

PREDICTIONS DE RUL PAR ANALYSE DE SIMILARITES

Laure SIRET^{1,2,3}, Bruno AGARD^{1,2}, Christophe DANJOU^{1,2}, Renan BONNARD³

¹Laboratoire en Intelligence des Données

²Département de Mathématiques et génie industriel, École Polytechnique de Montréal, Montréal (Québec) – Canada

³SENAI Innovation Institute for Embedded Systems– Florianópolis/SC – Brazil

laure.siret@polymtl.ca, bruno.agard@polymtl.ca, christophe.danjou@polymtl.ca, renan.bonnard@sc.senai.br

RESUME : *La prédiction du temps de vie résiduel – RUL (Remaining Useful Life) est un facteur clé de la maintenance prédictive. Ces prédictions étaient généralement basées sur une expertise métier et sur des procédés statistiques indépendants des conditions d'utilisation des machines en temps réel. Or avec la généralisation des concepts clés de l'Industrie 4.0, l'utilisation des nouvelles technologies permet dorénavant la supervision en tout temps des équipements de production. La recherche sur l'estimation des RUL devient de plus en plus populaire. Or seule une branche de la recherche utilise simultanément des données sur les états systèmes (vibrations, température, usure) et les contraintes qui lui sont appliquées (consignes cibles). La majorité de ces approches utilisent des modèles de prédiction d'un état théorique de la machine pour le calcul des RUL. Dans cet article, la méthodologie présentée a pour objectif la prédiction de RUL basée sur l'analyse des similarités des comportements en fonction des contraintes appliquées et des états du système.*

MOTS-CLES : *Maintenance industrielle, Politiques de maintenance, Data Mining, Fusion d'information*

1 INTRODUCTION ET CONTEXTE

Les équipements de production sont constamment sollicités de manières différentes en fonction du contexte et des objectifs de l'entreprise. Ces sollicitations induisent l'usure de l'équipement et ainsi des opérations de maintenance. La maintenance industrielle peut se définir comme la conception et la planification des opérations de réparations et d'entretien d'un équipement industriel en fonction des facteurs mécaniques, économiques et logistiques. Les stratégies de maintenance ont donc des répercussions fortes sur l'utilisation des équipements par la prévalence du respect des objectifs de sécurité, disponibilité, qualité du produit ou autre objectif de la compagnie. Les analyses économiques et logistiques permettent d'estimer si la maintenance de l'équipement doit être réalisée avant ou après la panne de l'équipement (Blanchard *et al.*, 1995). Les analyses fonctionnelles des équipements permettent d'identifier les causes possibles d'une panne (ISO 17359-2018-01). Des outils tels que les arbres d'anomalies – FTA (Fault Tree Analysis), ou encore les analyses de causalités – CNEA (Causal Network Event Analysis), sont des exemples de méthodologies utilisées (Dias *et al.*, 2012). Ces analyses permettent théoriquement d'identifier les sous-ensembles de l'équipement responsables des pannes.

Les politiques de gestion des pannes sont découpées en 2 catégories : correctives et préventives (NF EN 13306). La maintenance corrective convient généralement aux interventions rapides où les coûts générés par la panne sont inférieurs aux coûts d'interventions préventives. Les politiques de maintenances préventives se séparent en trois autres catégories : les maintenances **systématiques**, **conditionnelles** et **prédictive** (Høyland and Rausand, 2009).

Une maintenance systématique a lieu à un intervalle de temps fixe. Une parmi les nombreuses méthodes d'estimation de l'intervalle de temps est l'utilisation de la loi de Weibull (Moukhli, 2011). Le recensement des temps entre chaque panne TBF (Time Between Failures), leur tri puis la détermination des paramètres d'échelle et de forme sont des étapes connues de l'industrie de par leur simplicité, efficacité et faible besoin en données.

Une maintenance conditionnelle indique que les interventions de réparations seront réalisées selon l'occurrence de certains événements identifiés à l'avance ou selon l'état de fonctionnement de certains sous-ensembles de l'équipement défini par des paramètres clés (par exemple le niveau d'huile dans un moteur). Aucune condition temporelle n'est nécessairement associée à une maintenance conditionnelle. Des politiques de maintenance conditionnelles basées sur les réseaux Bayésiens ou des chaînes de Markov ont permis de grandement augmenter le temps entre les arrêts programmés (Sakib, et Wuest, 2018).

La maintenance prédictive est définie comme une extension d'une politique conditionnelle « suivant les prévisions extrapolées » (NF EN 13306) des mêmes paramètres significatifs. Avec l'arrivée des nouvelles technologies d'acquisition de données en temps réel, on observe maintenant une recrudescence des recherches portant sur les maintenances conditionnelles et prédictives (Carvalho *et al.*, 2019). Ces méthodes se basent, soit sur une expertise métier pour l'élaboration d'un modèle de dégradation, soit sur des données de fonctionnement pour l'entraînement de méthodes supervisées, semi-supervisées ou non-supervisées (Javed *et al.*, 2017).

Le temps de vie résiduel ou RUL (Remaining Useful Life) représente la différence de temps entre la date de fin de vie de l'équipement et l'instant t . L'utilisation de politiques de maintenances prédictives passe générale-

ment par une estimation du RUL d'un équipement (Lei *et al.*, 2018). Bien que l'utilisation du Data Mining ait beaucoup élargi les champs des possibilités de la maintenance prédictive et de la PHM (Prognostic Health Management), les approches statistiques sont encore utilisées et sujet de la recherche aujourd'hui (Carvalho *et al.*, 2019). Les procédés statistiques les plus connus sont certainement les procédés Markoviens (Kharoufeh *et al.*, 2010), de dégradation homogène de Wiener (Wang, 2010) ou les filtres de Kalman étendu (Batzel et Swanson, 2009). Le plus gros défaut de ces méthodes est leur caractère indépendant face aux conditions d'utilisation des équipements en temps réel et donc des lois de dégradation aussi indépendantes des conditions d'utilisation en temps réel (Si *et al.*, 2011).

L'arrivée de l'industrie 4.0 (Drath et Horch, 2014) encourage l'utilisation des nouvelles technologies et leur intégration aux systèmes d'information. Une bonne intégration permet de conserver une multitude de données sur les conditions de fonctionnement des équipements de production en temps réel (Danjou *et al.*, 2017). L'application des algorithmes de Machine Learning tels que les arbres de décisions, réseau de neurones, SVM (Support Vector Machine) et K-means sont alors de plus en plus fréquents dans la littérature (Carvalho *et al.*, 2019), (Vogl *et al.* 2019). Les chercheurs étudient dorénavant majoritairement des mesures de conditions dites indirectes (état de l'huile, vibration, température) aux conditions directes (fissure) (Si *et al.*, 2011).

L'analyse des conditions indirectes se distingue en deux grandes familles : celles reposant sur la fusion d'information provenant de différentes sources (vibration, huile, pression, etc.) (Goebel et Bonissone, 2005) et celles se concentrant sur l'analyse d'une information spécifique (uniquement les vibrations) (Knežević *et al.* 2019). Dans le contexte d'une entreprise ayant plusieurs types d'équipements différents, l'étude d'un seul paramètre restreint les cas d'application de la méthodologie. La méthodologie proposée s'oriente donc dans la famille des méthodologies portant sur la fusion d'information. Il existe également de plus en plus de méthodologies essayant de différencier les informations relatives aux consignes imposées par les utilisateurs (vitesse, pression) aux états physiques du système non contrôlable (vibration, huile) (Huang *et al.*, 2019). (Jardine *et al.*, 2005) recense de nombreuses méthodes de distinction et fusion d'information. L'objectif de ces méthodes est de pouvoir relier le cas d'utilisation présent de la machine avec les expériences passées selon un seul et même critère généralement nommé indicateur de survie. Les réseaux de neurones et autres algorithmes sont souvent entraînés sur cet indicateur et une panne est prédite lorsque l'indicateur tombe en dessous d'un certain seuil. Or certaines méthodes de fusion de l'information ne peuvent contraindre l'indicateur de survie à atteindre une valeur minimale de 0 (Jardine *et al.*, 2005). L'identification de la panne peut alors être estimée par l'étude des similarités des temps de fonctionnement. Soit l'analyse des patrons de comportement d'un ensemble de cas passés par

rapport au cas actuel étudié. Différentes méthodes d'analyse de similarités existent dans la littérature. Cependant une grande partie repose sur des conditions directes telles que (Kadry, 2012) qui analyse les similitudes de propagation de fissure par Pattern Fuzzy Similarity pour déterminer le RUL de roulement. (Wang *et al.*, 2008) propose une analyse sur des données indirectes - TSBP (Trajectory Similarity Based Prediction) et (Wang *et al.*, 2010) la complète par une intégration des contraintes appliquées au système. Cependant la sélection des données les plus pertinentes n'y est pas automatisée. Plusieurs méthodes de calcul du RUL sont abordées, mais l'estimation directe par une étude statistique des temps de vie des comportements voisins ni est pas étudiée. Enfin des comparaisons simples pour justifier l'analyse des similarités ainsi que l'apport des conditions appliquées ne sont pas spécifiées.

Dans le présent article, nous proposons une nouvelle méthodologie de prédiction de RUL se basant sur les conditions d'exploitation de l'équipement (contraintes imposées par les opérateurs et contraintes de mesure indirectes). La méthodologie a pour objectif une utilisation en industrie et se base sur l'étude des similarités des comportements passés. On note que la présente méthodologie a été développée afin de répondre à un besoin industriel de l'agroalimentaire de transformation de masse de céréales. Dans ce contexte, les paramètres de production sont optimisés pour la production et ne peuvent être modifiés. L'objectif du déploiement de cette méthodologie est d'améliorer l'ordonnancement des travaux de maintenance et de diminuer les temps d'arrêt dus à un contrôle systématique des équipements. Ainsi, on suppose qu'un équipement est **remis à neuf** après une réparation.

La structure de l'article est la suivante : la section 2 expose la méthodologie proposée. La section 3 expose un cas d'application qui montre la mise en œuvre de la méthodologie. Finalement, la section 4 présente les résultats et limites de la méthode.

2 MÉTHODOLOGIE

2.1 Définition des termes

On appelle **Contraintes d'Activité (CA)** l'ensemble des paramètres de la machine contrôlables par les opérateurs (vitesse, pression, alimentation); et **États du Système (ES)** les grandeurs physiques intrinsèques à l'équipement non contrôlable par l'opérateur (vibration, qualité d'huile, température résiduelle). Les **données** sont les mesures de ces paramètres provenant de différents capteurs de mesures, elles couvrent l'ensemble du temps de fonctionnement de la machine. Pour l'ensemble de la méthodologie, les données sont séparées aléatoirement en un **ensemble d'entraînement** (80% des données) et un ensemble de **tests** (20%). La variable **delta** est l'écart entre la valeur réelle et la prédiction du temps de fin de fonctionnement. Un delta négatif signifie que la prédiction est postérieure à la panne effective de l'équipement.

La méthodologie proposée repose sur 3 grandes étapes présentées par la suite (2.2) l'analyse des temps de fonctionnement (2.3) l'analyse mono-facteur caractérisant les ES et enfin (2.4) l'analyse multi-facteurs caractérisant les CA et les ES. La première étape de la méthodologie servira de témoin pour la comparaison des résultats.

2.2 Analyse des temps de fonctionnement

La loi de Weibull est utilisée comme référentiel. Seuls les temps de fonctionnement de l'équipement sont nécessaires, les paramètres d'échelle et de forme de la loi de Weibull sont estimés automatiquement sur l'ensemble d'entraînement (Kadry, 2012). Les prédictions de RUL étant dé-corrélées des données relatives aux ES et aux CA il est possible de réaliser pour chaque individu test plusieurs prédictions aléatoires pour rendre les résultats significatifs.

2.3 Analyse de similarités mono-facteur

Les données considérées dans cette section proviennent d'un capteur décrivant un ES pertinent.

2.3.1 Prétraitement

Les données relatives à la production sont généralement bruitées. Une normalisation en Z et un filtre sont appliqués aux données. Le but est de conserver uniquement la disparité des données au cours du temps afin d'identifier des comportements symptomatiques d'une fin de fonctionnement. Les données extrêmes doivent être conservées. Les données sont ensuite filtrées par la courbe enveloppe supérieure. On pose l'hypothèse que l'utilisation de la courbe enveloppe permet de garder les caractéristiques d'un filtre passe-bas par la sélection de la fenêtre d'échantillonnage des données tout en gardant les données extrêmes.

2.3.2 Analyse de similarités

On suppose que des machines ayant un même comportement auront également des temps de fonctionnement similaires. Cette sous-section va quantifier les comportements similaires entre l'ensemble des données d'entraînement et l'ensemble des données tests. Pour ce faire, plusieurs algorithmes de calculs de distance existent : le calcul des distances point à point ou Dynamic Time Warping (DTW) sont les plus communs (Goebel et Bonissone, 2005).

Pour vérifier les performances de la méthodologie, le calcul de RUL est réalisé à un instant t_c donné, inférieur à tout temps de fin de fonctionnement. L'ensemble des données de test sont tronquées de $t=0$ à $t = t_c$. Les tailles séries de données sont donc différentes entre elles. L'utilisation de l'algorithme de DTW présente plus d'avantages pour une analyse de structure sur des séries de tailles différentes. L'ensemble des distances est alors calculé et seuls les k (k étant posé arbitrairement) machines ayant les plus petits scores sont conservées pour la suite de la prédiction.

2.3.3 Prédiction de RUL

Chaque donnée de test possède dorénavant un ensemble de données d'entraînement avec des comportements proches que l'on appelle voisinage. Les temps de fonctionnement du voisinage sont extraits et la prédiction du RUL se base sur la médiane du voisinage. Le choix de la médiane nous vient (Wang, 2010)

2.4 Analyse de similarités multi-facteurs

Les données utilisées pour cette section sont l'ensemble des données relatives aux CA et ES. L'Analyse de Similarités Multi-Facteurs (ASMF) se base sur les 4 étapes suivantes.

2.4.1 Identification des régimes de fonctionnement

Les conditions de fonctionnement appliquées à un équipement industriel varient au cours de la production et ont été définies ci-dessus comme les CA d'un équipement. Un ensemble de CA est appelé **régime de fonctionnement**. Les données relatives aux ES varient d'un régime de fonctionnement à un autre (Lei, 2016). Afin de pouvoir comparer et analyser les données d'ES de manière équitable, la première étape de la méthodologie de l'ASMF consiste à normaliser les données selon leur régime de fonctionnement tel que réalisé dans la littérature (Jardine *et al.*, 2006). S'il existe un grand nombre de régimes de fonctionnement, il est alors pertinent de regrouper des régimes similaires entre eux. Cette étape s'organise par (1) le regroupement des régimes de fonctionnement en clusters, (2) la normalisation en Z des données sur les ES selon le régime de fonctionnement.

Le clustering des régimes de fonctionnement est réalisé par un algorithme de recherche des K plus proches voisins - KNN (K-Nearest Neighbors) et le nombre de clusters est sélectionné comme celui minimisant la distance entre les régimes et le centroïde du cluster selon un critère d'arrêt : la première distance inférieure ou égale à une valeur arbitraire.

2.4.2 Identification des paramètres prépondérants

Cette section consiste à l'identification des informations d'ES les plus pertinentes pour la détection de fin de fonctionnement. La partie s'organise selon 3 étapes :

(1) Sélection par corrélations. Dans un équipement, il est fréquent que plusieurs capteurs internes à la machine soient intimement corrélés. Afin d'éviter un surapprentissage, une analyse de covariance est réalisée. Deux capteurs sont dits corrélés entre eux si la valeur absolue de la covariance est supérieure à un seuil fixé arbitrairement.

Le tri des capteurs est présenté dans la figure 3.

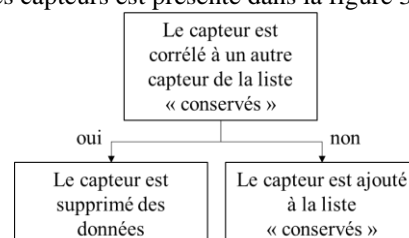


Figure 1 : Sous-étapes d'évaluation des corrélations

Ainsi, seuls les capteurs non corrélés (appartenant à la liste $L_{\text{conservés}}$) sont étudiés ce qui limite également les temps de calcul.

(2) Sélection par tendance. La sélection des indicateurs se base sur l'analyse de la tendance et deux méthodes de calcul différentes sont appliquées ; (a) Calcul du coefficient directeur de la régression linéaire issue de chaque capteur, (b) Calcul du rayon du plus petit cercle inscrit de la régression polynomiale d'ordre 3 ou plus.

La première méthode d'analyse met en évidence la tendance des données de manière globale. On prend l'hypothèse que la seconde méthode permet de quantifier la rapidité de modification de la tendance au cours du temps. Plus le rayon est petit, plus les données ont une tendance à évoluer rapidement dans les derniers cycles de fonctionnement de la machine. Chaque capteur est classé et obtient un score selon chacune des méthodes. La somme des scores permet de sélectionner les k meilleurs capteurs où k est un entier arbitraire respectant l'inégalité suivante :

$$k \leq \max\left(2, \left\lfloor \text{len}\left(\frac{L_{\text{conservés}}}{2}\right) \right\rfloor\right)$$

(3) Filtrage par enveloppe. Les capteurs sélectionnés sont ensuite filtrés en conservant leur courbe enveloppe. La méthodologie et les justifications utilisées pour chacun des capteurs sont les mêmes que pour la section de prétraitement des données de l'analyse à 1 facteur.

2.4.3 Fusion des données

Contrairement à l'analyse mono- facteur, il est nécessaire ici d'agrèger les informations des différents capteurs identifiés à la partie précédente dans un **indicateur de survie**. Pour rappel, l'équipement est supposé remis à neuf après chaque opération de maintenance ce qui implique que toutes les machines commencent avec la même valeur de survie. Cet indicateur vérifie les conditions suivantes :

- c1. Il est compris entre 0 et 1 pour tout instant t .
- c2. L'indicateur est décroissant
- c3. Il vaut 1 à $t = 0$ et 0 lorsque l'équipement tombe en panne

Une première estimation simple d'un tel indicateur est de l'estimer pour chaque machine par une droite dont les paramètres respectent les conditions c2.

La prochaine étape est de fusionner l'information des capteurs à cet indicateur de survie. La fusion de données est donc réalisée avec une régression linéaire multiple notée I . Pour garantir le caractère borné et décroissant, un nouvel indicateur \hat{I} est défini tel que $\hat{I}(t) = \min(I(t-1); I(t))$ où t est indice de temps.

2.4.4 Analyse de similarités et prédiction de RUL

La prédiction du RUL est basée sur la même méthodologie que pour l'analyse de similarités mono-facteur. Les données tests sont tronquées à un instant t arbitraire, mais antérieur à toute panne de l'équipement. Les étapes précédentes de l'ASMF sont ensuite appliquées à l'ensemble test soit (1) la l'identification des régimes de fonctionnement, (2) l'identification des paramètres prépondérants et (3) la fusion des données. À la fin de ces

trois sous étapes, une courbe tronquée semblable à une courbe de survie est associée à chaque machine test. L'algorithme de DTW (distance de Manhattan) va ensuite calculer l'ensemble des distances entre chaque courbe test avec chaque courbe d'entraînement. Les k plus proches voisins sont conservés, où k est fixé arbitrairement. La prédiction ponctuelle de RUL est la médiane de l'ensemble des k voisins. Finalement, l'écart δ est calculé selon la définition de donnée dans la section 2.1 définition des termes.

3 CAS D'APPLICATION

3.1 Contexte

Les données utilisées sont des données de fonctionnement de turbines fournies par la NASA en 2008 lors de la première conférence internationale Prognostics and Health Management (PHM08). Le cas d'application utilise une partition aléatoire du fichier contenant les données d'entraînement. Le set de données se compose d'un repère temporel, du numéro de machine, de 3 composantes relatives aux CA et de 21 composantes relatives aux ES. La composante de temps a pour unité un cycle de fonctionnement. Les informations obtenues par les capteurs peuvent être du domaine continu ou discret.

À noter que les dernières étapes de chaque sous-méthodologie sont présentées dans la partie 5 "résultat".

3.2 Analyse des temps de fonctionnement

Pour une meilleure estimation des paramètres de la loi de Weibull, l'offset de la série doit être nul. On retranche alors la valeur minimale de la série à l'ensemble des données. La figure 2 présente l'estimation de la loi de répartition.

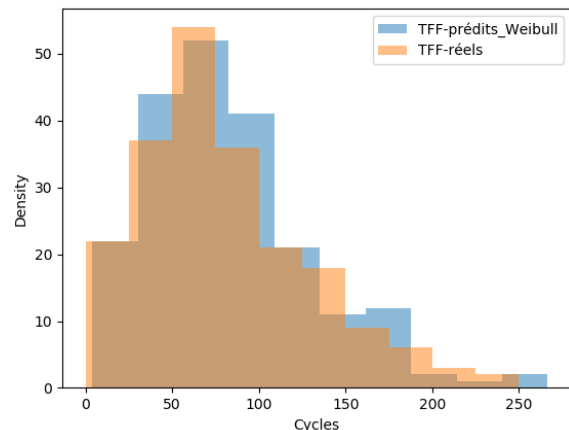


Figure 2 : Histogrammes des temps de fin de fonctionnement réels et prédits par Weibull

3.3 Analyse de similarités mono-facteur

L'analyse des similarités mono-facteur nécessite, comme son nom l'indique, les données relatives à un seul capteur. Les données détaillant les ES de la base PHM08 étant nombreuses, la sélection du capteur 8 a été effectuée aléatoirement.

3.3.1 Prétraitement

La figure 3 donne une visualisation après normalisation des données issues du capteur 8.

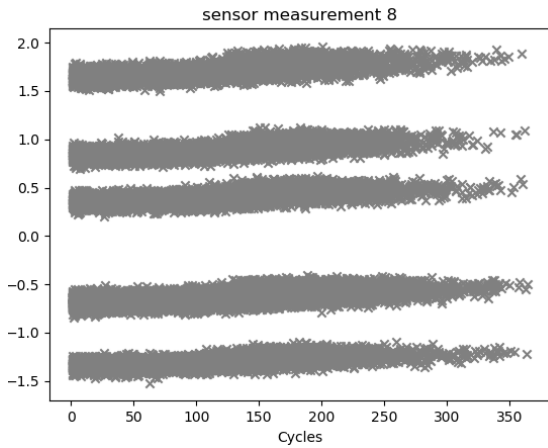


Figure 3 : Capteur 8 après normalisation

Le filtrage des données par la courbe enveloppe est ensuite réalisée et est présentée sur la figure 4.

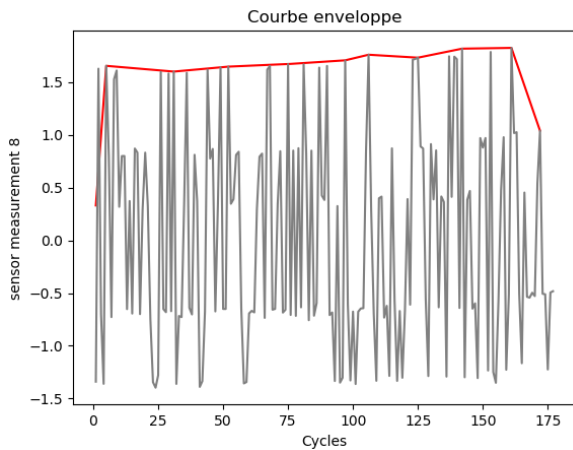


Figure 4 : Courbe enveloppe des données du capteur 8

3.3.2 Analyse de similarité

D'après les étapes présentées partie 2.3.2 et grâce à l'étude statistique on identifie un temps de coupure pour les individus test. Un $t_c = 100$ cycles est sélectionné pour ce cas d'application. L'ensemble des données test sont alors tronquées après t_c . Le vecteur des distances (analyse DTW) est calculé et seuls les $k=15$ premiers voisins sont conservés pour les prédictions de RUL. Les résultats seront présentés dans la partie suivante « Résultats ».

3.4 Multiples facteurs

3.4.1 Identification des régimes de fonctionnement

Les données d'étude contiennent trois informations relatives aux CA. L'analyse des régimes de fonctionnement portera donc sur ces trois colonnes. Les résultats du clustering des différents régimes sont présentés figure 5 et figure 6. La première figure présente les sommes cumulées des distances entre chaque élément du cluster à son centroïde et justifie la sélection de 3 clusters. La figure 6 permet de visualiser les régimes en bleu et les centroïdes

par les sphères rouges plus larges. Les données sont alors normalisées selon ces trois clusters.

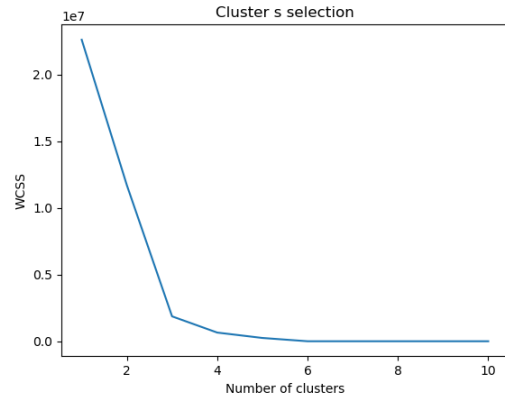


Figure 5 : Courbe des distances aux centroïdes

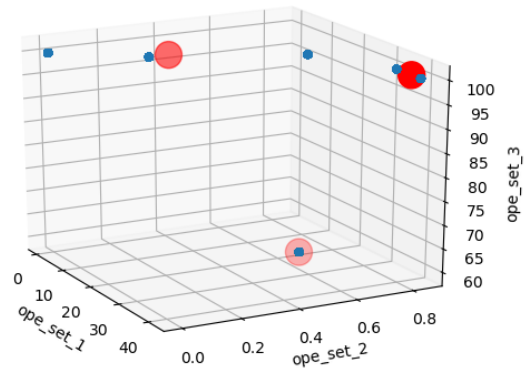


Figure 6 : Clusters des régimes de fonctionnement

3.4.2 Identification des paramètres prépondérants

Cette étape se focalise sur l'étude des ES soit 21 colonnes. Les données relatives aux ES étant nombreuses, on fixe le seuil de la covariance à 0.5 afin de limiter au maximum le surapprentissage dû à une redondance des données.

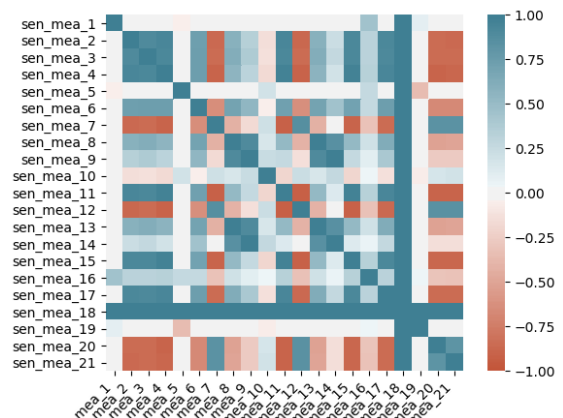


Figure 7 : Carte de chaleur de l'analyse des covariances

Seuls 12 sur les 21 capteurs initiaux sont conservés après l'analyse de covariance. Les tendances globale et locale de ces 12 capteurs sont alors analysées respectivement une régression linéaire et une régression polynomiale d'ordre 3. Les figures 8 et 9 suivantes présentent les résultats des deux régressions.

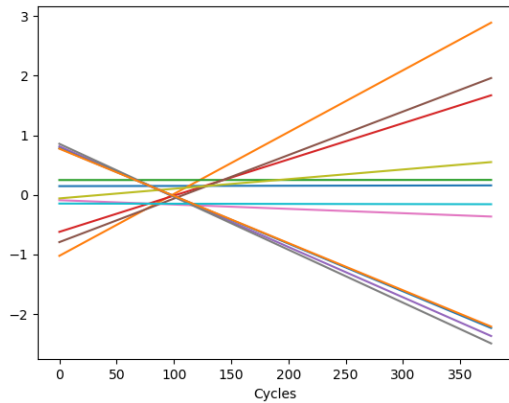


Figure 8 : Résultat analyse de la tendance globale

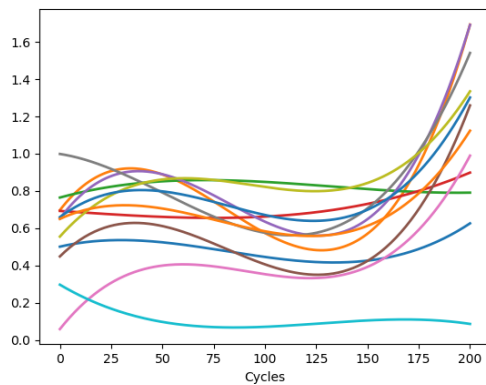


Figure 9 : Résultat analyse de la tendance locale

Chaque capteur est trié selon chacune des régressions avant qu'un score lui soit associé. Seuls les meilleurs capteurs sont conservés pour la suite de l'étude. Un filtrage par courbe enveloppe est appliqué avant de passer à la partie suivante de fusion d'information.

3.4.3 Fusion de données

Dans un premier temps, l'indicateur préliminaire : décroissant et linéaire est calculé. La figure 10 représente cet indicateur pour chaque machine de l'ensemble d'entraînement.

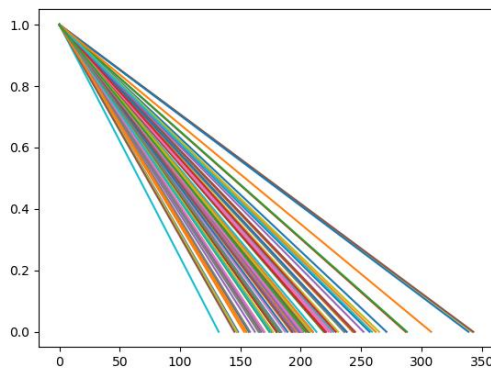


Figure 10 : Indicateur de survie préliminaire

La fusion d'information entre l'indicateur de survie et les capteurs de la partie précédente est réalisée par une régression linéaire. Cet indicateur est ensuite filtré par une fonction en escalier afin de respecter les contraintes de

bornes et de décroissance de l'indicateur. La figure 11 présente l'indicateur de survie final. La panne d'un équipement arrive dès qu'il n'y a plus de données et est caractérisée par une croix noire. On peut noter une forme de décroissance exponentielle caractéristique d'une dégradation.

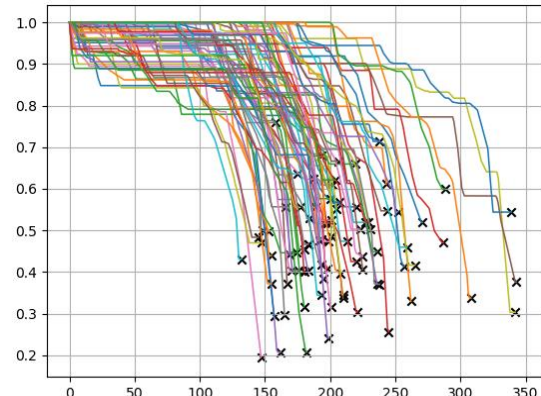


Figure 11 : Indicateur de survie set final

De cet ensemble de courbes de survie, une analyse de similarités avec des données tronquées permet de réaliser la prédiction du temps de vie.

4 RÉSULTATS

La partie suivante regroupe l'ensemble des prédictions réalisées pour le cas d'étude dans figure 12. Les résultats sont présentés via l'indicateur delta défini dans la partie 3.1.—Delta positif et plus proche de 0 caractérise de bonnes prédictions. Au contraire un delta négatif signifie que l'équipement est tombé en panne avant l'estimation ce que l'industriel cherche à éviter le plus possible. Si la prédiction est antérieure à la date de la panne effective, l'impact est moindre. En effet, comme expliqué dans la partie 3.3.2 les prédictions sont réalisées à des intervalles prédéfinis, il est donc possible pour l'opérateur de relancer des prédictions à une période postérieure et donc d'améliorer la prédiction sans prendre de risque supplémentaire sur une panne de la machine.

Pour l'ensemble des histogrammes présentés figure 12, la légende suivante est appliquée : (1-Vert) Vert : $0 < \text{delta} < 100$ cycles, (2-Orange) Orange : $-50 < \text{delta} < 0$ cycle ou $\text{delta} > 100$ cycles et (3-Rouge) Rouge : $\text{delta} < -50$ cycles

Les barres verticales représentent respectivement le premier quartile, la médiane et le 3^e quartile des prédictions.

4.1 Calcul et comparaison des prédictions

Les prédictions des trois sous partie de la méthodologie sont présentées sur la figure 12.

Le tableau 1 ci-dessous présente le nombre de prédictions postérieures au temps de fin de fonctionnement et la variance des prédictions pour les 3 sous-méthodologies. La variance des données d'origines est calculée et vaut 46.7 cycles.

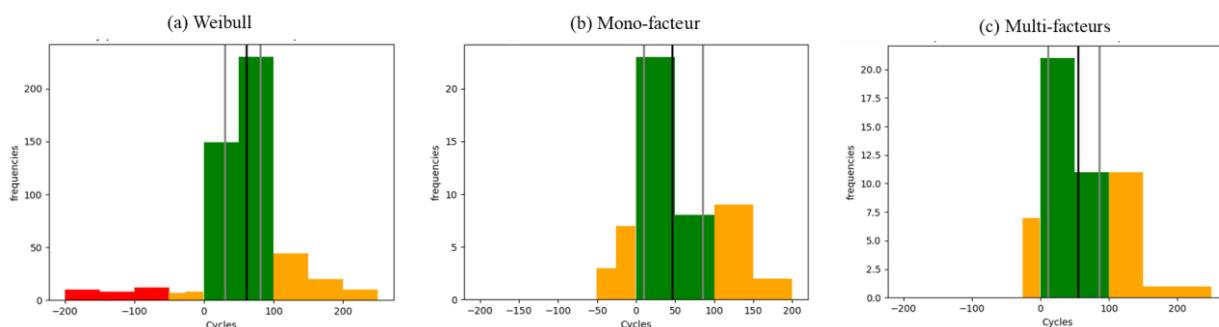


Figure 12 : Répartition delta méthode (a) Weibull, (b) mono-facteur, (c) multi-facteur

Le temps de calcul nécessaire pour 1 prédiction pour un ensemble test de 208 éléments est ajouté à titre indicatif pour un processeur Intel Core i5-3437U CPU, 1.9GHz, 2.4GHz.

	Prédiction après panne	Std (cycles)	Temps de calcul
Weibull	12.9 %	90.96	< 1 s
Mono	19.23 %	53.82	≈ 17 s
ASMF	13.46 %	56.76	≈ 30 s

Tableau 1 : Comparaison des résultats

Les résultats sont comparés selon quatre critères (1) le nombre de prédictions réalisées après que la panne ait réellement eu lieu (Jardin *et al.*, 2006), (2) la gravité des mauvaises prédictions, (3) la disparité de delta et (4) le temps de calcul (Huang *et al.*, 2019).

(1) On cherche à évaluer par ce critère quelle méthode donne le plus grand nombre de prédictions antérieures à la panne. On note que les prédictions réalisées sont antérieures à la panne réelle à plus de 80% pour les 3 méthodes. Le classement des méthodes, de la meilleure à la moins performante, est le suivant : (1) Weibull, (2) ASMF et (3) mono capteur, avec seulement 0.6% d'écart entre Weibull et ASMF.

(2) L'étude de la gravité des mauvaises prédictions (légende présentée partie 4 ci-dessus) permet de quantifier l'impact de mauvaises prédictions sur les décisions de maintenance. On obtient le classement (1) ASMF, (2) mono capteur et (3) Weibull. La présence de (3-Rouge) est ce qui place Weibull en dernier.

(3) On note une augmentation de quasi 2 fois l'écart type pour les prédictions par une loi de Weibull, mais de seulement 0.2 fois pour ASMF et mono capteur. Les intervalles de confiance liés aux prédictions sont donc plus petits pour ces deux dernières méthodes. Des prédictions plus précises permettent de plus facilement déterminer des intervalles de confiance sur l'état de la machine. La planification de la maintenance en est facilitée.

(4) Le temps de calcul nécessaire d'une prédiction pour chaque méthode reste inférieur à 1 minute. Weibull reste la méthode la plus rapide de par sa simplicité. L'ASMF nécessite le temps de calcul le plus long, mais une exécution de moins 1 minute ne représente pas d'inconvénients notables pour l'organisation de la maintenance.

5 CONCLUSION ET DISCUSSION

La méthodologie proposée permet de calculer et comparer des prédictions de RUL selon les similarités de comportement d'une machine. Une intégration des contraintes de fonctionnement et des états du système a donné de meilleurs résultats de prédiction. Les prédictions obtenues par la méthodologie présentent de meilleurs résultats que des analyses communément utilisées telles que Weibull. Bien que le nombre de prédictions en retard par rapport à la réalité soit plus grand de 0.6% pour l'ASMF vis-à-vis de Weibull, l'ASMF prédit de manière plus précise et les mauvaises prédictions impactent moins la production (gravité plus faible).

On remarque qu'une seule prédiction par machine est réalisée au cours de la méthodologie. Or, plus l'instant de la prédiction est proche de la fin de fonctionnement réelle, plus les prédictions seront précises. Afin d'améliorer les performances, il est donc possible de prévoir plusieurs points de calcul des prédictions.

On note cependant que cette méthodologie nécessite beaucoup de données de surveillance en continu et allant jusqu'à la panne de l'équipement (problème aussi identifié pour l'utilisation de la loi de Weibull). La problématique est d'obtenir ou de détecter automatiquement les anciennes pannes du système. Étendre la méthodologie avec un prétraitement par une analyse d'anomalies est une piste à étudier.

L'analyse des contraintes appliquées est réalisée ponctuellement au début de l'ASMF. Or de nouvelles conditions de fonctionnement peuvent impacter le regroupement des régimes de fonctionnement. Pour conserver les bonnes performances de la méthodologie, il serait donc nécessaire de recalculer les régimes de fonctionnement. Prévoir la mesure d'une déviance des centroïdes des régimes de fonctionnement par des méthodes de concept drift permettrait une plus grande autonomie.

Enfin la comparaison des résultats devrait évaluer la robustesse des prédictions. Des métriques telles que présentées par (Saxena *et al.*, 2008) pourraient être une piste d'amélioration.

REMERCIEMENTS

Merci à l'ensemble de l'équipe du centre de recherche du SENAI, SC Brésil et particulièrement au chercheur Herberth Birck Fröhlich pour leur soutien à l'élaboration de cette méthodologie.

REFERENCES

- Batzel, T.D., Swanson, D.C., 2009. Prognostic health management of aircraft power generators. *IEEE Transactions on Aerospace and Electronic Systems* 45 (2), p 473–483.
- Blanchard, B-S. Verma, D. and Peterson, E.-L., 1995. *Maintainability - A key to effective serviceability and maintenance management*, Wiley – Interscience
- Carvalho, T., Soares, F., Vita, R., Francisco, R., Basto, J., and Alcalá, S., 2019. A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & Industrial Engineering*, 137, 10602
- Danjou, C., Rivest, L., and Pellerin, R., 2017. Industrie 4.0 : des pistes pour aborder l'ère du numérique et de la connectivité. *PME 2.0 Le passage au numérique*.
- Dias, A. Totten, G. and DeNegri, V. J., 2012. *Handbook of Hydraulic Fluid Technology*. Second Edition. CRC Press.
- Drath, R., and Horch, A., 2014. Industrie 4.0: Hit or hype? *IEEE Industrial Electronics Magazine*, 8(2), p 56-58.
- Goebel, K., and Bonissone, P., 2005. Prognostic information fusion for constant load systems. *7th International Conference on Information Fusion*, vol. 2, p. 9 pp.
- Høyland, A., and Rausand, M., 2009. *System reliability theory: models and statistical methods* (vol. 420): John Wiley & Sons.
- Huang, C., Huang, H., and Li, Y., 2019. A Bidirectional LSTM Prognostics Method Under Multiple Operational Conditions. *IEEE Transactions on Industrial Electronics*, 66(11), p 8792-8802.
- James, G., Witten, D., Hastie, T., and Tibshirani, R., 2013. *An introduction to Statistical Learning: Springer*.
- Jardine, A. K. S., Lin, D., and Banjevic, D., 2006. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7), p 1483-1510.
- Javed, K., Gouriveau, R., and Zerhouni, N., 2017. State of the art and taxonomy of prognostics approaches, trends of prognostics applications and open issues towards maturity at different technology readiness levels. *Mechanical Systems and Signal Processing*, 94, p214-236.
- Lei, Y., 2016. *Intelligent fault diagnosis and remaining useful life prediction of rotating machinery*: Butterworth-Heinemann.
- Kadry, S., 2012. *Diagnostics and Prognostics of Engineering Systems: Methods and Techniques*, (Chapitre 7) p 135-150.
- Kharoufeh, J. P., Solo, C. J., and Ulukus, M. Y., 2010. Semi-Markov models for degradation-based reliability. *IIE Transactions*, 42(8), p 599-612
- Knežević, I., Zivkovic, A., Šejat, M., Rackov, M., Zeljković, M., and Cofaru, N., 2019. Analysis the amount of lubrication and roughness of raceways on dynamic behavior on the ball bearing. *MATEC Web of Conferences*, 290, 01005
- Kolar, D., Lisjak, D., Pajak, M., and Pavkovic, D., 2019. Fault Diagnosis of Rotary Machines using Deep Convolutional Neural Network with three axis signal input. *arXiv*, 1906.0244
- Lei, Y., Li, N., Guo, L., Li, N., Yan, T., and Lin, J., 2018 Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mechanical Systems and Signal Processing*, 104, p 799-834.
- Murthy, D.N. P., Rausand, M. and Østeras, T., 2008 *Product Reliability – Specification and performance*, Springer
- Moukhli, A., 2011. *Optimisation de la maintenance de roues de turbines hydroélectrique soumises à une dégradation par cavitation*. Mémoire de maîtrise ès science appliquées, École Polytechnique Montréal, Canada.
- Sakib, N., and Wuest, T., 2018. Challenges and Opportunities of Condition-based Predictive Maintenance: A Review. *Procedia CIRP*, 78, p 267-272.
- Saxena, A., Celaya, J., Balaban, E., Goebel, K., Saha, B., Saha, S., and Schwabacher, M. 2008. Metrics for evaluating performance of prognostic techniques. *2008 International Conference on Prognostics and Health Management*, Denver, CO, USA.
- Si, X., Wang, W., Hu, C., and Zhou, D., 2011. Remaining useful life estimation – A review on the statistical data driven approaches. *European Journal of Operational Research*, 213(1), p 1-14
- Vogl, G. W., Weiss, B. A., and Helu, M., 2019. A review of diagnostic and prognostic capabilities and best practices for manufacturing. *Journal of Intelligent Manufacturing*, 30(1), p79-95.
- Wang, Q., Zheng, S., Farahat, A., Serita, S., and Gupta, C., 2019. Remaining useful life estimation using functional data analysis. *IEEE International Conference on Prognostics and Health Management (ICPHM), 2019*, San Francisco, CA, USA
- Wang, T., Yu, J., Siegel, D., and Lee, J., 2008. A similarity based prognostic approach for remaining useful life estimation of engineered systems. *International Conference on Prognostics and Health Management*. Denver, USA.
- Wang, T., 2010. *Trajectory Similarity Based Prediction for Remaining Useful Life Estimation*. PhD, University of Cincinnati, USA.
- Wang, X., 2010. Wiener processes with random effects for degradation data. *Journal of Multivariate Analysis* 101 (2), p 340–351.