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



Une image contenant logo

Description générée automatiquement

# Equipe :

Diego Ruiz Ponsoda

Ben Silva

Valentin Gérard

Ayoub Ennajah

Mohammed Ahlal

# Encadrants :

Wassila Ouerdane (CentraleSupélec)

Jean-Philippe Poli (CentraleSupélec)

Lama Itani (MathWorks)

# **Sommaire :**

#add context of the problem

# **Introduction :**

#add information about the dataset such as format of data, plots of raw data, deliverables

# **État de l’art :**

**#add references/sources – RUL is not mentioned yet**

# 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 on 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 pièces.

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](https://doi.org/10.1109/IIC.2015.7150869)).

Surveillance vibratoire et maintenance prédictive

**Enjeux de la maintenance - some of this should be in introduction/context, not bibliography**

* **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.

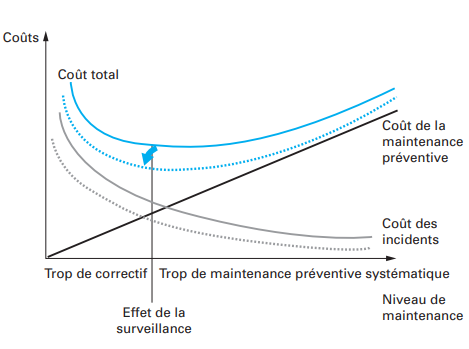


Figure - Coût en fonction de la politique de maintenance exploitée

# **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**

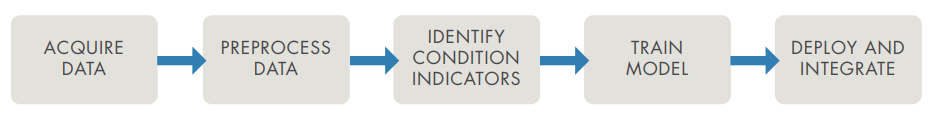
La maintenance prédictive nécessite un diagnostic minimal des défauts et de leur gravité. La première étape d'une action de surveillance consiste donc à déterminer les défauts susceptibles de survenir sur la machine à surveiller. La seconde étape concerne les manifestations de ces défauts. Il est important 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.

# 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.

# 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.

Une image contenant diagramme

Description générée automatiquement

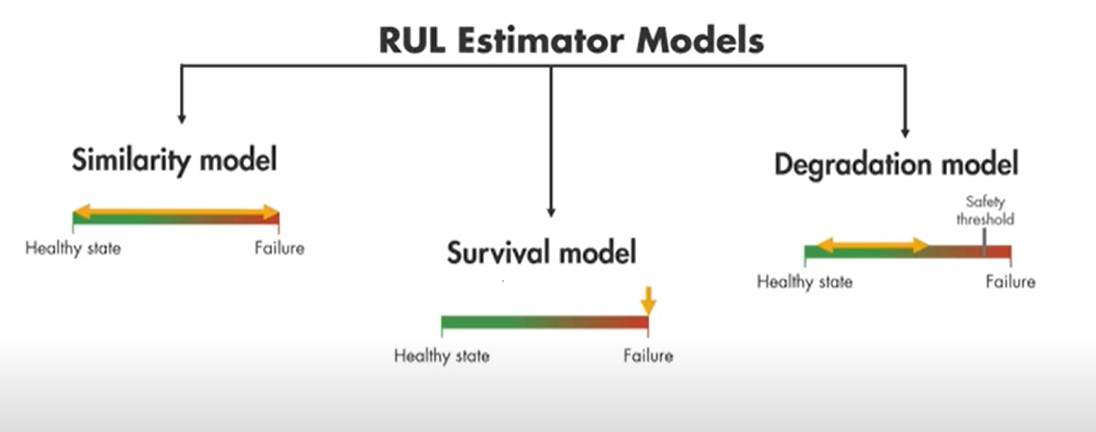
Les modèles de prédiction de la RUL peuvent être classés en trois catégories principales : les modèles de similarité, les modèles de survie et les modèles de dégradation. Chacun de ces modèles est utilisé pour estimer la RUL d'un objet en utilisant différentes sources de données et en utilisant des approches 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 connait 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 donnés 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 donnés sur l’état défectueux mais on connait un seuil qu’il faut pas dépasser notre indicateur de conditions



# **Description de travail réalisé :**

# Data Pre-Processing :

In all machine learning or AI projects, the data must be cleaned in order to make it compatible with the training model. Issues required to be rectified by pre-processing can include:

* Missing samples
* Incorrect or missing labels/formatting
* Outliers
* Scaling

The dataset provided to us by the client was already cleaned and so no processing was required. The format of the dataset was a list of tuples. The list was 50 entries long, signifying 50 days of data, and the tuple was length 2, representing one entry for the vibration signal and one entry for the tachometer. The two entries were arrays of length <>, as the measurements were sampled at rate of <> for 6 seconds.

The only data processing required was for the tachometer. The signal represented how many radians the turbine had rotated since the start of the sampling, as seen in the figure <>, it appeared as just a linear line starting from 0 each day. We therefore differentiated the signal to find the acceleration experienced by the blades.

# Feature extraction :

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. Il procède ensuite au classement des caractéristiques obtenues en fonction de la monotonicité ou d'autres critères qui peuvent être choisis.

Lors d'une première approche, le module a été utilisé sans modification préalable des échantillons, ce qui a conduit à un classement des caractéristiques qui n'était pas le meilleur que nous pouvions obtenir car les échantillons contenaient du bruit à haute fréquence. Ceci est clairement visible avec l'expression de la monotonicité :

Où M est le nombre de systèmes surveillés (M = 1), est le nombre de mesures dans le système et est le vecteur de mesures des caractéristiques sur le système. Il est facile de voir comment le bruit peut affecter la monotonicité de chaque caractéristique, de sorte qu'une moyenne mobile prenant les cinq valeurs précédentes est introduite dans les caractéristiques extraites avant de les classer selon la monotonicité.

Chart, line chart

Description automatically generated

Figure 2 - Comparison between Clearance Factor before and after smoothing

Après le lissage des données, celles-ci sont classées en fonction de la monotonicité, pour ensuite prendre les caractéristiques les plus importantes. Un seuil de 0,3 est choisi pour sélectionner ces caractéristiques. Une fois les principales caractéristiques choisies, une analyse PCA est envisagée pour réduire la dimensionnalité.

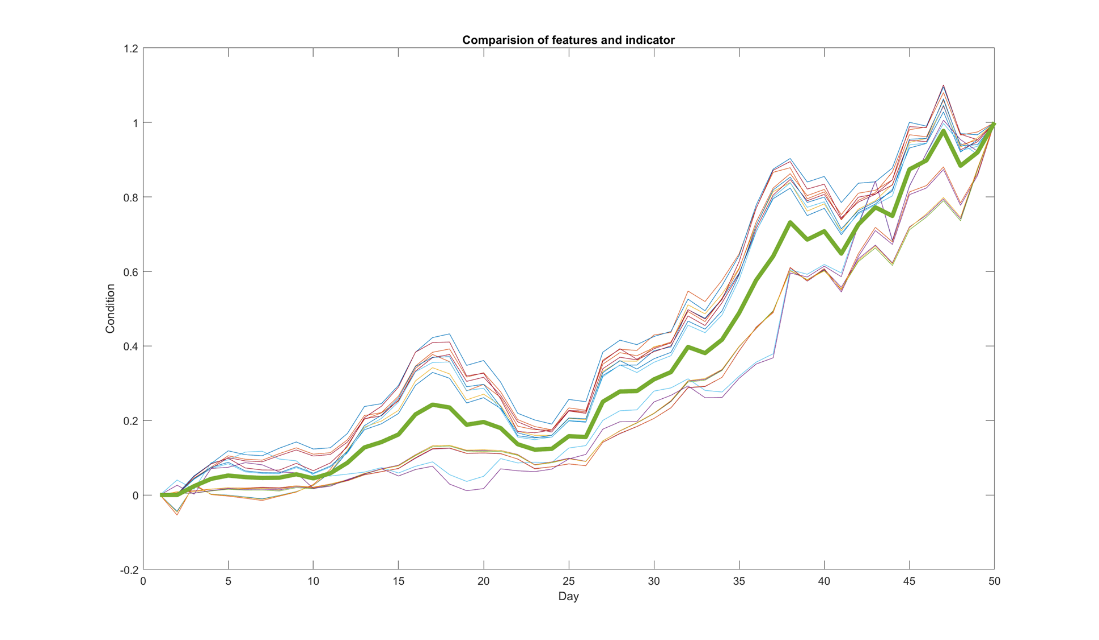
Une analyse de la quantité de caractéristiques prises en compte pour prédire le RUL devrait être effectuée. D'autres travaux concernant les caractéristiques du signal dans le domaine des fréquences restent à faire, car le module Matlab utilisé comporte moins de caractéristiques que les caractéristiques dans le domaine temporel (par exemple, spectral kurtosis).

# 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 figure 3 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 3: Indicateur de condition calculé par moyenne*

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.

# Modèle de Dégradation:

En raison de la limitation de notre ensemble de données, nous ne sommes pas en mesure d'utiliser des méthodes telles que celle du "plus proche voisin", qui compare le signal de vibration de l'éolienne à des signaux similaires provenant d'autres éoliennes, afin de tenter de prédire la progression future. Par conséquent, la seule méthode que nous pouvons utiliser est le modèle de dégradation exponentielle. Ce modèle postule qu'au fur et à mesure que la santé de l'éolienne se dégrade, le taux de dégradation augmente. L'utilisation de ce modèle se justifie de deux manières. La première est visuelle, nous pouvons voir dans la Figure <> que l'amplitude des vibrations n'augmente pas linéairement avec le temps, mais semble avoir une certaine augmentation exponentielle. La seconde provient d'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.

L'équation du modèle est la suivante:

Y(t) = β \* e^(-αt) + ε(t)

* Y(t) est le niveau de santé ou de dégradation de l'éolienne à l'instant (t)
* β est l'état de santé initial à t = 0
* α est le taux de dégradation de l'état de santé
* ε représente le bruit de mesure

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\* - tpresent Y(t\*) = Ythreshold

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, à mesure que l'état se détériore, alpha et bêta tendent à devenir plus significatifs par rapport au bruit du système.

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.

# Results

# Next Steps