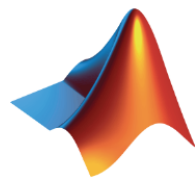


P10.S8.01 : Le Machine Learning au service de Maintenance Prédicative (le cas des éoliennes)



CentraleSupélec



MathWorks®

Equipe :

Diego Ruiz Ponsoda
Ben Silva
Valentin Gérard
Ayoub Ennajah
Mohammed Ahlal

Encadrants :

Wassila Ouerdane (CentraleSupélec)
Jean-Philippe Poli (CentraleSupélec)
Lama Itani (MathWorks)

Contents

Equipe	1
Encadrants	1
Sommaire	3
Structure du projet	3
État de l'art	4
A systematic literature review of machine learning methods applied to predictive maintenance	4
Surveillance vibratoire et maintenance prédictive	5
Predictive maintenance with Matlab.....	6
Définition de RUL et les différents modèles pour l'estimer	7
Description de travail réalisé	8
Prétraitement des données	8
Feature extraction	9
Indicateur de Condition	14
Modèle de Dégradation	14
Résultats initiaux.....	17
Tuning of the model.....	19
Déploiement et perspectives	19
Conclusion.....	20

1. Sommaire

Le client avec lequel nous travaillons est Mathworks, un éditeur de logiciel américain. Notre principal interlocuteur au sein de cette entreprise est Lama Itani, qui se concentre sur les problématiques de maintenance prédictive dans l'industrie. Pour mieux comprendre le lien entre Mathworks et la maintenance prédictive, il convient d'aborder brièvement Matlab, un logiciel développé par cette entreprise permettant de déployer rapidement du code pour résoudre des problèmes industriels. L'un des principaux atouts de Matlab est sa collection de boîtes à outils (toolboxes), des bibliothèques qui facilitent grandement la partie technique de la mise en œuvre. Dans le cadre de notre étude, nous nous intéressons principalement à la boîte à outils « predictive maintenance » qui rassemble des outils permettant de réaliser de la maintenance prédictive.

Le concept de maintenance prédictive sera développé dans l'état de l'art, mais pour le moment, il suffit de comprendre qu'il s'agit d'une méthode de maintenance pour un système qui consiste à surveiller différents indicateurs de dégradation et à remplacer les pièces lorsqu'ils dépassent un certain seuil (normalement avant la défaillance).

Le problème posé par Matlab est celui d'estimer le temps de vie restant d'une éolienne. Pour ce projet, nous avons reçu des signaux de tachymètre et de vibration représentant des données collectées à l'aide de capteurs sur une éolienne pendant 50 jours. L'objectif du projet est donc de réussir à estimer le moment de la défaillance d'une éolienne similaire en utilisant un signal analogue à celui de la tachymétrie et des vibrations, afin de planifier la maintenance et de réduire le temps d'arrêt du système. Pour mener à bien ce projet, le client nous encourage à utiliser Matlab et des méthodes de machine learning.

2. Structure du projet

Les livrables du projet consistent en une bibliographie qui nous permettra de prendre en main les problématiques de maintenance prédictive ainsi que les outils classiques utilisés dans ce domaine, ainsi qu'un code réalisé de préférence en Matlab permettant d'estimer le temps de vie restant d'une éolienne similaire à celle fournie dans le jeu de données. Le temps de vie restant de l'éolienne doit évoluer en permanence en fonction des nouveaux signaux collectés. Il est également attendu de notre algorithme qu'il donne des bornes inférieures et supérieures sur le temps de vie restant de l'éolienne.

Le MVP que nous avons proposé et celui d'un algorithme estimant le temps de vie restant d'une éolienne en se basant sur quelques features issues des signaux d'entrée.

Le groupe se rencontre pour échanger et travailler principalement sur les heures de projet dédiées dans l'emploi du temps, en outre nous rencontrons le client régulièrement afin d'échanger sur les solutions que nous mettons en place (1 fois par semaine au début du projet et 1 fois toutes les 2 semaines actuellement).

3. État de l'art :

i. A systematic literature review of machine learning methods applied to predictive maintenance

Cet article commence par rappeler les différentes terminologies et approches générales des procédures de maintenance. En bref, on distingue :

- **Run-to-Failure (R2F)** : On répare les pièces au moment où elles cassent.
- **Maintenance préventive** : Les casses passées ont permis de mettre au point une estimation de la durée de vie de chaque pièce de la machine. Cela permet de construire un planning de maintenance, les pièces sont alors en théorie changées avant la casse. Les problèmes de cette méthode viennent du fait qu'on va être amené dans certains cas à changer certaines pièces encore fonctionnelles ou à l'inverse se faire surprendre par les casses précoces.
- **Maintenance prédictive** : Ici, on utilise des capteurs qui mesurent certains signaux représentatifs de l'état des pièces et on déclenche l'opération de maintenance à partir du moment où leur dégradation est trop importante. L'avantage de cette procédure est qu'on optimise au maximum le nombre d'opérations et le temps de vie de chaque pièce.

La maintenance prédictive elle-même se sépare en différentes familles en fonction du modèle utilisé pour faire les prédictions :

- **Approche par modèle** : requière des connaissances mécaniques sur le système afin de le modéliser par des équations.
- **Approche statistique** : requière des connaissances mathématiques.
- **Approche par intelligence artificielle** : La plus prometteuse actuellement, d'après l'article, elle commence déjà à avoir des résultats supérieurs aux deux autres approches.

Cet article a pour but de faire un état de l'art des différentes méthodes de machine learning utilisées pour la maintenance prédictive.

- Le random forest (RF) est utilisé et a de bonnes performances lorsque le nombre de data est très inférieur au nombre d'échantillons (ce qui est notre cas dans ce projet). Il permet de faire de la classification et de la régression (donc est capable d'estimer le temps de vie restant). L'article nous redirige vers différents papiers qui ont réalisés de la maintenance prédictive avec un algorithme de RF (les résultats atteignent 85% de précision dans des cas réels). Finalement, l'article mentionne le fait que bien qu'il fut le modèle le plus utilisé, le RF souffre de problèmes de temps de calcul important.
- Les Artificial neural networks (ANN) ont des performances impressionnantes en comparaison aux RF. Ils ont notamment déjà été utilisés dans le cadre des turbines d'éoliennes avec des résultats de 92.6% en classification.
- Les Support Vectors Machine SVM sont utilisés a priori pour la classification mais l'article mentionne des modifications capables de lui faire effectuer une régression (voir Support Regression Vector SVR).
- L'article mentionne enfin les algorithmes K-means mais comme ces derniers ne sont pas capable de faire de la régression, nous ne les mentionnerons pas ici.

Pour conclure, cet article nous a permis de définir les différents algorithmes que nous pourrions mettre en place pour ce projet à savoir un RF, un ANN ou un SVR. Ici, on privilégiera les ANN car la

littérature nous indique qu'ils ont déjà été utilisé par le passé pour effectuer des tâches très semblables à la notre (cf Biswal and Sabareesh, 2015 Design and development of a wind turbine test rig for condition monitoring studies).

ii. Surveillance vibratoire et maintenance prédictive

• Maintenance

Même dans des conditions normales, le fonctionnement d'une installation entraîne un vieillissement des équipements et parfois des incidents. Afin d'assurer une exploitation correcte, il est essentiel de maintenir l'installation en bon état.

La maintenance a un coût, et il est crucial de trouver l'équilibre optimal pour maintenir l'installation à un coût minimal sans compromettre la sécurité des personnes. Comme le montre **la figure 1**, sans maintenance adéquate, le nombre d'incidents et leur coût augmenteront. À l'inverse, avec trop de maintenance préventive, le coût total augmentera également.

La maintenance optimale est donc un mélange harmonieux d'entretien préventif systématique et d'entretien correctif.

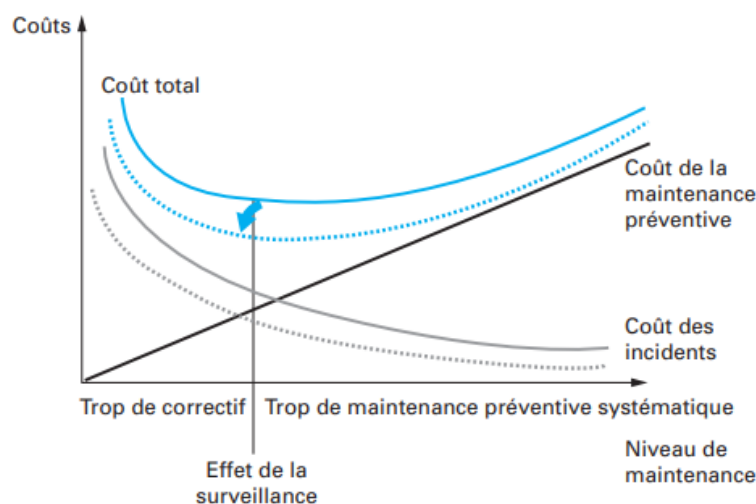


Figure 1 - Coût en fonction de la politique de maintenance exploitée

• Surveillance des matériels

La surveillance des équipements permet de réduire le niveau d'entretien préventif sans prendre de risques supplémentaires, car on s'assure toujours que le matériel ne présente pas d'anomalie. Par conséquent, les coûts peuvent être réduits si le coût de la surveillance reste limité. La surveillance s'intègre dans la stratégie de maintenance et doit assurer **la prévention** des risques majeurs (arrêt des machines lorsque les conditions de sécurité pour l'homme ou la machine ne sont pas respectées), **la détection** précoce des anomalies pour minimiser les dommages en remplaçant les composants défectueux si possible avant l'incident et au meilleur moment, et **l'analyse après**

incident pour remédier aux défauts constatés, éviter leur réapparition à l'avenir ou définir les modifications nécessaires.

- **Vibration : un bon indicateur d'état**

Le fonctionnement des machines génère des forces qui seront souvent à l'origine des défaillances ultérieures (forces tournantes, turbulences, chocs, instabilité). Ces forces provoquent à leur tour des vibrations qui endommagent les structures et les composants des machines. L'analyse de ces vibrations permet d'identifier les forces dès leur apparition, avant qu'elles n'aient causé des dommages irréversibles. Elle permet également, après analyse, d'en déterminer l'origine et d'estimer les risques de défaillance. **La maintenance prédictive repose sur ces concepts. Pour la mettre en œuvre, il est nécessaire de déterminer les causes de défaillance les plus fréquentes, d'évaluer leurs coûts et leur probabilité d'apparition, et de mettre en place une politique permettant de détecter le plus tôt possible les symptômes.**

- **Reconnaître les défauts**

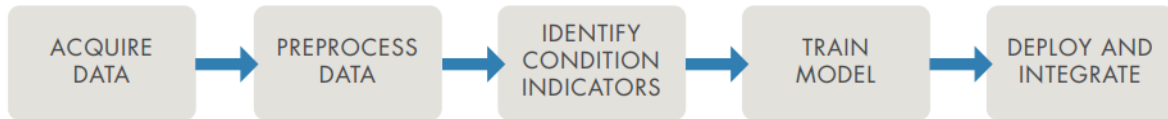
La maintenance prédictive nécessite un diagnostic minimal des défauts et de leur gravité. La première étape d'une action de surveillance consiste donc à déterminer les défauts susceptibles de survenir sur la machine à surveiller. La seconde étape concerne les manifestations de ces défauts. Il est important de déterminer et de mesurer les informations et les paramètres descripteurs du défaut pour disposer des informations appropriées. Celles-ci permettront de déterminer si la situation est normale ou non (détection d'anomalie) et d'identifier ultérieurement l'origine et la gravité des anomalies (diagnostic de l'origine et de la gravité des anomalies).

La maintenance prédictive dépend de la capacité à reconnaître et à diagnostiquer les défauts et leur gravité. Ainsi, la première étape d'une action de surveillance consiste à identifier les problèmes potentiels susceptibles de survenir sur la machine en question. La seconde étape porte sur la manière dont ces défauts se manifestent. Il est essentiel de déterminer et de mesurer les informations et les paramètres décrivant le défaut afin d'obtenir les bonnes informations. Ces informations permettront non seulement de déterminer si la situation est normale ou non (détection d'anomalies), mais aussi d'identifier ultérieurement la cause et la gravité des anomalies détectées (diagnostic de l'origine et de la gravité des anomalies).

En somme, la maintenance optimale consiste en un mélange équilibré d'entretien préventif systématique et d'entretien correctif, soutenu par la surveillance et l'analyse des équipements. La maintenance prédictive s'appuie sur l'analyse des vibrations et l'identification des défauts pour minimiser les risques de défaillance et réduire les coûts globaux. Cette approche nécessite une compréhension approfondie des machines et des défauts potentiels, ainsi qu'une stratégie de surveillance bien conçue pour assurer une gestion efficace et rentable des équipements.

iii. Predictive maintenance with Matlab

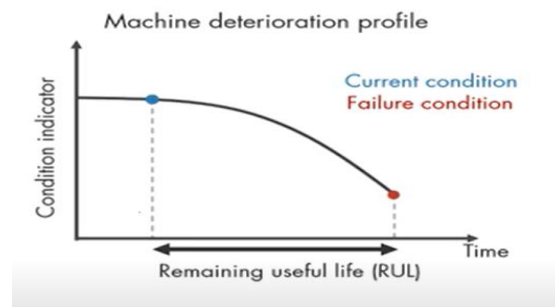
Ce livre électronique fournit par le client est une méthode pratique permettant d'implémenter à l'aide de Matlab un modèle de maintenance prédictive. Après avoir brièvement décrit l'intérêt de la maintenance prédictive, il identifie le workflow classique d'un projet d'implémentation de maintenance prédictive à travers le schéma suivant :



Pour ce projet, en accord avec le client, nous n'effectuerons que les tâches 3 et 4, c'est pourquoi je ne détaillerais que ces dernières.

iv. Définition de RUL et les différents modèles pour l'estimer

La "remaining useful life" (RUL) est une mesure du temps restant avant qu'un objet ne cesse de fonctionner ou ne nécessite une maintenance importante, on l'utilise dans le contexte de la maintenance prédictive. En utilisant les données sur la performance passée d'un objet, on peut prédire combien de temps il peut encore fonctionner avant de tomber en panne ou d'avoir besoin d'une maintenance.



Les modèles de prédiction de la RUL peuvent être classés en trois catégories principales : les modèles de similarité, les modèles de survie et les modèles de dégradation. Chacun de ces modèles est utilisé pour estimer la RUL d'un objet en utilisant différentes sources de données et en utilisant des approches différentes.

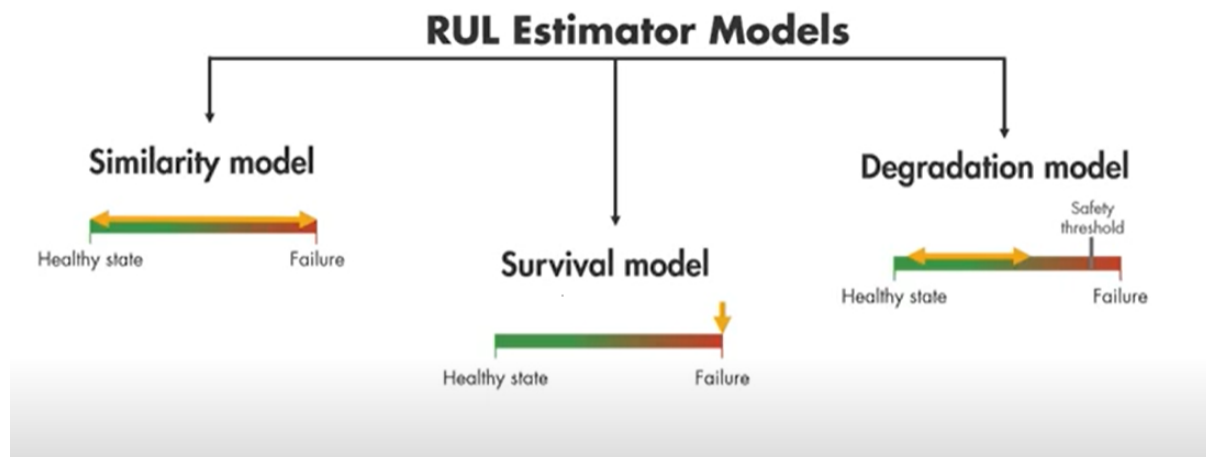
Les modèles de similarité sont des modèles qui estiment la RUL d'un objet en comparant ses caractéristiques actuelles avec celles d'objets similaires qui ont déjà atteint leur fin de vie. Ces modèles sont basés sur l'hypothèse que la durée de vie des objets similaires est corrélée, ce qui permet d'extrapoler la RUL de l'objet en question. Les modèles de similarité sont particulièrement utiles lorsque les données de performance passées de l'objet sont limitées, mais qu'il existe des données sur des objets similaires.

Les modèles de survie sont des modèles qui estiment la RUL d'un objet en fonction de sa durée de vie attendue, en utilisant des données sur la survie des objets similaires. Ces modèles sont basés sur la théorie de la survie, qui étudie la probabilité qu'un objet tombe en panne à un moment donné en fonction de ses caractéristiques. Les modèles de survie sont particulièrement utiles lorsque les données de performance passées de l'objet sont disponibles et que la distribution de la durée de vie de l'objet est connue ou peut être estimée.

Les modèles de dégradation sont des modèles qui estiment la RUL d'un objet en fonction de sa dégradation mesurée au fil du temps. Ces modèles sont basés sur l'hypothèse que la dégradation de l'objet est corrélée à sa durée de vie restante. Les modèles de dégradation sont particulièrement utiles lorsque les données de performance passées de l'objet sont disponibles et que la dégradation

de l'objet peut être mesurée de manière fiable c.-à-d. on connaît un seuil que l'indicateur de condition ne peut pas le dépasser .

Le choix du modèle à utiliser dépendra des données disponibles sur l'objet et de l'objectif de l'analyse de la RUL. Les modèles de similarité sont plus appropriés si on a des données qui varient entre l'état sain et défectueux. Les modèles de survie sont plus appropriés lorsque les données de performance passées sont disponibles à partir de moment de panes . Les modèles de dégradation sont plus appropriés lorsque les données de performance passées sont disponibles sur l'état sain et on n'a pas des données sur l'état défectueux mais on connaît un seuil qu'il ne faut pas dépasser notre indicateur de conditions.



4. Description de travail réalisé

i. Prétraitement des données

Dans tous les projets de machine learning ou d'IA, les données doivent être nettoyées afin de les rendre compatibles avec le modèle d'entraînement. Les problèmes qui nécessitent une rectification par prétraitement peuvent inclure :

- Des échantillons manquants
- Une étiquette/format incorrecte ou manquante
- Des valeurs aberrantes
- Une mise à l'échelle

L'ensemble de données fourni par le client avait déjà été nettoyé, de sorte qu'aucun traitement n'était nécessaire. Le format de l'ensemble de données était une liste de tuples. La liste comportait 50 entrées, ce qui signifiait 50 jours de données, et le tuple avait une longueur de 2, représentant une entrée pour le signal de vibration et une entrée pour le tachymètre. Les deux entrées étaient des tableaux de longueur 585936 pour vibration, et 2446 pour tachymètre, car les mesures étaient échantillonnées à une fréquence de 97658 sps pendant 6 secondes.

Le seul traitement de données requis était pour le tachymètre. Le signal représentait le nombre de radians que la turbine avait tourné depuis le début de l'échantillonnage, comme on peut le voir sur

la figure <>. Il apparaissait donc comme une ligne linéaire commençant à 0 chaque jour. Nous avons donc différencié le signal pour trouver l'accélération subie par les pales.

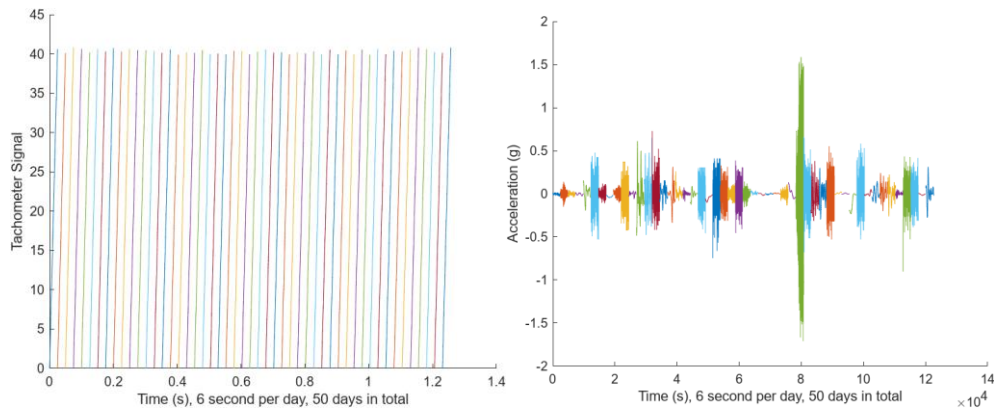


Figure 2: Tachometer signal and derived acceleration

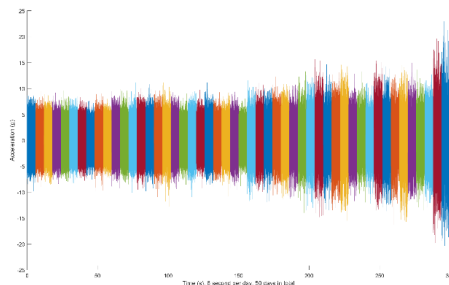


Figure 3: Vibration signal data

ii. Extraction des features

Pour l'extraction des caractéristiques, le Toolbox Diagnostic Feature Designer de Matlab a été utilisée. Ce Toolbox permet de saisir tous les échantillons dont nous disposons et de calculer les caractéristiques temporelles et fréquentielles, avec une estimation spectrale basée sur un modèle autorégressif dans le Toolbox. Il procède ensuite au classement des caractéristiques obtenues en fonction de la monotonie ou d'autres critères qui peuvent être choisis.

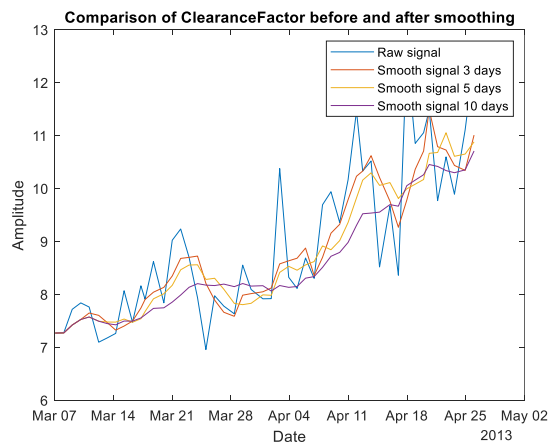


Figure 4: Moyenne mobile sur le Clearance

Une fois les caractéristiques extraites à l'aide du module, un script Matlab est généré, auquel les caractéristiques non incluses dans le module, telles que le spectral kurtosis ou l'approximation par ondelettes, sont ajoutées manuellement dans le script. D'autres features de la signal d'accélération sont ajoutés aussi. Après on procède au lissage des données avec une moyenne mobile prenant les cinq valeurs précédentes introduite dans les caractéristiques extraites. Dans la figure 4 on peut voir la différence entre prendre plus ou moins valeurs pour la moyenne mobile, les effets de changer ce valeur restent à être étudiés. On prend la moyenne de cinq jours et après on continue avec la classification des features selon différents critères.

5. Importance des caractéristiques

L'importance des caractéristiques est une technique qui assigne un score à chaque caractéristique d'un ensemble de données en fonction de la quantité d'informations que cette caractéristique apporte à l'ensemble de données. Il existe plusieurs méthodes pour déterminer l'importance des caractéristiques, mais dans ce rapport, nous allons nous concentrer sur deux : la monotonie et la corrélation de Pearson.

i. Monotonie

La Monotonie est une méthode permettant de quantifier le mérite des caractéristiques pour des fins de pronostic. Cette technique est particulièrement utile pour évaluer l'importance des caractéristiques lorsqu'il existe une relation ordonnée entre les caractéristiques et la variable cible. Par exemple, une caractéristique est dite avoir une forte monotonie si l'augmentation ou la diminution de la valeur de cette caractéristique conduit à une augmentation ou une diminution correspondante de la valeur de la variable cible.

La Monotonie de la i -ème caractéristique x_i est calculée comme suit :

$$\text{Monotonicity}(x_i) = \frac{1}{m} \sum_{j=1}^m \frac{|\text{nombre de différences positives}(x_{ji}) - \text{nombre de différences négatives}(x_{ji})|}{n - 1}$$

Où :

- n est le nombre de points de mesure. C'est la taille de l'échantillon ou le nombre total de mesures effectuées.
- m est le nombre de machines surveillées, c'est-à-dire le nombre d'observations ou de sujets étudiés.
- x_i est la i -ème caractéristique mesurée sur la j -ème machine. Chaque machine peut avoir plusieurs caractéristiques ou mesures.
- $\text{diff}(x_{ji}) = x_{ji}(t) - x_{ji}(t - 1)$ est la différence du signal x_{ji} . Il s'agit de la variation de la caractéristique au fil du temps.

Le terme 'nombre de différences positives' fait référence au nombre de fois où la valeur de la caractéristique augmente d'une mesure à la suivante, tandis que 'nombre de différences négatives' fait référence au nombre de fois où la valeur de la caractéristique diminue.

Cette formule prend la moyenne des différences absolues entre les différences positives et négatives de chaque caractéristique à travers toutes les machines surveillées, et la normalise par le nombre de points de mesure moins un. Plus la valeur de monotonie est élevée, plus la caractéristique est importante. On peut voir les résultats dans la figure 5 en bas.

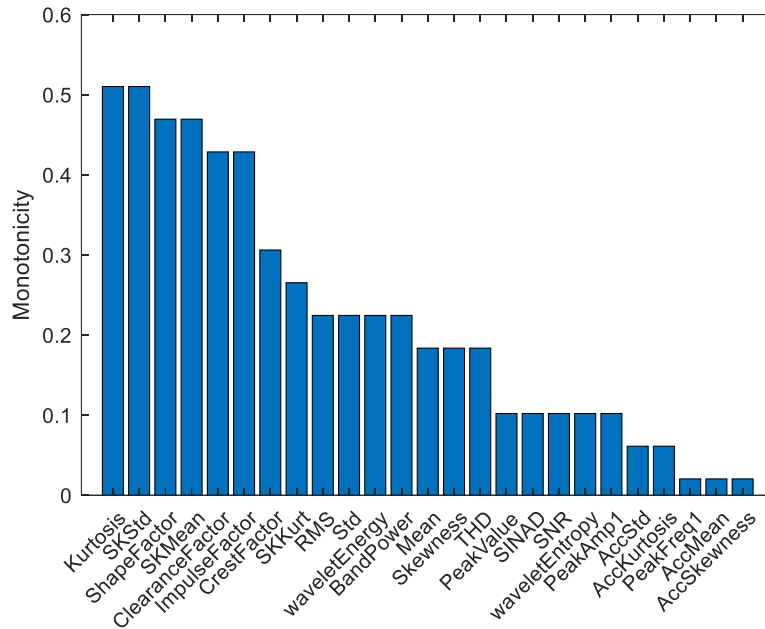


Figure 5 Monotonicity ranking

ii. Corrélation de Pearson

La corrélation de Pearson est une mesure de la relation linéaire entre deux variables. Elle varie entre -1 et 1, où 1 indique une corrélation positive forte, -1 une corrélation négative forte et 0 aucune corrélation.

La formule pour le coefficient de corrélation de Pearson est :

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 (y_i - \bar{y})^2}}$$

où:

- x_i et y_i sont les échantillons individuels indexés avec i ,
- \bar{x} et \bar{y} sont les moyennes d'échantillon de x et y , respectivement.

La corrélation de Pearson peut être utilisée pour déterminer si une caractéristique est importante en mesurant à quel point elle est corrélée avec la cible. Une corrélation plus forte suggère une caractéristique plus importante.

6. Indicateur de Condition :

Nous disposons actuellement d'une collection de caractéristiques lissées présentant une bonne monotonie. Nous cherchons à utiliser ces caractéristiques pour entraîner un modèle de dégradation exponentielle, comme expliqué dans la section suivante, mais nous cherchons d'abord à réduire la dimensionnalité du problème. Si nous cherchons à former un modèle qui prédit la durée de vie restante de l'éolienne (un simple scalaire), le fait d'avoir une entrée de données de 15 à 40 caractéristiques rendra probablement la formation plus longue et augmentera le risque de overfitting. Nous cherchons donc à créer un indicateur d'état unique, qui est un représentant numérique de l'état de santé de l'éolienne.

Lorsque nous combinons plusieurs caractéristiques, nous devons d'abord les mettre toutes à la même échelle, afin d'éviter que les caractéristiques ayant une faible valeur numérique ne soient ignorées. Pour ce faire, nous mettons à l'échelle toutes les caractéristiques entre 0 et 1.

Nous cherchons ensuite à combiner les caractéristiques ensemble en une seule valeur. La façon la plus simple de le faire est d'attribuer un poids à chaque caractéristique, correspondant à son importance perçue, puis de faire une somme pondérée. La **Erreur ! Source du renvoi introuvable.** montre l'indicateur de condition en vert, de 18 caractéristiques dérivées du signal de vibration, chacune avec un poids de 1/18. Ainsi, dans ce cas, c'est une simple moyenne.

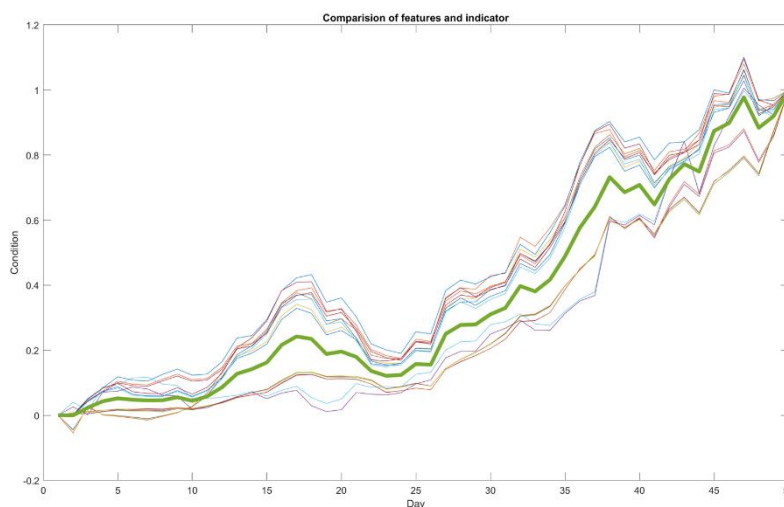


Figure 6: Condition Indicator calculated by a simple mean

Une moyenne n'est pas la meilleure façon de fusionner les données, une méthode que nous pourrions utiliser pour obtenir un indicateur plus précis serait l'analyse en composantes principales (PCA), qui cherche à projeter un ensemble de données de haute dimension dans un espace de dimension inférieure. Cependant, cette méthode n'a pas encore été implémentée dans notre code.

Voici la comparaison entre l'utilisation de Monotonie et corrélation de Pearson comme critère pour classer les caractéristiques :

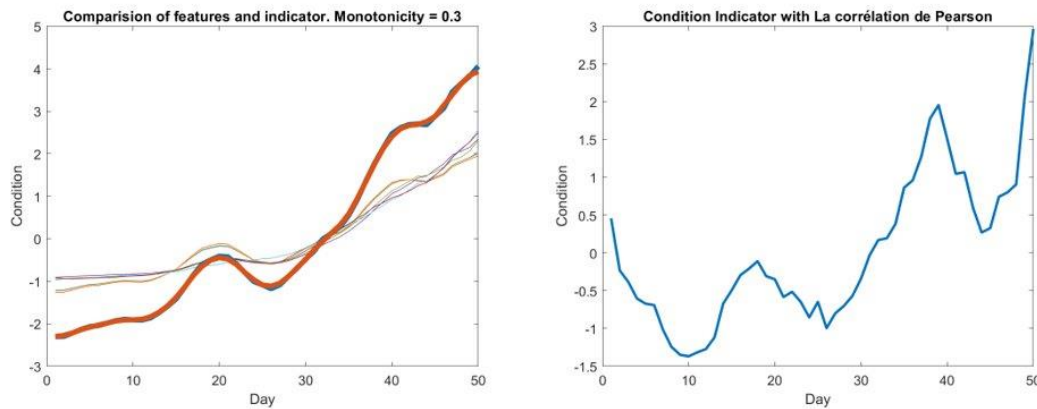


Figure 7 Comparaison du Monotonicity et correlation de Pearson

Il est observé dans les deux figures que les performances de notre indicateur de condition s'améliorent lorsque nous utilisons la Monotonie par rapport à l'utilisation de la corrélation de Pearson. Cette amélioration est attribuée à la nature croissante de notre indicateur de condition lorsqu'il est calculé à l'aide de la Monotonie. En revanche, l'utilisation de la corrélation de Pearson montre que notre indicateur de condition n'est pas monotone. Donc par la suite on se base sur la monotonie pour calculer l'importance des caractéristiques. On continue après avec une réduction de dimensionnalité des features.

iii. Réduction de Dimensionnalité

La réduction de dimensionnalité est une technique d'analyse de données qui est utilisée pour réduire le nombre de variables dans un ensemble de données. Il existe plusieurs méthodes pour effectuer une réduction de dimensionnalité, mais dans ce rapport, nous allons nous concentrer sur deux en particulier : l'analyse en composantes principales (PCA) et l'analyse factorielle (FA).

iv. Analyse en Composantes Principales (PCA)

L'Analyse en Composantes Principales (PCA) est une technique statistique utilisée pour examiner la structure globale d'un ensemble de données. L'idée derrière la PCA est de transformer un ensemble de variables possiblement corrélées en un nouvel ensemble de variables non corrélées appelées composantes principales.

La formule pour PCA est la suivante :

$$Y=XP$$

où:

- **Y** est la matrice de données transformées,
- **X** est la matrice de données d'origine,
- **P** est la matrice de chargement des composantes principales.

v. Analyse Factorielle (FA)

L'Analyse Factorielle est une méthode statistique utilisée pour décrire la variabilité entre des variables observées et corrélées en termes d'un nombre potentiellement inférieur de variables non observées appelées facteurs.

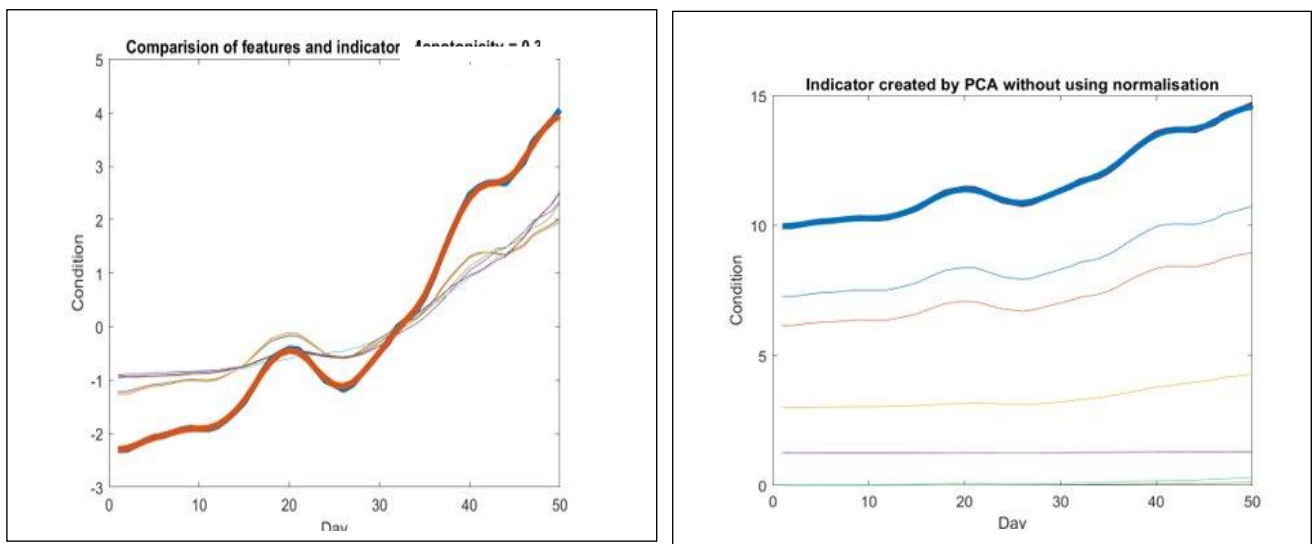
La formule pour l'Analyse Factorielle est la suivante :

$$X=LF+U$$

où:

- **X** est la matrice de données observées,
- **L** est la matrice de chargement des facteurs,
- **F** est la matrice de facteurs latents,
- **U** est la matrice de termes d'erreur.

Voici une comparaison entre l'utilisation des techniques de réduction de dimensionnalité et leur absence :



7. Modèle de Dégradation :

i. Description du modèle

Maintenant que nous disposons d'une valeur unique qui représente l'état de l'éolienne au fil du temps, nous devons concevoir un modèle de dégradation. Comme décrit dans la bibliographie, il existe de nombreux types de modèles, tels que le modèle de similarité, qui compare l'historique de dégradation actuel à tous les historiques pour voir lequel est le plus similaire, mais la plupart de ces modèles nécessitent un historique de nombreuses autres éoliennes. Cela limite les modèles que nous pouvons choisir au modèle de dégradation linéaire ou au modèle de dégradation exponentielle. Ces modèles fonctionnent en essayant d'ajuster une courbe linéaire ou exponentielle aux données, et n'ont donc besoin que des points de données de l'éolienne actuelle pour

fonctionner. Un historique de la flotte est toujours utile, car il détermine les valeurs de départ des coefficients de la courbe et peut donc accélérer la convergence. Nous avons décidé de nous concentrer sur le modèle exponentiel pour plusieurs raisons:

- Une connaissance a priori du système, qui nous indique qu'au fur et à mesure que les roulements sont exposés à davantage de vibrations, ils sont susceptibles de s'endommager plus rapidement, ce qui entraîne une augmentation plus rapide des vibrations.
- Dans notre analyse documentaire, nous avons constaté que les modèles exponentiels sont recommandés pour les systèmes présentant plusieurs modes de défaillance, tels qu'une éolienne.

L'équation du modèle est la suivante :

$$S(t) = \phi + \theta \exp\left(\beta t + \epsilon(t) - \frac{\sigma^2}{2}\right)$$

- $S(t)$ est la valeur de l'indicateur d'état au moment t
- ϕ est l'ordonnée à l'origine du modèle
- β et θ représentent la pente du modèle
- ϵ et σ représente le bruit de mesure

ii. Initialisation du modèle

Lors du choix de ces valeurs, deux options s'offrent à nous :

- Utiliser l'historique d'un parc d'éoliennes pour trouver les valeurs moyennes.
- Choisir les valeurs arbitrairement ou sur la base de l'intuition

Comme nous n'avons pas d'historique de la flotte, nous devons choisir ces valeurs à la main. Nous devons également spécifier la variance de ces valeurs, qui serait généralement calculée en prenant la variance statistique des paramètres dans l'historique de la flotte. ϕ est la valeur la plus facile à choisir car nous savons que la composante bruit a une valeur moyenne de 0.

$$\phi = S(0) - \theta \exp(0) = S(0) - \theta$$

Pour β et θ , nous les fixons tous deux à 1. C'est totalement arbitraire, mais nous n'avons aucun moyen de connaître une meilleure valeur. Pour contrer cela, nous spécifions une grande variance de 1000 pour indiquer au modèle qu'il est libre de changer beaucoup ces paramètres lorsqu'il reçoit de nouvelles données. D'autre part, nous fixons la composante bruit à $\epsilon_0 \hookrightarrow \mathcal{N}(0, \sigma_0)$ avec $\sigma_0 = 0.1\%$. Cela indique au modèle que le bruit du capteur est presque négligeable. Si nous fixons une valeur de bruit élevée, si le modèle reçoit une valeur de l'indicateur de condition qui ne correspond pas à sa prédiction, il n'ajustera pas beaucoup les valeurs de β et θ , car il pensera que l'écart est dû au bruit et que les valeurs de β et θ sont toujours correctes. Comme β et θ sont complètement arbitraires, nous ne voulons pas que le modèle pense qu'ils sont corrects, et nous voulons qu'ils soient ajustés rapidement, c'est pourquoi nous disons au modèle d'avoir plus confiance dans les données du capteur que dans les paramètres initiaux.

iii. Entraînement du modèle

Pour entraîner le modèle, nous lui donnons les valeurs de l'indicateur d'état pour chaque jour jusqu'à aujourd'hui, et la fonction matlab "exponentialDegradationModel" de la "Predictive Maintenance Toolbox" cherche à ajuster les paramètres de l'équation pour minimiser l'erreur. Nous

disposons désormais d'une expression capable de prédire l'évolution de l'état de santé de l'éolienne dans le futur. Pour prédire la RUL, nous devons savoir à quel état de santé la turbine se brisera. C'est là qu'il faut obtenir l'historique du parc de turbines ou une connaissance précise de la construction physique de la turbine. Comme nous ne disposons pas de ces informations dans l'ensemble de données actuel, nous supposons que notre turbine se casse le jour 50 (le dernier jour de l'ensemble de données), et nous prenons la valeur de l'indicateur d'état ce jour-là comme seuil de défaillance. Pour calculer le RUL, nous sous-intégrons donc cette valeur pour Y et calculons le t correspondant.

$$RUL = t^* - t_{\text{present}} \quad S(t^*) = S_{\text{threshold}}$$

Pour créer une prédiction dynamique de la RUL, il suffit de mettre à jour notre modèle avec les nouvelles données de vibration à chaque pas de temps, le modèle ajuste alors les paramètres de l'équation, et une nouvelle RUL est donnée. En règle générale, à mesure que la défaillance se rapproche, la prédiction RUL devient plus précise pour deux raisons. Premièrement, la prédiction est moins éloignée dans le temps, ce qui signifie que l'effet des paramètres incorrects est moindre, et deuxièmement, the parameters themselves become more precise, as they have been trained on more data.

Il est clair que ce modèle sera 'overfitted' à nos données, car nous utilisons en fait nos données d'entraînement pour valider le modèle (nous prenons le threshold de la même éolienne). Cependant, notre modèle pourrait être grandement amélioré en utilisant le même workflow avec même une petite amélioration, telle que la connaissance du threshold moyen de défaillance d'une centaine d'éoliennes différentes.

vi. Indicateur de performance

Maintenant que notre modèle est prêt à être testé, nous voulons définir une mesure numérique pour décrire sa performance. Un bon modèle doit présenter les caractéristiques suivantes:

- - une erreur minimale entre la durée de vie réelle et la durée de vie prévue
- - Une meilleure performance vers la fin de la vie de l'éolienne (c'est-à-dire qu'une erreur de 5 jours 2 mois avant la panne est moins importante qu'une erreur de 5 jours, 2 jours avant la panne)

Nous définissons donc un indicateur de performance:

$$score = \sum_{i=1}^n \frac{|RUL_{est}(i) - RUL_{actual}(i)|}{1 + RUL_{actual}(i)}$$

Le terme du dénominateur garantit que les jours les plus proches de la défaillance sont davantage pondérés. Comme il s'agit d'une fonction de perte, un score plus faible indique un meilleur modèle.

8. Résultats initiaux

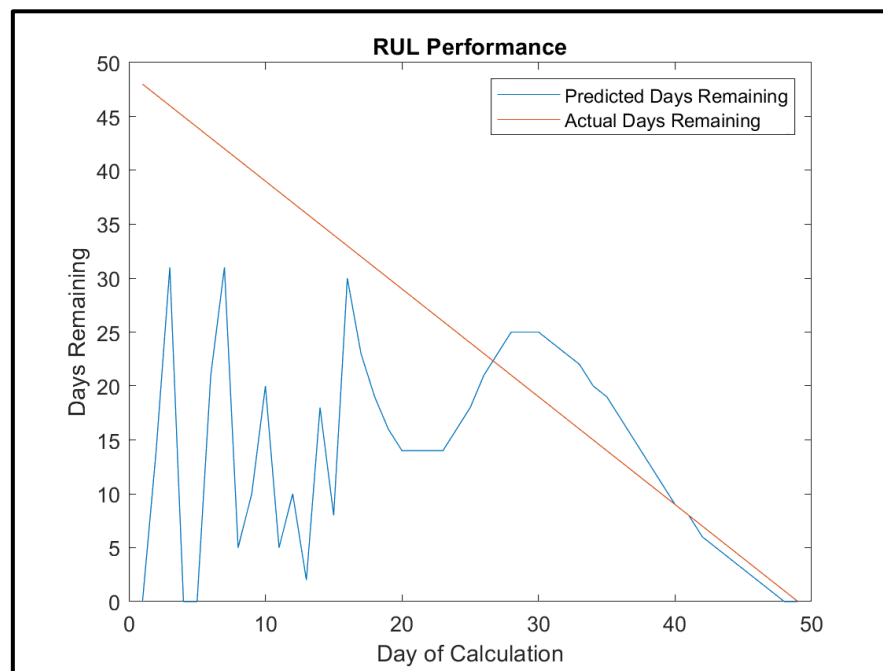


Figure 8: Modèle exponentiel premier prédiction RUL

Dans Figure 8, nous voyons les résultats de notre modèle utilisant l'ACP pour combiner les caractéristiques et utilisant un critère de sélection des caractéristiques de monotonie $= 0,3$. La ligne rouge représente le nombre réel de jours restants de l'éolienne, qui diminue de façon linéaire, puisque chaque jour la durée de vie restante diminue d'un jour, et la ligne bleue représente la RUL prédite par le modèle. Cela ne signifie pas qu'il faille utiliser un modèle de dégradation linéaire, car c'est l'indicateur d'état qui se comporte de manière exponentielle, et non la durée de vie restante.

Nous pouvons constater que pendant les 20 premiers jours, les performances du modèle sont très médiocres et que la RUL oscille fortement. Cela s'explique par le fait que les paramètres de départ n'étaient pas proches des paramètres de pente réels et qu'il faut donc un certain temps pour que les vraies valeurs convergent. Vers le 30e jour, les vraies valeurs ont été approximativement trouvées, et pour les 10 derniers jours, nous atteignons une précision de ± 1 jour. Ce comportement semble satisfaisant, car 10 jours devraient suffire à une entreprise de maintenance pour programmer un arrêt. Pour vérifier notre hypothèse selon laquelle un modèle linéaire se comporterait moins bien qu'un modèle exponentiel, nous testons également les performances d'un modèle linéaire.

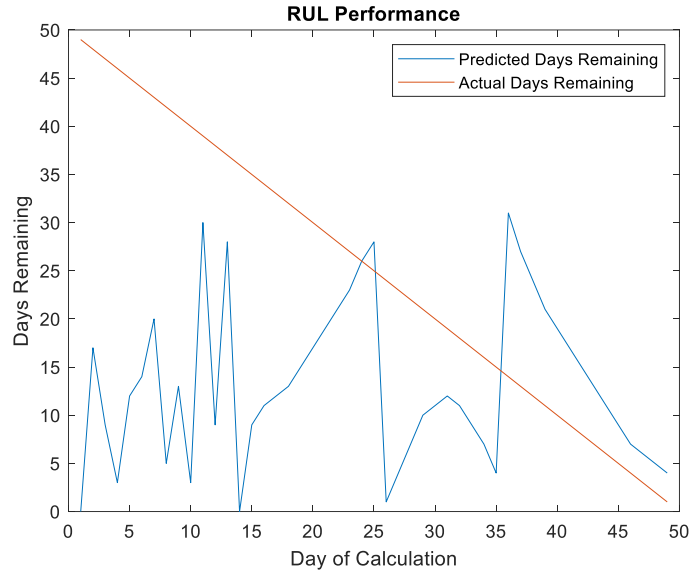


Figure 9: Performance d'un modèle de dégradation linéaire

Dans Figure 9, nous pouvons constater visuellement que la performance du modèle linéaire est moins bonne que celle du modèle exponentiel. Cependant, nous devons également valider cette constatation de manière quantitative. Le score de performance du modèle linéaire était de 42,3, alors que celui du modèle exponentiel était de 20,1, c'est pourquoi nous ne considérons que le modèle exponentiel lorsque nous passons à la mise au point du modèle.

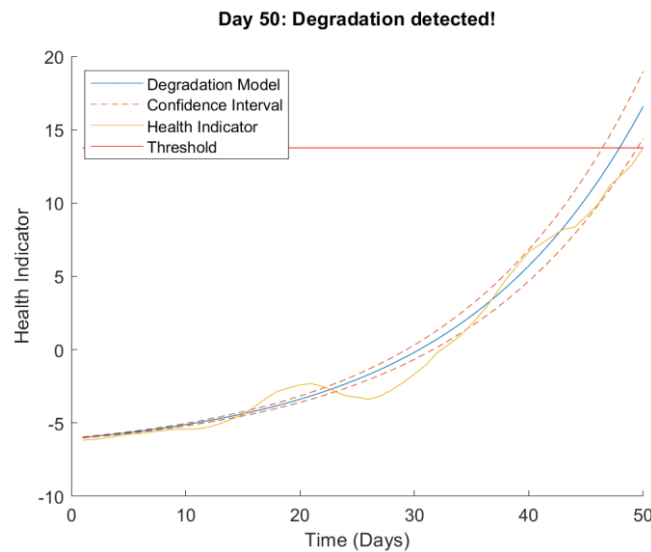


Figure 10: Évolution de l'indicateur de condition pour le premier modèle

Dans Figure 10, nous voyons comment le modèle se compare aux données réelles. La ligne bleue représente la courbe exponentielle sur laquelle le modèle a fait correspondre les données, la ligne jaune les données réelles et la ligne rouge le seuil qui, selon nous, signifie une défaillance.

L'intervalle de confiance est calculé sur la base de la variance des paramètres et du bruit du signal.

9. Tuning of the model

i. Résultats

Dans cette section, nous présentons les résultats obtenus en variant le paramètre de monotonie et en utilisant les méthodes de réduction de dimensionnalité PCA et FA. Il est important de noter que dans notre contexte, un score plus faible est meilleur.

ii. Variation de la Monotonie

Nous avons varié le paramètre de monotonie entre 0.2, 0.3, et 0.4. Pour chaque valeur de la monotonie, nous avons appliqué les méthodes PCA, FA, ainsi qu'une approche sans réduction de dimensionnalité (summation) pour comparer les performances. Voici les résultats obtenus :

Method	Monotonicity		
	0.2	0.3	0.4
PCA	39.8411	20.1490	20.8685
FA	24.7111	19.6523	18.3778
Summation	26.2291	20.8276	19.3125

Il est clair que la performance varie en fonction du paramètre de monotonie et de la méthode de réduction de dimensionnalité choisie. Pour une monotonie de 0.2, la méthode FA semble offrir de meilleures performances par rapport à la PCA. Cependant, pour des valeurs plus élevées de la monotonie (0.3 et 0.4), les performances des deux méthodes sont assez proches. Fait intéressant, l'approche sans réduction de dimensionnalité semble également offrir des performances comparables.

Ces résultats soulignent l'importance de la sélection des paramètres dans le processus de modélisation. Ils montrent également que différentes techniques de réduction de dimensionnalité peuvent être optimales pour différents paramètres de monotonie. Enfin, ils suggèrent que dans certains cas, une approche sans réduction de dimensionnalité peut être préférable.

10. Déploiement et perspectives

Maintenant que nous disposons d'un modèle performant, nous devons réfléchir à la manière dont il peut être déployé dans la réalité. Un modèle distinct sera utilisé pour chaque éolienne, car chaque profil de dégradation est unique. Ce modèle pourrait fonctionner sur un ordinateur hébergé dans chaque éolienne, mais comme la dérivation des caractéristiques nécessite une bonne dose d'analyse

informatique, il est plus probable que les données brutes du signal soient téléchargées vers un serveur central qui peut effectuer le calcul, mais cela dépend de la vitesse d'accès à l'internet disponible sur le site.

Lorsqu'une nouvelle éolienne est mise en service, le modèle est initialisé avec les paramètres moyens des éoliennes existantes, y compris leur variance, ce qui permet d'accélérer la convergence sur les paramètres spécifiques de cette éolienne. Le modèle fonctionne ensuite comme décrit précédemment dans le rapport, en calculant une valeur de RUL. Cette valeur RUL en temps réel peut ensuite être transmise à un optimiseur de programme, qui peut planifier le programme de travail optimal en tenant compte de la disponibilité de la main-d'œuvre, du prix futur prévu de l'énergie et des conditions météorologiques prévues. Ces facteurs sont importants, car il peut être plus économique pour l'entreprise d'effectuer la maintenance un jour calme, puisque la turbine ne pourrait de toute façon pas produire d'énergie, ainsi que les week-ends ensoleillés, lorsque la demande d'énergie est la plus faible et que le prix diminue en raison de l'excédent d'énergie solaire.

En cas de défaillance, la valeur de l'indicateur d'état au moment de la défaillance sera enregistrée, ce qui peut ensuite être utilisé pour créer une distribution de probabilité du risque de défaillance en fonction de la valeur du seuil, ce qui peut permettre à l'entreprise de décider quel est le niveau de risque acceptable. Par exemple, le choix d'un seuil en dessous duquel seulement 1/1000 turbines tombent en panne évitera les pannes, mais augmentera le coût de la maintenance. La RUL peut également être utilisée pour arrêter la turbine. Cela est dû à la différence de coût entre la maintenance corrective et la maintenance préventive, c'est-à-dire qu'il peut être beaucoup moins coûteux de remplacer un roulement usé que de laisser le roulement se casser et causer des dommages mécaniques à l'arbre et aux pales. Dans ce scénario, si la RUL prévoit que l'éolienne se brisera avant que la maintenance ne puisse être effectuée, il peut être préférable de désactiver l'éolienne depuis la salle de contrôle.

11. Conclusion

En conclusion, ce projet a été couronné de succès. En travaillant avec les signaux de vibration et de tachymétrie fournis par notre client, Mathworks, nous avons pu extraire des caractéristiques pertinentes à l'aide de l'outil Diagnostic Feature Designer de Matlab. Ces caractéristiques ont été agrégées pour créer des indicateurs conditionnels, permettant ainsi de choisir le modèle de dégradation approprié.

Grâce à l'entraînement du modèle et à l'analyse des résultats, nous avons pu développer un modèle d'estimation du temps de vie restant (RUL) d'une manière précise. En effet, notre modèle présente une précision élevée, avec une marge d'erreur de seulement +/- 1 jour. Cette précision accrue permet une planification plus efficace des opérations de maintenance, évitant les pannes inattendues et les coûts associés.

Un aspect essentiel de notre modèle est sa robustesse. Il nécessite peu de données pour fonctionner, ce qui le rend adaptable à différentes conditions et situations. De plus, notre modèle est rapide, ce qui permet une réponse rapide aux changements détectés dans les signaux de vibration et de tachymétrie. En utilisant des données existantes, il est même possible d'améliorer les prédictions initiales, ce qui réduit encore davantage le temps nécessaire pour obtenir des résultats fiables.

En termes d'économie de coûts, notre modèle se révèle être suffisamment précis et rapide pour prévenir les défaillances majeures. Cela permet d'éviter les coûts élevés liés aux réparations d'urgence, aux arrêts prolongés de production et aux pertes de revenus. Grâce à notre modèle, les opérations de maintenance peuvent être planifiées de manière proactive, ce qui entraîne des économies significatives à long terme.

La flexibilité de notre modèle est également un avantage majeur. Il est capable de fonctionner efficacement avec différents profils d'éoliennes, ce qui le rend adaptable à diverses configurations et environnements. Cette capacité à s'adapter facilite son déploiement dans des parcs éoliens de grande envergure, où les besoins de maintenance peuvent varier considérablement.

Enfin, notre modèle présente également un aspect pédagogique important. En fournissant un workflow clair et transparent pour l'intelligence artificielle appliquée à la maintenance prédictive, il offre une opportunité d'apprentissage et de compréhension approfondie des concepts sous-jacents. Cela permet aux utilisateurs de mieux appréhender les avantages et les limites de l'intelligence artificielle dans le domaine de la maintenance industrielle.

En résumé, le modèle de maintenance prédictive que nous avons développé offre une valeur ajoutée significative. Sa précision, sa robustesse, sa rapidité, son économie de coûts, sa flexibilité et son aspect pédagogique en font un outil puissant pour la prévention des pannes et l'optimisation des opérations de maintenance. Nous sommes convaincus que notre modèle contribuera à améliorer considérablement la fiabilité et la rentabilité des parcs éoliens et d'autres infrastructures industrielles similaires.