
3.4.1/ DISTANCE EUCLIDIENNE

La plupart des approches à base de l'expérience rapportées dans la littérature définissent la mesure de similarité à base d'une distance euclidienne de dimension n en suivant la définition 5, où l'expérience est représentée par une série temporelle de dimension n. En dimension 1, la distance entre deux points x et y est la valeur absolue de leur différence numérique. Ce qui désigne également la longueur de la ligne qui les relie $(\bar{xy})$

$$d(x, y) = \sqrt{(x - y)^2} = |x - y| \tag{3.7}$$

La distance euclidienne, pour un espace à n dimensions, est définie par :

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + ... + (x_N - y_N)^2} \tag{3.8}$$

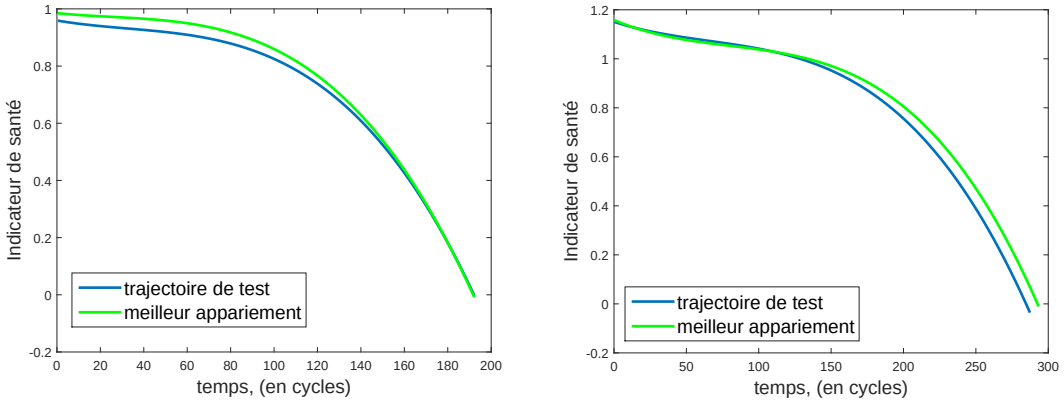


Distance euclidienne dans une dimension

Distance euclidienne dans n-dimension

FIGURE 3.11 – Illustration de la distance euclidienne.

La distance euclidienne dans la littérature des approches basées sur l'expérience pour la prédiction RUL a été utilisée pour comparer et récupérer les séries temporelles similaires [Mosallam et al., 2013, Ramasso et al., 2013]. La série temporelle de test est comparée à la bibliothèque de séries temporelles. Les longueurs de ces dernières sont ajustées à la longueur de l'élément de test en tronquant les trajectoires (séries temporelles). La distance minimale est utilisée comme critère pour retrouver la série temporelle la plus proche. La figure 3.12 montre un exemple de paires de séries temporelles similaires à partir de la bibliothèque en utilisant la distance euclidienne minimale en tant que critère.



(a) D(test, meilleur appariement)=0.4177 (b) D(test, meilleur appariement)=0.7881

FIGURE 3.12 – Exemple de paires de trajectoires similaires obtenues avec la distance euclidienne.

3.4.2/ SIMILARITÉ PONDÉRÉE

Data : - Les trajectoires de test et d'apprentissage.
- Valeurs de N,λ et σ
Result : - Score de similarité.
Initialisation : - Ajuster la longueur de la trajectoire d'apprentissage.;
- Diviser la trajectoire de test en N blocs.;
- Diviser la trajectoire d'apprentissage en N blocs.;
forall the blocs de test **do**
  Obtenir les poids, eq. 3.9;
  Calculer la similarité entre les blocs, eq. 3.10;
end
Calculer le score de similarité, eq. 3.11 ;

Algorithme 1 : Algorithme de mesure de similarité pondérée.

Dans la distance euclidienne, chaque point de la série est traité d'une manière égale. Cependant, lors de la surveillance de la dégradation, les observations tardives sont plus informatives et doivent être privilégiées en leur accordant un poids supérieur aux autres observations. Par conséquent, on propose une mesure de similarité par bloc pondérés sachant qu'une série est un ensemble de bloc concaténés.

L'algorithme 1 commence par ajuster la longueur de trajectoire d'apprentissage afin qu'elle ait la même longueur que la trajectoire de test. C'est-à-dire les valeurs de trajectoire d'apprentissage qui dépassent la longueur de la trajectoire de test sont tronquées (voir la surface grisée de la figure 3.13). Ensuite la similarité entre chaque deux bloc de même position est calculée (c.f. Équation 3.10) et le score final de la similarité est calculé comme une moyenne pondérée des similarités entre blocs (c.f. Équation 3.11).

Définition 7 : Un bloc

Un bloc BK_j^i est défini comme la jème sous-trajectoire de la ième trajectoire.

$$BK_j^i = \{y_i\}_{t_j}^{t_j+L-1}$$

Où « L » est la longueur du bloc et $t_j \in [1, ..., |y_i|]$est la position de départ du bloc.

Le poids associé à chaque bloc sera de la forme :

$$w(j) = \frac{exp(\frac{j}{2N^2\sigma^2})}{\sum_{i=1}^{N} exp(\frac{i}{2N^2\sigma^2})}, \tag{3.9}$$

où $w(j)$ est le poids associé au j^{ieme} bloc de la trajectoire et σ est le paramètre d'écart qui contrôle la façon dont les poids sont distribués, la figure 3.14 montre l'influence du paramètre sur cette distribution. On remarque que lorsque sigma vaut 1, les poids de chacun des blocs seront presque égaux. Une valeur de sigma inférieure à 1 donne plus d'importance aux derniers blocs. Cette importance est d'autant plus grande que sigma se rapproche de zéro.

Lorsque on compare deux trajectoires, T_{ts}, T_{tr} respectivement, les blocs de même position seront comparés entre eux.

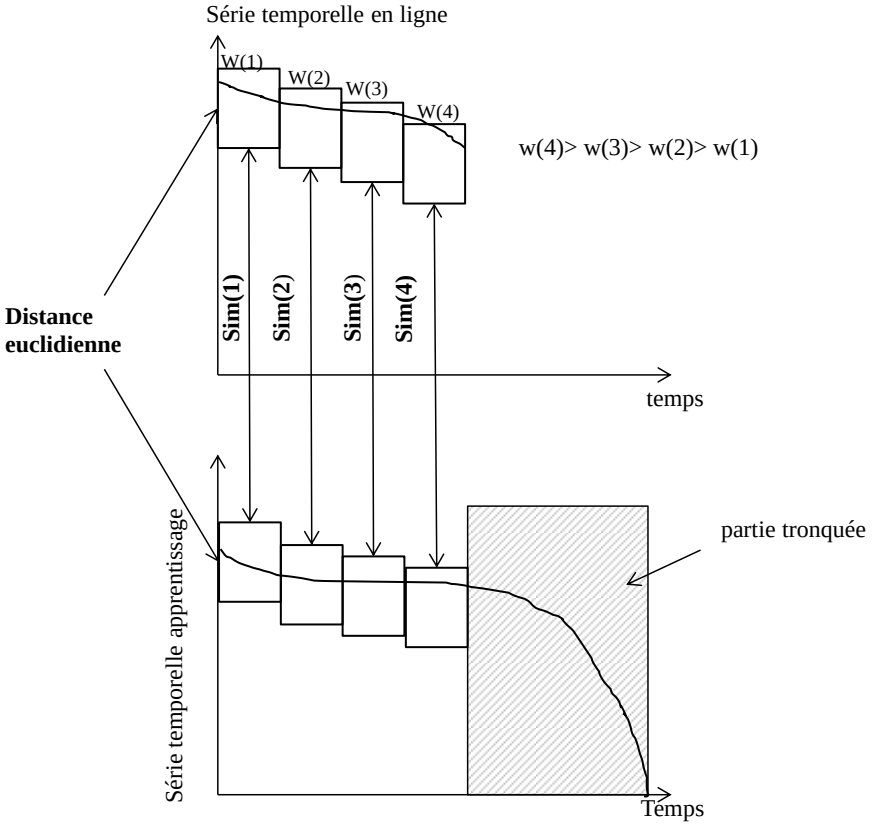


FIGURE 3.13 – Illustration de la similarité pondérée.

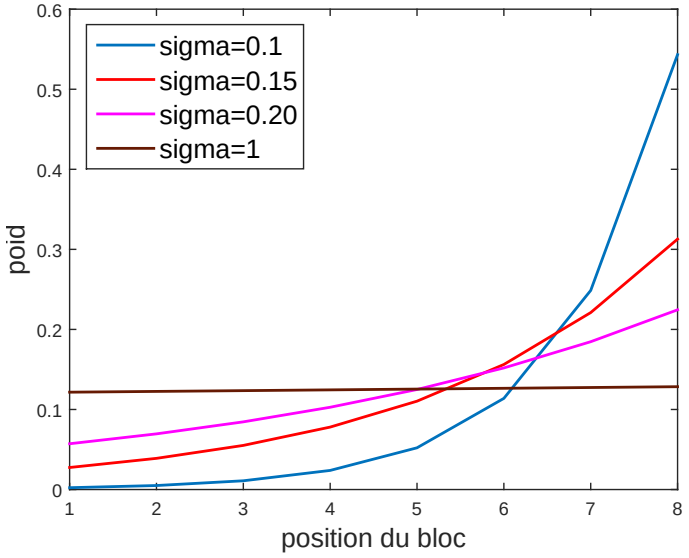


FIGURE 3.14 – Influence de sigma sur la distribution du poids.

On laisse une marge de dissemblance lors du calcul de la similarité entre les blocs en introduisant un facteur relaxant λ (cf. équation 3.10). La figure 3.15 montre l'influence de λ sur le score de similarité entre deux trajectoires de la même longueur pour un σ fixé à 0.15. Plus λ est grand, plus élevé est le score. La valeur de λ doit être choisie à bon escient. Des valeurs trop petites signifient une mesure de similarité très stricte alors que

de grandes valeurs signifient une similarité trop relaxée.

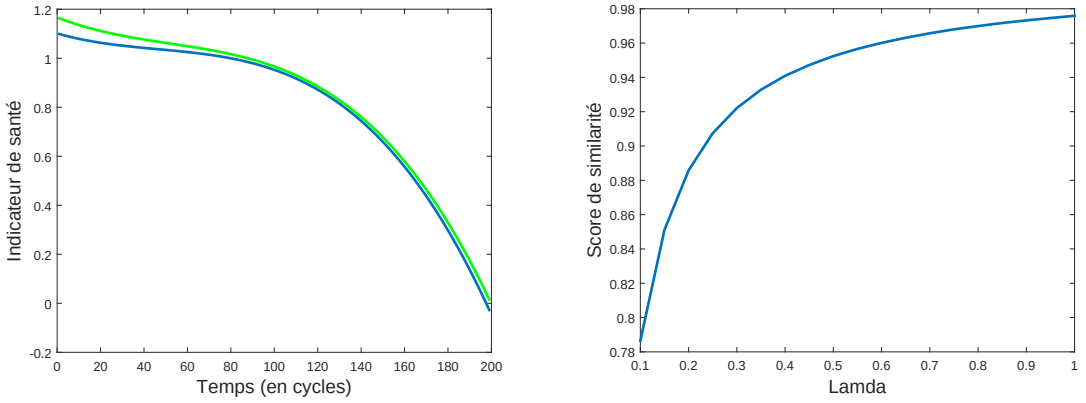
$$sim(j) = exp\{\frac{-D_j}{\lambda.L}\}, \tag{3.10}$$

où $sim(j)$ est la similarité bloc entre le j^{ieme} bloc de test et son bloc d'apprentissage associé. D_j est la distance euclidienne entre les éléments de ces deux blocs, L est la taille du bloc.

Le score de similarité est une valeur comprise entre 0 et 1, 0 pour des trajectoires complètement dissemblables et 1 pour des trajectoires identiques :

$$SC = \sum_{j=1}^{N} w(j)sim(j). \tag{3.11}$$

L'algorithme 1 commence par ajuster la longueur de la trajectoire d'apprentissage afin qu'elle ait la même longueur que la trajectoire de test. Les valeurs de trajectoire d'apprentissage qui dépassent la longueur de la trajectoire de test sont ainsi tronquées (voir figure 3.13). Ensuite la similarité entre chaque paire de blocs de même position est calculée (c.f. Équation 3.10) et le score final de la similarité est calculé comme une moyenne pondérée des similarités entre blocs (c.f. Équation 3.11)



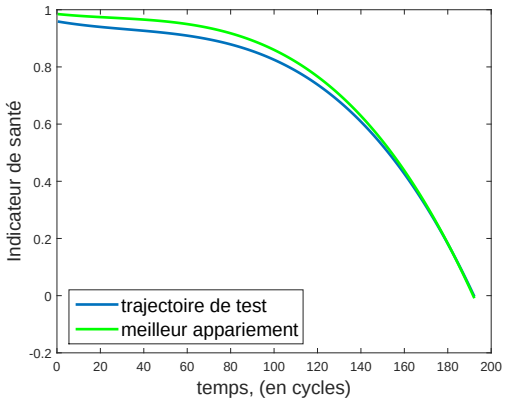
(a) Deux trajectoires similaires

(b)Score de similarité de ces deux trajectoires en fonction de λ

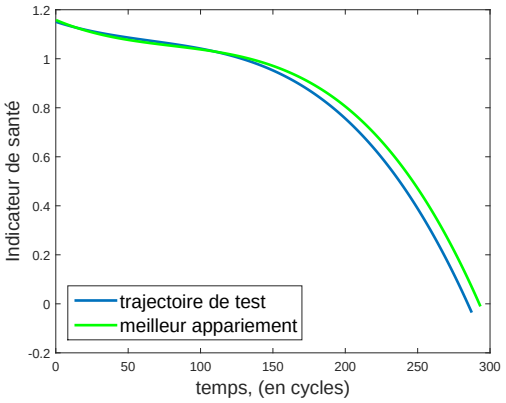
FIGURE 3.15 – Influence de λ sur le score de la similarité.

La figure 3.16 montre un exemple de paires de trajectoires les plus similaires en utilisant la similarité pondérée. Ces paires sont les mêmes que celles obtenues en utilisant la distance euclidienne minimale. Afin d'examiner la différence entre ces deux mesures, on prend des trajectoires plus petites. Cette fois, au lieu de considérer des trajectoires qui vont jusqu'à la défaillance, on les tronque 50% avant, λ et σ étant fixés à 0.15.

La figure 3.17 montre les trajectoires les plus similaires pour la même nouvelle trajectoire. Sur la gauche on voit la trajectoire obtenue en utilisant la distance euclidienne et sur la droite on voit la trajectoire obtenue en utilisant la mesure pondérée. On observe qu'avec la distance euclidienne les trajectoires sont similaires au début et s'éloignent vers la fin alors que la similarité pondérée favorise la similarité vers la fin. Cette dernière est plus adaptée au pronostic. Toutefois, en pronostic, la surveillance des composants ne commence pas au début de leur vie, et leurs états peuvent être plus ou moins dégradé. Leurs



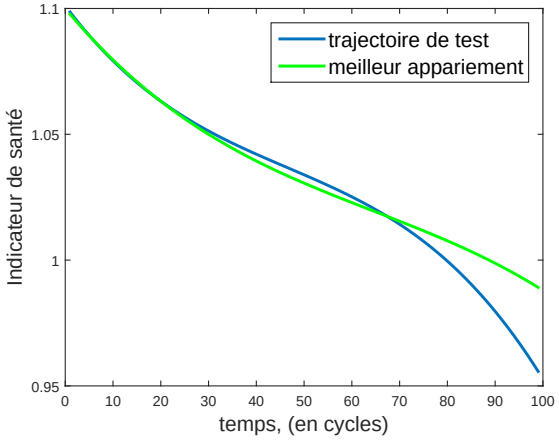
(a) score(test, meilleur appariement)=0.8664



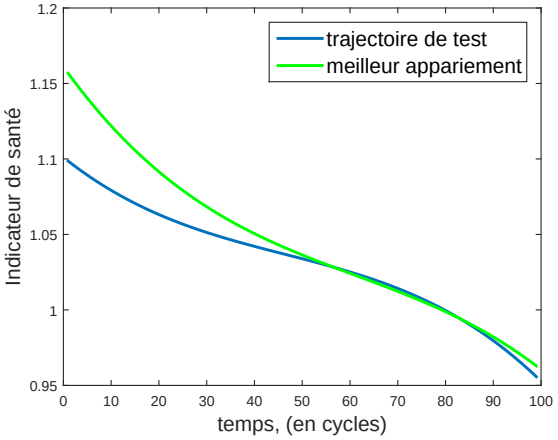
(b) Score(test, meilleur appariement)=0.7171

FIGURE 3.16 – Exemple de paires de trajectoires similaires avec la similarité pondérée.

profils de dégradation ne sont pas les mêmes pour tous les éléments. Certain composants peuvent se détériorer plus vite que d'autres. Ceci nous mène à développer une autre mesure de similarité qui abordera ces différents points.



(a) meilleur appariement en utilisant la distance euclidienne



(b) meilleur appariement en utilisant la similarité pondérée

FIGURE 3.17 – Comparaison entre la distance euclidienne et la similarité pondérée.

3.4.3/ SIMILARITÉ PONDÉRÉE AVEC UNE PROJECTION TEMPORELLE

Data : - Trajectoires de test et d'apprentissage
- Valeurs de L, λ, σ et seuil

Result : - Score de similarité entre les trajectoires
- Fin de la similarité
- NBS

Initialisation :
- Initialiser la Position de Début de Balayage, PDB=1
- Initialiser le nombre de succès, *hit*=0
- Initialiser le Nombre de Blocs Similaires, NBS=0
- Diviser la trajectoire de test en N blocs
- Diviser la trajectoire d'apprentissage en M blocs qui se chevauchent de 1 en 1

for $i \leftarrow 1$ **to** N **do**
  for $j \leftarrow 1$ **to** M **do**
    Calculer la similarité entre les blocs,
    $sim(i,j) = exp\{\frac{-D_{i,j}}{\lambda.L}\}$;
    if similarité $\geq$ seuil **then**
      PDB=j %Commencer le prochain balayage à la position actuelle du bloc d'apprentissage;
      incrémenter *hit* %un bloc similaire a été trouvé ;
    else
      PDB^{new}=PDB^{old} % Gardez la position précédente de début de balayage ;
    end
    if $hit \neq 0$ **then**
      incrémenter NBS
    end
  end
  Sauvegarder max $(sim_{i,j})$
end
Calculer le score de similarité,
$SC = \frac{NBS}{M}.\sum_{i=1}^{N} w(i).max(sim_{i,j})$, ;
Sauvegarder la fin de la similarité. % dernier bloc similaire de la trajectoire d'apprentissage.;

Algorithme 2 : Algorithme de la similarité pondérée avec une projection temporelle.

Bien que la similarité pondérée par bloc soit la mieux adaptée au pronostic, elle ne permet pas de superposer la nouvelle trajectoire sur celles d'apprentissage. Par conséquent, nous proposons de faire évoluer cette mesure en faisant une projection temporelle de la nouvelle trajectoire sur les trajectoires d'apprentissage. Cela nous permettra de localiser l'état actuel du nouveau composant en identifiant le bloc similaire translaté dans le temps comme le montre la figure 3.18.

Ce test de similarité mesure la ressemblance entre les évolutions par rapport au temps des séries temporelles considérées. Il est à noter que cette similarité fournit un moyen de comparaison entre des trajectoires de longueurs différentes tout en tenant compte de tout l'historique de la nouvelle trajectoire et en donnant plus d'importance aux cycles opérationnels tardifs.

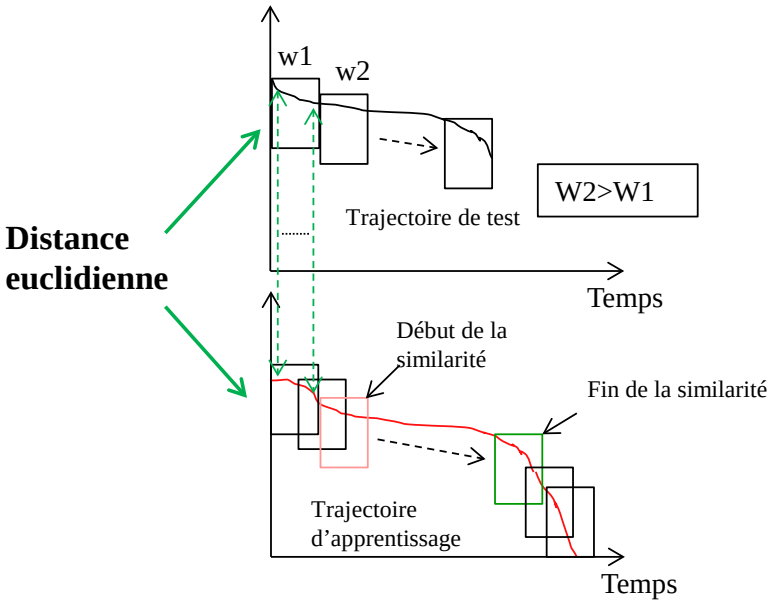


FIGURE 3.18 – Illustration de la similarité pondérée avec projection temporelle.

L'algorithme de similarité 2 schématisé à la figure 3.18 commence par diviser la trajectoire de test en N blocs. Chaque bloc est balayé sur toute la trajectoire d'apprentissage. La similarité des blocs est calculée comme dans la section 3.4.2. Le premier bloc similaire, c'est-à-dire le bloc d'apprentissage qui a la valeur de similarité la plus élevée, indique le début de la similarité et le dernier bloc similaire, c'est-à-dire le bloc d'apprentissage qui est similaire au dernier bloc de test indique la fin de la similarité et détecte ainsi la position actuelle sur l'axe de temps de l'élément d'apprentissage (la projection temporelle). Tout comme dans la section 3.4.2, chaque bloc est donné un poids d'une manière qui favorise les blocs tardifs.

Le score de similarité est donné par l'équation 3.12

$$SC = \frac{NBS}{M} . \sum_{i=1}^{N} w(i).max(sim_{i,j}), \tag{3.12}$$

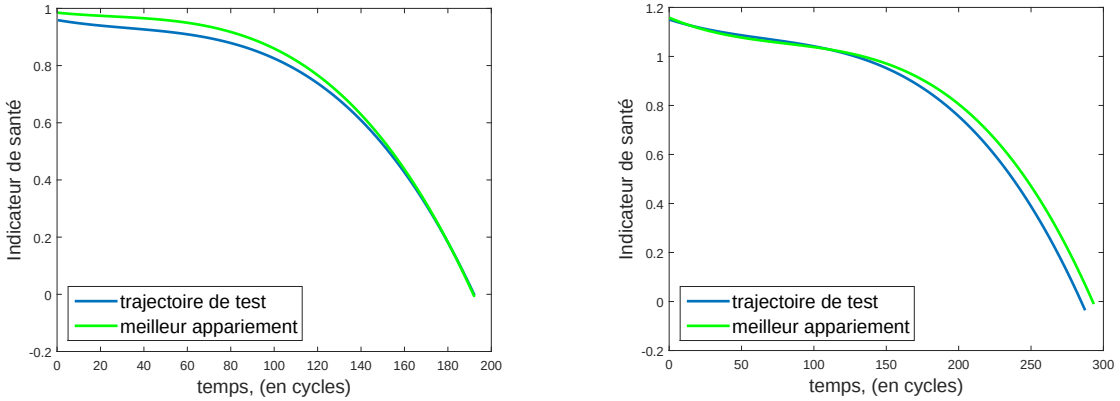
où NBS est le nombre de blocs similaires. M est le nombre total de blocs de trajectoire d'apprentissage, max $(sim_{i,j})$ est la valeur de similarité entre le i^{eme} bloc test et sont bloc d'apprentissage "j" le plus proche.

La figure 3.19 montre un exemple de paires de trajectoires les plus similaires en utilisant la similarité pondérée avec une projection temporelle. Ces paires sont les mêmes que celles obtenues en utilisant la distance euclidienne minimale. La performance de cette mesure de similarité se voit mieux sur des trajectoires tronquées, λ et σ sont fixés à 0.15.

La figure 3.20 montre les résultats obtenus en tronquant les trajectoires à 50%. À gauche, on trouve les paires obtenues avec la distance euclidienne, au milieu,pour les mêmes trajectoires de test, on retrouve les trajectoires similaires en utilisant la similarité pondérée et à droite on les retrouve en utilisant la similarité pondérée avec une projection temporelle. On constate qu'on est plus précis avec cette dernière.

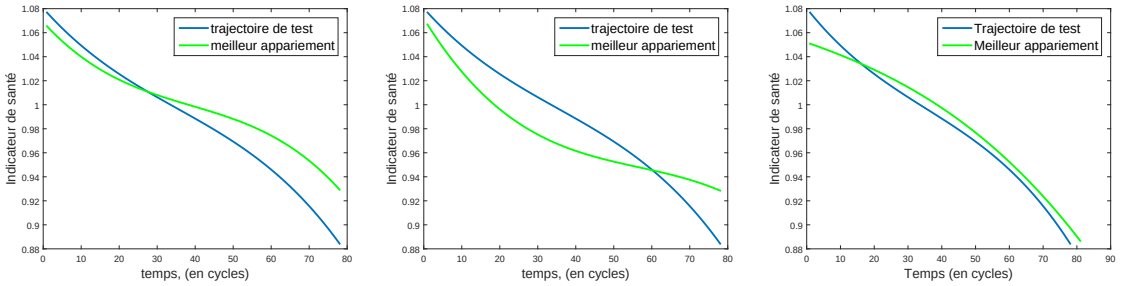
La figure 3.21 montre comment des sous-trajectoires de test sont localisés sur les exemples d'apprentissage identifiant ainsi la projection de la dégradation sur l'axe de

temps des éléments d'apprentissage.



(a) score(test, meilleur appariement)=0.9336 (b) Score(test, meilleur appariement)=0.7565

FIGURE 3.19 – Exemple de paires de trajectoires similaires avec la similarité pondérée avec projection temporelle.



(a) (b) (c)

FIGURE 3.20 – Les paires similaires obtenues (a) avec la distance euclidienne, (b) la similarité pondérée et (c) la similarité pondérée avec projection temporelle.

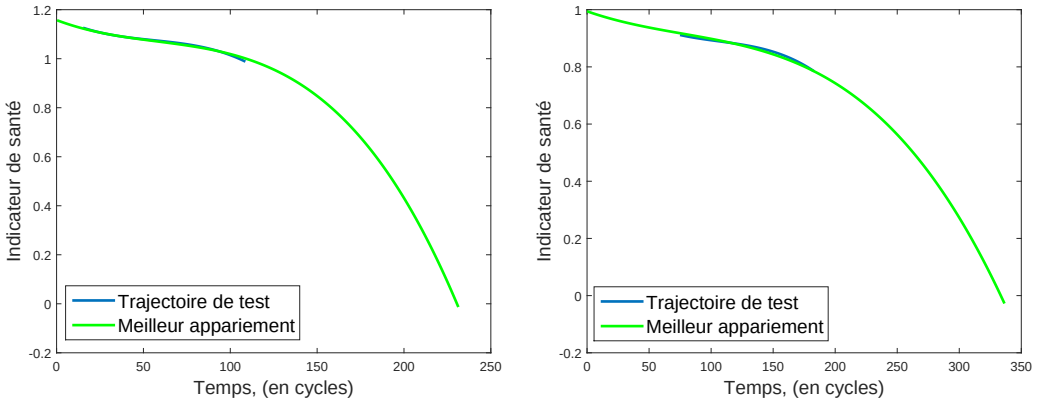


FIGURE 3.21 – Localisation de la trajectoire test sur la trajectoire d'apprentissage la plus similaire.

Nous vous présentons un tableau de comparaison entre ces différentes mesures de similarité (c.f. Tableau 3.1). Nous constatons que la mesure pondérée avec projection temporelle présente tous les atouts, qu'on est à même d'attendre pour des méthodes de

TABLE 3.1 – Comparaison entre les trois mesures de similarité.

caractéristiques	Distance euclidienne	Similarité pondérée	Similarité pondérée avec projection temporelle
Utilisation de l'ensemble entier des données	√	√	√
Plus d'importance aux observations tardives	×	√	√
Possibilité de comparer des séries temporelles de longueurs différentes (sans les tronquer)	×	×	√
Identification de l'état courant	×	×	√

pronostic, à savoir, l'identification de l'état courant, le traitement de séries temporelles de différentes longueurs et grâce aux jeux de pondération, on peut privilégier les observations se rapprochant de la défaillance.

3.5/ ESTIMATION DE L'ÉTAT COURANT ET DU RUL ASSOCIÉ

Définition 8 : RUL

Le RUL (Remaining useful Life) est défini comme la durée de vie résiduelle avant que le composant perde ses fonctionnalités et atteigne la défaillance, autrement dit la fin de vie (EOL : End Of Life).

Soit $I = \{I_i\}$, la librairie des instances . Soit $I_i = \{D_1^i, D_2^i, \ldots, D_{l(T_i)}^i, EOl_i\}$ l'expérience représentée par un vecteur de descripteurs extraites à partir de la série temporelle T_i composée d'échantillons indexés par leur numéro de cycle. EOL désigne la fin de vie et la solution à l'expérience. Il est égal à $l(T_i)$.

Soit $I_p = \{D_1^p, D_2^p, \ldots, D_{t_c}^p, ?\}$, le cas problème. t_c désigne le nombre du dernier cycle de mesure. Le problème d'estimation du RUL peut être défini comme le calcul du $RUL = EOL - t_c$ d'une instance de test, où EOL est la solution de K plus proches voisins, étant donné que :

1. L'historique mis-à-jour de vecteur de descripteur à partir de l'instance de test $I_p = \{D_1^p, D_2^p, \ldots, D_{t_c}^p\}$ est disponible.
2. L instances d'apprentissage du même type que le système avec un historique complet de vecteurs caractéristiques $I_k = \{D_1^k, D_2^k, \ldots, D_{l(T-i)}^k, EOl_k\}$, (k=1, . . .,L) sont disponibles.
   Pour une instance de test donnée, le RUL est prédit en utilisant les instances récupérées d'apprentissage. La bibliothèque contient des instances avec des durées de vie connues. Une fois que l'instance de test arrive, le critère de similarité pondérée avec projection temporelle nous aide à récupérer les k plus similaires instances et identifier l'état courant et t_c afin de calculer le RUL (voir figure 3.22).

Pour l'instance de test, le RUL est calculé comme la moyenne des RUL des K plus

proches instances d'apprentissage(cf. eq 3.13).

$$LeRULPredit = \frac{1}{k}\sum_{i=1}^{k} RUL(i) \tag{3.13}$$

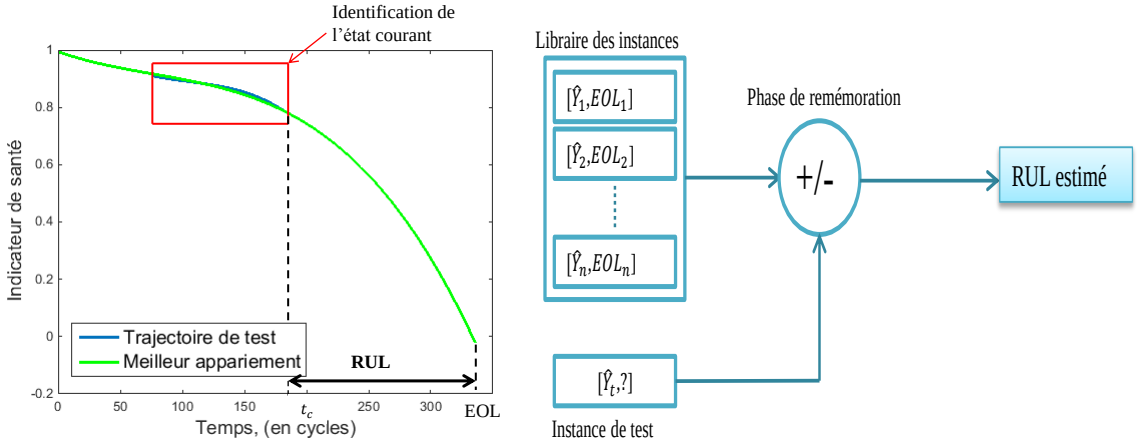


FIGURE 3.22 – Formalisation du RUL.

La bibliothèque des instances contient des expériences avec des EOL connus. Ce qui nous évite la question problématique de définition de seuil qui est l'un des verrous actuels du pronostic. Un seuil statique (le cas le plus commun) n'est pas pratique. Cette estimation du RUL nécessite d'être évaluée. Nous allons passé en revues les différentes métriques d'évaluation définies pour le pronostic à la prochaine section.

3.6/ LES MÉTRIQUES D'ÉVALUATION

[Saxena et al., 2008a] recensent les différents métriques utilisées afin d'évaluer les approches de pronostic. Il existe trois types de métriques : (i) des métriques de performance des algorithmes, (ii) des métriques de performance computationnelles, (iii) et des métriques de coûts-avantages (cost-benefit).

La plupart des métriques trouvées dans la littérature appartiennent à la catégorie des métriques de performance des algorithmes. Les publications du domaine de pronostic n'utilisent pas les critères de performance computationnelles et des coûts-avantages. Dans nos travaux, comme la majorité des travaux de pronostic, on s'intéresse aux métriques de performance des algorithmes et surtout aux métriques applicables dans notre cas (certaines nécessitent d'avoir accès aux valeurs exactes point par point, information qui n'est pas toujours disponible). Ces métriques peuvent être mesurées en évaluant l'erreur entre les RUL prédits et exacts.
Dn plus, on s'intéresse à des métriques adaptées au pronostic telles que l'horizon de pronostic [Saxena et al., 2008a] et le pourcentage des prédictions acceptables [Ramasso et al., 2013].
On note $\hat{RUL}_t$ la valeur prédite au temps t et RUL_t la valeur réelle du RUL au temps t. Les métriques que nous utiliserons sont détaillées ci-dessous.

3.6.1/ HORIZON DU PRONOSTIC

L'horizon de la prédiction tel que l'a défini [Saxena et al., 2008a] est la différence entre le premier indice de temps P qui satisfait un critère défini et la valeur EOL.

$$HP = EOL - P \tag{3.14}$$

L'exigence de performance peut être définie en permettant une borne d'erreur α autour de la valeur réelle d'EOL. On peut ainsi définir P par $P = min(t, telque : RUL_t - \alpha * EOL \leq \hat{RUL}_t \leq RUL_t + \alpha * EOL)$.

Plus l'horizon de pronostic est large, meilleure est la prédiction. Ce critère d'évaluation donne une idée sur le comportement du modèle prédictif par rapport au temps. Toutefois, ne considérer que P pour évaluer l'approche peut conduire à une évaluation partielle car on néglige la performance de ce qui vient après. Pour des raisons d'exhaustivité, d'autres critères sont utilisés.

3.6.2/ POURCENTAGE DE PRÉDICTIONS ACCEPTABLES

Une prédiction est considérée comme correcte si son erreur appartient à l'intervalle des erreurs acceptables. Pour évaluer le travail présenté ici, l'intervalle a été défini comme $I = [-10, 13]$. L'intervalle est asymétrique parce que les prédictions précoces sont plus tolérables par rapport aux prédictions tardives. L'intervalle est considéré comme une condition sévère par rapport à la littérature [Goebel et al., 2005].

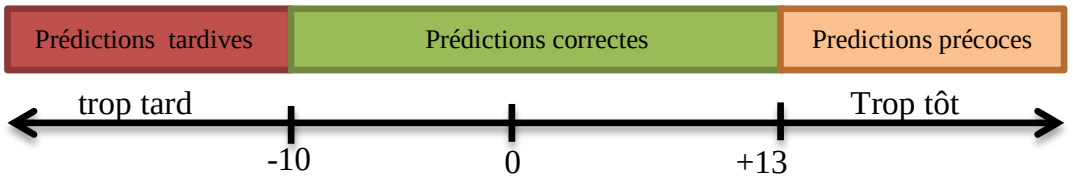


FIGURE 3.23 – Métrique de pourcentage de prédictions acceptables.

3.6.3/ ERREUR MOYENNE ABSOLUE (EMA)

L'erreur moyenne absolue est définie comme la moyenne des erreurs de prédictions pour plusieurs prédictions faites pour différents composants de test avec le même horizon de prédiction

$$EMA = \frac{1}{m} \sum_{i=1}^{m} |\hat{RUL}(i) - RUL(i)|, \tag{3.15}$$

où m est le nombre total des prédictions faites.

3.6.4/ Erreur de Pourcentage en Moyenne absolue

L'erreur de pourcentage en moyenne absolue est calculée pour toutes les prédictions de chaque composant de test. L'EPMA totale est calculée comme :

$$EPMA_t = \frac{1}{n}\sum_{i=1}^{n} MAPE_i, \tag{3.16}$$

où $L'EPMA_t$ est l'EPMA moyenne pour l'ensemble de composant de test, $MAPE_i = \frac{100}{L_i}\sum_{j=1}^{L_i}\frac{|EOL_j - \hat{EOL_j}|}{EOL_j}$ est l'EPMA pour le i^{eme} composant de test. EOL est la valeur réelle et $\hat{EOL}$ est celle prédite.

3.6.5/ Précision relative cumulative

Pour évaluer la précision relative à de multiples instances de temps, PRC est définie comme la moyenne des précisions relatives aux instances de temps données (cf. équations 3.17 et 3.18) :

$$PR(i) = 1 - \frac{|EOL(i) - \hat{EOL}(i)|}{EOL} \tag{3.17}$$

$$PRC = \frac{1}{n}\sum_{i=1}^{n} PR(i) \tag{3.18}$$

3.6.6/ L'écart type de l'échantillon

Ceci est une mesure de la dispersion de l'erreur par rapport à la moyenne d'échantillon de l'erreur, l'équation

$$ETE = \sqrt{\frac{\sum_{i=1}^{n}(\Delta(i) - m)^2}{n-1}}, \tag{3.19}$$

où $\Delta = EOL(i) - \hat{EOL}(i)$ et m est la moyenne de l'échantillon.

3.7/ Application

3.7.1/ Ensemble de données des turboréacteurs

Nous avons appliqué nos approches IBL à la prédiction de la durée de vie résiduelle des composants critiques sur une simulation de dégradation de turboréacteurs [Saxena et al., 2008b]. L'ensemble de données est une simulation de la propagation du dommage des moteurs à turbine à gaz des aéronefs et est disponible le site web de la NASA [2].

L'ensemble de données proposé correspond à 100 moteurs décrits chacun par 26 variables (features) évoluant du début de vie jusqu'à la défaillance. Chaque moteur est

2. http ://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository/ ()consulté dernièrement le 07/10/2015)

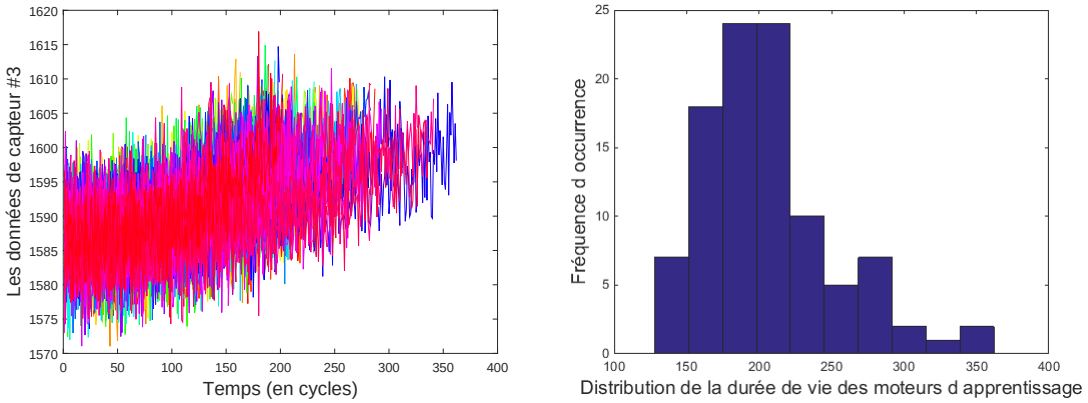


FIGURE 3.24 – Illustration des données turboréacteurs.

représenté par une série temporelle multivariée. Au début, les moteurs fonctionnent normalement et finissent par développer une faute préalable à la défaillance.

L'ensemble de données choisi est bruité et caractérisé par une seule condition de fonctionnement et un seul mode de défaillance. Il est constitué de trois fichiers : "TRAIN_FD001", "TEST_FD001" et "RUL_FD001".

- "TRAIN_FD001" est composé de 100 séries temporelles multivariées (26 features) représentant la propagation de 100 moteurs d'apprentissage où chaque moteur passe par le processus complet de dégradation.
- "TEST_FD001" est aussi compose de 100 séries temporelles. Toutefois, ces séries temporelles ne sont quune partie multivariées à 26 features, des séries temporelles.
- "RUL_FD001" contient les valeurs de durée de vie résiduelle (RUL) pour les éléments de test. L'objectif est de prédire ces valeurs pour chaque moteur de test.

La figure 3.24 montre un exemple des données capteurs utilisées ainsi que la distribution de durée de vie des moteurs. Comme on peut le voir sur la figure 3.24, les données sont corrompues par le bruit et les moteurs ont une large plage de durée de vie qui varie entre 128 et 362 cycles. Cela rend difficile la tâche de prédiction du RUL pour ces moteurs.

3.7.1.1/ SÉLECTION DES DONNÉES CAPTEURS

La sélection de capteurs utilisés pour construire les instances (indicateurs de santé et trajectoires de dégradation) est faite à partir des observations de chaque turboréacteur. Comme le montre la figure 3.25-a, quelques capteurs ont des valeurs constantes pour tous les composants d'apprentissage. D'autres capteurs ont des valeurs constantes tout au long du cycle de vie pour le même composant (figure 3.25-b).
Ce type de capteurs ne montrant aucun signe de dégradation n'est pas utile pour notre étude. Le reste présente une tendance monotone pendant la durée de vie du moteur. Cependant, parmi les moteurs servant à l'apprentissage, certains d'entre eux montrent des tendances de fin de vie incompatible. Ces tendances sont illustrées à la figure 3.25-c.

Dans ce travail, comme suggéré par [Wang et al., 2008], on ne retiendra que les capteurs ayant des tendances croissantes de façon monotone (figure 3.25-d). Ces capteurs sont indexés au fichier "TRAIN_FD001" par les numéros : $7, 8, 9, 13, 16$ [Ramasso et al., 2013].

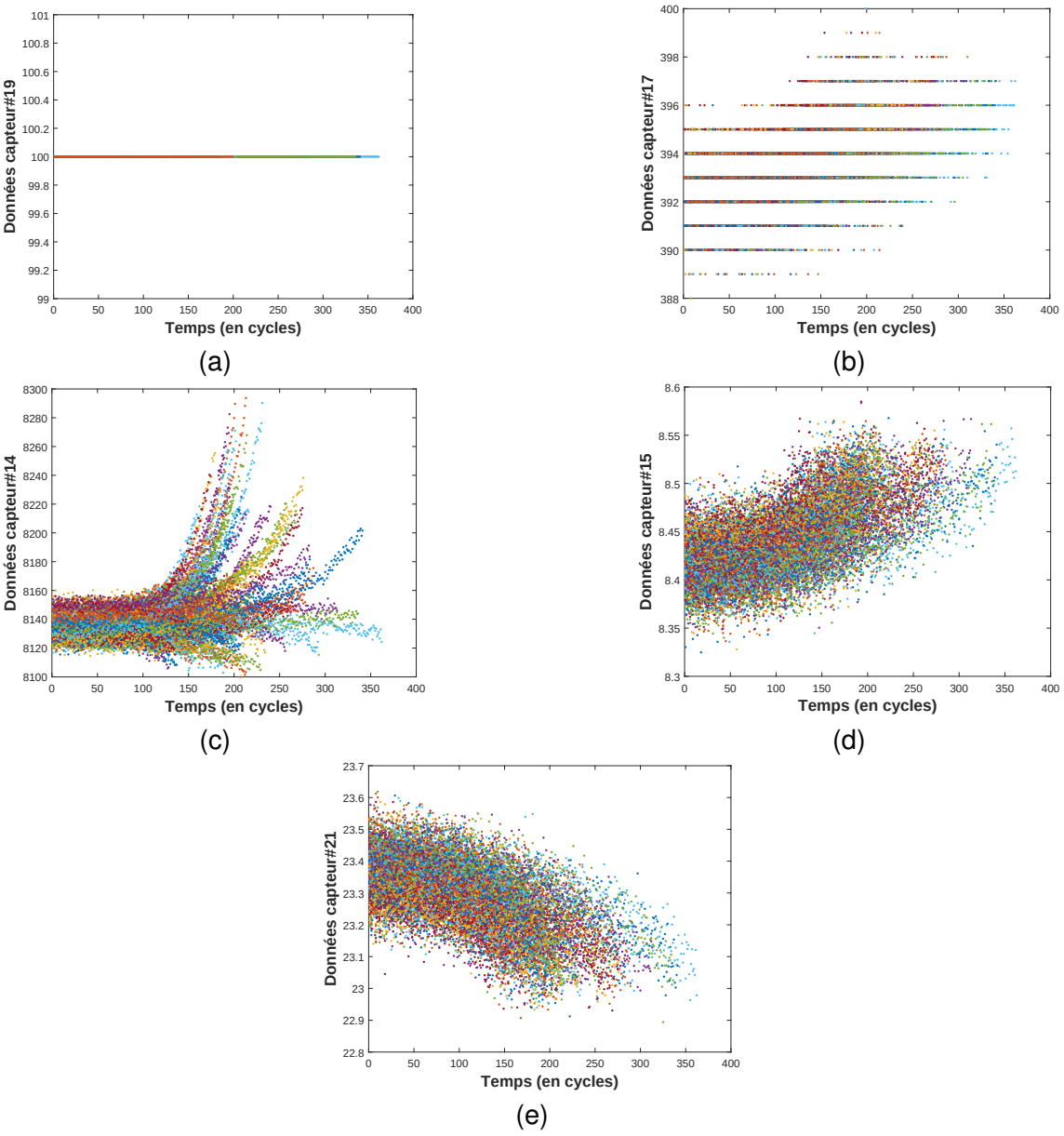


FIGURE 3.25 – Illustration de types de données capteurs.

3.7.1.2/ RÉSULTATS ET DISCUSSION

- *Paramètres*

Nous avons sélectionné les cinq capteurs ayant des tendances croissantes parmi les 21 disponibles afin de construire soit les indicateurs de santé ou les trajectoires de dégradation (méthode décrite à la section 3.3.1.2). Ces séries temporelles seront utilisées afin d'estimer le RUL en utilisant la mesure de similarité pondérée avec une projection temporelle (section 3.4.3). Cette similarité nécessite la définition des paramètres suivants :

— L : la taille du bloc, définition 7.
— σ : le paramètre d'écart, eq 3.9.
— λ : le facteur relaxant,eq 3.10.

— $seuil$: le seuil de similarité entre les blocs (algorithme 2).

La taille de bloc et le seuil sont fixés à 30 et 0.7 respectivement. Le reste des paramètres est plus important et sera réglé par une méthode essai-erreur.

1. **Application supervisée - Indicateurs de santé :** Dans cette section, on examinera la performance de l'approche proposée en utilisant les indicateurs de santé.
   - *Réglage des paramétres*

   Le tableau 3.2 présente les résultats obtenus en testant sur les 100 moteurs du fichier "test" et en appliquant le pourcentage de prédictions acceptables comme critère d'évaluation.

   Le choix de σ et λ est obtenu par essais-erreurs. On commence d'abord par fixer $\lambda = 0.25$ afin de chercher les meilleurs valeurs de σ. Selon le tableau, $\sigma = 0.7$. Une fois ce choix fait, on cherche à optimiser les valeurs de λ.
   - *La prédiction*

   La meilleure performance est obtenue en utilisant k=9, $\lambda = 1.25$ et $\sigma = 0.7$.
   La figure 3.26 détaille les résultats de cette prédiction. À partir de la figure 3.26-a, on constate que les valeurs prédites du RUL sont assez proches de valeurs réelles. La figure 3.26-b, illustre l'histogramme de l'erreur. Le pourcentage de prédictions correctes est de 58% et l'erreur est généralement entre -45 et 49 sauf pour trois moteurs où l'erreur dépasse -67. Il convient de mentionner que la prédiction pour ces moteurs est faite à moins de 50% du cycle de vie total.
2. **Application non supervisée- Trajectoires de dégradation :** Dans cette section, on examinera la performance de l'approche proposée en utilisant les trajectoires de dégradation.
   - *Réglage des paramètres*

   Le tableau 3.3 présente les résultats obtenus en testant sur les 100 moteurs du fichier "test" et en appliquant le pourcentage de prédictions acceptables comme critère d'évaluation. La prédiction du RUL est faite suivant la méthode décrite à la section 3.5 et en considérant k plus proches voisins.

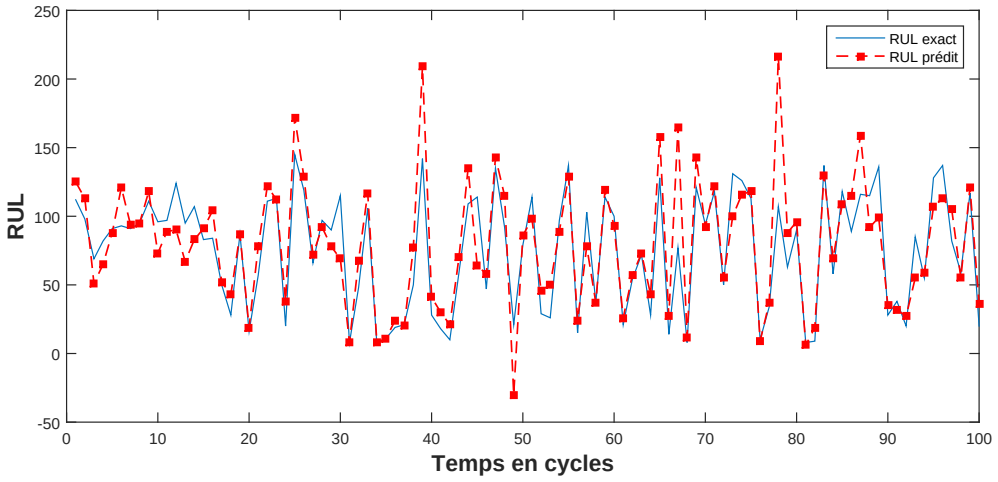
   Le choix de σ et λ est obtenu par essais-erreurs. On commence d'abord par fixer $\lambda = 1$ afin de chercher les meilleurs valeurs de σ. Selon le tableau, $\sigma = 0.1$. Une fois ce choix fait, on cherche à optimiser les valeurs de λ.
   - *La prédiction*

   La meilleure performance est obtenue en utilisant k=15, $\lambda = 1$ et $\sigma = 0.1$. La figure 3.28 montre les détails de cette prédiction. Le pourcentage de prédictions correctes est de 60% et l'erreur est généralement entre -65 et 54. On remarque que les erreurs qui dépassaient -67 en utilisant les indicateurs de santé ont disparu ce qui rend l'approche plus performante.

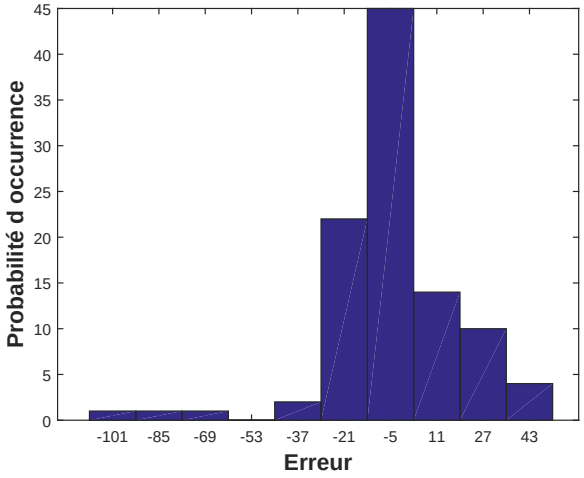
   Nous avons comparé cette approche UKR de réduction de dimension de données (par la fusion des données capteur en une courbe), à une autre méthode de réduction inspirée des travaux de [Mosallam et al., 2013].

   La réduction des données s'est faite par ACP (Analyse en composantes principales) en un signal unidimensionnel.

   Nous avons appliqué la méthode des k plus proches voisins et notre mesure de similarité, avec le même calcul du RUL. Nous avons comparé les performances de ces deux méthodes. La figure 3.27 montre les résultats obtenus. La performance de l'UKR est supérieure en tout point à celle de l'ACP. Ceci peut être dû, probablement au fait que la courbe obtenue avec l'ACP ne représente pas toutes les informations requises.



(a)



(b)

FIGURE 3.26 – L'erreur sur l'estimation du RUL en utilisant les indicateurs de santé, (a) l'histogramme de l'erreur, (b) RUL exact Vs RUL prédit.

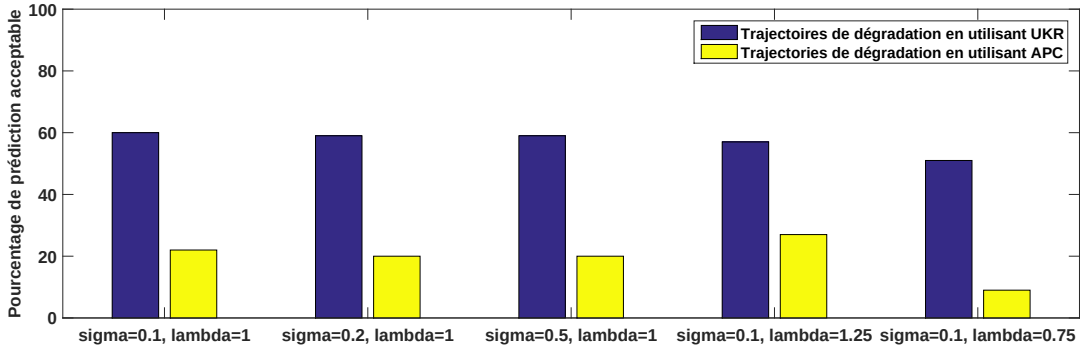


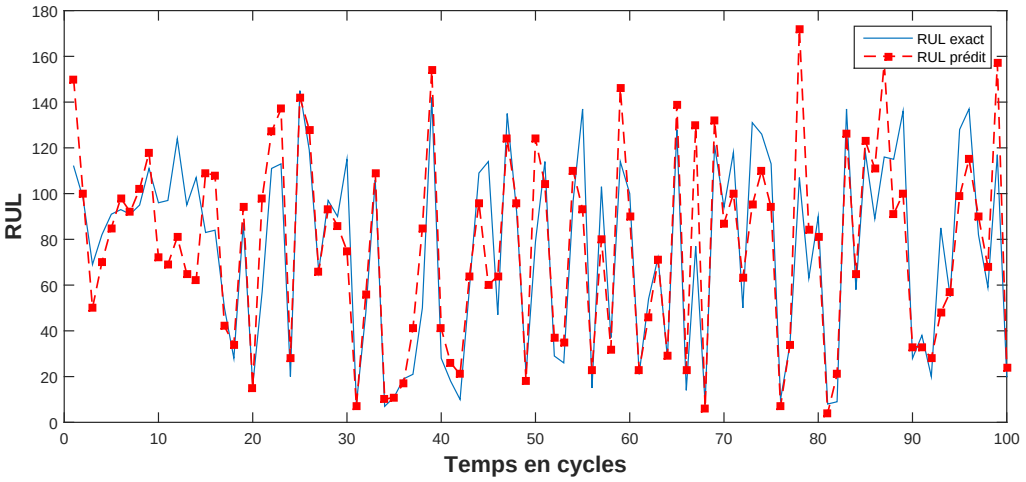
FIGURE 3.27 – Comparaison entre la prédiction en utilisant l'UKR (trajectoires de dégradation) et APC (réduction de dimension).

TABLE 3.2 – Résultats obtenus pour les indicateurs de santé.

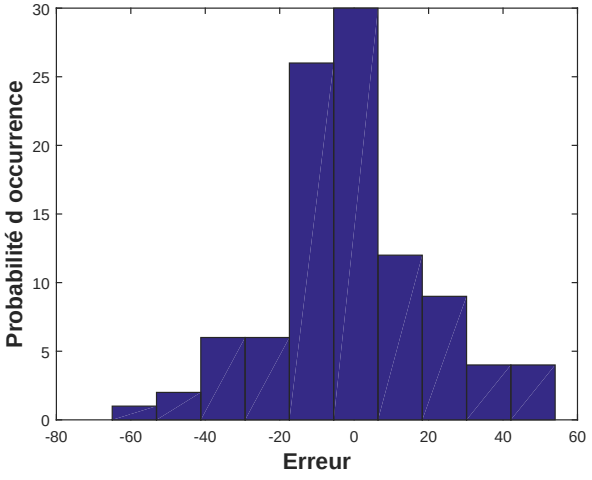
k	σ	λ	prédictions correctes %	Prédictions précoces %	Prédictions tardives%
5	0.5	0.25	45	35	20
	0.15	0.25	45	34	21
	0.7	0.25	50	29	21
	0.7	1.25	46	35	19
	0.7	1	46	35	19
	0.7	0.5	46	35	19
	0.7	0.15	48	36	16
7	0.5	0.25	49	31	20
	0.15	0.25	49	30	21
	0.7	0.25	55	24	21
	0.7	1.25	56	24	20
	0.7	1	56	24	20
	0.7	0.5	55	24	21
	0.7	0.15	50	24	23
9	0.5	0.25	56	24	20
	0.15	0.25	52	25	23
	0.7	0.25	57	25	18
	0.7	1.25	**58**	24	18
	0.7	1	**58**	24	18
	0.7	0.5	57	24	19
	0.7	0.15	**58**	22	20
11	0.5	0.25	**58**	19	23
	0.15	0.25	53	26	21
	0.7	0.25	**58**	20	22
	0.7	1.25	**58**	21	21
	0.7	1	57	21	22
	0.7	0.5	**58**	20	22
	0.7	0.15	56	18	26
13	0.5	0.25	**58**	19	23
	0.15	0.25	**58**	19	23
	0.7	0.25	56	22	22
	0.7	1.25	56	22	22
	0.7	1	56	22	22
	0.7	0.5	56	22	22
	0.7	0.15	56	19	25
15	0.5	0.25	54	21	25
	0.15	0.25	57	20	23
	0.7	0.25	53	22	25
	0.7	1.25	53	23	24
	0.7	1	53	23	24
	0.7	0.5	54	22	24
	0.7	0.15	53	20	27

TABLE 3.3 – Résultats obtenus pour les trajectoires de dégradation.

k	σ	λ	prédictions correctes %	Prédictions précoces %	Prédictions tardives%
5	0.1	1	43	33	24
	0.2	1	46	34	20
	0.5	1	45	34	21
	0.1	1.25	45	33	22
	0.1	0.75	43	36	21
7	0.1	1	48	27	25
	0.2	1	49	29	22
	0.5	1	51	28	21
	0.1	1.25	53	25	22
	0.1	0.75	50	27	23
9	0.1	1	53	25	22
	0.2	1	56	24	20
	0.5	1	54	27	19
	0.1	1.25	54	26	20
	0.1	0.75	52	27	21
11	0.1	1	54	23	23
	0.2	1	56	23	21
	0.5	1	**58**	23	19
	0.1	1.25	57	24	19
	0.1	0.75	56	26	18
13	0.1	1	57	21	22
	0.2	1	**58**	21	21
	0.5	1	**59**	21	20
	0.1	1.25	55	25	20
	0.1	0.75	54	26	20
15	0.1	1	**60**	17	23
	0.2	1	**59**	18	23
	0.5	1	**59**	19	22
	0.1	1.25	57	21	22
	0.1	0.75	51	25	24



(a)



(b)

FIGURE 3.28 – L'erreur sur l'estimation du RUL en utilisant les trajectoires de dégradation, (a) l'histogramme de l'erreur, (b) RUL exact Vs RUL prédit.

TABLE 3.4 – Comparaison entre les résultats obtenus par les indicateurs de santé et ceux trouvés dans la littérature.

Approche	Prédictions correctes %	Prédictions précoces	Prédictions tardives	Remarques
instances présentées par les indicateurs de santé	58%	24%	18%	testé sur les 100 moteurs de test
instances présentées par les trajectoires de dégradation	60%	17%	23%	testé sur les 100 moteurs de test
Approche de [Ramasso et al., 2013]	53%	36%	11%	testé sur les 100 moteurs de test
Approche de [Javed et al., 2013]	53%	27%	20%	testé sur seulement 15 moteurs de test
Inspiré par [Wang et al., 2008]	50%	19%	31%	testé sur les 100 moteurs de test

3. **Comparaison :** Les méthodes supervisées et non supervisées ont été testées en utilisant le fichier de test « test_FD001 » ce qui nous permet de comparer nos résultats avec la littérature. Le tableau 3.4 résume la performance de l'approche sur l'ensemble des moteurs (unités) de test ainsi que les résultats obtenus par d'autres approches.

   Nous nous sommes inspirés de la méthode proposée par [Wang et al., 2008] afin de réaliser une méthode IBL qui est comparée à notre méthode. La méthode a été testée sur le jeu de données en considérant à la fois la sélection de capteurs proposée par [Wang et al., 2008] et celle proposée par [Ramasso et al., 2013].

   Le tableau 3.4 recense les différentes performances obtenues par les méthodes développées et issues de la littérature scientifique. Les méthodes proposées dans ce chapitre montrent les meilleurs résultats. Elles dépassent les performances obtenues par des méthodes de réseaux de neurones [Javed et al., 2013], de fonction de croyance [Ramasso et al., 2013] et aussi des méthodes basées sur l'expérience inspirées par [Wang et al., 2008].

   La performance des trajectoires de dégradation conçues en utilisant l'approche non-supervisée (UKR) est meilleure par rapport aux indicateurs de santé construits à partir de l'approche supervisée (régression linéaire). Cela peut être expliqué par le fait que l'approche supervisée n'exploite pas l'ensemble de données. Uniquement 20% des données d'apprentissage sont utilisées durant la phase de modélisation des indicateurs de santé. Cela affecte la performance conduisant à des résultats moins bons que ceux obtenus par les trajectoires de dégradation.

3.7.2/ ENSEMBLE DES DONNÉES BATTERIES

La faisabilité de l'approche a également été montrée sur une application réelle et non pas simulée. L'ensemble de données Lithium-ion est fourni par le centre du pronostic à NASA ames [Saha et al., 2007].

Le jeu de données a été recueilli à partir des batteries « Li-ion » où l'opération à chaque cycle peut être de type : charge, décharge et impédance. Dans ce travail, cependant, seules les variables de charge et décharge sont considérées. Le vieillissement des batteries a été accéléré et l'expérience a été réalisée jusqu'à atteindre les critères de fin de vie en ayant l'impédance pour définir la défaillance.

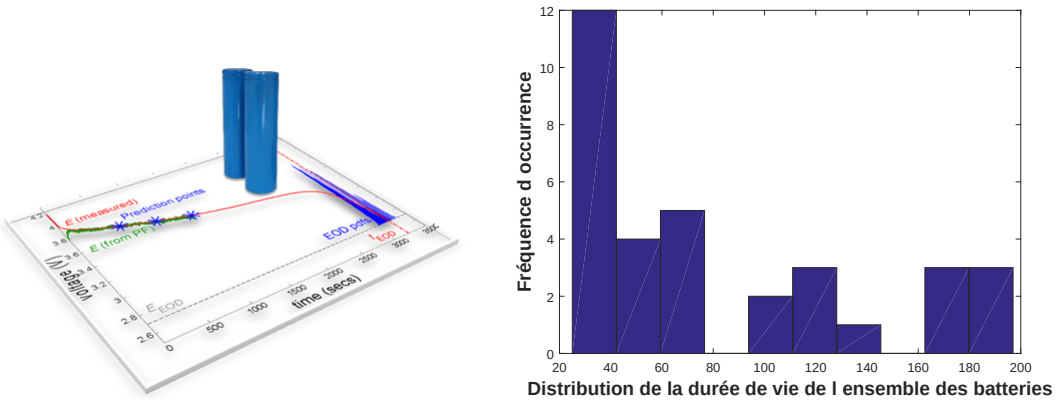


FIGURE 3.29 – Illustration des données des batteries.

Dans cette application, les données capteurs ne sont pas traitées directement. On extrait de ces données des caractéristiques (features) pour représenter au mieux le fonctionnement des batteries. Puis les caractéristiques les plus pertinentes sont sélectionnées. La figure 3.30 illustre le processus. À partir de données de chaque cycle, on extrait 6 caractéristiques : la moyenne, la moyenne quadratique, kurtosis, skewness, l'énergie et l'entropie. Les caractéristiques qui présentent des tendances décroissantes sont choisies ce qui nous a mené à sélectionner l'énergie comme feature.
L'extraction et la sélection des caractéristiques seront mieux expliquées à la section 4.4.2.1.

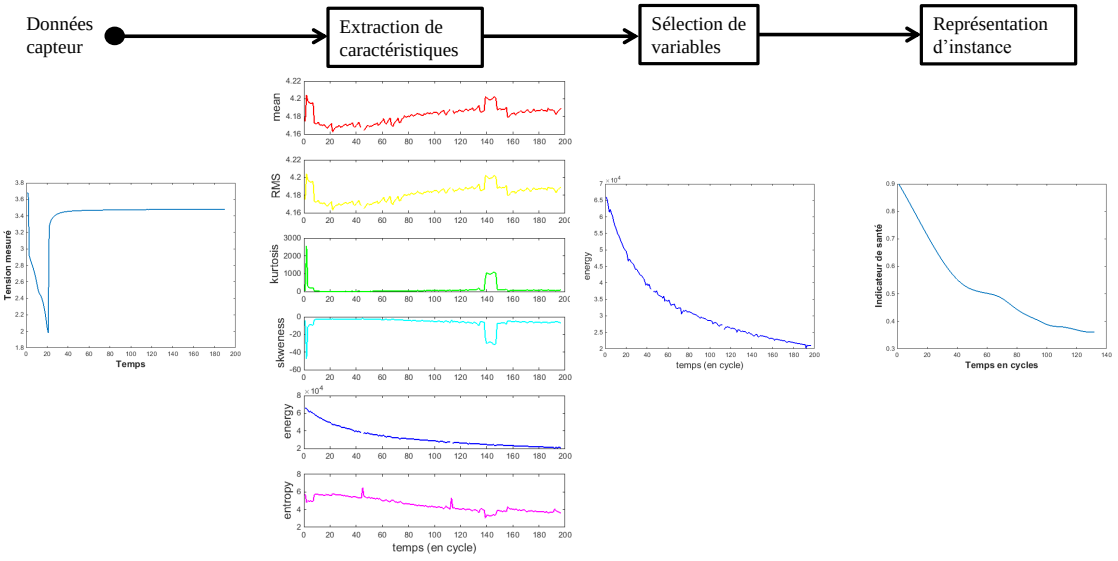
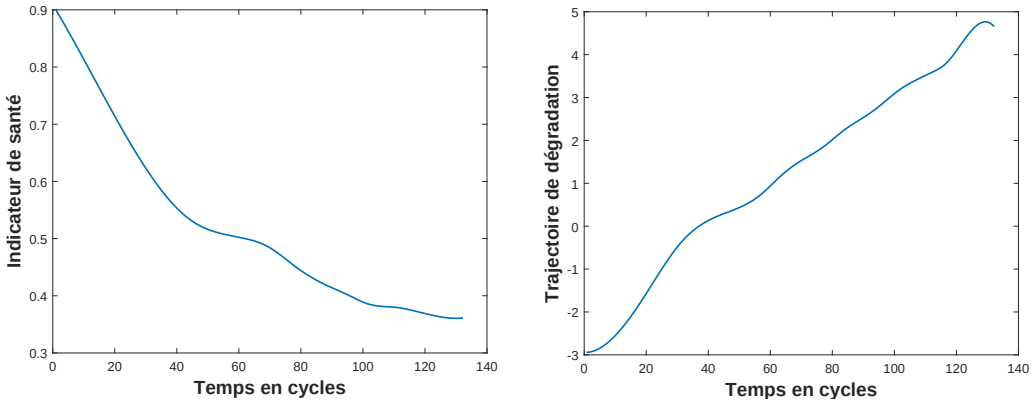


FIGURE 3.30 – Extraction et sélection de caractéristiques.

3.7.2.1/ RÉSULTATS ET DISCUSSION



(a) (b)

FIGURE 3.31 – (a) indicateur de santé, (b) trajectoire de dégradation.

La figure 3.31 montre un exemple d'instances obtenues. À gauche, l'instance est représentée par un indicateur de santé, et à droite, elle est représentée par une trajectoire de dégradation.

- Réglage des paramètres

Les séries temporelles obtenues présentent un degré de similarité élevé ce qui nous mène à choisir une valeur faible de λ (le facteur relaxant). Les composants de test, tout comme ceux d'apprentissage, commencent dans un bon état de santé. Donc on ne cherche pas à projeter l'état courant sur les éléments d'apprentissage, et par conséquent on utilise la similarité pondérée sans projection temporelle avec les paramètres suivants : $\sigma = 0.45$ et $\lambda = 0.1$, le nombre de blocs est fixé à 8, k=1 et le seuil est fixé à 0.7.

- Performance et comparaison

Le tableau 3.5 résume les résultats obtenus en terme d'erreur de pourcentage en moyenne absolue (EPMA), précision relative cumulative (PRC) et l'écart type de l'échantillon (ÉTÉ). On constate que la performance moyenne de la modélisation des trajectoires de dégradation est moins performante par rapport à la construction des indicateurs de santé, et elle est toujours en concurrence avec les résultats trouvés dans la littérature.

Nous ne pouvons comparer notre méthode IBL que par rapport aux travaux IBL de [Mosallam et al., 2014]. La méthode proposée extrait des tendances monotones à partir des signaux de capteurs et le RUL a été estimé sur la base des k plus proches voisins. Le meilleur EPMA rapporté était de 32.2086%, ce qui est encore plus grand que les erreurs moyennes de EPMA du tableau 3.5. Dans l'approche de [Mosallam et al., 2014], les données provenant de chaque composant sont traitées indépendamment à l'aide de l'analyse en composantes principales alors que la phase de modélisation présentée ici permet de mieux détecter les exemples similaires. Les instances traitées à l'aide du même modèle sont plus susceptibles de produire des résultats semblables rendant l'identification et la remémoration de l'instance la plus similaire plus précise et plus facile. Cependant, les performances de la méthode dépendent de la disponibilité d'exemples similaires. La généralité de nos approches a été montrée grâce à son application à deux types de composants de natures différentes : les turboréacteurs et les batteries.

TABLE 3.5 – Comparaison entre les résultats obtenus avec les indicateurs de santé et les trajectoires de dégradations.

Test	indicateur de santé			Trajectoire de dégradation		
	EPMA	PRC	ETÉ	EPMA	PRC	ETÉ
#1	7.4124	0.9259	2.0608	30.3534	0.6965	3.6011
#2	0	1.00	0	0	1.00	0
#3	17.4194	0.8258	1.0990	0	1.00	0
#4	0	1	0	0	1.00	0
#5	0	1	0	24.4060	0.7559	1.8163
#6	23.5636	0.7644	2.5701	30.0221	0.6998	3.4894
#7	41.0974	0.5890	7.3388	94.8488	0.0515	9.7244
#8	0	1	0	13.0721	0.8693	2.4689
#9	0	1	0	0	1.00	0
#10	0	1	0	0	1.00	0
#11	0	1	0	0	1.00	0
#12	77.00	1	0	88.00	0.12	0
#13	0	0.23	8.8912	18.9964	0.8100	2.4464
Moyenne	**12.8071**	**0.8719**	**1.6892**	23.0538	0.7695	1.8113

Les approches supervisées (indicateurs de santé) et non-supervisées (trajectoires de dégradation) ont été appliquées sur les deux composants. Pour les turboréacteurs l'approche non-supervisée a donné les meilleurs taux de prédictions acceptables. Par contre, c'est l'approche supervisée qui avait donné les meilleurs résultats pour les batteries. On conclut que le choix de l'approche dépend de l'application. La méthode en général, quelque soit le type de formalisation, reste compétitive par rapport à la littérature.

3.8/ CONCLUSION

Nous avons proposé deux approches basées sur l'expérience pour prédire le RUL d'un composant critique, à partir de son état courant, sans avoir à définir un seuil de défaillance. La méthode a été basée sur trois grandes phases, à savoir, la formalisation de l'expérience, la remémoration d'une expérience, et l'identification de l'état courant et l'estimation du RUL. Elle a été appliquée sur deux types de données. Une première approche tenant compte de l'état de santé et de fin de vie d'un composant a permis de créer un indicateur de santé. La formalisation de l'expérience a été faite d'une manière supervisée en développant un modèle de régression linéaire. Par contre l'application de cette modélisation n'a exploité que 20% des données capteurs. Nous avons ensuite proposé une deuxième approche ne tenant pas compte de l'état de santé,mais qui agrège les données capteurs en un trajectoire de dégradation. Cette méthode a été réalisée grâce au modèle UKR de régression non-supervisée (unsupervised kernel regression).

Le signal obtenu a été construit à partir de l'ensemble entier des données mais il n'est pas directement lié à l'état de santé, il est une représentation fidèle des données de dégradation et est appelé une trajectoire de dégradation . L'UKR peut être appréhendé comme un outil de réduction de dimensions qui construit une trajectoire de dégradation comme les méthodes trouvées dans la littérature, nommées abusivement indicateur de

santé, sans faire référence à l'état du composant, mais proposant une courbe de tendance qui est une agrégation des données capteurs par une approche non supervisée.

À la phase de remémoration, trois mesures de similarité ont été considérées. Les trois utilisent entièrement les données capteurs. Par contre la similarité pondérée et la similarité pondérée avec projection temporelle favorise la similarité vers la dégradation. La projection temporelle permet de détecter et positionner la dégradation actuelle des composants à pronostiquer sur les instances d'apprentissage. Le RUL a été estimé comme une moyenne des RUL des éléments d'apprentissage.

L'approche a été appliquée sur les turboréacteurs et les batteries. Les résultats obtenus sont compétitifs par rapport à la littérature et montrent une bonne généralisation de l'algorithme proposé.
L'application de la formalisation non-supervisée a donné de meilleurs résultats. Cette formalisation a exploité la totalité des données capteurs mais sans lier les trajectoires obtenues à l'état de santé. Elle a par contre donné de moins bon résultats sur le jeu de données des batteries. Par conséquent, le type de formalisation dépend de la nature des données. Les données turboréacteurs montrent de franches tendances par rapport aux données des batteries ce qui peut expliquer l'état de santé comme une information sous-jacente qui est plus présente dans les turboréacteurs même si elle n'est pas directement prise en compte.

L'approche d'expérience formalisée par les données (approche à base d'instances) présentée au chapitre 3 a donnée des résultats prometteurs, ce qui nous amène à continuer sur cette voie afin d'étudier des pistes d'amélioration.

4

PRONOSTIC ORIENTÉ EXPÉRIENCE FORMALISÉ PAR LA CONNAISSANCE

4.1/ INTRODUCTION

On se propose de développer une démarche prenant appui sur la méthode présentée au chapitre 3 et de définir de la connaissance à partir des expériences traitées pour enrichir les instances et les différentes phases de la méthode IBL. Nous faisons ainsi évoluer la méthode basée sur les instances en une méthode de raisonnement à partir de cas (RàPC), où l'on définira à la section 4.2.3.2 les différents conteneurs composant cette méthode.

Nous proposons dans ce chapitre, premièrement d'affiner la définition de l'indicateur de santé défini dans le chapitre 3, en exploitant toutes les données capteurs contrairement à la démarche précédente qui n'utilisait que 20% des données observées. Les 80% autres serviront à l'extraction de règles de connaissance, qui permettront de modifier la formalisation de l'indicateur de santé. Deuxièmement, de compléter la connaissance du comportement du système surveillé en extrayant la connaissance fréquentielle associée aux expériences par une méthode de décomposition en mode empirique. Cette méthode sera utilisée conjointement avec les mesures de similarité pour améliorer la phase de remémoration.

La section 4.2 présente un état de l'art sur le raisonnement à partir de cas, la section énumère les différents conteneurs de connaissance qui caractérisent les systèmes RàPC et recense des travaux de prédiction et d'analyse des séries temporelles par RàPC. La section 4.3 est consacrée à la formalisation des cas. Cette formalisation est renforcée par l'injection des deux types de connaissance (temporelle et fréquentielle), à la phase de remémoration et à la détermination de l'état de santé par le biais des indicateurs de santé et le calcul de RUL par l'adaptation des cas récupérés à partir de la base des cas au cas problème en utilisant la mesure de similarité. La section 4.4 évalue la méthode sur les mêmes jeux de données utilisés précédemment. On conclut le chapitre par une synthèse et quelques remarques à la section 4.5.

4.2/ LE RAISONNEMENT À PARTIR DE CAS

Le RàPC se situe entre deux communautés de recherche : l'intelligence artificielle et les sciences cognitives. L'approche constitue un pont naturel entre elles [Mille, 1999].

Le paradigme du RàPC doit son évolution aux travaux faits en sciences cognitives et plus particulièrement sur la théorie de mémoire dynamique de R. Schank qui constitue un réseau dense d'expériences connu comme "Memory Organization Packets" (MOPs). [Minsky, 1975] présente un réseau de nœuds et de relations entre ces nœuds ainsi que la notion de "framework (script, schéma)" qui correspond à une structure remémorée qui doit être adaptée afin de correspondre à la nouvelle situation rencontrée. [Schank, 1983] reprend ces travaux et formule pour la première fois le paradigme de raisonnement à partir des cas.

Le RàPC est une approche utilisant un raisonnement par analogie.

4.2.1/ DÉFINITION DU CAS

Un cas est une expérience qui résume une leçon permettant au système de RàPC de résoudre des problèmes de différentes natures. Les informations contenues dans le cas varient en fonction du domaine d'application et des objectifs fixés.

Selon [Fuchs et al., 2006], un cas est la description informatique d'un épisode de résolution de problème.

Tout d'abord, un cas en RàPC est généralement composé de deux espaces disjoints : l'espace des problèmes et l'espace des solutions.

Il existe deux types de cas : les cas sources qui sont les expériences enregistrées dans une base de cas et les cas cibles qui sont les nouveaux problèmes qui se posent. Les parties "problèmes" et "solution" sont renseignées pour le cas source. C'est-à-dire, qu'on va s'inspirer de ce cas pour résoudre un nouveau problème (cas cible, où seulement la partie problème est renseignée). Le cas source peut aussi contenir une partie appelée "information de qualité" [Reinartz et al., 2000]. Cette partie sert à donner des informations sur l'utilisation du cas dans le système.

On peut distinguer trois représentations de cas en fonction du problème à traiter
- La représentation textuelle ;
- La représentation semi structurée (vecteur de composants) ;
- La représentation structurée.

La présentation structurée est largement utilisée par la majorité des travaux. Le cas dans cette situation est représenté sous la forme d'un ensemble de descripteurs.

Un cas source est représenté par un couple $(srce, Sol(srce))$ et le cas cible par le couple $(cible, Sol(cible))$ où la partie solution $Sol(cible)$ est inconnue, elle est à déterminer à partir de l'ensemble $(srce, Sol(srce))$.

Dans nos travaux sur le pronostic, les descripteurs *problème* sont les valeurs de la courbe de régression des indicateurs de santé où les valeurs de la courbe de tendance et le descripteur $solution$ est le EOL.

4.2.2/ CARRÉ D'ANALOGIE

[Mille et al., 1996] présente le raisonnement à partir de cas comme un carré d'analogie (cf. Figure 4.1) qui décrit :

— Le lien entre la description d'un cas et sa solution (la trace du raisonnement menant à la solution).

— Les liens entre la description et la solution du cas de l'expérience dans la bibliothèque et du cas cible à résoudre (similarité entre deux problèmes). Ce lien est utilisé afin d'adapter la solution du cas cible en s'appuyant sur la similarité ainsi que les descripteurs des cas sources similaires de la base de cas qui sont adaptés au cas cible.

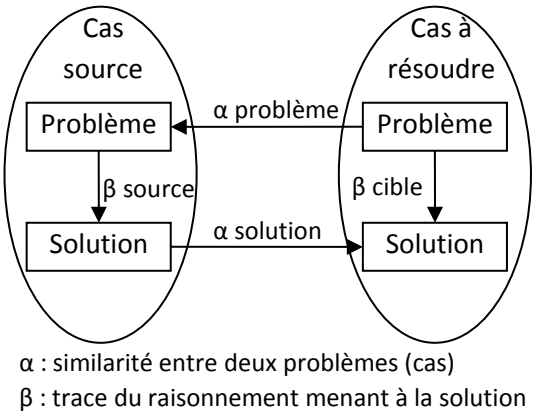


FIGURE 4.1 – Carré d'analogie [Mille et al., 1996].

α est la mesure de similarité qui détermine le degré de similarité entre le cas source sélectionné et les valeurs de descripteurs du problème cible. Il sert à sélectionner un cas source, dit similaire à partir des valeurs de descripteurs du problème cible.

β représente la relation de dépendance entre les valeurs de descripteurs de problème et les valeurs de descripteurs de la solution. Les descripteurs de solution qui doivent être adaptés sont mis en évidence.

Si une valeur de descripteur source dépend d'une valeur de descripteur de problème, une modification de la valeur du descripteur de problème entraînera une modification « analogue » à la dépendance du descripteur de solution correspondant. Cette connaissance est nécessaire pour l'adaptation. En fonction de ces dépendances et des écarts α constatés à corriger, l'adaptation permet de proposer une solution cible candidate qui pourra être vérifiée par rapport à sa conformité aux dépendances particulières qui pourraient exister entre problème et solution cible.

4.2.3/ LE SYSTÈME RÀPC

4.2.3.1/ CYCLE DU RÀPC

Selon les différentes sources bibliographiques, le cycle peut contenir trois, quatre ou cinq étapes. [Aamodt et al., 1994] ont été les premiers à décrire le cycle RàPC. Ils le décomposent en quatre phases à savoir : la remémoration (recherche du cas, *retrieve*), l'adaptation (la réutilisation du cas retrouvé, *reuse*), la validation (la révision du cas

sélectionné, *revise*) et la mémorisation (l'apprentissage, *retain*). D'autre part, [Mille, 1999] complète le processus proposé par Aamodt et Plaza en ajoutant une phase préliminaire d'élaboration au début du cycle. La figure 4.2 montre le cycle RàPC avec ces cinq phases.

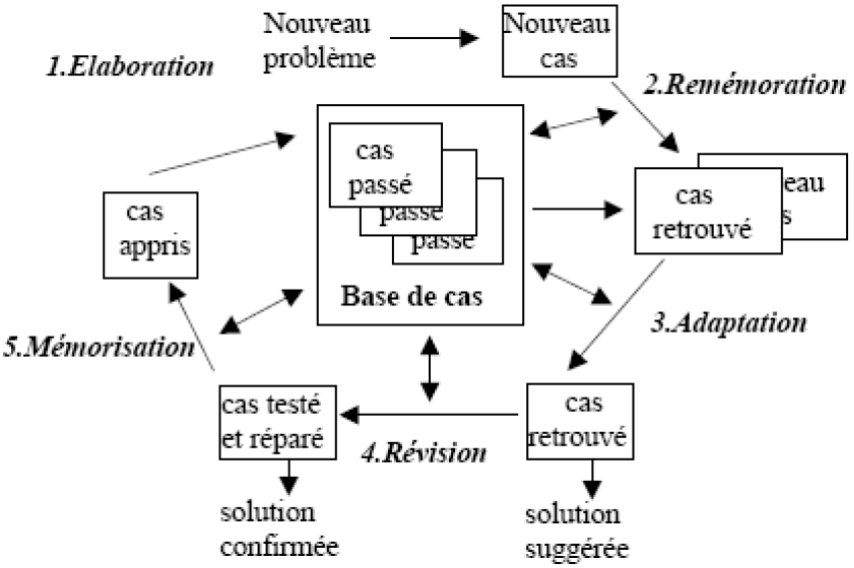


FIGURE 4.2 – Le cycle de raisonnement à partir de cas selon [Mille, 1999].

Le cycle de RàPC de [Mille, 1999] est composé des cinq phases suivantes :

1. **L'élaboration** d'un cas cible représente la formalisation des informations décrivant le nouveau problème. Le cas doit être représenté d'une manière similaire à un cas source. Dans le cas du pronostic, cela concerne la phase de représentation des données capteurs en courbe.
2. **La remémoration** des cas sources les plus similaires à partir de la base de cas. Ceci se fait en cherchant des correspondances entre les descripteurs des cas sources et du cas cible en calculant un degré d'appariement entre ces descripteurs.
3. **L'adaptation** des cas afin de résoudre un nouveau problème consiste à construire une solution en réutilisant la partie "*solution*" du (des) cas source(s).
4. **La révision** de la solution se fait dans le cas d'une éventuelle solution insatisfaisante afin de la corriger. La solution est vérifiée dans le monde réel par une introspection de la base de cas.
5. **La mémorisation** d'un nouveau cas consiste à ajouter éventuellement le cas cible résolu dans la base de cas si ce stockage enrichit la mémoire du système.

4.2.3.2/ LES CONTAINERS DE CONNAISSANCE

Selon [Richter, 2003], dans un système RàPC, la connaissance est repartie sur quatre containers : celui de vocabulaire, de mesure de similarité, de la base de cas et d'adaptation (c.f. Figure 4.3).

1. **Le vocabulaire :** L'une des premières questions à être traitées dans un système de représentation de connaissance est la définition de structures de données et quels éléments de structure sont utilisés pour représenter les notions primitives.

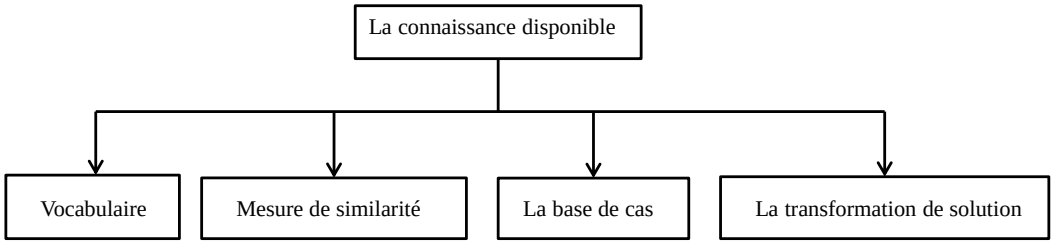


FIGURE 4.3 – Les conteneurs de connaissances [Richter, 2003].

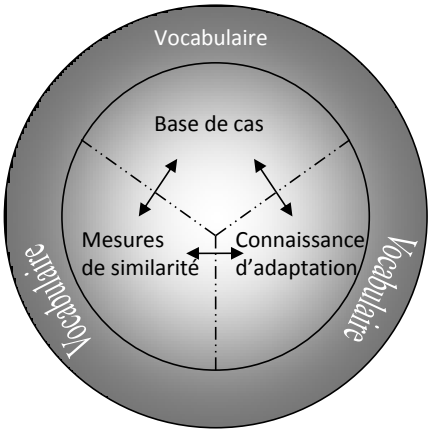


FIGURE 4.4 – Les containers de connaissances et leurs relations [Roth-Berghofer, 2003].

[Lieber et al., 2004] définissent le container de vocabulaire comme étant l'ensemble des éléments de représentation atomiques utilisés pour représenter les cas dans la base de cas. La représentation la plus connue est du type attribut-valeur.

2. **La mesure de similarité :** contient les mesures nécessaires pour la remémoration des cas. Ce container fait appel à ces mesures pour le calcul de la similarité ou la dissimilarité afin de déterminer les cas sources les plus appropriés au nouveau cas cible. La phase de remémoration et même d'adaptation des cas utilisent le raisonnement de similarité. La conception d'une mesure de similarité perspicace suit la représentation des cas et doit être adaptée au domaine d'application. La similarité est évaluée en fonction de cette représentation.
3. **La base de cas :** représente le contenu et l'organisation de la base de cas. C'est le container qui stocke toutes les expériences (cas). Il contient autant de cas que possible parce que chaque cas supplémentaire peut éventuellement élargir la compétence du système.
4. **La transformation de solution :** contribue à la modification des cas sources pour s'adapter au mieux au cas cible. Il aborde le fait que la solution obtenue à partir de la base de cas peut-être encore insuffisante. La transformation de la solution repose sur la mesure de similarité. Une fois que les cas les plus semblables sont récupérés, la mesure de similarité avec la projection temporelle est utilisée afin d'adapter les cas récupérés (trajectoires) au cas « problème » actuel.

Les trois premiers containers sont conçus avant que le système ne fonctionne (phase off-line) alors que la base de cas est mise à jour à la phase on-line.

Selon [Roth-Berghofer, 2003], les quatre containers de connaissance sont liés entre eux et le container de vocabulaire représente la base des trois autres containers. La figure 4.4 illustre les relations qui existent entre les quatre containers. Les flèches représentent le transfert de connaissance d'un container à un autre.

4.2.3.3/ MODÈLE GÉNÉRIQUE DE RÀPC

[Lamontagne et al., 2002] décrivent le système RàPC comme une combinaison de processus et de connaissances ('Knowledge containers") permettant de préserver et d'exploiter les expériences passées. Afin de simplifier la présentation, les auteurs proposent un modèle générique (c.f. Figure 4.5) qui se compose de deux aspects : (i) le processus où les principales phases sont la remémoration, l'adaptation, la mémorisation et la construction (authoring). (ii) Les containers de connaissance tels que le vocabulaire d'indexation, la base de cas, les métriques de similarité et les connaissances d'adaptation de la solution.

Le modèle est découpé en deux parties : on-line et off-line.

La partie on-line comporte le processus de RàPC, tandis que la partie off-line implique l'acquisition et représentation des connaissances ainsi que les containers de connaissances. La partie construction du cas et acquisition de connaissances guide la structuration initiale de la base de cas et des autres connaissances telles que la base de données, les expertises du domaine et la base de données.

Dans nos études, on s'est intéressé à ce modèle dans lequel nous avons exploité les phases du cycle de RàPC en développant l'approche à base d'instances dans le chapitre 3. Oon s'intéressera plus particulièrement aux containers de connaissance dans ce chapitre.

4.2.4/ ÉTAT DE L'ART SUR LES SÉRIES TEMPORELLES ET LE RÀPC

Il y a peu de travaux sur le RàPC et le pronostic. On ouvre notre recherche sur le RàPC et l'analyse et prédiction de séries temporelles.

Le RàPC a été utilisé afin de prédire les coûts de construction des projets à des stades précoces [Jin et al., 2012, Koo et al., 2011, Kim et al., 2004]. Le choix de raisonnement à partir de cas pour résoudre ce type de problème selon [Kim et al., 2004] vient du fait que ce dernier offre un compromis entre l'utilisation à long terme (horizon de prédiction), l'information disponible à partir des résultats (contrairement au réseaux de neurones par exemple, le RàPC a la capacité d'expliquer la solution) et la précision de la prédiction.
[Jin et al., 2012] proposent de prédire le coût de construction des projets aux premiers stades de construction en utilisant le RàPC. Les cas sont représentés comme des attributs (ex : surface, nombre de pièces, nombre d'étages, etc). Le coût est ensuite exprimé, pour chaque cas source, en fonction des attributs en utilisant l'analyse en régression multiple. Les coefficients de la régression sont ensuite convertis en poids pour chaque attribut. Lorsqu'un nouveau cas cible arrive, les attributs sont assortis et une similarité pondérée entre les attributs "$cible$" et "$source$" est calculée afin de récupérer le cas source le plus similaire à partir de la base de cas. La solution obtenue est révisée en prenant en considération la différence entre les attributs "$cible$" et "$source$".
[Koo et al., 2011] ont utilisé le RàPC pour résoudre le même problème. Le travail se

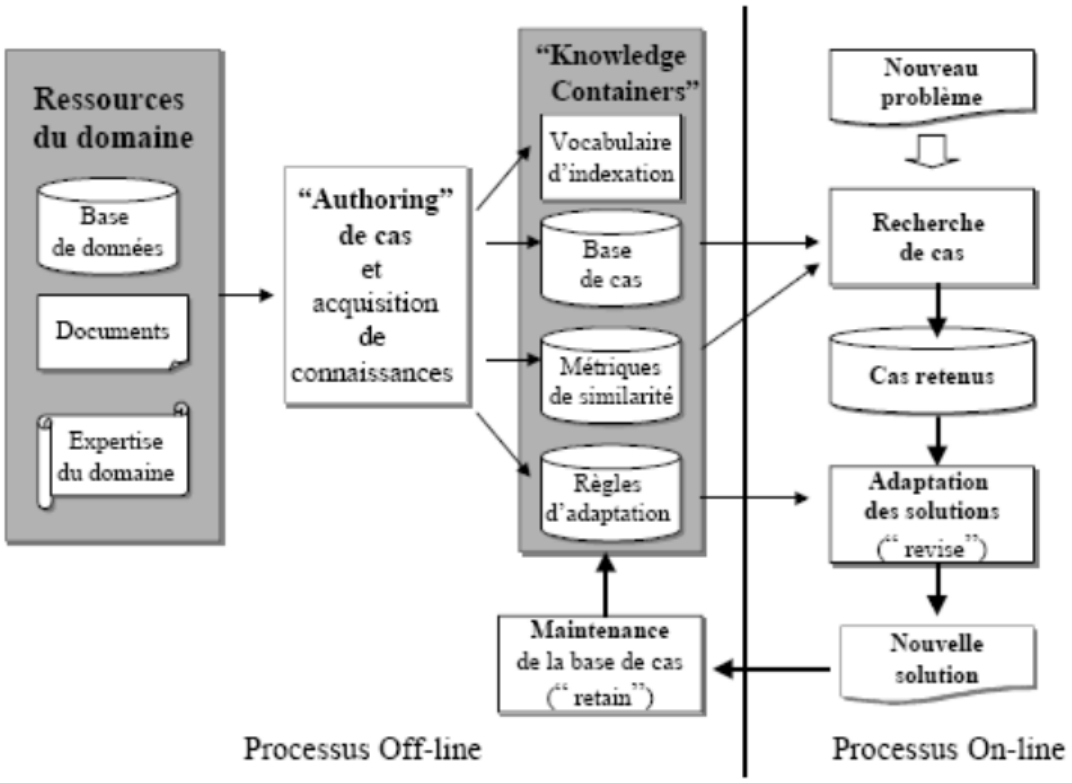


FIGURE 4.5 – Modèle générique d'un système de RàPC [Lamontagne et al., 2002].

concentre sur l'amélioration de la phase de remémoration en définissant quatre paramètres d'optimisation : le critère minimal pour définir le score de similarité entre les attributs, MCAS (le score est défini seulement si la similarité calculée comme la différence entre les attributs est supérieur à MCAS. Autrement le score est égale à zéro), la plage des poids des attributs, la plage de similarité des cas sélectionnés, (seulement les cas ayant une similarité comprise dans cette plage sont sélectionnés), et la plage de tolérance calculée à partir de RNA et l'analyse en régression multiple (l'intersection entre les plages de valeurs produites par RNA et l'analyse en régression multiple). Ces paramètres sont utilisés comme un filtre afin de choisir le cas "$source$" le plus adapté au cas "$cible$".

[De Paz et al., 2012] ont combiné le RàPC et la régression avec les séparateurs à vecteurs de support (SVR) afin de prédire le taux de CO_2 échangé entre l'atmosphère et l'océan. Le système RàPC développé reçoit des images satellitaires, les sauvegarde en tant que descripteurs de cas et prédit le taux de CO_2 en utilisant les SVR. Les SVR sont en mesure de fournir des modèles de régression pour les ensembles de données non linéaires. Une fois que les cas les plus similaires ont été récupérés, le modèle de régression est généré et utilisé pour estimer le taux de CO_2 du nouvel cas.

[Xing et al., 2012] ont utilisé le RàPC comme un algorithme d'estimation de paramètres d'un modèle qui calcule les coefficients de transfert de chaleurs, la conductivité thermique, et la diffusivité thermique. Les cas étaient représentés par un ensemble de descripteurs qui sont les conditions d'utilisation pratiques utilisées pour calculer les différents coefficients. La partie solution contient les paramètres du modèle. Une fois le cas source le plus similaire sélectionné, sa solution (paramètres du modèle) est utilisée pour estimer les coefficients.

[Fdez-Riverola et al., 2003] ont proposé un modèle RàPC hybride, le modèle est dit hybride car il combine le RàPC avec un modèle neuro-symbolique utilisé par les différentes phase du processus RàPC. La méthode proposée était utilisée afin de prévoir les paramètres d'un environnement complexe et dynamique dans le but de prédire les marées rouges qui apparaissent dans la zone côtière. L'approche RàPC proposée est hybride : la phase de remémoration utilise une structure RNA de cellules croissantes et la phase de réutilisation est effectuée à l'aide d'une fonction RNA de base radiale, qui génère une solution initiale créant un modèle qui estime la concentration des micro algues appelées Pseudo-nitzschia responsable de la production des marées rouges une semaine à l'avance.
[Ping et al., 2015] ont utilisé le RàPC pour une application dans le domaine médical. Le RàPC été utilisé afin de construire des modèles prédictifs de récurrence de cancer de foie. Le travail étudie le couplage dynamique des cas de patients selon l'information temporelle durant le développement du cancer. Les cas sont appariés afin de ne pas ignorer aucun cas.

[Pecar, 2002] a proposé une méthode RàPC afin d'analyser et prédire l'évolution des séries temporelles. Le cas *cible* était comparé à tous les cas *source* de la base de cas afin d'identifier le meilleur appariement. Les différences entre le cas "*cible*" et le meilleur cas "*source*" ont été utilisées afin de prédire la première valeur dans l'horizon de prédiction. Les cas étaient représentés par des motifs symboliques qui décrivent la dynamique de la série temporelle. La série est d'abord décomposée en intervalle de "*patterns*" comprenant des observations séquentielles. Ensuite, ces *patterns* sont sauvegardés comme des cas. Le cas cible (le dernier intervalle de la série temporelle) est comparé à la base de cas et tous les cas "*source*" ayant des *patterns* identiques sont remémorés.
[Elsayed et al., 2011] ont combiné le RàPC avec des techniques venant du domaine médical. L'approche était appliquée pour le dépistage de l'image rétinienne pour la dégénérescence maculaire liée à l'âge et la classification de la résonance magnétique des images scannées du cerveau. Les auteurs définissent deux bases de cas ; une principale (contenant les informations générales) et une deuxième secondaire (contenant les détails). Quand une nouvelle image arrive (cas cible), son histogramme est comparé aux histogrammes de la base de cas en calculant la similarité par une déformation temporelle dynamique. Si les cas remémorés sont cohérents (donnent la même réponse), la solution est claire et le cas "*cible*" est étiqueté de la même manière. Autrement (réponses contradictoires), le même processus est répété mais cette fois en considérant la base de cas secondaire et un nouveau descripteur formé par les pixels représentant le disque optique des images.
[Kurbalija et al., 2008] d'autre part, ont crée une structure RàPC qui génère des systèmes d'aide à la décision. La structure présentée intègre deux types de shell RàPC [1] ; le *CaBaGe* (Case Base Generator) qui est utilisé si les cas sont représentés par des attributs et *CuBaGe* (Curve Base Generator) qui est utilisé si les cas sont représentés par des séries temporelles. La mesure de similarité dans ce cas est vue comme un calcul de surface entre deux courbes (séries temporelles). La surface est obtenue en calculant l'intégral définitif.
[Kurbalija, 2009] a utilisé le *CuBaGe* afin de comparer et récupérer les courbes (séries temporelles) les plus semblables à la courbe *cible*. Les courbes récupérées sont utilisées afin de prédire le comportement futur de problème courant. L'approche était appliquée pour prédire le rythme de l'émission de facture et la réception des paiements à la société

1. une application générique conçue spécifiquement pour le raisonnement à partir de cas

" Novi Sad Fair". Le RàPC a été utilisé dans différentes applications et plus précisément pour l'estimation des variables, indexation et prédiction des séries temporelles dans la base de cas. Par contre son utilisation se fait rare dans le domaine du pronostic. Ce travail propose l'utilisation de RàPC pour la prédiction de durée de vie résiduelle des composants critiques. Une approche de raisonnement à partir des instances a déjà été proposée au chapitre 3 et elle sera renforcée dans ce chapitre par l'injection de la connaissance.

4.3/ APPROCHE PROPOSÉE

4.3.1/ FORMALISATION DE CAS

Les cas typiques de RàPC sont généralement supposés avoir un degré de richesse d'informations par rapport aux instances construites par les approches de raisonnement à partir des instances. Dans cette section on rajoute deux types de connaissance. Une connaissance temporelle utilisée pour améliorer la construction des indcateurs de santé (les cas) et une connaissance fréquentielle qui complète l'information présente dans les cas.

4.3.1.1/ CONNAISSANCE TEMPORELLE

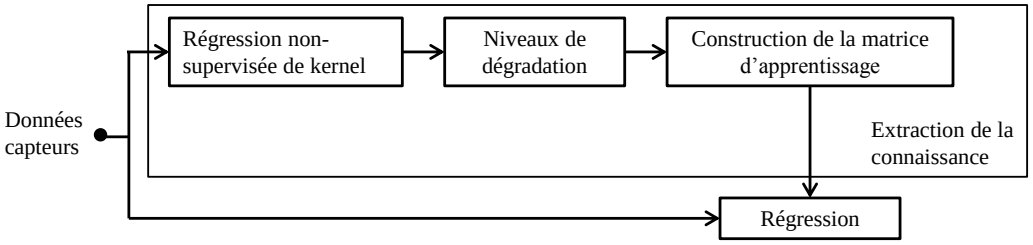


FIGURE 4.6 – Construction d'indicateur de santé avec des règles de connaissance.

La modélisation de l'indicateur de santé présentée dans le chapitre précédent ne prend en compte que 20% des données d'apprentissage. La construction des des trajectoires de dégradation d'autre part utilise 100% des données d'apprentissage, mais les trajectoires obtenues ne sont pas directement liées à l'état de santé. Pour remédier à ces lacunes on combinera les deux approches pour définir un indicateur de santé construit à partir de plus de données d'apprentissage. Les trajectoires de dégradation obtenues par la méthode UKR définiront les variations des données capteurs tout au long de cycle de vie du composant. Ces variations seront injectées dans la matrice d'apprentissage de la première méthode (modélisation d'indicateur de santé). La méthode exploitera ainsi toutes les données du cycle de vie.
Cet indicateur de santé sera donc issu de la modélisation de tendance (trajectoire de dégradation) combinée avec une droite de régression et l'EMD qui permettra d'obtenir un indicateur de santé lisse et d'exploiter l'information fréquentielle présente dans les fonctions intrinsèques.

Les instances obtenues en utilisant la régression non supervisée des noyaux (UKR) sont des signaux mono-dimensionnels fidèles aux données capteurs.

Ces trajectoires sont découpées en intervalle afin d'obtenir une connaissance sur l'évolution de la dégradation (c.f. Figure 4.7). À partir des niveaux obtenus, une matrice d'apprentissage plus complète est construite pour chaque ensemble d'apprentissage. La matrice comprend des données associées au début, au milieu et à la fin de la dégradation. La matrice d'apprentissage finale est une concaténation des matrices individuelles et elle est utilisée pour construire le modèle de régression linéaire (c.f. Figure 4.7).

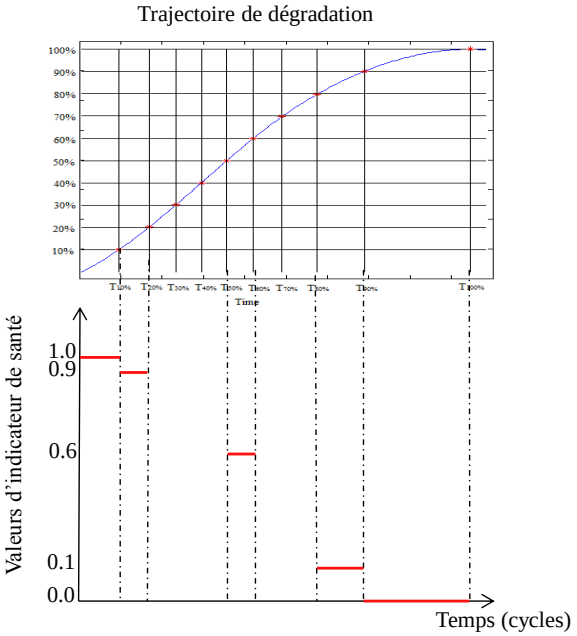


FIGURE 4.7 – Construction de la matrice d'apprentissage à partir de trajectoire de dégradation.

Pour chaque composant d'apprentissage, les valeurs de l'indicateur de santé (IS) sont affectées selon ce qui suit :

$$
\begin{aligned}
IS &= 1.0 \text{ pour } t \in [1, T_{10\%}] \\
IS &= 0.9 \text{ pour } t \in [T_{10\%}, T_{20\%}] \\
IS &= 0.6 \text{ pour } t \in [T_{50\%}, T_{60\%}] \\
IS &= 0.1 \text{ pour } t \in [T_{80\%}, T_{90\%}] \\
IS &= 0.0 \text{ pour } t \in [T_{90\%}, T_{100\%}]
\end{aligned}
$$

4.3.1.2/ CONNAISSANCE FRÉQUENTIELLE

Jusqu'à présent, les cas étaient représentés sous forme de séries temporelles $T_i = T_1^i, T_2^i, \ldots, T_{l(T_i)}^i$. Les trajectoires obtenues à partir de la régression étaient perturbées par des oscillations considérées comme du bruit et ont été filtrées en lissant les courbes. Cependant, ces fluctuations peuvent ne pas être du bruit et exprimer une information sur la défaillance. Nous avons décidé d'exploiter cette information en extrayant leurs fréquences par la méthode de décomposition en modes empiriques. Le signal est décomposé en fonctions intrinsèques (c.f. Définition 9) et un résidu.

Définition 9 : Fonctions intrinsèques

Pour une série temporelle T_i, soit $\Psi_i = \{\psi(t)_i^j\}$, $1 \leq l \leq n$ l'ensemble de fonctions intrinsèques qui y correspondent. Une fonction intrinsèque (IMF) $\psi(t)_i^j$ est un signal à un seul composant à partir duquel on peut extraire des fréquences instantanées.

$$\psi(t) = r(t)sin\theta(t)$$

Où $r(t) = \sqrt{\psi(t)^2 + \phi(t)^2}$ et $\theta(t) = arctan\frac{\phi(t)}{\psi(t)}$

Le résiduel qui n'est autre que la trajectoire lissée détient l'information temporelle d_i (descripteur de cas à l'instance i). Tandis que les fonctions intrinsèques sont à la base des informations fréquentielles qui enrichissent la représentation des cas. Les fonctions intrinsèques sont converties en domaine de fréquence en calculant la densité spectrale de puissance (DSP, c.f. définition 10). La figure 4.9 illustre les fonctions intrinsèques d'un composant donné et les signaux DSP associés.

Définition 10 : Densité spectrale de puissance (DSP)

La densité spectrale de puissance est définie comme la transformée de Fourrier de la fonction d'auto-corrélation d'un Signal X
La transformée de Fourrier :

$$\mathcal{F}[x(t)] = X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-jwt}dt$$

L'auto-corrélation :

$$R(\tau) = x(\tau) \times x(-\tau) = \int_{-\infty}^{\infty} x(t)x(t+\tau)dt$$

DSP :

$$\mathcal{F}[R(\tau)] = X(\omega)X^*(\omega) = |X(\omega)|^2$$

la DSP représente la répartition de la puissance d'un signal X suivant les fréquences ω.

Pour chaque composant, un vecteur **F** des fréquences qui ont les valeurs de puissance les plus élevées est déduit à partir des signaux de densité spectrale de puissance.

Le vecteur est constitué de deux maxima locaux du premier et deuxième signal DSP (Densité Spectrale de Puissance).

$$\mathbf{F}_i = \{f_{1,1}^i, f_{1,2}^i, f_{2,1}^i, f_{2,2}^i\}, \tag{4.1}$$

où $\mathbf{F}_i$ est le vecteur de fréquence du ième composant, $f_{1,k}^i$ est le k-ième maximum local du premier signal DSP, et $f_{2,k}^i$ est le k-ième maximum local du deuxième signal DSP, k=1,2. Ces deux signaux ont été choisis parce qu'ils portent plus d'information (c.f. Figure 4.9-b).

La partie *"problème"* des cas est donc représentée par le couple descripteurs temporels et information fréquentielle (voir figure 4.8 et définition 11).

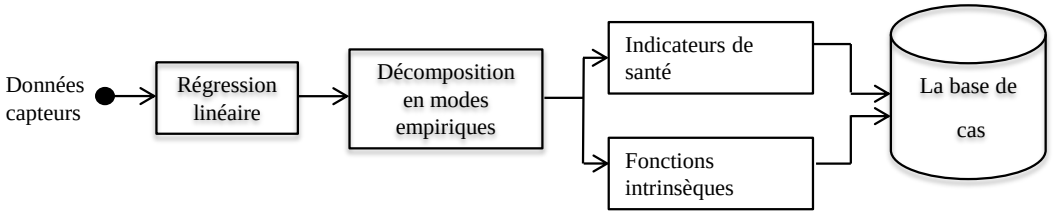


FIGURE 4.8 – Formalisation des cas avec la connaissance fréquentielle.

Définition 11 : Un cas

Un cas est un couple composé d'une série temporelle $T_i = d^i_{s,1}, d^i_{s,2}, \ldots d^i_{s,l(T_i)}$ et de l'information fréquentielle représentant l'historique de défaillance du composant.

$$C_i = \{T_i, \mathbf{F}^i, EOL_i\},$$

où $d^i_{s,j}$ est un descripteur du cas i à l'instant j , F^i est le vecteur d'information fréquentielle et EOL_i est la durée de vie du composant.

4.3.2/ LA PHASE DE REMÉMORATION

On fait évoluer la mesure de similarité en fonction de la représentation du cas. L'analyse fréquentielle permet de comparer les valeurs de fréquences dominantes des fonctions intrinsèques et d'inclure l'effet de la différence entre les valeurs de fréquence en tant que facteur de pénalité qui ajuste le score de similarité initialement calculé à la section 3.4.3.

Pour chaque cible et source, la différence entre les vecteurs de fréquence de forme : $\mathbf{F}_p = \{f^p_{1,1}, f^p_{1,2}, f^p_{2,1}, f^p_{2,2}\}$ est utilisée pour calculer un facteur de pénalité qui ajuste le score final de similarité.

$$\Delta\mathbf{F}_{i,k} = |\mathbf{F}_i - \mathbf{F}_k|, \tag{4.2}$$

$$PF(i)_k = \frac{-\Delta f_{i,k}}{\sum_{j=1}^{N} \Delta f_{j,k}}, \tag{4.3}$$

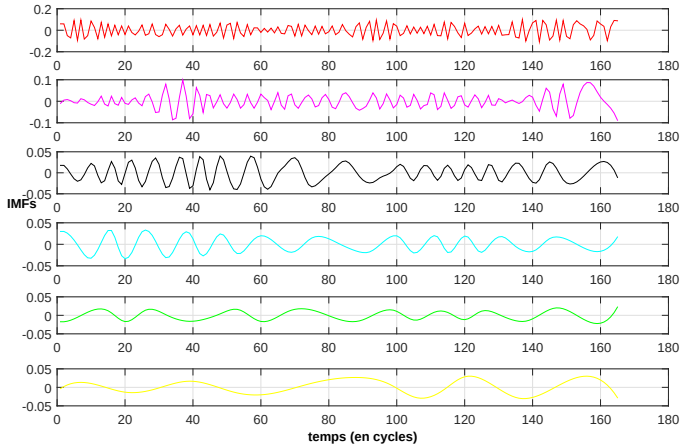
où $PF(i)_k$ est le facteur de pénalité entre le cas cible i et le cas source k. Ce facteur appartient à l'intervalle $[0, 1]$ avec 1 indiquant l'absence de différence entre les fréquences des cas. Le score de similarité finale ajusté devient :

$$SC = \frac{NBS}{M}. \sum_{i=1}^{N} w(i).max(sim_{i,j}), \tag{4.4}$$

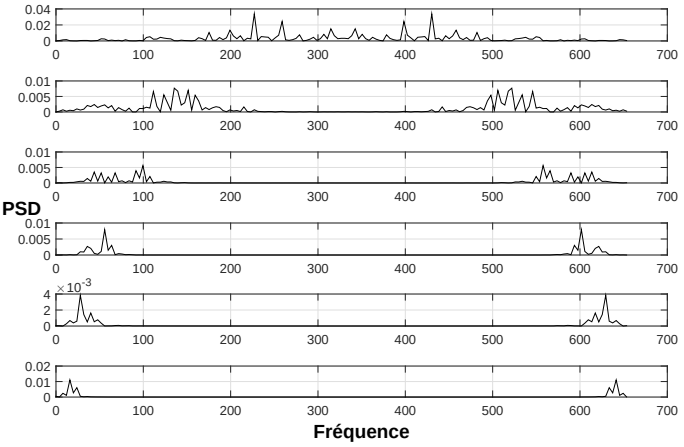
$$F.S.C(i)_k = PF(i)_k * S.C(i)_k, \tag{4.5}$$

où $FSC(i)_K$ et $SC(i)_k$ sont les scores de similarité final, et initial entre le ième cas de test et le kème cas d'apprentissage.

La figure 4.10 est une illustration du meilleur appariement entre trajectoire cible et trajectoire source. Comme on peut le voir sur la figure les trajectoires partagent la même valeur de fin vie et ont tendance à être plus semblables aux cycles tardifs.



(a)



(b)

FIGURE 4.9 – La décomposition d'une trajectoire en (a) fonctions intrinsèques et (b) les correspondants signaux de densité spectrale de puissance.

4.3.3/ DÉTERMINATION DE L'ÉTAT DE SANTÉ ET LE RUL

La construction des indicateurs de santé permet d'avoir une idée sur l'état de santé du nouveau composant. L'intervalle des valeurs est entre « 0 » et « 1 », zéro pour indiquer l'arrêt de fonctionnement du composant et un pour indiquer le bon fonctionnement. Les états intermédiaires sont exprimés par des valeurs entre 0 et 1 [2]. La figure 4.11 est un exemple d'indicateurs de santé pour des cas cibles. Selon la description de jeux de données, les éléments de test ne débutent pas dans le même état. Certains composants sont dans un bon état de santé. Pour d'autres composants, la dégradation commence même avant de sauvegarder l'historique , et est représentée sur les indicateurs de santé. Par exemple, à la figure 4.11-(a), les signes de dégradation sont déjà présents dès les premiers cycles (indicateur de santé=0.72). Une fois l'acquisition de données arrêtée, le composant est déjà dans un état dégradé avancé (valeur d'indicateur de santé égale à 0.4). Pour l'exemple à la figure 4.11.b, le composant est dans un bon état au début (indicateur de santé= 1.1). La dégradation commence à apparaitre juste avant la fin de l'acquisition.

2. À cause des résidus, les valeurs peuvent dépasser légèrement la plage [0,1]

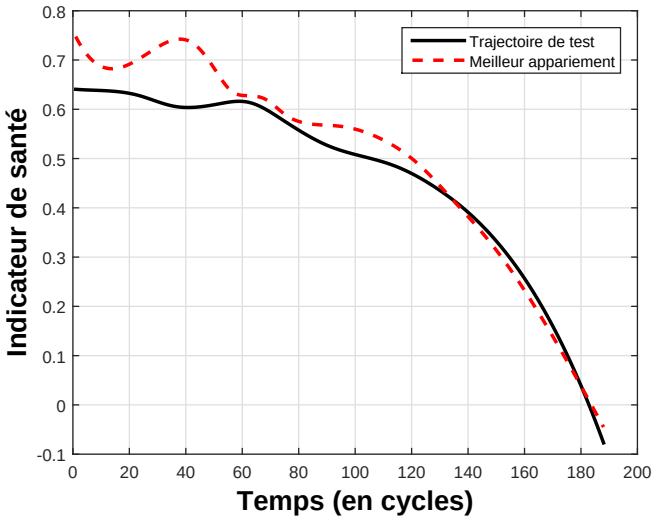
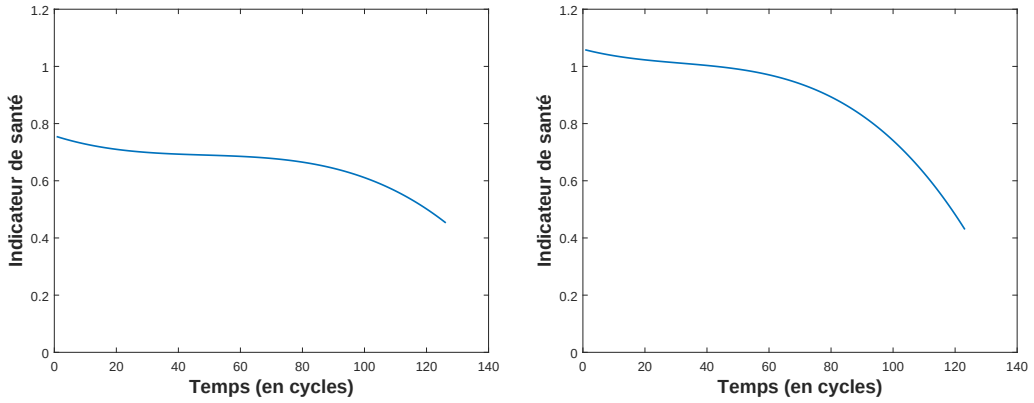


FIGURE 4.10 – Une trajectoire de test et son meilleur appariement.



(a) (b)

FIGURE 4.11 – Un exemple d'indicateurs de cas cible.

L'estimation de la durée de vie résiduelle reste inchangée. Une fois que l'ensemble des cas les plus similaires est retrouvé, les cas sources sélectionnés sont transformés afin d'identifier l'état courant et calculer la durée de vie résiduelle (RUL) des cas cibles.

Soit $C_c = \{T_c, \mathbf{F}^c, ?\}$ le cas cible à résoudre et $C_s = [C_{s,i} = \{T_{s,i}, \mathbf{F}^{i,s}, EOL_{i,s}\}.i = 1, \ldots, k]$ l'ensemble des cas les plus similaires. La solution de chaque $C_{s,i} \in C_s$ est adaptée au cas cible en identifiant la localisation du dernier cycle de mesure (t_c) du cas cible sur les cas sources en utilisant la similarité pondérée avec projection temporelle qui permet aussi d'identifier l'état courant du composant de test sur les trajectoires qui représentent les cas sources (c.f. Figure 4.12) .

Le RUL final du cas cible est calculé comme la moyenne des RUL des cas les plus similaires.

Exemple : (c.f. Figure 4.12)

$RUL_{s,1} = EOL_{c,1} - t_{c,1} = 148\,cycles$ (la solution transformée en utilisant le premier cas source similaire).

$RUL_{s,2} = EOL_{c,2} - t_{c,2} = 140\,cycles$ (la solution transformée en utilisant le deuxième cas

source similaire).

$RUL_r = moyenne(RUL_{s,1}, RUL_{s,2}) = 144\, cycles$ (la solution du cas cible)

Le cas résolu peut être maintenu à la base de cas $C_r = \{T_r, \mathbf{F}^r, EOL_r\}$ où $EOL_r = RUL_r + T_r$ (t_r est le dernier cycle de mesure)

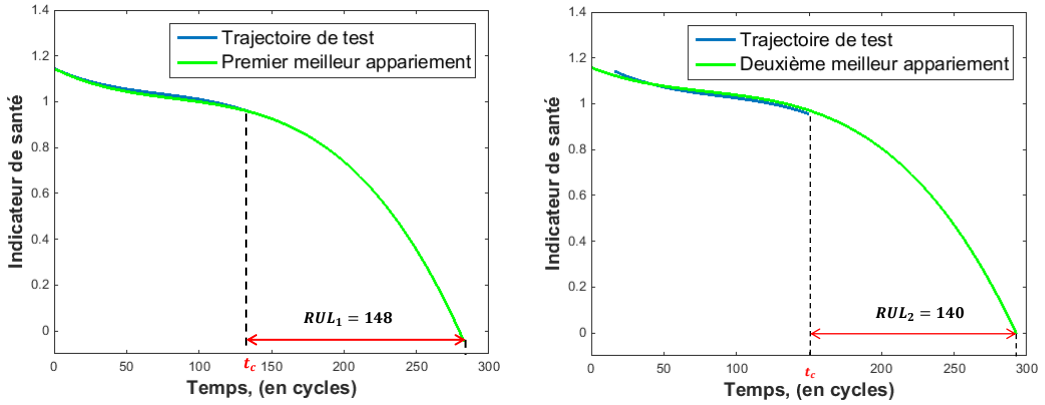


FIGURE 4.12 – Exemple d'adaptation de calcul du RUL selon la position de cas problème sur la trajectoire d'apprentissage.

4.4/ ÉVALUATION DE LA MÉTHODE

Les modifications apportées à l'approche décrite précédemment sont évaluées sur les mêmes jeux de données (section 3.7).

4.4.1/ ENSEMBLE DE DONNÉES DES TURBORÉACTEURS

Pour évaluer l'approche proposée ici, on considère deux tests. Le premier consiste à appliquer une validation croisée (leave-one-out cross validation) sur le fichier d'apprentissage « TRAIN_FD001 »[3]. 99 moteurs ont été utilisés comme des données d'apprentissage et le 100ème est laissé pour faire le test. La procédure est répétée 100 fois.

Le deuxième test consiste à prédire la durée de vie résiduelle des moteurs de fichier « TEST_FD001 »[4].

La validation croisée permet une évaluation plus riche (ex : calcul de l'horizon de pronostic) car on a accès aux valeurs exacte à chaque cycle. Tandis que le fichier test « TEST_FD001 », simule les situations réelle où l'algorithme doit être appliqué sur des données qui n'ont pas été vues auparavant. De plus le fichier offre un moyen uniforme d'évaluer et comparer les résultats de l'approche proposée par rapport aux résultats trouvés dans la littérature.

3. Rappel :le fichier constitue l'historique de défaillance de 100 moteurs.
4. Le fichier est fourni par les créateurs du jeu de données, il est constitué de l'historique de 100 moteurs arrêtés avant l'arrêt de la défaillance.

4.4.1.1/ LA VALIDATION CROISÉE

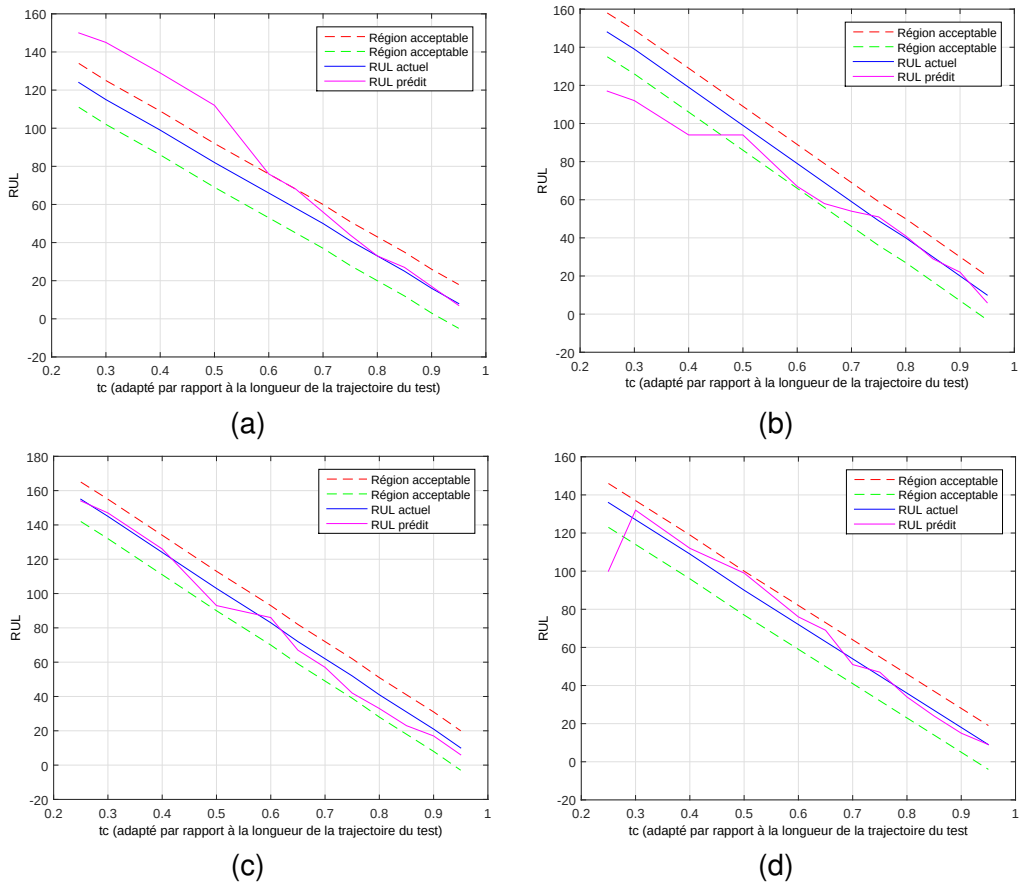


FIGURE 4.13 – Détails de quelques prédictions (a) moteur 28, (b) moteur 50, (c) moteur 15, (d) moteur 35.

Pour chaque moteur de test, on commence la prédiction à diffèrents t_c (le temps qui indique le début de prédiction c'est-à-dire le dernier cycle de mesure). Les valeurs de t_c sont adaptées en fonction de la durée de vie des moteurs de test. Elles sont exprimées comme des pourcentages des longueurs de trajectoires de test. Plus grand est t_c, meilleure est la prédiction. La figure 4.13 montre les résultats obtenus pour certains moteurs de test.

Pour évaluer l'effet de l'injection de la connaissance, on compare les résultats de l'approche expérience formalisée par l'ajout de la connaissance présenté ici avec l'approche expérience formalisée par les données (IBL) présentée dans le chapitre 3.

Le tableau 4.1 indique l'horizon de pronostic (équation 3.14) en mettant $\alpha = 0.3$. Les valeurs reportées dans ce tableau sont les valeurs moyennes obtenues pour les 100 moteurs de test de validation croisée. Le tableau 4.2 donne les valeurs EMA à différent t_c et la figure 4.14 illustre le pourcentage de prédictions acceptable.

TABLE 4.1 – L'horizon de pronostic.

Méthode	Horizon du pronostic
Expérience formalisée par la connaissance	**152**
Expérience formalisée par les données	143

À en juger à partir de la figure 4.14, les performances globales des deux approches prises comme les moyennes des performances à chaque t_c et dénotées par une ligne continue pour l'approche orientée expérience formalisée par l'ajout de la connaissance et par une ligne discontinue pour l'approche à partir d'instances sont approximativement égales. Par contre l'approche orientée expérience formalisée par l'ajout de la connaissance avait une meilleure performance à des valeurs t_c de début de vie. Cela est confirmé par un horizon de pronostic plus élevé pour l'approche orientée connaissance. Avec l'injection de la connaissance, on augmente l'horizon de presque dix cycles (tableau 4.1). Cela se manifeste aussi au niveau des valeurs d'erreurs moyennes absolues (tableau 4.2). Les valeurs d'EMA obtenues avec la connaissance sont inférieures à celles obtenues avec l'approche à base d'instances au début de tableau. La valeur moyenne est aussi inférieure à celle obtenue par l'approche basée sur les instances, ce qui indique une sorte de supériorité de l'approche orientée expérience formalisée par l'ajout de la connaissance.

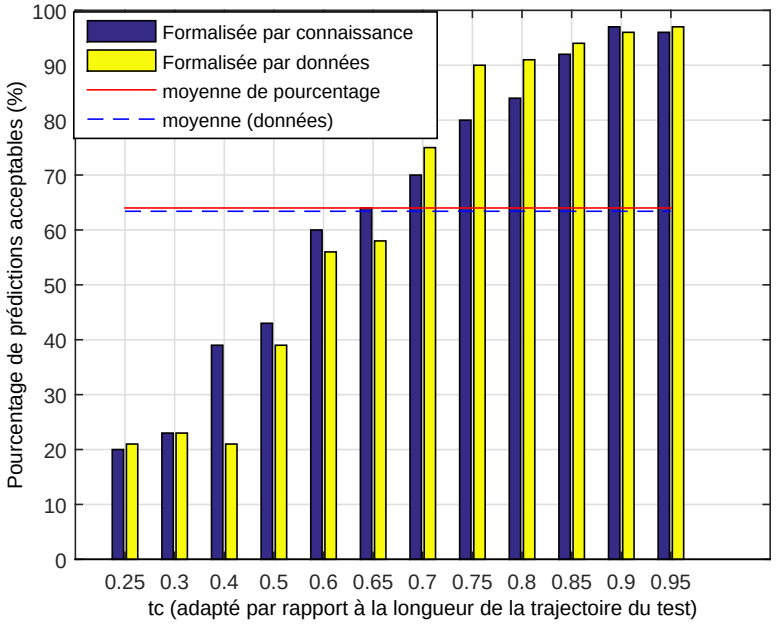


FIGURE 4.14 – La performance globale du prédiction du RUL à différents valeurs de t_c.

4.4.1.2/ RÉSULTATS EN UTILISANT LES DONNÉES DE TEST

Le jeu de test contient 100 moteurs avec des RUL inconnus. Ces derniers se trouvent dans le fichier « RUL_FD001.text » et sont utilisés pour évaluer la méthode.

Figure 4.15 montre que les RUL estimés sont suffisamment proches des valeurs actuelles. Les valeurs des erreurs faites sur ces prédictions sont illustrées à la figure 4.16-a. La figure 4.16-b présente l'histogramme de l'erreur des résultats obtenus précédemment avec l'approche à base d'instances (indicateur de santé). À partir de cette figure, on constate que la limite inférieure de l'intervalle est réduite par l'injection de la connaissance. Les erreurs inférieures à -69 sont éliminées. Par contre la limite supérieure de l'intervalle est augmentée ce qui veut dire que le nombre des prédictions précoces est plus élevé sachant qu'en pronostic les prédictions précoces sont mieux tolérées par rapport au prédictions tardives.

TABLE 4.2 – Erreur Moyenne Absolue (EMA).

$t_c\%$	Expérience formalisée par la connaissance	Expérience formalisée par les données
25	35.34	35.10
30	31.24	34.16
40	19.56	32.30
50	17.27	22.80
60	11.77	14.38
65	10.96	11.23
70	8.99	7.9
75	7.43	5.56
80	6.35	5.84
85	5.49	4.25
90	4.08	4.21
95	5.26	5.33
Valeur moyenne	**13.64**	15.25

TABLE 4.3 – L'erreur Moyenne Absolue pour les éléments de test.

Méthode	Erreur Moyenne Absolue (EMA)
Expérience formalisée par l'ajout de la connaissance	**16.55**
Expérience formalisée par les données	17.8099

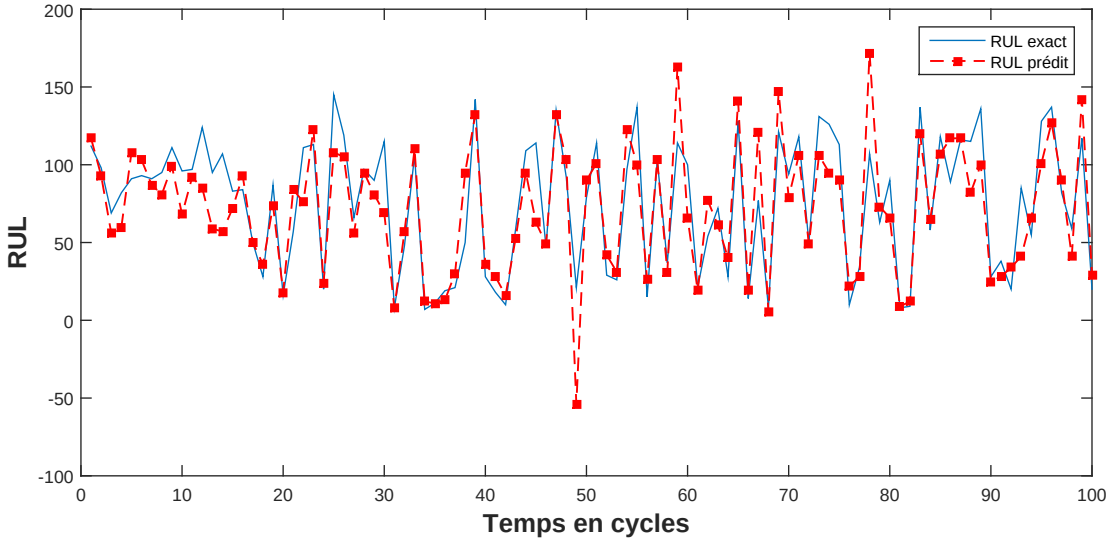
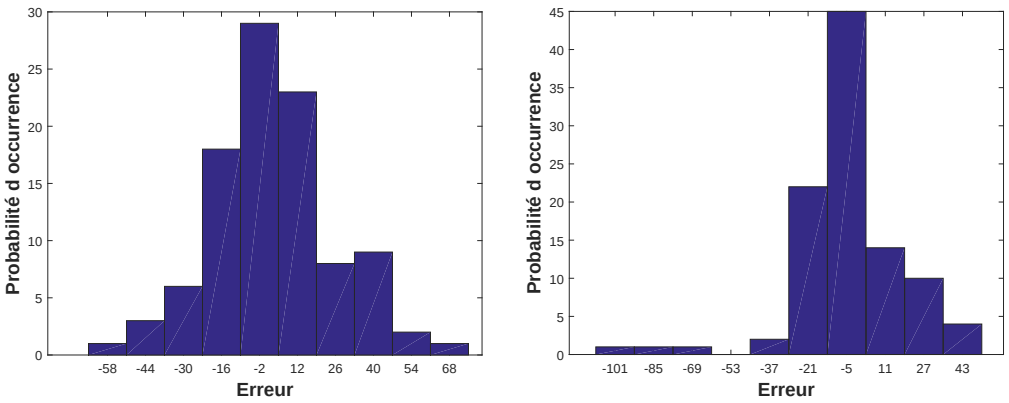


FIGURE 4.15 – RUL exact vs RUL prédit.

Le tableau 4.4 présente les résultats reportés dans la littérature pour le même jeu de données. Pour des raisons de comparaison, une méthode inspirée par le travail de [Wang et al., 2008] et suivant la même procédure a été réalisée et testée sur le même jeu de données.

Afin d'offrir un outil uniforme d'évaluation des approches présentées dans le tableau 4.4, on propose de quantifier les coûts de fausses prédictions. Le coût est inversement proportionnel à la performance et comme les prédictions précoces sont moins sévères par



(a) (b)

FIGURE 4.16 – L'histogramme de l'erreur (a) expérience formalisée par l'ajout de la connaissance, (b) expérience formalisée par les instances.

rapport aux prédictions tardives, elles sont moins pénalisées. (équation 4.6).

$$cout = 0.75 \times (predictions\,tardives) + 0.25 \times (predictions\,precoces) \tag{4.6}$$

TABLE 4.4 – Comparaison entre les résultats obtenus par les indicateurs de santé et ceux trouvés dans la littérature.

Approche	Prédictions correctes %	Prédictions précoces	Prédictions tardives	Coût
expérience formalisée par l'ajout de la connaissance	56%	32%	12%	17
expérience formalisée par les données (IS)	58%	24%	18%	19.5
Approche de[Ramasso et al., 2013]	53%	36%	11%	17.25
Approche de[Javed et al., 2013]	50%	40%	31%	19
Inspiré par [Wang et al., 2008]	50%	19%	31%	25

À partir du tableau de comparaison, on peut constater que l'approche décrite dans ce chapitre donne de bons résultats par rapport aux approches présentes dans le tableau 4.4. L'injection de la connaissance n'améliore pas forcement le taux de prédictions correctes par contre le coût de mauvaises prédictions est le plus bas. L'injection de la connaissance rend l'approche plus prudente, ce qui est traduit par un coût de mauvaise prédictions moins élevé pour l'approche formalisée par l'ajout de la connaissance. Basé sur ces deux critères (coût de mauvaises prédictions et le taux de correctes prédictions), on peut conclure que l'approche a une bonne performance.

TABLE 4.5 – Les caractéristiques extraites.

caractéristiques (features)	Formule
Moyenne	$\frac{\sum_{i=1}^{n} x_i}{n}$
Moyenne quadratique	$\sqrt{\frac{1}{n}(x_1^2 + x_2^2 + ... + x_n^2)}$
Kurtosis	$\frac{E(x-\mu)^4}{\sigma^4}$
Skewness	$\frac{\sum_{i=1}^{N}(x_i-\bar{x})^3}{(N-1)\sigma^3}$
Énergie	$e = \sum_{i=0}^{n} x_i 2$
Entropie	$H(x) = -\sum_x p(x).\log p(x)$

4.4.2/ ENSEMBLE DE DONNÉES BATTERIES

4.4.2.1/ PRÉPARATION DES DONNÉES

Contrairement aux données des turboréacteurs qui ont été directement utilisées dans l'approche propose, les données "batteries" exigent un certain traitement avant d'être exploitables par notre approche. On commence d'abord par l'extraction des caractéristiques pertinentes à partir des données capteurs. Ces caractéristiques peuvent être temporelles où fréquentielles. On propose d'extraire les caractéristiques temporelles du tableau 4.5. Parmi celles là, on sélectionnera seulement les caractéristiques qui sont le mieux adaptées à notre cas d'étude. La figure 4.17 illustre les caractéristiques de la variable "tension de charge" pour toutes les batteries.

La construction des indicateurs de santé requiert la disponibilité des données qui reflètent l'évolution de la dégradation. C'est-à-dire, des données qui présentent de tendances claires. À partir de la figure 4.17, on constate que seulement l'énergie satisfait cette condition. Il est difficile de tracer la dégradation des batteries en utilisant les autres caractéristiques. La moyenne et la moyenne quadratique (RMS) sont presque plates tandis que le kurtosis, skweness et l'entropie ne présentent pas de tendances claires. L'énergie est un signal complet qui considère la tension ainsi que l'intensité du signal.Elle reflète le comportement globale des batteries et semble être la caractéristique la plus adéquate pour notre cas d'étude. Par conséquent, les énergies du courant mesuré et de tension de cycle de charge ont été sélectionnées pour construire les indicateurs de santé developpés dans ce travail (c.f. Figure 4.18).

4.4.2.2/ RÉSULTATS ET DISCUSSION

L'approche présentée ici (expérience formalisée par l'ajout de la connaissance) a été appliquée sur le jeu de données des batteries. Les deux variables issues de la préparation des données (énergies du courant mesuré et de tension de cycle de charge) sont fusionnées afin de construire les indicateurs de santé tout en prenant en compte la connaissance obtenue à partir des trajectoires de dégradation. Le test a été effectué de la même manière que dans la section 3.7.2. Les parties suivantes décrivent les résultats obtenus et les comparent par rapport à la littérature et par rapport à l'approche basée sur les

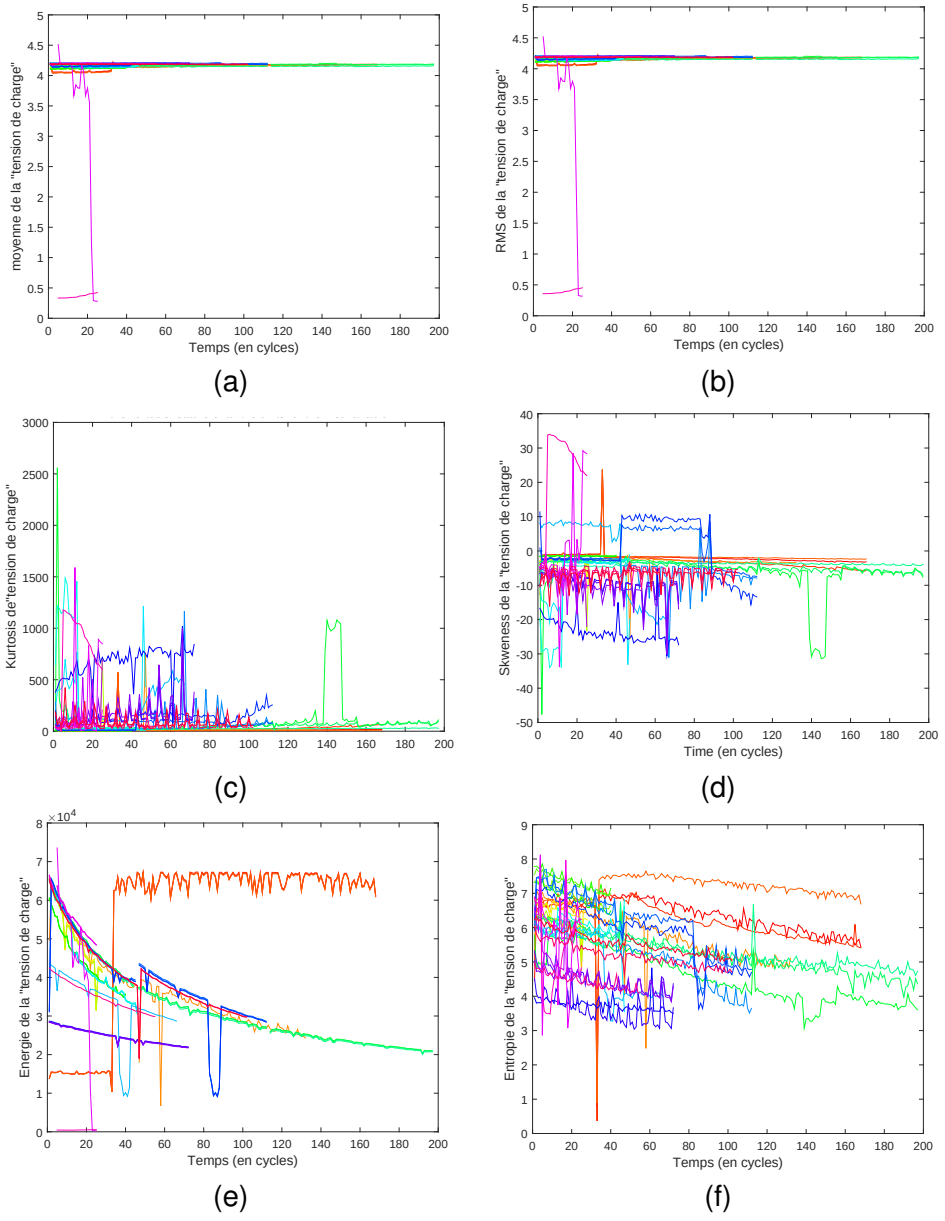
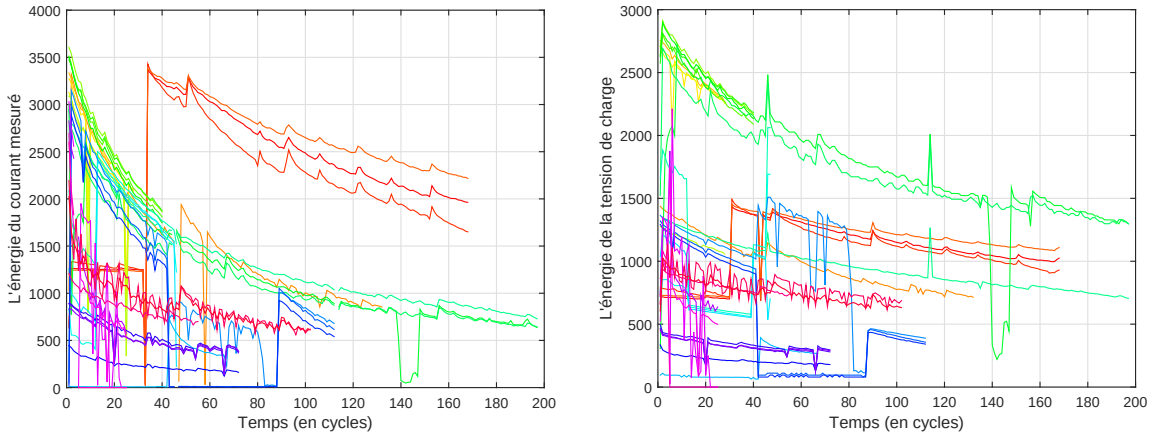


FIGURE 4.17 – Exemple de caractéristiques extraites à partir de la tension de charge (a) la moyenne, (b) RMS.(c) kurtosis, (d) skewness, (e) énergie, (f) entropie.

instance présentée au chapitre 3.

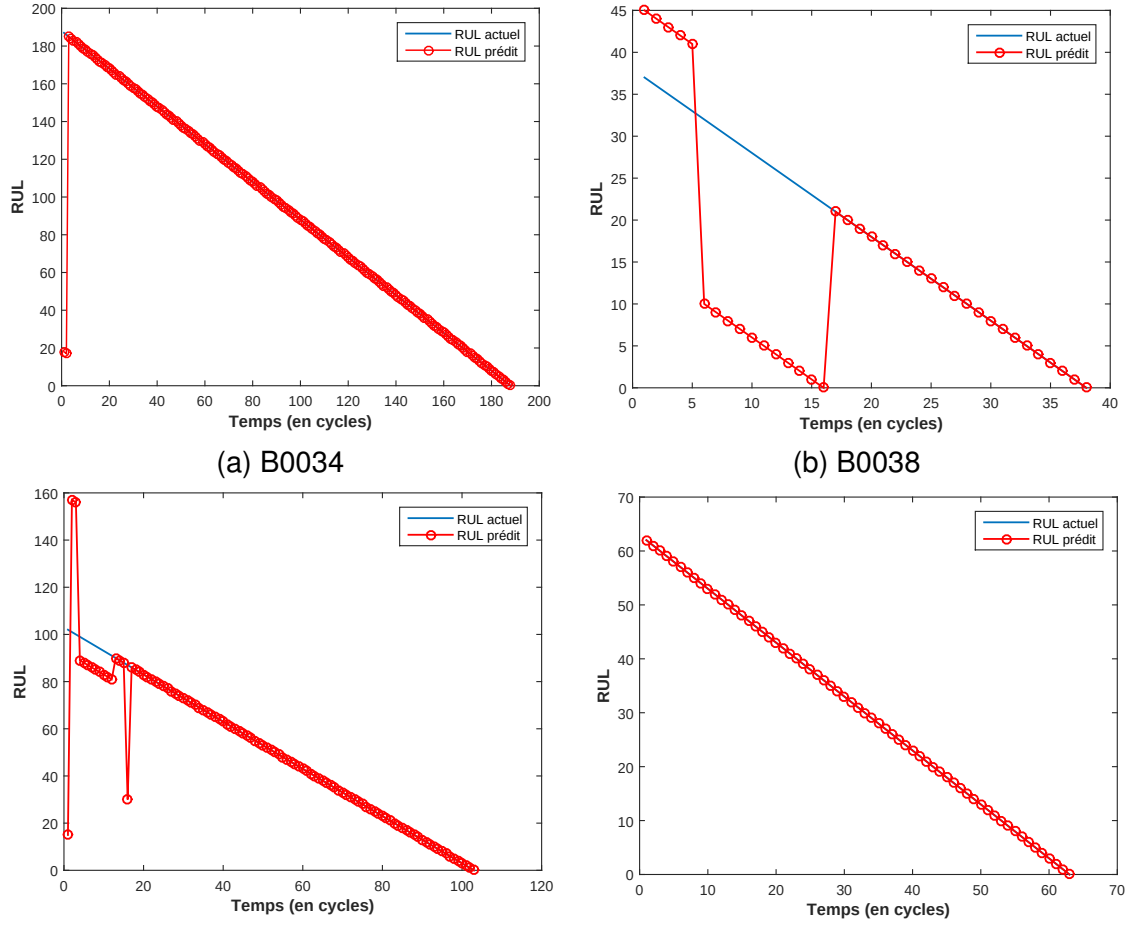
La figure 4.19 montre un exemple des résultats de prédiction du RUL faite pas à pas. Pour certaines batteries, le RUL a été correctement prédit à partir des premiers cycles. Pour d'autres (ex : B0038), les prédictions s'améliorent avec le temps (cycles). Ceci peut être expliqué par le fait que le cycle de vie de ces composants a été très court ce qui nous conduit à croire que l'expérience a mal tourné dès le début.

Comme on le voit à partir du tableau 4.6, l'injection de la connaissance temporelle et l'amélioration de construction des indicateurs de santés (cas) améliorent l'erreur de la prédiction et donnent la plus grande précision avec l'écart type le plus petit.



(a) (b)

FIGURE 4.18 – Les caractéristiques sélectionnées(a) l'énergie du courant mesuré de toutes les batteries, (b) l'énergie de la tension de toutes les batteries.



(a) B0034 (b) B0038

(c) B0042 (d) B0045

FIGURE 4.19 – Prédiction du RUL pour certaines batteries.

TABLE 4.6 – Les résultats de prédiction obtenus.

Test	IS sans connaissance			Trajectoire de dégradation			IS avec connaissance		
	EPMA	PRC	ETÉ	EPMA	PRC	ETÉ	EPMA	PRC	ETÉ
#1	7.412	0.925	2.060	30.353	0.696	3.601	11.073	0.889	3.4773
#2	0	1.00	0	0	1.00	0	0	1.00	0
#3	17.419	0.825	1.099	0	1.00	0	18.387	0.816	1.084
#4	0	1	0	0	1.00	0	0	1.00	0
#5	0	1	0	24.406	0.755	1.816	0.912	0.990	1.271
#6	23.563	0.764	2.570	30.022	0.699	3.489	26.090	0.739	2.387
#7	41.097	0.589	7.338	94.848	0.051	9.724	15.789	0.842	1.827
#8	0	1	0	13.072	0.869	2.468	2.999	0.970	1.309
#9	0	1	0	0	1.00	0	0	1.00	0
#10	0	1	0	0	1.00	0	0	1.00	0
#11	0	1	0	0	1.00	0	0	1.00	0
#12	77.00	1	0	88.00	0.12	0	37.500	0.625	3.708
#13	0	0.23	8.891	18.996	0.810	2.446	22.791	0.772	2.463
Moyen.	12.807	0.871	1.689	23.053	0.769	1.811	**10.426**	**0.895**	**1.3484**

Pour comparer les résultats obtenus par l'approche expérience formalisée par l'ajout de la connaissance avec d'autres approches du pronostic qui ont été testées sur ce jeu de données, on fait référence à [Mosallam et al., 2014]. Au meilleur de la connaissance de l'auteur, [Mosallam et al., 2014] ont été les seuls à aborder ce jeu de données de la même manière présentée dans ce chapitre. Le tableau 5.6 résume les moyennes de valeurs d'Erreur de Pourcentage en Moyenne Absolue (EPMA) obtenues par les approches comparées.

TABLE 4.7 – Comparaison entre les valeurs d'erreur de pourcentage en moyenne absolue.

Approche	EPMA
Approche à base d'instances (indicateurs de santé)	12.8071%
Approche à base d'instances (trajectoires de dégradation)	23.0538%
Expérience formalisée par la connaissance	**10.4265%**
[Mosallam et al., 2014]	32.2086%

L'approche orientée connaissance a donné les meilleurs résultats, l'inclusion de la connaissance temporelle lors de la construction des cas (indicateurs de santé) a été fructueuse. Elle permet de réduire l'erreur par plus de la moitié pour les batteries de petites valeurs de fin de vie (tableau 4.6) mais il est à noter que l'utilisation de trop de niveaux de dégradation pendant la construction de la matrice d'apprentissage pour la modélisation des indicateurs de santé peut conduire à un problème de sur-apprentissage. Dans ce travail, cinq niveaux de dégradation ont été utilisés (niveau 0%, 10%, 60%, 90% et 100%).

D'après l'évaluation de la méthode formalisée par l'ajout de la connaissance, on remarque que les deux types de connaissance (fréquentielle et temporelle) injectés à la phase de formalisation des cas et utilisés aussi pour raffiner la phase de remémoration améliorent les résultats de prédiction pour les deux types d'application, c'est-à-dire les turboréacteurs et les batteries.

4.5/ CONCLUSION

On s'est intéressé à améliorer particulièrement la construction précédente des indicateurs de santé qui n'utilisait que 20% des données et la phase de remémoration par l'injection de deux types de connaissance : (i) une connaissance temporelle qui a permis d'exploiter toutes les données capteur afin d'obtenir des règles utilisées pour construire les indicateurs de santé, (ii) une connaissance fréquentielle qui enrichit la présentation des cas et raffine la phase de remémoration. Cette connaissance a été extraite en utilisant la décomposition en modes empiriques.

Par conséquent, avec l'injection de la connaissance, les cas ont été présentés par deux dimensions : une dimension temporelle et l'autre fréquentielle qui sert à développer la mesure de similarité dans le temps et dans la fréquence.

Ces modifications apportées à la présentation de l'expérience ont constitué un pas vers l'évolution de l'approche à base d'instance (3) vers une approche typique de raisonnement à partir de cas. Le RàPC a été utilisé dans plusieurs domaines tels que l'estimation des variables et l'analyse des séries temporelles. Notre cas d'étude combine les deux aspects en estimant la durée de vie résiduelle à partir des séries temporelles.

L'approche a été appliquée sur les deux composants critiques utilisés tout au long de ce travail. Les résultats obtenus ont montré l'intérêt de l'injection de la connaissance. L'horizon de pronostic a été amélioré et l'erreur faite sur la prédiction a été réduite. La considération de la connaissance a rendu l'approche plus prudente. C'est-à-dire, que le taux des prédictions précoces était supérieur aux prédictions tardives ce qui est préféré en pronostic.

Dans le prochain chapitre en étudiera d'autres pistes de prédiction de durée de vie résiduelle par l'application d'une approche complémentaire basée sur la régression par les séparateurs à vaste marge.

5

ESTIMATION DIRECTE DU RUL VIA LES SVR

5.1/ INTRODUCTION

Dans ce chapitre, nous explorons une autre démarche, qui comme les approches précédentes ne nécessitent pas de définir un seuil de défaillance. On s'intéresse à l'estimation directe du RUL à partir des données capteur observées lors des expériences de suivi d'état du fonctionnement du composant. Pour une expérience donnée, à chaque valeur d'entrée, connaissant la durée de vie de l'expérience on associe le temps restant avant défaillance (le RUL). La représentation de l'expérience qu'elle soit mono-dimensionnelle, ou multidimensionnelle s'exprimera non pas en fonction du nombre de cycle de fonctionnement de la machine, mais du nombre de cycle restant avant défaillance. Cette approche est donc de type supervisée, sans qu'on fasse intervenir dans la représentation de l'expérience, l'état du composant définie au chapitre 3. Nous développons ainsi une technique d'apprentissage supervisée destinée à résoudre des problèmes de régression basée sur les machines à vecteurs de support ou séparateurs à vaste marge(en anglais Support Vector Machine, SVM) qui sont une généralisation des classifieurs linéaires.

À partir de cette nouvelle représentation de l'expérience, des caractéristiques sont extraites sur une fenêtre temporelle ayant comme entrée la valeur moyenne et la pente de la courbe et comme sortie (le RUL). Ces caractéristiques sont soumises à un modèle de régression à vecteurs de support (SVR) bien connu pour son traitement de données non-linéaires et présentant un grand pouvoir de discrimination sur d'importante quantité de données. Le RUL est ainsi exprimé en fonction de ces caractéristiques. Nos travaux sont les premiers à proposer une approche directe d'estimation du RUL.

Cette approche a été appliquée sur des séries temporelles multidimensionnelles, c'est-à-dire directement sur les données de surveillance capteurs et mono-dimensionnelles c'est-à-dire sur les indicateurs de santé mis en place au chapitre 3. Nous avons ainsi pu comparer (i) le pouvoir discriminant des SVR par rapport à la phase de remémoration définie dans les approches précédentes. (ii) les résultats obtenus via un indicateur de santé ou directement sur les données capteurs, pour décider s'il est vraiment utile de définir une courbe de tendance pour l'estimation du RUL.

De plus dans un souci d'amélioration de nos résultats, nous avons fait une sélection de données capteurs, en appliquant une méthode enveloppe, ce qui nous a permis d'obtenir

de meilleurs résultats.

Le chapitre est organisé de la manière suivante : la section 5.2 décrit les différentes types d'estimation du RUL dans le but de positionner le travail présenté dans ce chapitre. Après un rapide rappel sur les SVM à la section 5.3, un état de l'art sur les travaux de prédiction des séries temporelle à base de SVR est mené à la section 5.4. Puis à la section 5.5, nous proposons notre méthode à base de SVR. La section 5.6 présente la méthode de sélection de variable associée à l'approche proposée. La section 5.7 détaille les résultats obtenus après l'application de l'approche sur les mêmes jeux de données.

5.2/ L'ESTIMATION DU RUL

L'objectif principal d'une approche de pronostic est d'obtenir une estimation exacte du temps restant avant défaillance, (RUL). Il existe trois types de prévisions du RUL qui peuvent être réalisées par des modèles de pronostic basés sur soit les états de dégradation soit la régression ou sur l'appariement de formes.

— **Modèles de prédiction du RUL basés sur des états de dégradation** qui estiment l'état de santé actuel du système (niveau de dégradation) et prédisent le comportement du système dans le futur. Le modèle est souvent constitué de deux modules : un classifieur qui détermine l'état de dégradation le plus probable et un module prédictif qui continue à prévoir les futures observations jusqu'à ce qu'un seuil de défaillance soit atteint. Le RUL est obtenu comme la durée des transitions nécessaires pour atteindre l'état défaillant (figure 5.2), [Ramasso et al., 2013, Javed et al., 2015, Tamilselvan et al., 2013].

— **Modèles de prédiction du RUL basés sur la régression** qui surveillent et prévoient l'évolution d'un signal dégradant et ensuite estiment le RUL comme le temps nécessaire pour que le signal atteigne le critère de fin de vie (seuil de dégradation). Ces approches n'exigent pas un historique de données de défaillance, mais le seuil doit être fixé (figure 5.1), [Wu et al., 2007, Tse et al., 1999, Wang, 2007].

— **Modèles de prédiction du RUL basés sur l'appariement de formes** comparent les observations actuelles à une bibliothèque d'exemples de tendance qui contiennent une information à propos de valeurs de fin de vie utilisée pour estimer le RUL (figure 5.3) [Malinowski et al., 2014, Wang et al., 2008, Khelif et al., 2014].

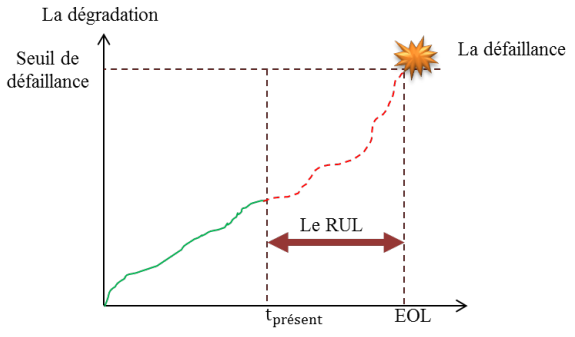


FIGURE 5.1 – Modèle de prédiction du RUL basé sur la régression.

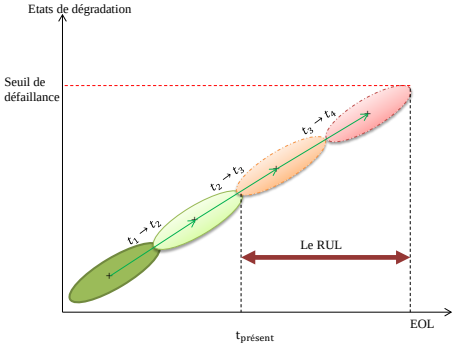


FIGURE 5.2 – Modèle de prédiction du RUL basé sur les états de dégradation.

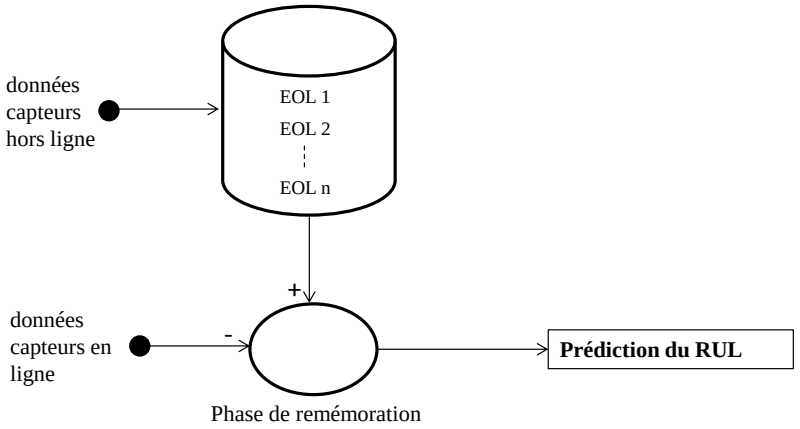


FIGURE 5.3 – Modèle de prédiction du RUL basé sur l'appariement de formes.

Les modèles de prédiction basés sur les états de dégradation et régression nécessitent une connaissance sur le seuil de dégradation. Par contre ce seuil est un des verrous de pronostic. Afin d'éviter la question problématique de définition de seuil de défaillance, qui doit être fixé pour estimer le RUL, on a développé des modèles de prédiction du RUL basés sur l'appariement de formes qui prennent appui sur les méthodes des K plus proches voisins telles que les méthodes à base d'instance (c.f. Chapitre 3), raisonnement à partir de cas (c.f. Chapitre 4) et dans ce chapitre on proposera un autre modèle prédictif qui modélise le RUL directement par une régression à vecteurs de support (SVR) en fonction de caractéristiques extraites à partir des données de surveillance.

5.3/ LES MACHINES À VECTEURS DE SUPPORT

Les machines à vecteurs de support appartiennent à une famille de classifieurs linéaires généralisés. En d'autres termes, les SVM sont un outil de prédiction de classification et de régression qui utilise la théorie d'apprentissage automatique pour maximiser la précision prédictive tout en évitant automatiquement le sur-apprentissage des données. Les machines à vecteurs de support peuvent être définies comme des systèmes qui cartographient une fonction non-linéaire dans un espace de dimension plus élevé.

Les fondations des SVM ont été développées par [Vapnik, 1995] et ont gagné une population due à leurs nombreuses caractéristiques prometteuses comme la meilleure performance empirique. La formulation utilise le principe de Minimisation de Risque Structurel (MRS) qui a été montré supérieur au principe traditionnel de Minimisation de Risque Empirique (MRE) utilisé par les réseaux de neurones classiques [Burges, 1998].

Le MRS minimise une borne supérieure sur le risque attendu, alors que le MRE minimise l'erreur sur les données d'apprentissage. C'est cette différence qui équipe les SVM avec une meilleure généralisation [Jakkula, 2006].

Les SVM ont été initialement développés pour résoudre le problème de classification, mais récemment, ils ont été étendus à la résolution de problèmes de régression.

5.3.1/ LES MACHINES À VECTEURS DE SUPPORT LINÉAIRES

On se base sur les travaux de [Burges, 1998] pour définir les bases des SVM. Les lecteurs intéressés par plus de détails sont invités à consulter les travaux de Burges.

5.3.1.1/ LE CAS SÉPARABLE

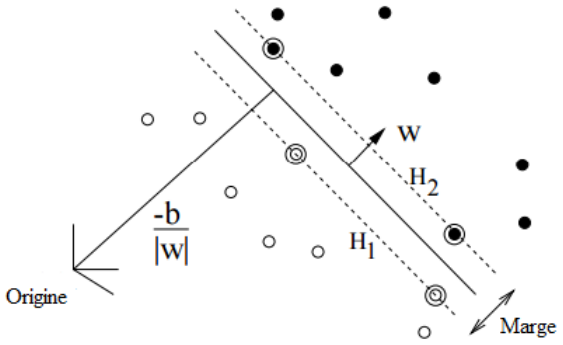


FIGURE 5.4 – Des hyperplans linéaires séparateurs pour le cas séparable. Les vecteurs de support sont encerclés [Burges, 1998] .

Les machines sont entrainées sur des données séparables et les données d'apprentissage sont d'abord labellisées : $\{\mathbf{X}_i, y_i\}$, $i = 1, \ldots, l$, $y_i \in \{-1, 1\}, \mathbf{X}_i \in \mathbf{R}^d$.

On suppose que la construction des hyperplans qui séparent les exemples positifs des exemples négatifs est faisable.

Les points $\mathbf{X}$ qui se trouvent sur l'hyperplan satisfont l'équation $\mathbf{W}.\mathbf{X} + b = 0$, où $\mathbf{W}$ est perpendiculaire à l'hyperplan, $|b|/\|\mathbf{W}\|$ est la distance perpendiculaire à partir de l'hyperplan à l'origine, et $\|\mathbf{W}\|$ est la norme euclidienne de $\mathbf{W}$.
Soit $d_+(d_-)$ la distance la plus courte entre "l'hyperplan séparateur" et l'exemple positif (négatif) le plus proche. On définit la "marge d'un hyperplan, séparateur comme $d_+ + d_-$.

Pour le cas séparable, les vecteurs de support recherchent simplement l'hyperplan séparateur ayant la plus grande marge. En supposant que toutes les données d'apprentissage satisfont certaines contraintes, le problème peut être formulé comme :

$$\mathbf{X}_i \,.\, \mathbf{W} + b \geq +1 \; pour \; y_i = +1, \tag{5.1}$$

$$\mathbf{X}_i \,.\, \mathbf{W} + b \leq -1 \; pour \; y_i = -1, \tag{5.2}$$

ceux-si peuvent être combinés en un seul ensemble d'inégalités :

$$y_i(\mathbf{X}_i \,.\, \mathbf{W} + b) - 1 \geq +1 \; pour \; \forall i, \tag{5.3}$$

on considère seulement les points pour lesquels l'égalité dans l'équation 5.1 est satisfaite. Ces points se trouvent sur l'hyperplan $H_1 : \mathbf{X}_i \,.\, \mathbf{W} + b = +1$ avec un perpendiculaire $\mathbf{W}$ et une distance perpendiculaire à partir de l'origine $|1 - b|/\|\mathbf{W}\|$.

De même, les points pour lesquels l'égalité dans l'équation 5.2 est satisfaite, se trouvent sur l'hyperplan H_2 : $textbfX_i$. $\mathbf{W} + b = -1$ avec une distance perpendiculaire à partir de l'origine $|-1-b|/\|\mathbf{W}\|$.

Par conséquent, $d_+ = d_- = 1/\|\mathbf{W}\|$ et la marge est simplement $2/\|\mathbf{W}|$.

Il est à noter que H_1 et H_2 sont parallèles et qu'il n'y a pas de données d'apprentissage comprises entre les deux hyperplans. Ainsi, on peut trouver la paire d'hyperplans qui donne la marge maximale en minimisant $\|\mathbf{W}\|^2$n sujet aux contraintes de l'égalité 5.3.

La solution attendue pour un cas typique à deux dimensions devrait avoir la forme présentée dans la figure 5.4. Les vecteurs de support sont les points d'apprentissage pour lesquels l'égalité l'équation 5.3 est satisfaite et dont l'enlèvement changerait la solution trouvée. Ces vecteurs sont indiqués sur la figure 5.4 par les cercles supplémentaires.

Dans la suite, nous représentons le problème avec une formulation Lagrangienne pour deux raisons : (i) les contraintes de l'équation 5.3 sont remplacées par des multiplicateurs de Lagrange qui sont beaucoup plus facile à manipuler. (ii) Dans cette re-formulation, les données d'apprentissage apparaîtront seulement sous la forme de produit scalaires entre les vecteurs. Ceci est une propriété cruciale qui permettra de généraliser la procédure aux cas non-linéaires.

Les multiplicateurs positifs de Lagrange $\alpha_i, i = 1, \ldots, l$, un pour chaque inégalité de la contraintes 5.3 sont introduits [1].

Pour des contraintes d'égalité, les multiplicateurs de Lagrange sont sans contraintes. Cela donne le Lagrangien :

$$L_p = \frac{1}{2}\|\mathbf{W}\|^2 - \sum_{i=1}^{l} \alpha_i y_i (\mathbf{X}_i \,.\, \mathbf{W} + b) + \sum_{i=1}^{l} \alpha_i, \tag{5.4}$$

L_p est minimisé par rapport à $\|\mathbf{W}\|$, b tout en exigeant que les dérivées de L_p par rapport à α_i disparaissent, tous soumis aux contraintes $\alpha_i \geq 0$.
On peut résoudre le problème dual : maximiser L_p, soumis également aux contraintes que $\alpha_i \geq 0$. Cette formulation a la propriété que la maximisation de L_p se produit aux mêmes valeurs de $\mathbf{W}$, b, et α, comme la minimisation de L_p.

Les conditions d'annulation de gradient de L_p par rapport à $\mathbf{W}$ et b sont :

$$\mathbf{W} = \sum_i \alpha_i y_i \mathbf{X}_i, \tag{5.5}$$

$$\sum_i \alpha_i, \tag{5.6}$$

Puisque ce sont des contraintes d'égalité dans la formulation duale, on peut les remplacer dans l'équation- 5.4 :

$$L_D = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \mathbf{X}_i \,.\, \mathbf{X}_j, \tag{5.7}$$

Les vecteurs de support d'apprentissage s'élèvent à maximiser L_D par rapport à α_i soumis aux contraintes 5.6 et la positivité de α_i avec la solution données dans l'équation 5.5.

1. La règle est que pour les contraintes de la forme $cr_i \geq$, les équations de contraintes sont multipliées par des multiplicateurs de Lagrange positifs et soustraits de la fonction objective, pour former le Lagrangien.

Dans la solution, les points pour lesquels $\alpha_i > 0$ sont appelés "vecteurs de support" et se trouvent dans un des hyperplans H_1, H_2.

5.3.1.2/ LE CAS NON SÉPARABLE

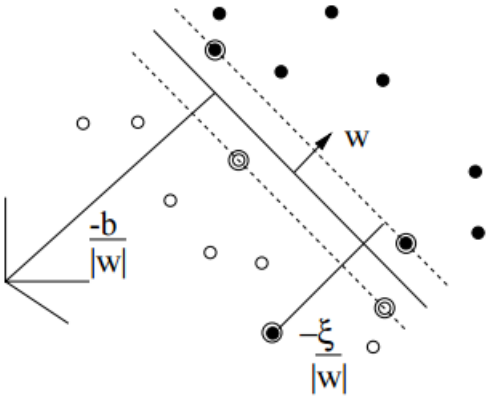


FIGURE 5.5 – Des hyperplans linéaires séparateurs pour le cas non séparable [Burges, 1998] .

Lorsque les observations ne sont pas séparables par un plan, il est nécessaire de relaxer les contraintes des équations 5.1, 5.2 quand c'est nécessaire. On veut représenter un coût supplémentaire en introduisant des variables d'écart positives ζ_i, $i = 1, \ldots, l$ dans les contraintes qui deviennent :

$$\mathbf{X}_i \,.\, \mathbf{W} + b \geq +1 - \zeta_i \; pour \; y_i = +1, \tag{5.8}$$

$$\mathbf{X}_i \,.\, \mathbf{W} + b \leq -1 + \zeta_i \; pour \; y_i = -1, \tag{5.9}$$

$$\zeta_i \geq 0 \forall i \tag{5.10}$$

Le modèle attribue ainsi une réponse fausse à un vecteur $\mathbf{X}_i$ si le ζ_i correspondant dépasse l'unité. La somme de tous les ζ_i représente donc une limite du nombre d'erreurs. Le problème de minimisation est réécrit en introduisant une pénalisation pour le dépassement de la contrainte :

$$\frac{1}{2}\|\mathbf{W}\|^2 + C(\sum_i \zeta_i)^k, \tag{5.11}$$

où Cest un paramètre choisi par l'utilisateur. Plus il est grand et plus cela revient à attribuer une pénalité plus élevées aux erreurs.

5.3.2/ LES MACHINES À VECTEURS DE SUPPORT NON LINÉAIRES

L'idée est d'utiliser un classifieur linéaire pour résoudre un problème non-linéaire, en transformant l'espace de caractéristiques d'entrées en un espace de plus grande dimension, où le classifieur linéaire est alors utilisé.

Les données sont d'abord cartographiées en utilisant la cartographie Φ :

$$\Phi : \mathbf{R}^d \rightarrow \mathcal{H}, \tag{5.12}$$

où $\mathcal{H}$ est l'espace euclidien.

L'algorithme d'apprentissage dépend des données par le biais de produits scalaires dans $\mathcal{H}$, à savoir sur les fonctions de la forme :

$$\Phi(\mathbf{X}_i) \,.\, \Phi(\mathbf{X}_j) \tag{5.13}$$

Ce produit scalaire est effectué dans un espace de grande dimension, ce qui conduit à des calculs impraticables. L'idée est donc de remplacer ce calcul par une fonction noyau telle que :

$$K(\mathbf{X}_i, \mathbf{X}_j) = \Phi(\mathbf{X}_i) \,.\, \Phi(\mathbf{X}_j) \tag{5.14}$$

Les fonctions noyaux à utiliser sont définies par la condition de Mercer, qui montre qu'étant donné une fonction noyau continue, symétrique, semi-définie positive $K(\mathbf{X}_i, \mathbf{X}_j)$, la fonction peut s'exprimer comme un produit scalaire dans un espace de grande dimension.

Plus précisément, il existe une cartographie Φ et une expansion

$$K(\mathbf{X}, \mathbf{Y}) = \sum_i \Phi(\mathbf{X})_i \Phi(\mathbf{Y})_i \tag{5.15}$$

Si et seulement si, pour toute $g(\mathbf{X})$ tel que

$$\int g(\mathbf{X})^2 d\mathbf{X} \text{ est fini} \tag{5.16}$$

Donc

$$\int K(\mathbf{X}, \mathbf{Y}) g(\mathbf{X}) g(\mathbf{Y}) d\mathbf{X} d\mathbf{Y} \geq 0, \tag{5.17}$$

cette condition est satisfaite pour des puissances entières positives du produit scalaire : $K(\mathbf{X}, \mathbf{Y}) = (\mathbf{X} \,.\, \mathbf{Y})^p$

5.3.2.1/ LES FONCTIONS NOYAUX

Les fonctions de noyau cartographient les attributs de l'espace d'entrées à l'espace de caractéristiques. Elles jouent un rôle critique dans les SVM et leurs performances.
Les différentes fonctions noyaux sont énumérées ci-dessous :

1. **polynôme :** une cartographie polynomiale est une méthode populaire pour la modélisation non linéaire.
$$K(\mathbf{X}_i, \mathbf{X}_j) = \langle \mathbf{X}_i, \mathbf{X}_j \rangle^d,$$
$$K(\mathbf{X}_i, \mathbf{X}_j) = (\langle \mathbf{X}_i, \mathbf{X}_j \rangle + 1)^d,$$
le deuxième noyau est généralement préférable car il évite des problèmes de type le hessien devenant zéro [Jakkula, 2006].

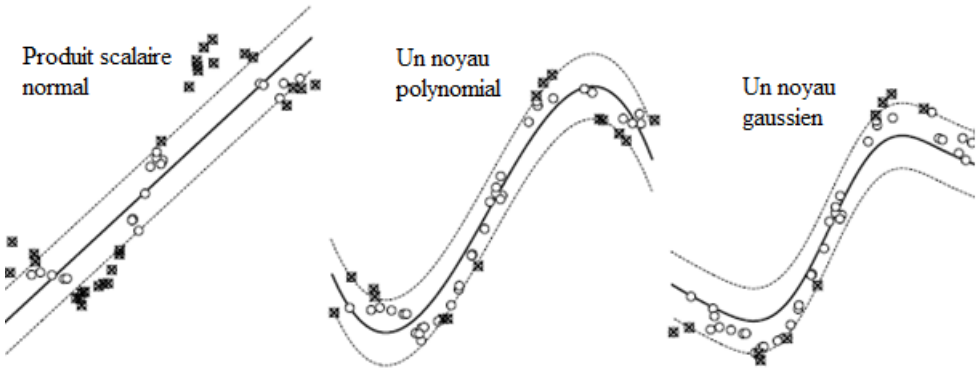


FIGURE 5.6 – Régression sur des vecteurs 1D [Frezza-Buet, 2012].

2. **fonction gaussienne de base radiale :** le plus souvent utilisée avec une forme gaussienne

$$K(\mathbf{X}_i, \mathbf{X}_j) = exp(-\frac{\|\mathbf{X}_i - \mathbf{X}_j\|^2}{2\sigma^2})$$

3. **fonction exponentielle de base radiale :** la fonction produit une solution linéaire par morceau, ce qui peut être intéressant lorsque mes discontinuités sont acceptables.

$$K(\mathbf{X}_i, \mathbf{X}_j) = exp(-\frac{\|\mathbf{X}_i - \mathbf{X}_j\|}{2\sigma^2})$$

5.3.3/ LES MACHINES À VECTEURS DE SUPPORT POUR LA RÉGRESSION

Les SVM peuvent être également appliqués à des problèmes de régression par l'introduction d'une variante de la fonction de perte.

Les mêmes principes décrits précédemment sont utilisés , avec quelques différences. Dans le cas de la régression, la sortie est un nombre réel et pas un label, il devient très difficile de prédire la variable, le nombre de possibilités est infini. Une marge de tolérance (epsilon) est alors fixée en approximation du SVM. l'objectif est de trouver une fonction $f(\mathbf{X})$ qui produit au maximum une déviation ϵ de la valeur cible $\mathbf{Y}_i$ pour tout l'ensemble de données d'apprentissage, et en même temps est aussi plate que possible. Autrement dit, nous ne nous soucions pas des erreurs tant qu'il sont moins de ϵ. En revanche, toute erreur supérieure à ϵ ne sera pas acceptée.
l'idée principale est toujours la même tout en gardant à l'esprit qu'une partie de l'erreur est tolérée (ϵ-SVR). La Figure 5.6 illustre un exemple de différents noyaux pour la régression ; à gauche il y a l'utilisation d'un noyau linéaire, au milieu d'un noyau polynomial de degré 3, et à droite, d'un noyau gaussien. Les vecteurs supports sont marqués d'une croix. La tolérance $+\epsilon$, $-\epsilon$ est représentée en pointillés.

Plus de détail sur ϵ-SVR seront donnés à la section 5.5.3

5.4/ ÉTAT DE L'ART SUR LES SVR

De plus en plus de travaux s'intéressent au modèle régressif des SVM, lors de la résolution de problèmes de prédiction. Les SVM ont été employés avec succès dans différentes applications de prédiction de séries temporelles comme la prévision de valeurs de productions de machines industrielles, prédiction de la vitesse du vent, la prévision de séries temporelles financières, et la prédiction de radiation solaire, [Mohandes et al., 2004, Ince et al., 2006, Cao, 2003].

La performance des SVR pour la prédiction des séries temporelles a été comparée à plusieurs méthodes basée sur les réseaux de neurones[Cao et al., 2001, Tay et al., 2001, Thissen et al., 2003]. Les SVR ont présentés de meilleurs résultats par rapport aux réseaux neuronaux qu'ils soient perceptron multicouche,à retour de propagation(back-propagation) [Cao et al., 2001, Tay et al., 2001] ou bien qu'ils utilisent des réseaux récurrents des Elman [Thissen et al., 2003]. La supériorité des SVR peut être expliquée par la façon dont l'optimisation est faite. Les réseaux de neurones traditionnels appliquent le principe de minimisation de risque empirique (Empirical Risk Minimization) tandis que les SVM/ SVR se basent sur la minimisation du risque structurel (Structural Risk Minimization) qui induit une meilleure généralisation. La minimisation du risque structurel [Vapnik et al., 1974] est un principe d'induction pour la sélection du modèle utilisé pour l'apprentissage à partir d'ensembles de données d'apprentissage finis. Il décrit un modèle général de contrôle de la capacité et offre un compromis entre la complexité et la qualité de l'ajustement des données d'apprentissage (erreur empirique).

Parmi les exemples d'application réussie des SVR dans la prédiction des séries-temporelles on compte le travail de [Lu et al., 2009]. La méthode proposée combine les SVR avec l'analyse en composantes indépendantes permettant de prédire les indices d'ouverture et de clôture de trésorerie de Nikkei 225. L'analyse en composantes indépendantes est utilisée pour filtrer le bruit et les données aberrantes. Les variables de prévision filtrées sont les entrées du SVR pour la construction du modèle prédictif. [Cao, 2003] associent les SVR avec les cartes auto-adaptatives. Ces dernières sont utilisées pour classer les données de tâches solaires en différentes régions. Au lieu de construire un seul modèle à partir de l'ensemble de données (modèle unique), ces différentes régions sont apprises séparément par les modèles SVR (les experts) les plus appropriés. Les résultats ont montré que les SVR experts atteignent une amélioration significative de la performance de généralisation en comparaison au modèle SVR simple et unique. [Pai et al., 2005] ont étudié l'applicabilité des SVR dans la prévision de valeurs de production de machine industrielles à Taïwan sachant que ces données montrent une forte saisonnalité et des tendances croissantes. La performance des SVR a été comparée avec une approche de réseaux de neurones et une approche auto-régressive. Les résultats expérimentaux indiquent que les SVR dépassent les autres approches en termes de précision et de prédiction.

Dans nos travaux on ne s'intéresse pas seulement à la prédiction de séries temporelles mais à la prédiction de durée de vie résiduelle (RUL) des composants critiques, autrement dit au pronostic. Dans ce domaine, on compte les travaux de [Soualhi et al., 2015] qui ont extrait des indicateurs de santé à l'aide de la transformée de Hilbert. Les signaux d'entrée traités ont été classés en états de dégradation grâce aux SVMs. Le RUL a été obtenu à l'aide d'un algorithme de SVR destiné à faire une prédiction pas à pas de la série temporelle étudiée. Le RUL a été calculé comme le temps minimum nécessaire pour atteindre

le seuil de dégradation, seuil au préalable posé par l'expert. Dans [Wang et al., 2014], les SVR ont été utilisés pour prédire itérativement les valeurs des batteries « Li-ion » en utilisant les valeurs du passé. Les données d'apprentissage ont été divisées en plusieurs régions en fonction de la complexité de la distribution des données.Différents paramètres ont été fixés à chaque région. Le RUL a été calculé comme le temps nécessaire pour atteindre le seuil de défaillance. Dans [Li et al., 2006], un filtre à particules a été utilisé pour estimer l'état de santé d'un moteur à induction triphasé et les SVR ont été utilisés pour estimer la condition à venir sur moteur afin de prédire la disponibilité du composant. Le temps de disponibilité a été mesuré comme le temps requis pour atteindre le niveau défaillant. [Qin et al., 2015] ont saisi la tendance globale de la dégradation de l'état de santé des batteries « Li-ion ». Les paramètres des SVR ont été définis avec une optimisation par essaims et le RUL a été estimé grâce à un seuil de défaillance. [Benkedjouh et al., 2013] ont construit des indicateurs de santé des roulements grâce à l'outil de réduction de dimension cartographique isométrique (ISOMAP). Le calcul de RUL est fait par la prédiction de futures valeurs des indicateurs de santé et par la définition du temps nécessaire pour atteindre la défaillance.[Widodo et al., 2011] ont combiné les SVR avec l'analyse de survie. L'analyse de survie est une collection de techniques statistique utilisées afin d'estimer la probabilité de survie et le temps de défaillance d'un composant et les SVR ont servis à prédire cette probabilité pour l'unité étudiée.

Les SVR ont une bonne généralité et la capacité de traiter des données non linéiares mais leur application dans le pronostic a deux limitations majeures :

— Dans le cas d'une prédiction à long terme (multi-step ahead prediction), l'erreur est accumulée à chaque pas. La valeur à prédire à T_{i+n} est obtenue en fonction de la valeur déjà prédite et des valeurs du passé
$$Var(T_{i+n}) = f(Var(T_1), \ldots, Var(T_i), Var(T_{i+1}), \ldots, Var(T_{i+n-1})).$$
Cette accumulation d'erreur, limite l'applicabilité à des prédictions faites un pas à l'avance ce qui n'est pas très utile en pronostic et n'offre pas aux agents de maintenance le temps nécessaire pour intervenir.
— La nécessité de définir un seuil de défaillance afin de calculer le RUL, qui est un des verrous du pronostic. Un seuil statique (le cas le plus commun) n'est pas pratique à définir et est posé empiriquement. En réalité le seuil de défaillance dépend du profil d'utilisation du composant mais malheureusement, les algorithmes de pronostic ne sont pas adaptés à des seuils dynamiques.
— Afin d'éviter ces écueils, nous utiliserons les SVR sans avoir à définir de seuil de défaillance, et nous ne définirons pas de prédiction pas à pas.

5.5/ PROPOSITION D'UNE APPROCHE D'ESTIMATION DU RUL PAR LES SVR

5.5.1/ CADRE GÉNÉRAL DE L'APPROCHE

La figure 5.7 illustre le cadre général de notre approche, qui est composée de deux phases : une hors-ligne relative à l'étape d'apprentissage de l'algorithme et une en ligne.

Le module de traitement de données dans la figure 5.7 peut par exemple correspondre au dé-bruitage des signaux d'entrée, à la transformation des mesures capteurs multidimensionnelles en indicateurs de santé, ou à la sélection de variables. À partir des signaux

d'entrée traitées, les caractéristiques de tendance sont extraites. Ces caractéristiques et les valeurs associées des RUL constituent les entrées de l'algorithme de SVR. Les SVR dans la phase offline sont entrainés à apprendre la relation liant les caractéristiques extraites et le RUL pour l'ensemble de l'apprentissage. En ligne, on extrait des caractéristiques de façon similaire sur les données tests. On donne ces caractéristiques au modèle SVR déjà appris, qui va estimer le RUL.

Dans ce chapitre, le filtrage des signaux et du bruit les accompagnant est obtenu par une méthode de lissage qui sera brièvement décrite dans la partie application. La construction des indicateurs de santé reprend les travaux faits au chapitre 3.

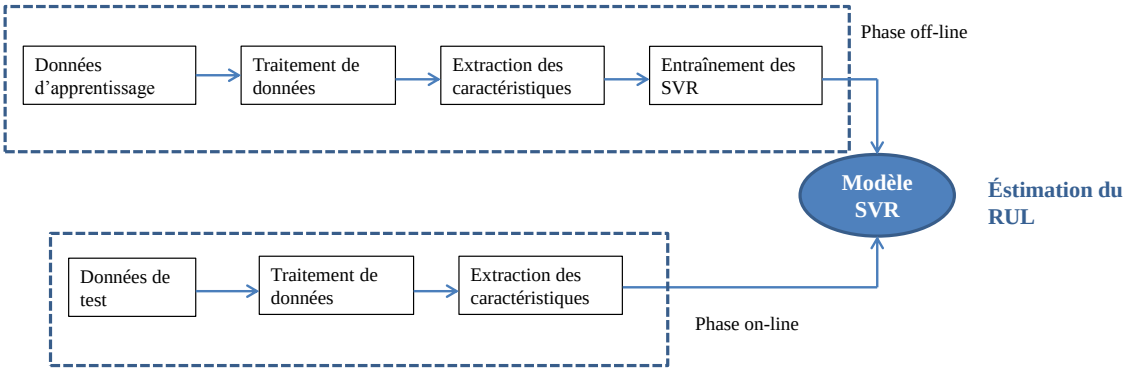


FIGURE 5.7 – Le cadre général de l'approche proposée.

Soit $\mathcal{T}$ un ensemble de données d'apprentissage composé de $|\mathcal{T}|$ séries temporelles (expériences), $T_1, \ldots, T_{|\mathcal{T}|}$ qui sont le résultat de l'étape traitement de données. Chaque série temporelle de cet ensemble représente la surveillance d'un équipement jusqu'à la défaillance. Par conséquent, les séries temporelles dans cet ensemble peuvent avoir des longueurs différentes. La longueur de T_i est dénotée par$l(T_i)$. Avec cette notation, la série temporelle T_i appartenant à $\mathcal{T}$ peut-être écrite comme $T_i = t_1^i, t_2^i, \ldots, t_{(l(T_i))^i}$. En fonction de l'application, la série temporelle peut-être uni-variée ou multivariée, c'est-à-dire $t_j^i \in \mathbb{R}^d, d \geq 1, 1 \leq i \leq |\mathcal{T}|, 1 \leq j \leq l(T_i)$.

Dans cette section, nous cherchons d'abord à préparer l'ensemble de séries temporelles $|\mathcal{T}|$ (soit par une sélection de variable, soit par la construction d'indicateurs de santé), à extraire des caractéristiques (features) à partir de cet ensemble et d'apprendre une relation entre les caractéristiques extraites et la durée de vie résiduelle. Cette relation est apprise par les SVR. La méthode proposée est basée sur quatre étapes, qui sont détaillées ci-après :

1. Pré-traitement des données (filtrage du bruit) .
2. Extraction de caractéristiques et association avec la valeur du RUL.
3. Apprentissage du modèle SVR entre les caractéristiques et le RUL.
4. Prédiction du RUL d'un nouvel équipement en utilisant ce modèle.

5.5.2/ EXTRACTION DE CARACTÉRISTIQUES DE TENDANCE À PARTIR DES SÉRIES TEMPORELLES

Nous considérons ici des caractéristiques qui décrivent à la fois la valeur et la tendance de la série temporelle sur une fenêtre donnée. La série temporelle T_i est d'abord décomposée en fenêtres concaténées de taille L (où L est un des paramètres de la méthode). La fenêtre k de la série temporelle T_i est composée de $t^i_{(k-1)L+1}, \ldots, t^i_{kL}$. Deux paramètres par dimension sont extraits à partir de chaque fenêtre : la valeur moyenne "a" sur la fenêtre et le coefficient de tendance s de la régression linéaire sur la fenêtre. Par conséquent, un vecteur de caractéristiques de taille $2 \times d$ est calculé à partir de chaque fenêtre. On associe ensuite à chaque vecteur de caractéristiques sa durée de vie résiduelle. Cette durée pour la fenêtre k de T_i est : $(T_i) - kL$.

En répétant ces opérations pour la totalité de séries temporelles $\mathcal{T}$ on obtient en un ensemble d'apprentissage $\Psi = (x_1, y_1), \ldots, (x_N, y_N)$ où $x_i \in \mathbb{R}^{2d}$ correspond aux caractéristiques extraites sur chaque fenêtre de chaque série temporelle et y_i correspond à la durée de vie résiduelle de la série temporelle de la dernière instante de la fenêtre associée. Cette étape est illustrée à la figure 5.8.

5.5.3/ APPRENTISSAGE D'UN MODÈLE SVR ENTRE LES CARACTÉRISTIQUES ET LE RUL

Après l'étape d'extraction de caractéristiques décrite ci-dessus, l'ensemble d'apprentissage Ψ est disponible. Nous visons dans cette étape à prédire la variable cible $\ll y \gg$ en utilisant les variables prédictives $\ll x \gg$ grâce à la régression.

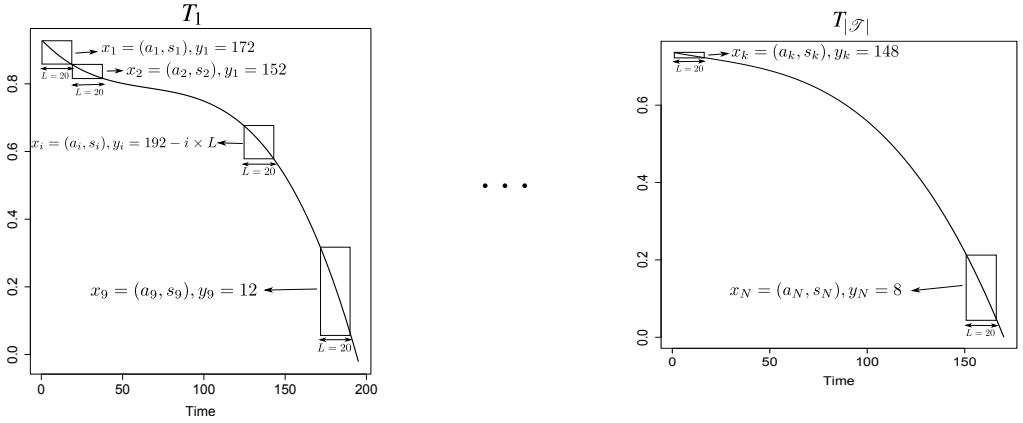


FIGURE 5.8 – Description schématique d'extraction de caractéristiques de tendance à partir des séries temporelles et leurs valeurs du RUL associées. Dans cet exemple, $L = 20$, $l(T_1) = 192$, et $l(T_{|\mathcal{T}|}) = 158$.

On est donc à la recherche d'une fonction f qui modélise la relation entre les variables prédictives et cibles : $y \simeq f(x)$. Les SVR ont montré leur efficacité dans la modélisation des relations non linéaires entre une variable cible et un ensemble de variables prédictives.

Nous considérons dans la suite le $\epsilon - SVR$ [Vapnik, 1995]. L'objectif de $\epsilon - SVR$ est de trouver la fonction $f(x)$ qui a au maximum une déviation $\ll \epsilon \gg$ de la valeur cibles $\ll y \gg$ pour tous les échantillons d'apprentissage, et qui en même temps est le plat possible.

Nous rappelons dans ce qui suit les principes basique de $\epsilon - SVR$. Nous supposons d'abord que nous cherchons une relation linéaire entre x et y :$y \simeq \langle w, x\rangle$+b, où $\langle w, x\rangle$ représente le produit scalaire et $w \in \mathbb{R}^{2d}$. La planéité ici signifie qu'on est à la recherche de petites valeurs $\ll w \gg$ en termes de norme euclidienne.

Le problème d'optimisation de $\epsilon - SVR$ est le suivant :

$$\text{Minimiser } \|w\|^2, \text{ sujet à } \begin{cases} y_i - \langle w, x_i\rangle - b & \leq \epsilon \\ \langle w, x_i\rangle + b - y_i & \leq \epsilon \end{cases} \tag{5.18}$$

L'équation suppose que cette optimisation est possible, c'est-à-dire qu'il est possible de trouver une telle fonction linéaire. Cependant, ce n'est pas toujours le cas : on ne peut pas toujours être en mesure de trouver une fonction qui satisfait la marge ϵ. Pour faire face à ce problème, des variables d'écart ζ_i, ζ_i^* sont introduites dans le problème d'optimisation qui devient :

$$\text{Minimiser } \|w\|^2 + C\sum_{i=1}^{N}(\zeta_i + \zeta_i^*), \text{ sujet à}$$

$$\begin{cases} y_i - \langle w, x_i\rangle - b & \leq \epsilon + \zeta_i \\ \langle w, x_i\rangle + b - y_i & \leq \epsilon + \zeta_i^* \\ \zeta_i, \zeta_i^* \geq 0 \end{cases} \tag{5.19}$$

Le paramètre du coût $C > 0$ contrôle la tolérance des erreurs supérieures à ϵ. Pour modéliser les relations non linéaires entre x et y, on utilise un noyau comme dans le cas des machines à vecteurs de support : les données d'entrées (input data) sont cartographiés dans un espace de caractéristiques de dimension supérieur. Un $\epsilon - SVR$ linéaire est appliqué sur ce nouvel espace de grande dimension. Plus de détails sur les SVR peuvent être trouvés dans [Smola et al., 2004]. Les SVR retournent en sortie un modèle $\ll m \gg$ que nous allons utiliser plus tard pour prédire le RUL de nouveaux composants.

5.5.4/ PRÉDIRE LE RUL DE NOUVEAUX COMPOSANTS À L'AIDE DES SVR

Une fois que le modèle SVR est appris offline, on l'utilise pour prédire le RUL de nouveaux composants. La surveillance d'un nouveau composant est modélisée par une série temporelle $U = u_1, ..., u_{l(U)}$. Aucune information n'a été donnée sur la dernière instance de surveillance $l(U)$: Elle peut être à un stade précoce de la vie du composant, un stade tardif ou n'importe quel stade entre les deux. On désire estimer la différence de temps entre $l(U)$ et la défaillance du composant du test.

La série temporelle U est d'abord divisée en fenêtres de taille L. il est à noter qu'ici nous permettons aux fenêtres de se chevaucher : une nouvelle fenêtre est obtenue en décalant la précédente d'une unité de temps. Il existe par conséquent $n_u = l(U) - L + 1$ de telles fenêtres dans U. De chacune de ces fenêtres, nous extrayons des caractéristiques de tendance (section 5.5.2), pour obtenir un vecteur de caractéristiques de taille $2d$. Lorsque le vecteur de caractéristiques x_k de la fenêtre k de U est donné au modèle SVR $\ll m \gg$, ce modèle prédit le temps (à partir du dernier instant de la fenêtre) auquel une défaillance est censée se produire.

Par conséquent, si m prédit une valeur y_k pour une fenêtre k, cela signifie que le modèle prédit qu'une panne se produira à l'instant $\hat{f}_k = L + k - 1 + y_k$. Par conséquent, nous obtenons n_U prédictions : $\hat{f}_1, \ldots, \hat{f}_{n_U}$ pour le moment auquel la défaillance va se produire. Une moyenne de ces valeurs est considérée comme la prédiction finale du RUL du composant $\ll U \gg$. Cette étape est illustrée à la figure 5.9.

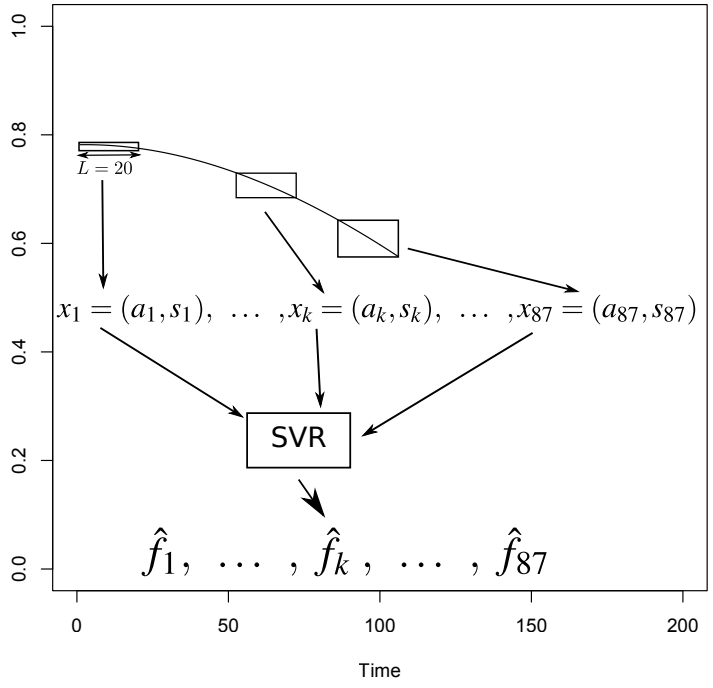


FIGURE 5.9 – Estimation du RUL d'un nouvel composant U en utilisant le modèle SVR construit au préalable : Les caractéristiques de tendance sont extraites de fenêtres coulissantes et le modèle SVR qui estime le RUL de chaque fenêtre. Dans cet exemple, $L = 20$ et $l(u) = 106$.

Figure 5.10 présente un exemple d'estimation du RUL faite en suivant cette méthode décrite. Le RUL est exprimé en fonction de valeurs d'indicateur de santé et t_c désigne le début de la prédiction.

5.6/ SÉLECTION DE VARIABLES (DONNÉES CAPTEURS)

Il est nécessaire de distinguer, avant tout apprentissage, l'information pertinente de l'information inutile. Cette distinction peut se faire à l'aide d'une méthode appelée « processus de sélection de variables » lors de la phase de traitement de données.

La sélection de variables est un processus permettant de « sélectionner » et de maintenir à partir de l'ensemble de variables, seulement celles qui sont les plus pertinentes. Différentes stratégies de sélection de variables ont été proposées et utilisées dans le contexte de PHM [Liu et al., 1998, Liu et al., 2005]. Ces stratégies sont caractérisées par quatre points majeurs :

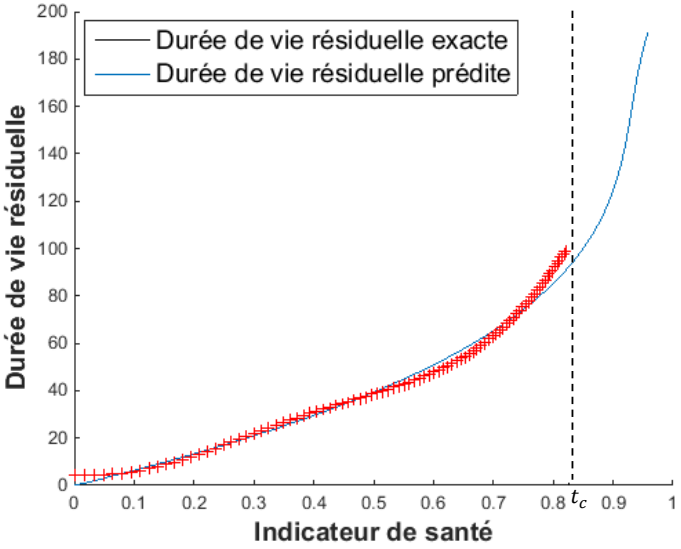


FIGURE 5.10 – Exemple d'estimation du RUL en utilisant les SVR.

— **L'ensemble initial de variables sélectionnées** : le point de départ dans l'espace de variables pourrait commencer soit par un ensemble vide où la direction de recherche est « forward ». Les variables sont ajoutées une par une à chaque itération, ou par un ensemble qui contient au départ toutes les variables. Avec une élimination « backward», à chaque itération, une variable est enlevée de l'ensemble tel que cette suppression donne un meilleur sous-ensemble. Sinon, l'espace pourrait commencer par un sous-ensemble aléatoire de variables avec une recherche bidirectionnelle. Par conséquent, les variables peuvent être ajoutées ou retirées successivement suivant une certaine procédure.
— **La stratégie de recherche** : la stratégie de recherche des sous-ensembles de variables peut être faite avec des heuristiques aléatoires ou par une procédure complète [Chebel-Morello et al., 2015].
— **Le critère d'évaluation** : est un élément important dans la sélection de variables comme c'est une mesure de la qualité d'un sous-ensemble. D'après [Liu et al., 2005], on distingue deux groupes de critères selon leur dépendance d'algorithmes sur lesquels les variables sélectionnées sont appliquées. Ces critères peuvent être indépendants ou dépendants. Les critères indépendants évaluent la qualité de la sélection en exploitant les caractéristiques intrinsèques des données d'entraînement sans impliquer les algorithmes. D'autres part, les critères dépendants nécessitent que l'algorithme soit prédéterminé dans la sélection et utilisent la performance de l'algorithme appliqué sur le sous-ensemble choisi afin de déterminer quelles variables sont à être gardées. Ce critère donne généralement de meilleurs résultats car il sélectionne les variables qui sont les mieux adaptées à l'algorithme. Par conséquent, dans notre travail, nous utilisons un critère dépendant afin d'évaluer les sous-ensembles.
— **Le critère d'arrêt** : détermine quand la procédure devrait être arrêtée. Parmi les critères les plus utilisés, on trouve la qualité des sous-ensembles sélectionnés c'est-à-dire la procédure peut être arrêtée si la performance est assez bonne. Un autre critère est le nombre d'itérations. Le processus est arrêté soit car le nombre maximal est atteint, soit si la recherche est compète.

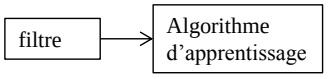
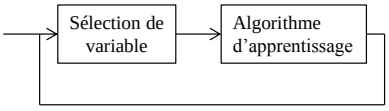
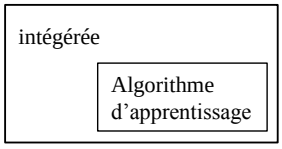
Il existe trois types de méthodes de sélection de variables : les méthodes «filtre», les méthodes «enveloppe» et les méthodes «intégrées». Les filtres fonctionnent indépendamment du tout algorithme d'apprentissage contrairement aux méthodes enveloppes qui utilisent la performance de l'algorithme d'apprentissage. Les méthodes intégrées sont construites sur un algorithme d'apprentissage (ex : algorithmes d'exploration, algorithmes de classification . . .) pour effectuer la sélection de variables.

— **Les méthodes "filtre"** : la sélection de variables avec les méthodes « filtre» est un processus séparé qui se produit avant l'étape d'induction ou d'apprentissage. Elles filtrent les variables pertinentes avant la phase d'apprentissage. La phase de pré-traitement utilise les caractéristiques générales de l'ensemble d'apprentissage afin de sélectionner certaines variables et en exclure d'autres [Blum et al., 1997]. Le schéma le plus simple est d'évaluer individuellement chaque variable en considérant une fonction d'évaluation et de sélectionner les variables possédant les plus grandes valeurs. La fonction d'évaluation est donc considérée comme le critère de sélection. Il existe deux critères de sélection : (i) les critères myopes qui sont des mesures de la qualité d'une variable indépendamment du contexte et des autres variables. Ces critères ne détectent pas les corrélations entre les variables. Par conséquent, ils ne sont pas adaptés aux algorithmes traitant des données corrélées. (ii) Les critères contextuels qui sont des critères de consistance. Ils estiment la qualité d'une variable dans le contexte d'autres variables et ils permettent de traiter les situations où les variables sont corrélées.
  Parmi les méthodes "filtre" connues, on liste : La sélection des fonctionnalités en fonction de corrélation (Correlation-based Feature Selection) qui est un algorithme filtre multivarié simple qui classe les sous-ensembles de caractéristiques selon une fonction d'évaluation heuristique basée sur la corrélation [Hall, 1999]. La fonction cherche à classer les sous-ensembles qui contiennent des caractéristiques qui sont fortement corrélées avec la classe et non corrélées entre elles. Et le filtre basé sur la cohérence (consistency-based feature selection) qui évalue la valeur d'un sous-ensemble de caractéristique par son niveau de cohérence [Dash et al., 2003].
— **Les méthodes "enveloppe"** : ces approches sont introduites par John, Kohavi et Plfeger [John et al., 1994]. La motivation principale derrière telles approches réside dans le fait que les méthodes filtre ignorent l'influence de l'ensemble de variables sélectionnées sur la performance de l'algorithme d'apprentissage utilisé. Les méthodes «enveloppe» typiques cherchent le même sous-ensemble de variables comme les méthodes filtre, mais elles évaluent les sous-ensembles en exécutant l'algorithme sur les données sélectionnées et la performance de l'approche est utilisée comme critère (critère dépendant). Le point fort de ces méthodes est que le choix de variable est fait en se basant sur la performance de l'algorithme d'apprentissage, ce qui donne de meilleurs résultats par rapport aux méthodes de sélection qui choisissent le sous-ensemble indépendamment de la performance de l'algorithme d'apprentissage. Cependant le coût calculatoire est élevé ce qui est le résultat de l'utilisation d'algorithme d'apprentissage pour chaque sous-ensemble considéré puisque ce dernier est une sous-routine du processus de sélection [Kohavi et al., 1997, El Akadi et al., 2011].
  "WrapperSubsetEval" [Witten et al., 2005] est une des méthodes enveloppe qui évalue l'ensemble des attributs en utilisant un schéma d'apprentissage. La validation croisée est utilisée pour estimer la précision du système d'apprentissage pour un ensemble d'attributs et l'algorithme commence avec l'ensemble vide d'attributs

et applique une stratégie de recherche ”forward” en ajoutant des attributs jusqu'à atteindre une performance stable qui ne s'améliore plus.

— **Les méthodes ”intégrées”** : le processus de sélection de variables de ces méthodes est intégré à la phase d'entraînement de l'approche d'apprentissage. Elles évitent d'entraîner l'algorithme d'apprentissage à partir de zéro à chaque fois un sous ensemble est étudié ce qui les rend moins exigeant en terme de calcul par rapport aux méthodes enveloppe.
La méthode LASSO [Tibshirani, 1996] est un algorithme intégré. Il construit un modèle et pénalise les coefficients en rétrécissant beaucoup d'entre eux à zéro. le principe de LASSO est lié à une technique de régularisation qui vise à pénaliser les modèles complexes sur lesquels l'approche ”Elastic Net” est également construite [Zou et al., 2005].
SVM-REF [Guyon et al., 2002] est aussi l'une des plus célèbres approches intégrées. Les poids sont affectés à des fonctions pendant la construction du modèle. Les variables avec des petits poids sont récursivement éliminées.

TABLE 5.1 – Avantage et inconvénients des approches de sélection de variables.

Méthodes	avantages	inconvénients	Références
	Faible coût calculatoire.	Ignorent l'influence de l'ensemble sélectionné sur l'algorithme d'apprentissage.	[Bo et al., 2002, Dash et al., 2003, Seth et al., 2010]
	Capturent les dépendances entre les variables sélectionnées et la performance de l'algorithme.	Un coût calculatoire élevé.	[Maroño et al., 2011, Li et al., 2014, Erguzel et al., 2015]
	Capturent les dépendances entre les variables sélectionnées et la performance de l'algorithme.	Elles sont spécifique à un algorithme d'apprentissage automatique.	[Rakotomamonjy, 2003, M.Lavalle et al., 2006, Peralta et al., 2014]

Le tableau 5.1 résume les avantages et inconvénients de chaque type d'approche de sélection de variables avec des références à des travaux qui ont été réalisés dans ce cadre. Les méthodes « enveloppe » considèrent la performance de l'algorithme d'apprentissage appliqués sur les données sélectionnées ce qui les rend les plus puissantes mais ce qui augmente aussi leur coût en terme de calcul. Les méthodes intégrées sont moins coûteuses mais elles sont spécifiques aux algorithmes utilisés. La taille de l'ensemble de variables des applications considérées ici est petite ce qui fait que le coût est tolérable c'est pourquoi, dans ce chapitre on applique une méthode de sélection de variables « enveloppe » illustrée à la figure 5.11. L'ensemble de départ contient tous les capteurs. À chaque itération, un sous-ensemble est généré avec une élimination « bidirectionnelle » et la performance de l'algorithme d'estimation du RUL est évaluée en utilisant le sous-ensemble étudié. Cette procédure est répétée jusqu'à atteindre une performance satisfaisante.

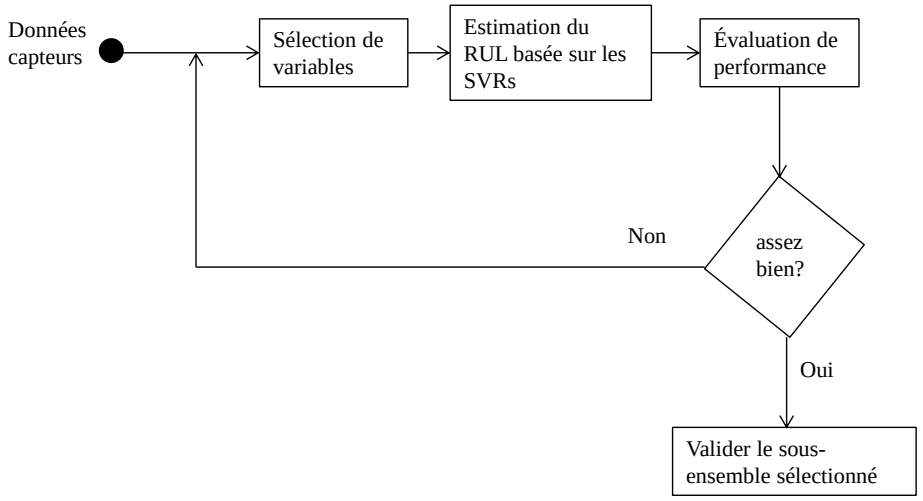


FIGURE 5.11 – Méthode de sélection de variables.

5.7/ APPLICATION ET RÉSULTATS

L'approche présentée ici a été appliquée sur les deux jeux de données : turboréacteurs et batteries. L'approche est développée sur des signaux mono-dimensionnels (les indicateurs de santé), figure 5.12 et avec des signaux multidimensionnels (directement sur les données capteurs), figure 5.13.

Les mesures de capteurs disponibles dans les applications sont soit constantes, où croissantes où décroissante d'une façon monotone. Selon l'application, nous avons d'abord choisi manuellement les capteurs ayant des tendances monotones. Puis, nous avons appliqué une méthode de sélection de variables à traiter, ce qui a permis d'améliorer les résultats obtenus.

5.7.1/ ENSEMBLE DE DONNÉES DES TURBORÉACTEURS

- **Sélection de données**

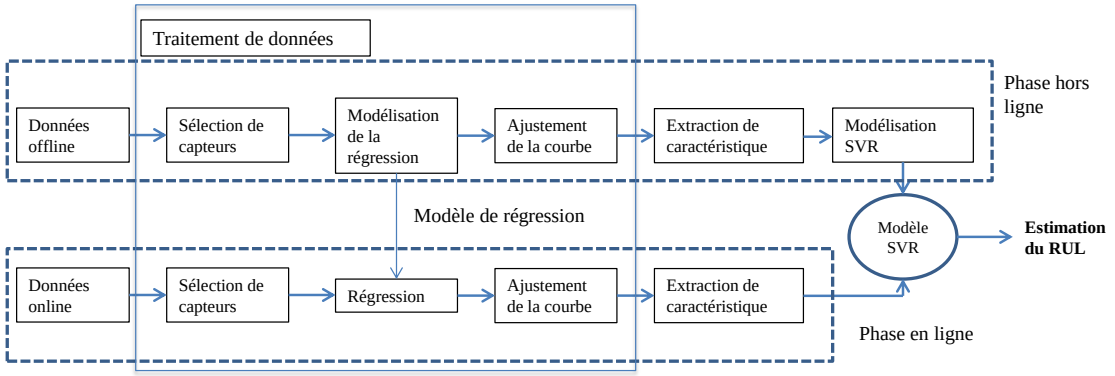


FIGURE 5.12 – L'approche développée sur des signaux mono-dimensionnels.

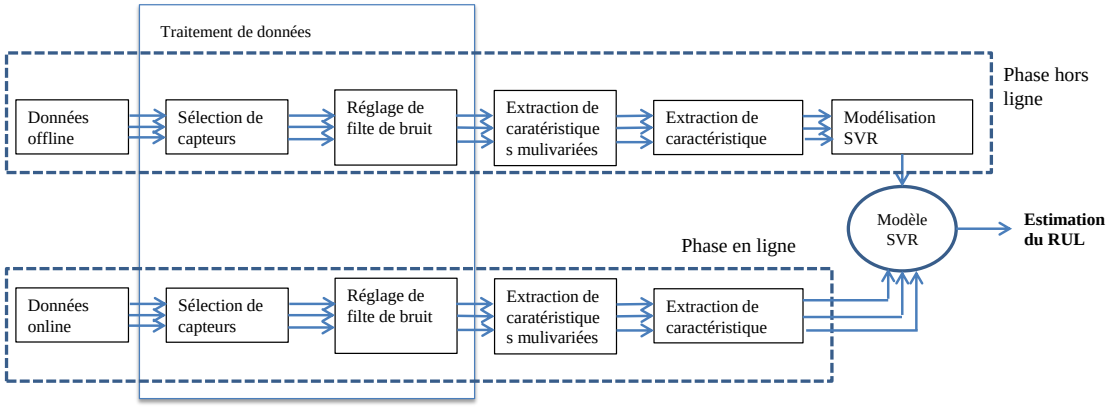


FIGURE 5.13 – L'approche développée sur des signaux multidimensionnels.

L'approche a été appliquée sur les indicateurs de santé construits au le chapitre 3, sur les données capteurs avec tendances croissantes. Cela se traduit par l'utilisation de seulement 5 capteurs parmi les 21 disponibles (c.f. Section 3.7.1.1).

Ces données bruitées sont lissées à l'aide d'une régression linéaire par les moindres carrées pondérées et un modèle polynomial. Chaque capteur a été traité indépendamment des autres. la figure 5.14 montre un exemple d'un signal sans lissage et avec différents paramètres de lissage. Le tableau 5.2, donne les paramètres utilisés pour chaque capteur. Le paramètre de la "durée" est le pourcentage de points de données total.

TABLE 5.2 – Le paramètre du durée utilisé pour chaque capteur.

Capteur	Paramètre du « durée »
Capteur 1	0.5
Capteur 2	0.8
Capteur 3	0.6
Capteur 4	0.5
Capteur 5	0.999

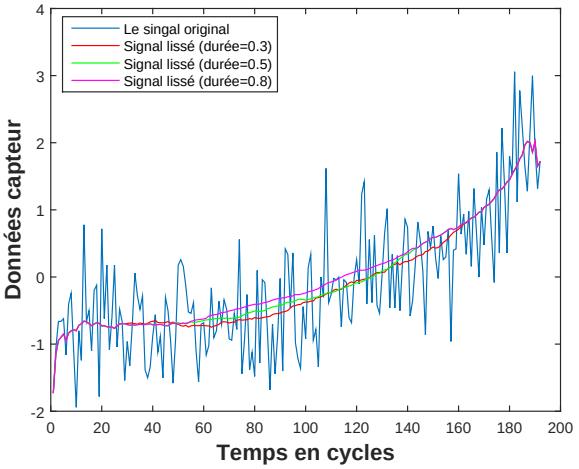


FIGURE 5.14 – Exemple d'un capteur lissé.

• **Résultats et discussion**

Le tableau 5.3 résume les résultats obtenus en choisissant une fonction noyau gaussienne et en faisant varier les paramètres ϵ et c (coût) des SVR pour différentes types d'entrées (c'est-à-dire, les indicateurs de santé, les données capteurs et les données capteurs sélectionnées par une méthode enveloppe). Les résultats dépendent de ces paramètres. Au départ, on ne savait pas lesquels choisir. Par conséquent, on a procédé par coups d'essais et erreurs. Mais on aurait pu aussi les sélectionner par une validation croisée sur l'ensemble d'apprentissage.

L'application de l'approche sur les indicateurs de santé avait donné de moins bon résultats. Cela est du au phénomène d'accumulation des erreurs. Les SVR sont entraînés avec des données approximées en utilisant la régression linéaire. L'application de l'approche sur les données capteurs donne de meilleurs résultats mais ces derniers sont encore améliorés avec la sélection de variables. Le processus a été répété jusqu'à l'obtention de résultats satisfaisants. L'ensemble qui donne les meilleurs résultats c'est capteur 1, capteur 3, capteur 5 avec un taux de 70% de correctes prédictions (selon le critère défini à la section 3.6).

La Figure 5.15 illustre les résultats obtenus avec l'ensemble de capteurs sélectionné en utilisant la méthode « enveloppe », l'erreur de l'estimation appartient à l'intervalle [-31, 58] mais 70% des erreurs sont entre -10 et 13. Les plus grandes erreurs sont dues au fait que l'horizon de prédiction diffère d'un cas à un autre. Pour certaines unités de test (turboréacteurs de test) on est amené à faire la prédiction à 20% de la durée de vie totale ce qui entraîne de grande différences entre les valeurs exactes et prédites. La figure 5.16 détaille le RUL exact et estimé pour chaque moteur de test. On constate que ces valeurs sont assez proches.

• **Comparaison**

[Ramasso et al.,], ont présenté une revue qui donne quelques statistiques sur ce jeu de données. Selon la revue, les données ont été utilisées dans plus de 22 publications. Par contre seulement cinq ont employé pleinement l'ensemble apprentissage/test. Le reste des travaux ont seulement utilisé les données d'apprentissage. Cela permet une évaluation plus riche (ex : Calculer l'horizon du pronostic) mais dans les situations réelles,

TABLE 5.3 – Comparaison entre les trois types d'entrée des SVR.

Données d'entrées aux SVR	c	ϵ	préd. Exactes	préd. Précoces	préd. Tardive
Indicateurs	1	0.04	51%	14%	35%
	1	0.03	51%	14%	35%
de	1	0.02	51%	14%	35%
	1	0.01	51%	14%	35%
santé	1.5	0.01	55%	13%	32%
	0.9	0.01	50%	15%	35%
Données d'entrées aux SVR	c	ϵ	préd. Exactes	préd. Précoces	préd. Tardive
Toutes	1	0.04	60%	20%	20%
les	1	0.03	60%	20%	20%
données	1	0.02	60%	21%	19%
capteurs	1	0.01	61%	20%	19%
	1.5	0.01	61%	20%	19%
	0.9	0.01	61%	20%	19%
Données d'entrées aux SVR	c	ϵ	préd. Exactes	préd. Précoces	préd. Tardive
	1	0.04	49%	25%	26%
	1	0.03	48%	25%	27%
	1	0.02	48%	25%	27%
capteurs$\{1,2,3,4\}$	1	0.01	51%	25%	24%
	1.5	0.01	48%	25%	27%
	0.9	0.01	48%	25%	27%
Données d'entrées aux SVR	c	ϵ	préd. Exactes	préd. Précoces	préd. Tardive
	1	0.04	57%	19%	24%
	1	0.03	58%	18%	24%
	1	0.02	59%	17%	24%
capteurs$\{1,2,4,5\}$	1	0.01	57%	19%	24%
	1.5	0.01	53%	23%	24%
	0.9	0.01	59%	17%	24%
Données d'entrées aux SVR	c	ϵ	préd. Exactes	préd. Précoces	préd. Tardive
	1	0.04	44%	23%	24%
	1	0.03	43%	24%	24%
	1	0.02	43%	24%	24%
capteurs$\{1,2,4\}$	1	0.01	43%	24%	24%
	1.5	0.01	47%	23%	24%
	0.9	0.01	44%	23%	24%
Données d'entrées aux SVR	c	ϵ	préd. Exactes	préd. Précoces	préd. Tardive
	1	0.04	**70**%	10%	20%
	1	0.03	**70**%	10%	20%
	1	0.02	**70**%	10%	20%
capteurs$\{1,3,5\}$	1	0.01	69%	11%	20%
	1.5	0.01	69%	11%	20%
	0.9	0.01	68%	19%	11%

l'algorithme, une fois entraîné, doit être appliqué sur des données qui n'ont pas été vu auparavant. L'évaluation de l'approche avec les données de test est une simulation de ce scénario.

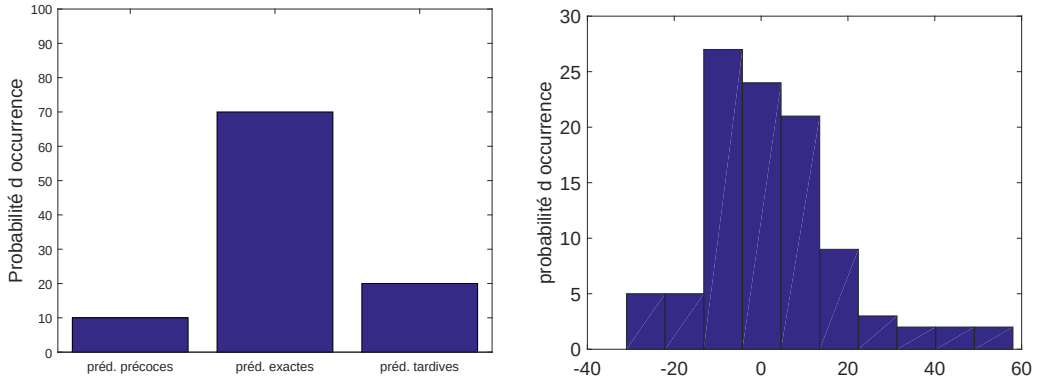


FIGURE 5.15 – Détails de résultats de prédiction obtenus avec la sélection de variables capteur 1, 3 et 5 (a) l'histogramme de l'erreur, (b) taux de prédictions correctes, précoces et tardives.

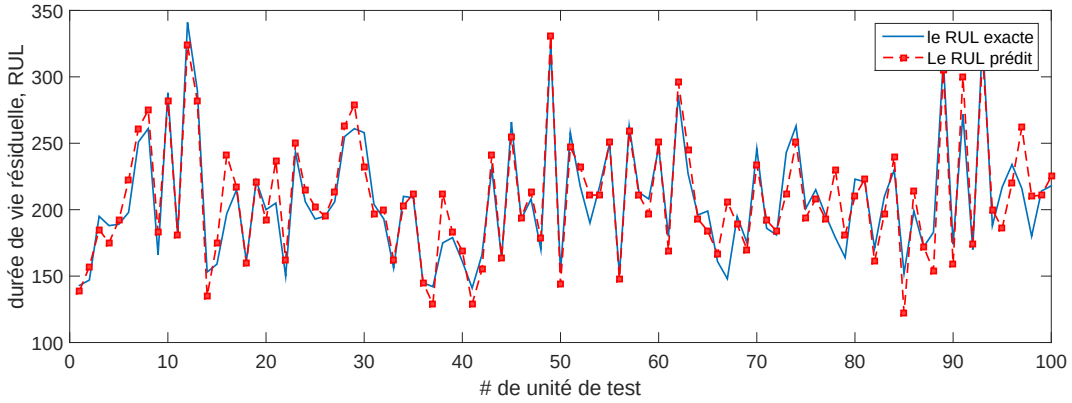


FIGURE 5.16 – Estimation du RUL.

Les résultats obtenus ici sont compétitifs (c.f. Tableau 5.4). L'approche a donné de meilleurs résultats par rapport aux réseaux de neurones [Javed et al., 2015], et les fonctions de croyance combinées avec des méthodes d'appariement de formes [Ramasso et al., 2013], des méthodes qui nécessitent la définition d'un seuil de défaillance, une condition qui n'est pas exigée dans ce travail. L'approche avait de meilleures performances par rapport aux approches basées sur l'expérience développées précédemment dans ce travail. D'autre part, le travail basé sur la similarité proposé par [Ramasso, 2014] présentait une performance proche à nos travaux. Toutefois nous nous sommes affranchis d'une part de la définition de seuil de défaillance et d'autre part de la définition de règles empiriques sur le RUL qui ont été spécifiquement dérivées et utilisées pour améliorer les résultats.

Le paramètre coût de prédiction qui a été introduit dans le chapitre 4 (équation 4.6) sert à quantifier les mauvaises prédictions (les prédictions qui ne sont pas considérées « correctes »). Plus faible est ce coût, meilleure est la performance de l'algorithme. À en juger par ce critère, l'approche basée sur l'expérience formalisée par l'ajout de la connaissance donne le coût le plus bas. Cela est du au fait que les prédictions précoces sont moins pénalisées par rapport aux prédictions tardives. Le pourcentage de ces prédictions obtenu par la méthode basée sur les SVR est de 20% est et plus grand que celui de l'ap-

TABLE 5.4 – Comparaison de résultats obtenus par différentes approches pour les 100 moteurs de test.

Approche	Prédictions correctes %	Prédictions précoces	Prédictions tardives	Coût
Approche de[Ramasso et al., 2013]	53%	36%	11%	17.25
Inspiré par [Wang et al., 2008]	50%	19%	31%	25
Approche de[Javed et al., 2015]	48%	40%	12%	19
Approche de [Ramasso, 2014]	[56%,64%]	N/A	N/A	N/A
expérience formalisée par les données (IS)	58%	24%	18%	19.5
expérience formalisée par les données (TD)	60%	17%	23%	21.5
expérience formalisée par l'ajout de la connaissance	56%	32%	12%	17
Approche proposée ici, basée sur les SVR	70%	10%	20%	17.5

proche basée sur l'expérience formalisée par l'ajout de la connaissance qui est de 12%. Tout de même, le taux de prédictions correctes reste le plus élevé avec l'approche basée sur les SVR et le coût de mauvaise prédictions des SVR ne s'éloigne pas trop de la valeur minimal. À partir de ces deux critères, on peut conclure que la prédiction directe du RUL en utilisant les SVR donne les meilleurs résultats pour ce type de données.

5.7.2/ ENSEMBLE DES DONNÉES BATTERIES

La deuxième application considérée tout au long de ces travaux de thèse est la batterie « Li-ion ». Ce jeu de données était présenté à la section 3.7.2 et le test est conduit de la même manière.

- **Sélection des données**

Les données capteurs sont d'abord sélectionnées par rapport à leur tendance. On cherche des signaux qui présentent un comportement monotone clair en liaison avec la dégradation des composants. Cela se traduit par l'utilisation de deux signaux : l'énergie du courant mesuré et de tension de charge. Tout comme dans la première partie de cette section, on applique l'approche basée sur les SVR en choisissant une fonction noyau gaussienne. Après plusieurs essais, ϵ a été fixé à 0.01 et c était variée entre 2 et 10.

- **Résultats et discussion**

Le tableau 5.5 résume les résultats obtenus en appliquant les indicateurs de santé, les données capteurs et les données capteurs sélectionnées par une méthode enveloppe.

TABLE 5.5 – Comparaison entre les trois types d'entrée des SVR pour les batteries.

Données d'entrées aux SVR	c	EPMA	PRC	ETÉ
Indicateurs	2	31.2957	0.687	1.1467
	2.5	33.2461	0.6675	1.4919
de	3.5	32.0438	0.6796	1.4784
santé	5	32.317	0.6768	1.5005
	10	35.2609	0.6474	1.5196
Données d'entrées aux SVR	c	EPMA	PRC	ETÉ
l'énergie du courant mesuré	2	44.1367	0.5586	1.5956
	2.5	47.6934	0.5231	1.5875
et	3.5	45.3675	0.5463	1.5706
énergie de tension de charge	5	45.5866	0.0.5441	1.562
	10	46.5839	0.0.5342	0.6192
Données d'entrées aux SVR	c	EPMA	PRC	ETÉ
	2	42.6188	0.5738	1.7646
	2.5	4.088	0.586	1.7507
l'énergie du courant mesuré	3.5	41.87	0.584	1.7914
	5	42.26	0.577	1.7792
	10	41.51	0.583	1.7852
Données d'entrées aux SVR	c	EPMA	PRC	ETÉ
	2	31.2972	0.687	1.887
	2.5	31.8929	0.6811	1.8618
énergie de tension de charge	3.5	30.8772	0.6912	1.8841
	5	31.2168	0.6878	1.8798
	10	30.6181	0.6938	1.8364

À partir de tableau 5.5, on constate que les meilleurs résultats sont obtenus avec la variable « énergie de tension de charge ». Sans l'utilisation de cette sélection, l'erreur est multipliée par 1.5.

L'utilisation des indicateurs de santé a donné de bons résultats par rapport aux données capteurs mais la performance de la variable « énergie de tension de charge » reste meilleure.

TABLE 5.6 – Comparaison entre les valeurs d'erreur de pourcentage en moyenne absolue.

Approche	EPMA
Expérience formalisée par les données (indicateurs de santé)	12.8071%
Expérience formalisée par les données (trajectoires de dégradation)	23.0538%
Expérience formalisée par la connaissance	**10.4265%**
Approche présentée ici (variable énergie de tension de charge)	30.6181%
[Mosallam et al., 2014]	32.2086%

• **Comparaison des résultats**

Les résultats mentionnés dans le tableau sont illustrés aux figures 5.17, 5.19 et 5.19 qui montrent l'évolution de l'erreur de pourcentage en moyenne absolue, la précision relative

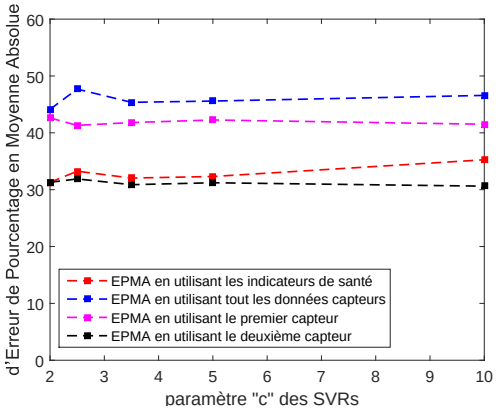


FIGURE 5.17 – L'évolution de l'erreur de pourcentage en moyenne absolue par rapport au paramètre « coût » des SVR.

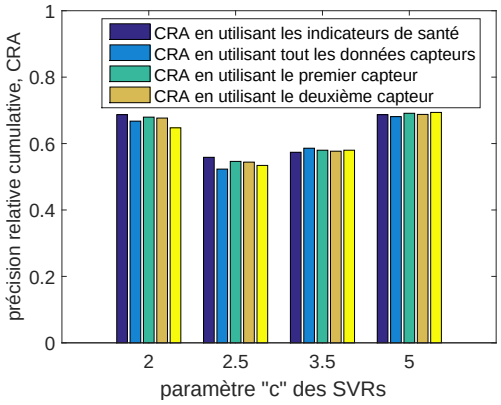


FIGURE 5.18 – La précision relative cumulative par rapport au paramètre « coût » des SVR.

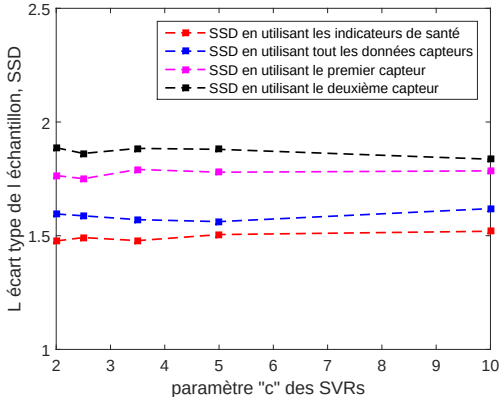


FIGURE 5.19 – L'écart type de l'échantillon par rapport au paramètre « coût » des SVR.

cumulative et l'écart type de l'échantillon par rapport au paramètre « coût » des SVR. La précision de l'algorithme est aux alentours de 0.6 et l'écart type aux alentours de 1.8 pour la variable sélectionnée. Ces résultats sont moins bons par rapport aux résultats trouvés précédemment, (c.f. Tableau 5.6).

Selon le tableau 5.6, l'approche présentée dans le chapitre 4, donne toujours les

meilleurs résultats. Par contre ces résultats sont compétitifs avec ceux trouvés par [Mosallam et al., 2014].

Un algorithme idéal qui peut être appliqué avec la même efficacité sur tous types de composants critiques est difficile à développer. La performance varie d'une application à une autre. L'approche présentée ici a donné de très bons résultats pour les turboréacteurs et de moins bons résultats lorsqu'elle est appliquée sur les batteries mais qui reste tout de même compétitifs par rapport aux résultats trouvés dans la littérature.

5.8/ CONCLUSION

Le travail présenté dans ce chapitre propose une nouvelle approche du problème de pronostic. On s'est intéressé à l'estimation directe du RUL en associant à chaque expérience le temps restant avant la défaillance. Deux types de représentation de l'expérience ont été testés. Une représentation par une série temporelle mono-dimensionnelle et une deuxième par des données capteurs. Ce qui a permis de transformer un problème à la base non supervisé, en un problème supervisé comme dans le cas d'un indicateur de santé, mais en s'affranchissant de la contrainte d'évaluation de l'état du composant, et du problème de définition de seuil.

Après débruitage de ces séries temporelles, des caractéristiques ont été extraites sur fenêtres glissante, et ont été associé au RUL.

La méthode supervisée de pronostic proposé concerne les méthodes de régression à base de vecteurs de support (SVR). Ce sont des méthodes présentant des performances supérieures aux méthodes de réseaux de neurones, en terme de précision et de prédiction.

L'élaboration de la méthode a nécessité la mise en place de paramètres, tels que le type de fonction et les paramètres SVR (Coût, ϵ) qui influencent le résultats de la méthode.

Après le réglage de ces paramètres, notre approche sur les deux types de représentation a été testé sur deux applications (Turboréacteurs et batteries).

De plus, nous avons dans le cas de la représentation de séries multidimensionnelles (les données capteurs) appliqué une méthode de sélection de données, qui a permis d'identifier l'ensemble de données le mieux adapté à notre méthode et a amélioré les résultats des SVR.

La performance de l'approche a été comparée aux méthodes de la littérature et aux méthodes développées aux chapitre 3 et 4. Les résultats obtenus ont montré que les SVR combinés avec la sélection de variable ont la meilleure performance quand appliqués sur le benchmark "turboréacteurs". Le pouvoir discriminant des SVR se montre supérieur à la phase de remémoration définie au chapitre 3. Par contre, l'application de la même méthode sur le jeu de données "batteries", a prouvé l'utilité de la formalisation et étape de remémoration proposées au chapitre 4.

III

CONCLUSION GÉNÉRALE

6

CONCLUSION GÉNÉRALE

La conception de PHM est devenue un élément important dans la réalisation d'une stratégie de maintenance prédictive efficace. On entend par efficace une maintenance qui assure la fiabilité des systèmes tout en réduisant les coûts engendrés par les interventions et les arrêts forcés dus à des pannes soudaines.
Les travaux de recherche présentés dans ce mémoire ont porté sur le développement des approches de pronostic orientées données et plus spécifiquement basées sur l'expérience. Nous nous sommes focalisés sur les composants les plus critiques de systèmes industriels en faisant l'hypothèse que le pronostic d'un système complexe est équivalent à celui réalisé sur les composants critiques

Au chapitre 2, nous avons introduit les notions de PHM, présenté un rapide inventaire sur les différents modules le composant. Un état de l'art rapide a été fait sur les modules évaluation de l'état de santé et diagnostic, suivi d'une étude plus poussée sur le pronostic ce qui nous a permis de positionner nos recherches par rapport aux méthodes existantes que nous avons classées et de définir notre problématique de recherche. On a décrit les algorithmes de pronostic, verrous et hypothèses. On a traité ces verrous par la réalisation d'une approche orientée données et basée sur l'expérience ou une expérience est constituée à partir de l'historique de défaillance d'un composant.

Dans le troisième chapitre, nous avons proposé des formalisations de l'expérience qui ont été utilisées dans le cadre d'une approche de pronostic à partir d'instances. Deux types de formalisation ont été proposées : une supervisée qui génère des signaux directement liés à l'état de santé en développant un modèle de régression linéaire. Cette formalisation a l'avantage de refléter l'état de santé des composants, par contre la phase de modélisation n'exploite que 20% des données capteurs (les données de bornes où l'état de santé est sûrement soit "bon" au début ou "défectueux" à la fin de vie). Nous avons proposé une deuxième formalisation non-supervisée qui agrège l'ensemble des données capteur en trajectoires de dégradation grâce au modèle UKR (Unsupervised Kernel Regression). Le signal obtenu est une représentation fidèle de l'ensemble des données capteur mais n'est pas directement lié à l'état de santé.
Afin de réutiliser l'expérience, nous avons défini deux mesures de similarité adaptées au pronostic qui privilégient les dernières observations. La première mesure est faite sur des fenêtres fixes tandis que la deuxième est faite sur des fenêtres glissantes qui ont permis de déterminer l'état courant d'un nouveau composant ainsi que d'estimer le RUL. L'approche a été testée sur deux types de composants critiques ; "les turboréacteurs" et "les batteries".
Les résultats ont montré que la formalisation non-supervisée appliquée à des turbo réacteurs a donné les meilleurs résultats. Cette formalisation a exploité la totalité des

données capteur mais sans lier les trajectoires obtenues à l'état de santé. Par contre,elle a donné de moins bon résultat sur le jeu de données "batteries". Par conséquent, le type de formalisation dépend de la nature des données. Les données turboréacteurs montrent des tendances plus claires car cette information sous-jacente sur l'état de santé est plus présente même si elle n'est pas directement exploitée lors de la formalisation non-supervisée de l'expérience.

Dans le quatrième chapitre, on a fait évoluer l'approche à base d'instances vers une approche de raisonnement à partir de cas par l'élaboration de containers de connaissance. Cette connaissance a permis de compléter la phase de modélisation de l'indicateur de santé en injectant une connaissance temporelle. La connaissance a été extraite à partir de l'approche UKR et a permis d'exploiter la totalité de données capteurs en combinant l'approche supervisée et non-supervisée. Cette dernière a été utilisée afin d'obtenir des règles temporelles à partir des trajectoires de dégradation et a été utilisée pour construire les indicateurs de santé.
Une deuxième connaissance fréquentielle a été extraite à partir des données grâce à la décomposition en modes empiriques. Cette connaissance a enrichi la représentation des cas en ajoutant un aspect fréquentiel à l'aspect temporel déjà existant des cas.
Les résultats obtenus ont montré l'utilité de l'injection de la connaissance. L'horizon de pronostic a été élargi et les erreurs faites sur les prédictions ont été diminuées. La considération de la connaissance a rendu l'approche plus prudente. Le taux des prédictions précoces est supérieur à celui des prédictions tardives ce qui est préféré en pronostic.

Dans le cinquième chapitre, nous avons représenté l'expérience par des séries temporelles soit mono-dimensionnelles (indicateurs de santé) soit multidimensionnelles. À partir de cette nouvelle représentation, on a extrait des caractéristiques qu'on a directement lié à la durée de vie résiduelle des composants. Le pouvoir discriminant des SVR a été comparé à la phase de remémoration développée précédemment et aux méthodes proposées dans le troisième et quatrième chapitres. De plus ces résultats ont été améliorés par une méthode de sélection de variables enveloppe, appliquées aux données capteurs.
La performance de l'approche décrite dans ce chapitre appliquée sur les données turboréacteurs a donné les meilleurs résultats Les résultats obtenus dépendent du type de données traitées.

Les méthodes développées dans ces travaux de thèse se sont basées sur la représentation de l'expérience à partir d'un historique de dégradation et estiment le RUL sans la nécessité de définir un seuil de défaillance qui est une question problématique en pronostic. Un seuil statique (le cas le plus commun) n'est pas pratique à définir et est posé empiriquement. En réalité le seuil de défaillance dépend du plusieurs facteurs tels que le profil d'utilisation par exemple.
Il est difficile de développer un algorithme de pronostic applicable avec la même efficacité sur tous types de composant. La performance varie d'une application à une autre. Un compromis doit être fait entre la performance et l'applicabilité. Si on juge les méthodes proposées dans ces travaux de thèse, on peut conclure que l'approche orientée expérience formalisée par l'ajout de la connaissance offre le meilleur compromis. Les résultats obtenus par cette approche quelque soit l'application développée sont supérieurs aux résultats des méthodes existantes dans la littérature scientifique.

• **Perspectives**

Les résultats que nous avons obtenus sont prometteurs. Mais il est important de noter que ces travaux sont résolus suivant quelques hypothèses restrictives et peuvent être améliorés en explorant d'autres pistes de recherche.

Le choix des paramètres tels que le facteur relaxant de la mesure de similarité, le nombre de voisins, et les paramètres des SVR est fait empiriquement. Nous envisageons d'améliorer cette procédure en développant des algorithmes d'optimisation telles que les algorithmes génétiques et les algorithmes de fourmis. Nous pourrons ainsi obtenir des résultats plus satisfaisant en trouvant le meilleur ensemble de paramètres.

Une autre piste consisterait à améliorer la sélection de capteurs faite manuellement dans le deuxième chapitre, par une autre méthode de sélection de variables telles que celle utilisée dans le quatrième chapitre, la méthode de sélection enveloppe combinée avec un modèle de SVR qui a donné de très bon résultats. Ceci nous encourage à développer d'autre type de méthode de sélection.

Nous envisageons également de combiner l'estimation du RUL faite via les SVR et l'approche à base d'expériences formalisées d'une manière supervisée. L'idée est de classer les composants par rapport à leur durée de vie et d'obtenir un modèle de dégradation appris pour chaque classe. Ce modèle sera utilisé afin de générer les trajectoires qui seront liées à la durée de vie résiduelle par un modèle de régression à vecteurs de support. Les cas seront représentés par les classes et la phase de remémoration consistera à réutiliser les bons modèles de dégradation et des SVR. La difficulté sera de classer en ligne les nouveaux composants sachant que ces derniers n'ont pas un historique de dégradation complet. Autrement dit, comment attribuer le nouveau composant à une des classes établies.

Une des hypothèses restrictives est que la représentation de l'expérience repose sur sur la disponibilité des tendances monotones de la dégradation. Ces derniers sont soit directement traitées à partir de données capteurs (ex : jeu de données turboréacteur) soit obtenues par des méthodes d'extraction de caractéristiques (batteries). Par contre, rien ne prouve que ces tendances seront toujours disponibles. Cela dépend de la nature des données et de leur degré de réflexion de la dégradation du composant. Une des perspectives est donc d'étudier la nature des données et leur degré de réflexion de la dégradation du composant.

Une autre hypothèse est la disponibilité d'expériences similaires. L'absence de ces dernières rend la méthode incapable de prédire avec exactitude le RUL des nouveaux composants. Par conséquent, nous comptons ajouter une phase d'adaptation qui permettrait d'adapter les expériences existantes au nouveau problème de prédiction de RUL.

Enfin, le travail repose sur la disponibilité d'un historique de défaillance. Cependant, les communautés industrielles ne permettent que très rarement à leurs équipements de fonctionner jusqu'à la défaillance. Nous sommes amenés à simuler cette dégradation d'une manière accélérée ou simulée à l'aide de la mise en place de bancs d'études. Dans notre travail, on a utilisé des benchmarks issus de données simulées et issus de bancs d'essais. Par contre,se pose la question du degré d'adéquation entre le modèle réel et le modèle simulé ou accéléré ? Afin d'éviter cette question, nous envisageons de développer des méthodes robustes basées sur une connaissance partielle de la dégradation.

BIBLIOGRAPHIE

[Aamodt et al., 1994] Aamodt, A., et Plaza, E. (1994). **Case-based reasoning : Foundational issues, methodological variations, and system approaches**. *AI communications*, 7(1) :39–59.

[Aghabozorgi et al., 2015] Aghabozorgi, S., Shirkhorshidi, A. S., et Wah, T. Y. (2015). **Time-series clustering–a decade review**. *Information Systems*, pages 16–38.

[Agrawal et al., 1993] Agrawal, R., Faloutsos, C., et Swami, A. (1993). **Efficient similarity search in sequence databases**. Springer.

[An et al., 2013] An, D., Choi, J.-H., et Kim, N. H. (2013). **Prognostics 101 : A tutorial for particle filter-based prognostics algorithm using matlab**. *Reliability Engineering & System Safety*, 115 :161–169.

[Aref et al., 2004] Aref, W. G., Elfeky, M. G., et Elmagarmid, A. K. (2004). **Incremental, online, and merge mining of partial periodic patterns in time-series databases**. *Knowledge and Data Engineering, IEEE Transactions on*, 16(3) :332–342.

[Bagnall et al., 2005] Bagnall, A., et Janacek, G. (2005). **Clustering time series with clipped data**. *Machine Learning*, 58(2-3) :151–178.

[Bartlett et al., 2009] Bartlett, L. M., Hurdle, E., et Kelly, E. M. (2009). **Integrated system fault diagnostics utilising digraph and fault tree-based approaches**. *Reliability Engineering & System Safety*, 94(6) :1107–1115.

[Baruah et al., 2005] Baruah, P., et Chinnam*, R. B. (2005). **Hmms for diagnostics and prognostics in machining processes**. *International Journal of Production Research*, 43(6) :1275–1293.

[Bechhoefer et al., 2006] Bechhoefer, E., Bernhard, A., He, D., et Banerjee, P. (2006). **Use of hidden semi-markov models in the prognostics of shaft failure**. *HIP*, 1(1) :3.

[Benkedjouh et al., 2013] Benkedjouh, T., Medjaher, K., Zerhouni, N., et Rechak, S. (2013). **Remaining useful life estimation based on nonlinear feature reduction and support vector regression**. *Engineering Applications of Artificial Intelligence*, 26(7) :1751–1760.

[Biagetti et al., 2004] Biagetti, T., et Sciubba, E. (2004). **Automatic diagnostics and prognostics of energy conversion processes via knowledge-based systems**. *Energy*, 29(12) :2553–2572.

[Blum et al., 1997] Blum, A. L., et Langley, P. (1997). **Selection of relevant features and examples in machine learning**. *Artificial intelligence*, 97(1) :245–271.

[Bo et al., 2002] Bo, T., et Jonassen, I. (2002). **New feature subset selection procedures for classification of expression profiles**. *Genome biology*, 3(4) :1–0017.

[Box et al., 2011] Box, G. E., Jenkins, G. M., et Reinsel, G. C. (2011). **Time series analysis : forecasting and control**, volume 734. John Wiley & Sons.

[Burges, 1998] Burges, C. J. (1998). **A tutorial on support vector machines for pattern recognition**. *Data mining and knowledge discovery*, 2(2) :121–167.

[Butler, 1996] Butler, K. L. (1996). **An expert system based framework for an incipient failure detection and predictive maintenance system**. Dans *Intelligent Systems Applications to Power Systems, 1996. Proceedings, ISAP'96., International Conference on*, pages 321–326.

[Byington et al., 2002] Byington, C. S., et Roemer, M. J. (2002). **Prognostic enhancements to diagnostic systems for improved condition-based maintenance [military aircraft]**. Dans *Aerospace Conference Proceedings, 2002. IEEE*, volume 6, pages 6–2815. IEEE.

[Cadini et al., 2009] Cadini, F., Zio, E., et Avram, D. (2009). **Model-based monte carlo state estimation for condition-based component replacement**. *Reliability Engineering & System Safety*, 94(3) :752–758.

[Caesarendra et al., 2011] Caesarendra, W., Widodo, A., et Yang, B.-S. (2011). **Combination of probability approach and support vector machine towards machine health prognostics**. *Probabilistic Engineering Mechanics*, 26(2) :165–173.

[Camci et al., 2006] Camci, F., et Chinnam, R. B. (2006). **Hierarchical hmms for autonomous diagnostics and prognostics**. Dans *Neural Networks, 2006. IJCNN'06. International Joint Conference on*, pages 2445–2452. IEEE.

[Cao, 2003] Cao, L. (2003). **Support vector machines experts for time series forecasting**. *Neurocomputing*, 51 :321–339.

[Cao et al., 2001] Cao, L., et Tay, F. E. (2001). **Financial forecasting using support vector machines**. *Neural Computing & Applications*, 10(2) :184–192.

[Cao, 1989] Cao, X.-R. (1989). **The predictability of discrete event systems**. *Automatic Control, IEEE Transactions on*, 34(11) :1168–1171.

[Chebel-Morello et al., 2015] Chebel-Morello, B., Malinowski, S., et Senoussi, H. (2015). **Feature selection for fault detection systems : application to the tennessee eastman process**. *Applied Intelligence*, pages 1–12.

[Choi et al., 1995] Choi, S. S., et Chang, S. H. (1995). **Development of an on-line fuzzy expert system for integrated alarm processing in nuclear power plants**. *Nuclear Science, IEEE Transactions on*, 42(4) :1406–1418.

[Chu et al., 2002] Chu, S., Keogh, E. J., Hart, D. M., Pazzani, M. J., et others (2002). **Iterative deepening dynamic time warping for time series.** Dans *SDM*, pages 195–212. SIAM.

[Dash et al., 2003] Dash, M., et Liu, H. (2003). **Consistency-based search in feature selection**. *Artificial intelligence*, 151(1) :155–176.

[De Paz et al., 2012] De Paz, J. F., Bajo, J., González, A., Rodríguez, S., et Corchado, J. M. (2012). **Combining case-based reasoning systems and support vector regression to evaluate the atmosphere–ocean interaction**. *Knowledge and information systems*, 30(1) :155–177.

[Dong et al., 2007] Dong, M., et He, D. (2007). **Hidden semi-markov model-based methodology for multi-sensor equipment health diagnosis and prognosis**. *European Journal of Operational Research*, 178(3) :858–878.

[Dong et al., 2006] Dong, M., He, D., Banerjee, P., et Keller, J. (2006). **Equipment health diagnosis and prognosis using hidden semi-markov models**. *The International Journal of Advanced Manufacturing Technology*, 30(7-8) :738–749.

[Dong et al., 2008] Dong, M., et Yang, Z.-b. (2008). **Dynamic bayesian network based prognosis in machining processes**. *Journal of Shanghai Jiaotong University (Science)*, 13 :318–322.

[El Akadi et al., 2011] El Akadi, A., Amine, A., El Ouardighi, A., et Aboutajdine, D. (2011). **A two-stage gene selection scheme utilizing mrmr filter and ga wrapper**. *Knowledge and Information Systems*, 26(3) :487–500.

[Elsayed et al., 2011] Elsayed, A., Hijazi, M. H. A., Coenen, F., García-Fiñana, M., Sluming, V., et Zheng, Y. (2011). **Time series case based reasoning for image categorisation**. Dans *Case-Based Reasoning Research and Development*, pages 423–436.

[Erguzel et al., 2015] Erguzel, T. T., Tas, C., et Cebi, M. (2015). **A wrapper-based approach for feature selection and classification of major depressive disorder–bipolar disorders**. *Computers in biology and medicine*, 64 :127–137.

[Fdez-Riverola et al., 2003] Fdez-Riverola, F., et Corchado, J. M. (2003). **Cbr based system for forecasting red tides**. *Knowledge-Based Systems*, 16(5) :321–328.

[Ferreiro et al., 2012] Ferreiro, S., Arnaiz, A., Sierra, B., et Irigoien, I. (2012). **Application of bayesian networks in prognostics for a new integrated vehicle health management concept**. *Expert Systems with Applications*, 39(7) :6402–6418.

[Fink et al., 2014] Fink, O., Zio, E., et Weidmann, U. (2014). **Predicting component reliability and level of degradation with complex-valued neural networks**. *Reliability Engineering & System Safety*, 121(0) :198 – 206.

[Frezza-Buet, 2012] Frezza-Buet, H. (2012). **Machines à vecteurs supports didacticiel**. page 64.

[Fuchs et al., 2006] Fuchs, B., Lieber, J., Mille, A., et Napoli, A. (2006). **Une première formalisation de la phase d'élaboration du raisonnement à partir de cas**. *Actes du 14ième atelier du raisonnement à partir de cas, Besançon*.

[Galati et al., 2006] Galati, F. A., Forrester, B. D., et Dey, S. (2006). **Application of the generalised likelihood ratio algorithm to the detection of a bearing fault in a helicopter transmission**. Dans *Engineering Asset Management*, pages 400–405. Springer.

[Gebraeel et al., 2005] Gebraeel, N. Z., Lawley, M. A., Li, R., et Ryan, J. K. (2005). **Residual-life distributions from component degradation signals : A bayesian approach**. *IiE Transactions*, 37(6) :543–557.

[Genc et al., 2006] Genc, S., et Lafortune, S. (2006). **Predictability in discrete-event systems under partial observation**. Dans *IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes, Beijing, China*.

[Georgoulas et al., 2015] Georgoulas, G., Karvelis, P., Loutas, T., et Stylios, C. D. (2015). **Rolling element bearings diagnostics using the symbolic aggregate approximation**. *Mechanical Systems and Signal Processing*, 60 :229–242.

[Goebel et al., 2005] Goebel, K., et Bonissone, P. (2005). **Prognostic information fusion for constant load systems**. Dans *Information Fusion, 2005 8th International Conference on*, volume 2, pages 9–pp. IEEE.

[Guyon et al., 2002] Guyon, I., Weston, J., Barnhill, S., et Vapnik, V. (2002). **Gene selection for cancer classification using support vector machines**. *Machine learning*, 46(1-3) :389–422.

[Hall, 1999] Hall, M. A. (1999). **Correlation-based feature selection for machine learning**. PhD thesis, The University of Waikato.

[He et al., 2006] He, D., Wu, S., Banerjee, P., et Bechhoefer, E. (2006). **Probabilistic model based algorithms for prognostics**. Dans *Aerospace Conference, 2006 IEEE*, pages 10–pp.

[Heng et al., 2009] Heng, A., Zhang, S., Tan, A. C., et Mathew, J. (2009). **Rotating machinery prognostics : State of the art, challenges and opportunities**. *Mechanical Systems and Signal Processing*, 23(3) :724–739.

[Hu et al., 2011] Hu, J., Zhang, L., Ma, L., et Liang, W. (2011). **An integrated safety prognosis model for complex system based on dynamic bayesian network and ant colony algorithm**. *Expert Systems with Applications*, 38(3) :1431–1446.

[Huang et al., 1998] Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., et Liu, H. H. (1998). **The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis**. Dans *Proceedings of the Royal Society of London A : Mathematical, Physical and Engineering Sciences*, volume 454, pages 903–995. The Royal Society.

[Hurdle et al., 2009] Hurdle, E., Bartlett, L. M., et Andrews, J. (2009). **Fault diagnostics of dynamic system operation using a fault tree based method**. *Reliability Engineering & System Safety*, 94(9) :1371–1380.

[Ince et al., 2006] Ince, H., et Trafalis, T. B. (2006). **A hybrid model for exchange rate prediction**. *Decision Support Systems*, 42(2) :1054–1062.

[ISO, 2004] ISO (2004). **ISO133811 : Condition monitoring and diagnostics of machines, prognostics part 1 : General guidelines**. International Organization for Standardization.

[Jakkula, 2006] Jakkula, V. (2006). **Tutorial on support vector machine (svm)**. *School of EECS, Washington State University*.

[Jardine et al., 2006] Jardine, A. K., Lin, D., et Banjevic, D. (2006). **A review on machinery diagnostics and prognostics implementing condition-based maintenance**. *Mechanical systems and signal processing*, 20(7) :1483–1510.

[Javed et al., 2013] Javed, K., Gouriveau, R., et Zerhouni, N. (2013). **Novel failure prognostics approach with dynamic thresholds for machine degradation**. Dans *Industrial Electronics Society, IECON 2013-39th Annual Conference of the IEEE*, pages 4404–4409. IEEE.

[Javed et al., 2015] Javed, K., Gouriveau, R., et Zerhouni, N. (2015). **A new multivariate approach for prognostics based on extreme learning machine and fuzzy clustering**.

[Jéron et al., 2007] Jéron, T., Marchand, H., Genc, S., et Lafortune, S. (2007). **Predictability of sequence patterns in discrete event systems**.

[Jiang et al., 2004] Jiang, S., et Kumar, R. (2004). **Failure diagnosis of discrete-event systems with linear-time temporal logic specifications**. *Automatic Control, IEEE Transactions on*, 49(6) :934–945.

[Jin et al., 2012] Jin, R., Cho, K., Hyun, C., et Son, M. (2012). **Mra-based revised cbr model for cost prediction in the early stage of construction projects**. *Expert Systems with Applications*, 39(5) :5214–5222.

[John et al., 1994] John, G. H., Kohavi, R., Pfleger, K., et others (1994). **Irrelevant features and the subset selection problem**. Dans *Machine Learning : Proceedings of the Eleventh International Conference*, pages 121–129.

[Kalpakis et al., 2001] Kalpakis, K., Gada, D., et Puttagunta, V. (2001). **Distance measures for effective clustering of arima time-series**. Dans *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on*, pages 273–280. IEEE.

[Keogh, 2006] Keogh, E. (2006). **A decade of progress in indexing and mining large time series databases**. Dans *Proceedings of the 32nd international conference on Very large data bases*, pages 1268–1268. VLDB Endowment.

[Keogh et al., 2003] Keogh, E., et Kasetty, S. (2003). **On the need for time series data mining benchmarks : a survey and empirical demonstration**. *Data Mining and knowledge discovery*, 7(4) :349–371.

[Khelif et al., 2014] Khelif, R., Malinowski, S., Chebel-Morello, B., et Zerhouni, N. (2014). **Rul prediction based on a new similarity-instance based approach**. Dans *Industrial Electronics (ISIE), 2014 IEEE 23rd International Symposium on*, pages 2463–2468. IEEE.

[Kim et al., 2004] Kim, G.-H., An, S.-H., et Kang, K.-I. (2004). **Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning**. *Building and environment*, 39(10) :1235–1242.

[Kim et al., 2012] Kim, H.-E., Tan, A. C., Mathew, J., et Choi, B.-K. (2012). **Bearing fault prognosis based on health state probability estimation**. *Expert Systems with Applications*, 39(5) :5200–5213.

[Kohavi et al., 1997] Kohavi, R., et John, G. H. (1997). **Wrappers for feature subset selection**. *Artificial intelligence*, 97(1) :273–324.

[Koo et al., 2011] Koo, C., Hong, T., et Hyun, C. (2011). **The development of a construction cost prediction model with improved prediction capacity using the advanced cbr approach**. *Expert Systems with Applications*, 38(7) :8597–8606.

[Kumar et al., 2010] Kumar, R., et Takai, S. (2010). **Decentralized prognosis of failures in discrete event systems**. *Automatic Control, IEEE Transactions on*, 55(1) :48–59.

[Kurbalija, 2009] Kurbalija, V. (2009). **TIME SERIES ANALYSIS AND PREDICTION USING CASE BASED REASONING TECHNOLOGY**. PhD thesis, University of Novi Sad.

[Kurbalija et al., 2008] Kurbalija, V., et Budimac, Z. (2008). **Case-based reasoning framework for generating decision support systems**. *Novi Sad J. Math*, 38(3) :219–226.

[Lall et al., 2010] Lall, P., Lowe, R., et Goebel, K. (2010). **Prognostics using kalman-filter models and metrics for risk assessment in bgas under shock and vibration loads**. Dans *Electronic Components and Technology Conference (ECTC), 2010 Proceedings 60th*, pages 889–901. IEEE.

[Lamontagne et al., 2002] Lamontagne, L., et Lapalme, G. (2002). **Raisonnement à base de cas textuels : Etat de l'art et perspectives**. *Revue d'intelligence artificielle*, 16(3) :339–366.

[Lebold et al., 2001] Lebold, M., et Thurston, M. (2001). **Open standards for condition-based maintenance and prognostic systems**. Dans *Maintenance and Reliability Conference (MARCON)*, pages 6–9. May.

[Lee et al., 2006] Lee, J., Ni, J., Djurdjanovic, D., Qiu, H., et Liao, H. (2006). **Intelligent prognostics tools and e-maintenance**. *Computers in industry*, 57(6) :476–489.

[Lee et al., 2014] Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L., et Siegel, D. (2014). **Prognostics and health management design for rotary machinery systems—reviews, methodology and applications**. *Mechanical Systems and Signal Processing*, 42(1) :314–334.

[Lewis et al., 1994] Lewis, F. L., et Kamen, E. (1994). **Applied optimal control and estimation**. *IEEE Transactions on Automatic Control*, 39(8) :1773–1773.

[Li et al., 2005] Li, C. J., et Lee, H. (2005). **Gear fatigue crack prognosis using embedded model, gear dynamic model and fracture mechanics**. *Mechanical systems and signal processing*, 19(4) :836–846.

[Li et al., 2014] Li, H., Li, C.-J., Wu, X.-J., et Sun, J. (2014). **Statistics-based wrapper for feature selection : An implementation on financial distress identification with support vector machine**. *Applied Soft Computing*, 19 :57–67.

[Li et al., 2000] Li, Y., Kurfess, T., et Liang, S. (2000). **Stochastic prognostics for rolling element bearings**. *Mechanical Systems and Signal Processing*, 14(5) :747–762.

[Li et al., 2006] Li, Y., et Yu, H. (2006). **Three-phase induction motor operation trend prediction using support vector regression for condition-based maintenance**. Dans *Intelligent Control and Automation, 2006. WCICA 2006. The Sixth World Congress on*, volume 2, pages 7878–7881. IEEE.

[Lieber et al., 2004] Lieber, J., D'aquin, M., Brachais, S., et Napoli, A. (2004). **Une étude comparative de quelques travaux sur l'acquisition de connaissances d'adaptation pour le raisonnement à partir de cas**. *12ème Atelier de Raisonnement à Partir de Cas-RàPC'04*, pages 53–60.

[Lin et al., 2007] Lin, J., Keogh, E., Wei, L., et Lonardi, S. (2007). **Experiencing sax : a novel symbolic representation of time series**. *Data Mining and knowledge discovery*, 15(2) :107–144.

[Liu et al., 1998] Liu, H., et Motoda, H. (1998). **Feature extraction, construction and selection : A data mining perspective**. Springer Science & Business Media.

[Liu et al., 2005] Liu, H., et Yu, L. (2005). **Toward integrating feature selection algorithms for classification and clustering**. *Knowledge and Data Engineering, IEEE Transactions on*, 17(4) :491–502.

[Liu et al., 2012] Liu, Q., et Dong, M. (2012). **Online health management for complex nonlinear systems based on hidden semi-markov model using sequential monte carlo methods**. *Mathematical Problems in Engineering*, 2012.

[Liu et al., 2013] Liu, Z., Cao, H., Chen, X., He, Z., et Shen, Z. (2013). **Multi-fault classification based on wavelet svm with pso algorithm to analyze vibration signals from rolling element bearings**. *Neurocomputing*, 99 :399–410.

[Lu et al., 2009] Lu, C.-J., Lee, T.-S., et Chiu, C.-C. (2009). **Financial time series forecasting using independent component analysis and support vector regression**. *Decision Support Systems*, 47(2) :115–125.

[Mahmoud et al., 2013] Mahmoud, M. S., et Khalid, H. M. (2013). **Expectation maximization approach to data-based fault diagnostics**. *Information Sciences*, 235 :80–96.

[Maio et al., 2012] Maio, F. D., Tsui, K. L., et Zio, E. (2012). **Combining relevance vector machines and exponential regression for bearing residual life estimation**. *Mechanical Systems and Signal Processing*, 31(0) :405 – 427.

[Malinowski et al., 2014] Malinowski, S., Chebel-Morello, B., et Zerhouni, N. (2014). **Shapelet-based remaining useful life estimation**. Dans *Automation Science and Engineering (CASE), 2014 IEEE International Conference on*, pages 794–799. IEEE.

[Maroño et al., 2011] Maroño, N., et Alonso-Betanzos, A. (2011). **Combining functional networks and sensitivity analysis as wrapper method for feature selection**. *Expert Systems with Applications*, 38(10) :12930–12938.

[Meinicke et al., 2005] Meinicke, P., Klanke, S., Memisevic, R., et Ritter, H. (2005). **Principal surfaces from unsupervised kernel regression**. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(9) :1379–1391.

[Memisevic et al., 2003] Memisevic, R., Ritter, H., et Meinicke, P. (2003). **Unsupervised kernel regression for nonlinear dimensionality reduction**. *Master's th., U. Bielefeld*.

[Miao et al., 2007] Miao, Q., et Makis, V. (2007). **Condition monitoring and classification of rotating machinery using wavelets and hidden markov models**. *Mechanical systems and signal processing*, 21(2) :840–855.

[Mille, 1999] Mille, A. (1999). **Tutorial cbr : Etat de l'art de raisonnement à partir de cas**. *Plateforme AFIA*, 99.

[Mille et al., 1996] Mille, A., Fuchs, B., et Herbeaux, O. (1996). **A unifying framework for adaptation in case-based reasoning**. Dans *Workshop on Adaptation in Case-Based Reasoning, ECAI-96*, pages 22–28. Citeseer.

[Minsky, 1975] Minsky, M. (1975). **A framework for representing knowledge-in the psychology of computer vision, p**. *NY : McGraw-Hill*.

[M.Lavalle et al., 2006] M.Lavalle, M., Sucar, E., et Arroyo, G. (2006). **Feature selection with a perceptron neural net**. Dans *Proceedings of the international workshop on feature selection for data mining*, pages 131–135.

[Mohandes et al., 2004] Mohandes, M., Halawani, T., Rehman, S., et Hussain, A. A. (2004). **Support vector machines for wind speed prediction**. *Renewable Energy*, 29(6) :939–947.

[Moosavi et al., 2015] Moosavi, S., Djerdir, A., Ait-Amirat, Y., et Khaburi, D. (2015). **Ann based fault diagnosis of permanent magnet synchronous motor under stator winding shorted turn**. *Electric Power Systems Research*, 125 :67–82.

[Mosallam et al., 2013] Mosallam, A., Medjaher, K., et Zerhouni, N. (2013). **Bayesian approach for remaining useful life prediction.** *Chemical Engineering Transactions*, 33 :139–144.

[Mosallam et al., 2014] Mosallam, A., Medjaher, K., et Zerhouni, N. (2014). **Integrated bayesian framework for remaining useful life prediction**. Dans *Prognostics and Health Management (PHM), 2014 IEEE Conference on*, pages 1–6. IEEE.

[Myötyri et al., 2006] Myötyri, E., Pulkkinen, U., et Simola, K. (2006). **Application of stochastic filtering for lifetime prediction**. *Reliability Engineering & System Safety*, 91(2) :200 – 208.

[Oppenheimer et al., 2002] Oppenheimer, C. H., et Loparo, K. A. (2002). **Physically based diagnosis and prognosis of cracked rotor shafts**. Dans *AeroSense 2002*, pages 122–132. International Society for Optics and Photonics.

[Pai et al., 2005] Pai, P.-F., et Lin, C.-S. (2005). **Using support vector machines to forecast the production values of the machinery industry in taiwan**. *The International Journal of Advanced Manufacturing Technology*, 27(1-2) :205–210.

[Pecar, 2002] Pecar, B. (2002). **Case-based algorithm for pattern recognition and extrapolatio**.

[Peel, 2008] Peel, L. (2008). **Data driven prognostics using a kalman filter ensemble of neural network models**. Dans *Prognostics and Health Management, 2008. PHM 2008. International Conference on*, pages 1–6. IEEE.

[Pencolé, 2004] Pencolé, Y. (2004). **Diagnosability analysis of distributed discrete event systems**. Dans *ECAI*, volume 16, page 43.

[Peng et al., 2010] Peng, Y., Dong, M., et Zuo, M. J. (2010). **Current status of machine prognostics in condition-based maintenance : a review**. *The International Journal of Advanced Manufacturing Technology*, 50(1-4) :297–313.

[Peng et al., 2012] Peng, Y., Wang, H., Wang, J., Liu, D., et Peng, X. (2012). **A modified echo state network based remaining useful life estimation approach**. Dans *Prognostics and Health Management (PHM), 2012 IEEE Conference on*, pages 1–7.

[Peralta et al., 2014] Peralta, B., et Soto, A. (2014). **Embedded local feature selection within mixture of experts**. *Information Sciences*, 269 :176–187.

[Ping et al., 2015] Ping, X.-O., Tseng, Y.-J., Lin, Y.-P., Chiu, H.-J., Lai, F., Liang, J.-D., Huang, G.-T., et Yang, P.-M. (2015). **A multiple measurements case-based reasoning method for predicting recurrent status of liver cancer patients**. *Computers in Industry*, 69 :12–21.

[Qin et al., 2015] Qin, T., Zeng, S., et Guo, J. (2015). **Robust prognostics for state of health estimation of lithium-ion batteries based on an improved pso–svr model**. *Microelectronics Reliability*.

[Rabiner et al., 1986] Rabiner, L. R., et Juang, B.-H. (1986). **An introduction to hidden markov models**. *ASSP Magazine, IEEE*, 3(1) :4–16.

[Rakotomamonjy, 2003] Rakotomamonjy, A. (2003). **Variable selection using svm based criteria**. *The Journal of Machine Learning Research*, 3 :1357–1370.

[Ramasso, 2014] Ramasso, E. (2014). **Investigating computational geometry for failure prognostics in presence of imprecise health indicator : Results and comparisons on c-mapss datasets**. Dans *2nd Europen Confernce of the Prognostics and Health Management Society.*, volume 5, pages 1–13.

[Ramasso et al., 2013] Ramasso, E., Rombaut, M., et Zerhouni, N. (2013). **Joint prediction of continuous and discrete states in time-series based on belief functions**. *Cybernetics, IEEE Transactions on*, 43(1) :37–50.

[Ramasso et al.,] Ramasso, E., et Saxena, A. **Review and analysis of algorithmic approaches developed for prognostics on cmapss dataset**. Dans *Annual Conference of the Prognostics and Health Management Society 2014.*

[Ratanamahatana et al., 2005] Ratanamahatana, C., Keogh, E., Bagnall, A. J., et Lonardi, S. (2005). **A novel bit level time series representation with implication of similarity search and clustering**. Dans *Advances in knowledge discovery and data mining*, pages 771–777. Springer.

[Reinartz et al., 2000] Reinartz, T., Iglezakis, I., et Roth-Berghofer, T. (2000). **On quality measures for case base maintenance**. Dans *Advances in Case-Based Reasoning*, pages 247–260. Springer.

[Ricci et al., 2011] Ricci, R., et Pennacchi, P. (2011). **Diagnostics of gear faults based on emd and automatic selection of intrinsic mode functions**. *Mechanical Systems and Signal Processing*, 25(3) :821–838.

[Richter, 2003] Richter, M. M. (2003). **Knowledge containers**. *Readings in Case-Based Reasoning. Morgan Kaufmann Publishers*.

[Rigamonti et al., 2015] Rigamonti, M., Baraldi, P., Zio, E., Astigarraga, D., et Galarza, A. (2015). **Particle filter-based prognostics for an electrolytic capacitor working in variable operating conditions**.

[Rilling et al., 2003] Rilling, G., Flandrin, P., Goncalves, P., et others (2003). **On empirical mode decomposition and its algorithms**. Dans *IEEE-EURASIP workshop on nonlinear signal and image processing*, volume 3, pages 8–11. NSIP-03, Grado (I).

[Rintanen et al., 2007] Rintanen, J., et Grastien, A. (2007). **Diagnosability testing with satisfiability algorithms.** Dans *IJCAI*, pages 532–537.

[Roth-Berghofer, 2003] Roth-Berghofer, T. (2003). **Developing maintainable cbr systems : Applying siam to empolis orenge.** Dans *Wissensmanagement*, pages 305–306.

[Saha et al., 2007] Saha, B., et Goebel, K. (2007). **Battery data set**. *NASA AMES prognostics data repository*.

[Saha et al., 2011] Saha, B., et Goebel, K. (2011). **Model adaptation for prognostics in a particle filtering framework**. *International Journal of Prognostics and Health Management Volume 2 (color)*, page 61.

[Sampath et al., 1995] Sampath, M., Sengupta, R., Lafortune, S., Sinnamohideen, K., et Teneketzis, D. (1995). **Diagnosability of discrete-event systems**. *Automatic Control, IEEE Transactions on*, 40(9) :1555–1575.

[Saxena et al., 2008a] Saxena, A., Celaya, J., Balaban, E., Goebel, K., Saha, B., Saha, S., et Schwabacher, M. (2008a). **Metrics for evaluating performance of prognostic techniques**. Dans *Prognostics and health management, 2008. phm 2008. international conference on*, pages 1–17. IEEE.

[Saxena et al., 2008b] Saxena, A., Goebel, K., Simon, D., et Eklund, N. (2008b). **Damage propagation modeling for aircraft engine run-to-failure simulation**. Dans *Prognostics and Health Management, 2008. PHM 2008. International Conference on*, pages 1–9. IEEE.

[Schank, 1983] Schank, R. C. (1983). **Dynamic memory : A theory of reminding and learning in computers and people**. Cambridge University Press.

[Selak et al., 2014] Selak, L., Butala, P., et Sluga, A. (2014). **Condition monitoring and fault diagnostics for hydropower plants**. *Computers in Industry*, 65(6) :924–936.

[Seth et al., 2010] Seth, S., et Principe, J. C. (2010). **Variable selection : A statistical dependence perspective**. Dans *Machine learning and applications (icmla), 2010 ninth international conference on*, pages 931–936. IEEE.

[Sheppard et al., 2005] Sheppard, J. W., Kaufman, M., et others (2005). **Bayesian diagnosis and prognosis using instrument uncertainty**. Dans *Autotestcon, 2005. IEEE*, pages 417–423.

[Sikorska et al., 2011] Sikorska, J., Hodkiewicz, M., et Ma, L. (2011). **Prognostic modelling options for remaining useful life estimation by industry**. *Mechanical Systems and Signal Processing*, 25(5) :1803–1836.

[Sinha, 2002] Sinha, B. (2002). **Trend prediction from steam turbine responses of vibration and eccentricity**. *Proceedings of the Institution of Mechanical Engineers, Part A : Journal of Power and Energy*, 216(1) :97–103.

[Smola et al., 2004] Smola, A., et Schölkopf, B. (2004). **A tutorial on support vector regression**. *Statistics and Computing*, 14(3) :199–222.

[Smyth et al., 1997] Smyth, P., et others (1997). **Clustering sequences with hidden markov models**. *Advances in neural information processing systems*, pages 648–654.

[Soualhi et al., 2015] Soualhi, A., Medjaher, K., et Zerhouni, N. (2015). **Bearing health monitoring based on hilbert–huang transform, support vector machine, and regression**. *Instrumentation and Measurement, IEEE Transactions on*, 64(1) :52–62.

[Subrahmanya et al., 2013] Subrahmanya, N., et Shin, Y. C. (2013). **A data-based framework for fault detection and diagnostics of non-linear systems with partial state measurement**. *Engineering Applications of Artificial Intelligence*, 26(1) :446–455.

[Tamilselvan et al., 2013] Tamilselvan, P., et Wang, P. (2013). **Failure diagnosis using deep belief learning based health state classification**. *Reliability Engineering & System Safety*, 115 :124–135.

[Tay et al., 2001] Tay, F. E., et Cao, L. (2001). **Application of support vector machines in financial time series forecasting**. *Omega*, 29(4) :309–317.

[Thissen et al., 2003] Thissen, U., Van Brakel, R., De Weijer, A., Melssen, W., et Buydens, L. (2003). **Using support vector machines for time series prediction**. *Chemometrics and intelligent laboratory systems*, 69(1) :35–49.

[Tibshirani, 1996] Tibshirani, R. (1996). **Regression shrinkage and selection via the lasso**. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288.

[Tsay, 2000] Tsay, R. S. (2000). **Time series and forecasting : Brief history and future research**. *Journal of the American Statistical Association*, 95(450) :638–643.

[Tse et al., 1999] Tse, P., et Atherton, D. (1999). **Prediction of machine deterioration using vibration based fault trends and recurrent neural networks**. *Journal of vibration and acoustics*, 121(3) :355–362.

[Uckun et al., 2008] Uckun, S., Goebel, K., et Lucas, P. J. (2008). **Standardizing research methods for prognostics**. Dans *Prognostics and Health Management, 2008. PHM 2008. International Conference on*, pages 1–10. IEEE.

[Vachtsevanos et al., 2006] Vachtsevanos, G., Lewis, F., Roemer, M., Hess, A., et Wu, B. (2006). **Intelligent fault diagnosis and prognosis for engineering systems, 2006**. *Usa 454p Isbn*, 13 :978–0.

[Vapnik et al., 1996] Vapnik, V., Golowich, S. E., et Smola, A. (1996). **Support vector method for function approximation, regression estimation, and signal processing**. Dans *Advances in Neural Information Processing Systems 9*. Citeseer.

[Vapnik, 1995] Vapnik, V. N. (1995). **The Nature of Statistical Learning Theory**. Springer-Verlag New York, Inc., New York, NY, USA.

[Vapnik et al., 1974] Vapnik, V. N., et Chervonenkis, A. J. (1974). **Theory of pattern recognition**.

[Venkatasubramanian et al., 2003a] Venkatasubramanian, V., Rengaswamy, R., et Kavuri, S. N. (2003a). **A review of process fault detection and diagnosis : Part ii : Qualitative models and search strategies**. *Computers & Chemical Engineering*, 27(3) :313–326.

[Venkatasubramanian et al., 2003b] Venkatasubramanian, V., Rengaswamy, R., Kavuri, S. N., et Yin, K. (2003b). **A review of process fault detection and diagnosis : Part iii : Process history based methods**. *Computers & chemical engineering*, 27(3) :327–346.

[Venkatasubramanian et al., 2003c] Venkatasubramanian, V., Rengaswamy, R., Yin, K., et Kavuri, S. N. (2003c). **A review of process fault detection and diagnosis : Part i : Quantitative model-based methods**. *Computers & chemical engineering*, 27(3) :293–311.

[Wang et al., 2014] Wang, G., Yang, Y., Zhang, Y., et Xie, Q. (2014). **Vibration sensor based tool condition monitoring using ν support vector machine and locality preserving projection**. *Sensors and Actuators A : Physical*, 209 :24–32.

[Wang et al., 2001] Wang, P., et Vachtsevanos, G. (2001). **Fault prognostics using dynamic wavelet neural networks**. *AI EDAM*, 15(04) :349–365.

[Wang, 2010] Wang, T. (2010). **Trajectory similarity based prediction for remaining useful life estimation**. PhD thesis, University of Cincinnati.

[Wang et al., 2008] Wang, T., Yu, J., Siegel, D., et Lee, J. (2008). **A similarity-based prognostics approach for remaining useful life estimation of engineered systems**. Dans *Prognostics and Health Management, 2008. PHM 2008. International Conference on*, pages 1–6. IEEE.

[Wang, 2007] Wang, W. (2007). **An adaptive predictor for dynamic system forecasting**. *Mechanical Systems and Signal Processing*, 21(2) :809–823.

[Wang et al., 2006] Wang, X., Smith, K., et Hyndman, R. (2006). **Characteristic-based clustering for time series data**. *Data mining and knowledge Discovery*, 13(3) :335–364.

[Wei et al., 2009] Wei, X., et Yingqing, G. (2009). **Aircraft engine sensor fault diagnostics based on estimation of engine's health degradation**. *Chinese Journal of Aeronautics*, 22(1) :18–21.

[Widodo et al., 2011] Widodo, A., et Yang, B.-S. (2011). **Machine health prognostics using survival probability and support vector machine**. *Expert Systems with Applications*, 38(7) :8430–8437.

[Witten et al., 2005] Witten, I. H., et Frank, E. (2005). **Data Mining : Practical machine learning tools and techniques**.

[Wu et al., 2007] Wu, W., Hu, J., et Zhang, J. (2007). **Prognostics of machine health condition using an improved arima-based prediction method**. Dans *Industrial Electronics and Applications, 2007. ICIEA 2007. 2nd IEEE Conference on*, pages 1062–1067.

[Wu et al., 2009] Wu, X., Li, Y., Lundell, T. D., et Guru, A. K. (2009). **Integrated prognosis of ac servo motor driven linear actuator using hidden semi-markov models**. Dans *Electric Machines and Drives Conference, 2009. IEMDC'09. IEEE International*, pages 1408–1413.

[Xing et al., 2012] Xing, G., Ding, J., Chai, T., Afshar, P., et Wang, H. (2012). **Hybrid intelligent parameter estimation based on grey case-based reasoning for laminar cooling process**. *Engineering Applications of Artificial Intelligence*, 25(2) :418–429.

[Xiong et al., 2002] Xiong, Y., et Yeung, D.-Y. (2002). **Mixtures of arma models for model-based time series clustering**. Dans *Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on*, pages 717–720. IEEE.

[Xue et al., 2008] Xue, F., Bonissone, P., Varma, A., Yan, W., Eklund, N., et Goebel, K. (2008). **An instance-based method for remaining useful life estimation for aircraft engines**. *Journal of Failure Analysis and Prevention*, 8(2) :199–206.

[Yam et al., 2001] Yam, R., Tse, P., Li, L., et Tu, P. (2001). **Intelligent predictive decision support system for condition-based maintenance**. *The International Journal of Advanced Manufacturing Technology*, 17(5) :383–391.

[Yan et al., 2004] Yan, J., Koc, M., et Lee, J. (2004). **A prognostic algorithm for machine performance assessment and its application**. *Production Planning & Control*, 15(8) :796–801.

[Yeo et al., 2003] Yeo, S.-M., Kim, C.-H., Hong, K., Lim, Y., Aggarwal, R., Johns, A., et Choi, M. (2003). **A novel algorithm for fault classification in transmission lines using a combined adaptive network and fuzzy inference system**. *International journal of electrical power & energy systems*, 25(9) :747–758.

[Yu et al., 2006] Yu, G., Qiu, H., Djurdjanovic, D., et Lee, J. (2006). **Feature signature prediction of a boring process using neural network modeling with confidence bounds**. *The International Journal of Advanced Manufacturing Technology*, 30(7-8) :614–621.

[Yu, 2011] Yu, J. (2011). **A hybrid feature selection scheme and self-organizing map model for machine health assessment**. *Applied Soft Computing*, 11(5) :4041–4054.

[Zaidan et al., 2015] Zaidan, M. A., Harrison, R. F., Mills, A. R., et Fleming, P. J. (2015). **Bayesian hierarchical models for aerospace gas turbine engine prognostics**. *Expert Systems with Applications*, 42(1) :539–553.

[Zhang et al., 2015] Zhang, X., Chen, W., Wang, B., et Chen, X. (2015). **Intelligent fault diagnosis of rotating machinery using support vector machine with ant colony algorithm for synchronous feature selection and parameter optimization**. *Neurocomputing*.

[Zhang et al., 2005] Zhang, X., Xu, R., Kwan, C., Liang, S. Y., Xie, Q., et Haynes, L. (2005). **An integrated approach to bearing fault diagnostics and prognostics**. Dans *American Control Conference, 2005. Proceedings of the 2005*, pages 2750–2755.

[Zhang et al., 2006] Zhang, Z., Huang, K., et Tan, T. (2006). **Comparison of similarity measures for trajectory clustering in outdoor surveillance scenes**. Dans *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 3, pages 1135–1138. IEEE.

[Zio, 2012] Zio, E. (2012). **Prognostics and health management of industrial equipment**. *Diagnostics and Prognostics of Engineering Systems : Methods and Techniques*, pages 333–356.

[Zio et al., 2010] Zio, E., Di Maio, F., et Stasi, M. (2010). **A data-driven approach for predicting failure scenarios in nuclear systems**. *Annals of Nuclear Energy*, 37(4) :482–491.

[Zou et al., 2005] Zou, H., et Hastie, T. (2005). **Regularization and variable selection via the elastic net**. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 67(2) :301–320.

Document réalisé avec LaTeX et :
le style LaTeX pour Thèse de Doctorat créé par S. Galland — http://www.multiagent.fr/ThesisStyle
la collection de paquets `tex-upmethodology` — http://www.arakhne.org/tex-upmethodology/

Résumé :

Nos travaux de thèse s'intéressent au pronostic de défaillance de composant critique et à l'estimation de la durée de vie résiduelle avant défaillance (RUL). Nous avons développé des méthodes basées sur l'expérience. Cette orientation nous permet de nous affranchir de la définition d'un seuil de défaillance, point problématique lors de l'estimation du RUL.
Nous avons pris appui sur le paradigme de Raisonnement à Partir de Cas (RàPC) pour assurer le suivi d'un nouveau composant critique et prédire son RUL. Une approche basée sur les instances (IBL) a été développée en proposant plusieurs formalisations de l'expérience: une supervisée tenant compte de l'état du composant sous forme d'indicateur de santé et une non-supervisée agrégeant les données capteurs en une série temporelle mono-dimensionnelle formant une trajectoire de dégradation.
Nous avons ensuite fait évoluer cette approche en intégrant de la connaissance à ces instances. La connaissance est extraite à partir de données capteurs et est de deux types : temporelle qui complète la modélisation des instances et fréquentielle qui, associée à la mesure de similarité permet d'affiner la phase de remémoration. Cette dernière prend appui sur deux types de mesures: une pondérée entre fenêtres parallèles et fixes et une pondérée avec projection temporelle. Les fenêtres sont glissantes ce qui permet d'identifier et de localiser l'état actuel de la dégradation de nouveaux composants.
Une autre approche orientée donnée a été testée. Celle-ci est se base sur des caractéristiques extraites des expériences, qui sont mono-dimensionnelles dans le premier cas et multidimensionnelles autrement. Ces caractéristiques seront modélisées par un algorithme de régression à vecteurs de support (SVR). Ces approches ont été évaluées sur deux types de composants : les turboréacteurs et les batteries « Li-ion ». Les résultats obtenus sont intéressants mais dépendent du type de données traitées.

Mots-clés : Pronostic de défaillance basée sur l'expérience, durée de vie résiduelle avant défaillance, indicateurs de santé, trajectoires de dégradation, IBL, RàPC, connaissance, similarité, SVR.

Abstract:

Our thesis work is concerned with the development of experience based approachesfor critical component prognostics and Remaining Useful Life (RUL) estimation. This choice allows us to avoid the problematic issue of setting a failure threshold.
Our work was based on Case Based Reasoning (CBR) to track the health status of a new component and predict its RUL. An Instance Based Learning (IBL) approach was first developed offering two experience formalizations. The first is a supervised method that takes into account the status of the component and produces health indicators. The second is an unsupervised method that fuses the sensory data into degradation trajectories.
The approach was then evolved by integrating knowledge. Knowledge is extracted from the sensory data and is of two types: temporal that completes the modeling of instances and frequential that, along with the similarity measure refine the retrieval phase. The latter is based on two similarity measures: a weighted one between fixed parallel windows and a weighted similarity with temporal projection through sliding windows which allow actual health status identification.
Another data-driven technique was tested. This one is developed from features extracted from the experiences that can be either mono or multi-dimensional. These features are modeled by a Support Vector Regression (SVR) algorithm. The developed approaches were assessed on two types of critical components: turbofans and "Li-ion" batteries. The obtained results are interesting but they depend on the type of the treated data.

Keywords: Experience based prognostics, RUL, health indicators, degradation trajectories, IBL, CBR, knowledge, similarity, SVR.

École doctorale SPIM 16 route de Gray F - 25030 Besançon cedex
■ tél. +33 (0)3 81 66 66 02 ■ ed-spim@univ-fcomte.fr ■ www.ed-spim.univ-fcomte.fr