

# Table des figures

Figure 1:Les Types de Maintenance.....	7
Figure 2:moteur à turbines .....	10
Figure 3:composants d'un turbo-fan .....	10
Figure 4: captures d'écran des données .....	12
Figure 5:Fonction de défaillance $F(t)$ .....	14
Figure 6:les fonction de fiabilité $R(t)$ et de défaillance $F(t)$ .....	14
Figure 7:Distribution des temps de défaillance (failure time) par moteur.....	18
Figure 8:les fonctions de densités (pdf) ajustées pour chaque distribution .....	19
Figure 9:Les résultats du test de Kolmogorov-Smirnov (KS).....	20
Figure 10:résultats de test chi-deux .....	21
Figure 11:résultats AIC et BIC pour chaque loi ajusté.....	22
Figure 12:Comparaison des données simulées et réelles.....	24
Figure 13:densité de probabilité $f(t)$ - Log Normal .....	25
Figure 14:Fonction de défaillance $F(t)$ et de fiabilité $R(t)$ - Log Normal .....	25
Figure 15:Taux de défaillance $\lambda(t)$ - Log Normal .....	26
Figure 16:Évolution des erreurs en fonction de la fiabilité souhaitée.....	29
Figure 17:Évolution du $R^2$ en fonction de la fiabilité souhaitée .....	29
Figure 18:Comparaison des données simulées (par distribution Gamma) et réelles.....	31
Figure 19:Fonction de densité de probabilité $f(t)$ (pour la loi Gamma ) .....	31
Figure 20:Fonction de défaillance $F(t)$ et de fiabilité $R(t)$ (pour la loi Gamma ) .....	32
Figure 21:Taux de défaillance ( $\lambda(t)$ ) ( pour la loi Taux de défaillance ( $\lambda(t)$ ) ( pour la loi Gamma) .....	32
Figure 22:l'évolution de ces métriques en fonction de la fiabilité choisie pour la distribution Gamma.....	33
Figure 23:Comparaison entre la durée de vie restante estimée et réelle (RUL) - loi Log-Normal .....	33
Figure 24:Évolution de la durée de vie restante estimée et réelle par moteur - loi Log-Normal.....	34
Figure 25:Figure 23:Comparaison entre la durée de vie restante estimée et réelle (RUL) - loi Gamma .....	34
Figure 26:Figure 24:Évolution de la durée de vie restante estimée et réelle par moteur - loi Gamma.....	34
Figure 27:métriques d'erreur pour les lois Gamma et Log-Normale à une fiabilité de 0.57 .....	35
Figure 28:Valeurs Manquantes, Unicité et Types de données .....	37
Figure 29:aperçu des colonnes restantes .....	37
Figure 30:Scores de Monotonie par Colonne .....	39
Figure 31:Scores de Tendabilité par Colonne .....	39
Figure 32:Scores de Prognosabilité par Colonne .....	40
Figure 33:la la matrice de corrélation entre les colonnes .....	41
Figure 34:corrélations élevées (seuil > 0.9) .....	41
Figure 35:Température à la sortie du compresseur basse pression (LPC).....	42
Figure 36:Température à la sortie du compresseur haute pression (HPC) .....	42
Figure 37:Température à la sortie de la turbine basse pression (LPT) .....	42
Figure 38:Pression à la sortie du compresseur haute pression (HPC) .....	42
Figure 39:Vitesse physique du ventilateur.....	43
Figure 40:Vitesse physique du noyau .....	43
Figure 41:Pression statique à la sortie du compresseur haute pression (HPC) .....	43
Figure 42:Rapport du débit de carburant à Ps30.....	43
Figure 43:Vitesse corrigée du ventilateur .....	44
Figure 44:Rapport de dérivation.....	44
Figure 45:Enthalpie du prélèvement d'air .....	44

Figure 46:Flux d'air de refroidissement des turbines haute pression .....	44
Figure 47:Flux d'air de refroidissement des turbines basse pression.....	45
Figure 48:composantes principales .....	47
Figure 49:les composants pC1 et PC2 en fonction de RUL .....	47
Figure 50:Évolution de l'indicateur de santé (PC1) par cycle .....	48
Figure 51:Indicateurs de Santé pour Plusieurs Moteurs .....	48
Figure 52:Distribution de l'Indicateur de Santé (pc1) selon les Niveaux de RUL.....	49
Figure 53:les paramètres A, B, et $\lambda$ pour chaque moteur .....	50
Figure 54:l'indicateur de santé observé et prédit pour un moteur donné (moteur 23) .....	50
Figure 55: Comparaison des Prédictions de RUL pour Quelques Moteurs de Test .....	52
Figure 56:Comparaison entre l'Indicateur de Santé Réel et la Dégradation Prédictive pour le Moteur 84 .....	52
Figure 57: Comparaison entre la durée de vie restante((RUL) Prédit et réelle .....	53
Figure 58: Comparaison entre la durée de vie restante (RUL) Prédit et réelle pour Chaque Moteur .....	53
Figure 59: Évaluation des Performances du Modèle : RMSE, RAE et R2R^2R2 .....	53
Figure 60:RUL Prédit et Réel pour un Moteur avec Plus de 200 Cycles.....	54
Figure 61:Prédit et Réel pour un Moteur avec moins de 30 Cycles.....	54
Figure 62:Indicateur de Santé Initial pour les moteur 70 à 100 .....	56
Figure 63:Indicateur de Santé Final pour les Unités 71 à 100 .....	57
Figure 64:Indicateur de Santé Final et Points de Défaillance .....	57
Figure 65:Comparaison de l'Indicateur de Santé Réel et Ajusté pour différentes moteurs.....	58
Figure 66 :5 Modèles les Plus Similaires pour un moteur de test .....	59
Figure 67:évolution de <b>R2</b> et RMSE en fonction de K .....	60
Figure 68: Comparaison entre la durée de vie restante((RUL) Prédit et réelle .....	60
Figure 69: Comparaison entre la durée de vie restante (RUL) Prédit et réelle pour Chaque Moteur .....	61
Figure 70: Évaluation des Performances du Modèle : RMSE, RAE et R2 .....	61
Figure 71: Support Vector Machine (SVM) .....	64
Figure 72:Illustration de l'algorithme Random forest .....	65
Figure 73:comparaison des Performances des trois Modèles.....	66
Figure 74: Comparaison entre la durée de vie restante(RUL) Prédit par SVM et réelle.....	66
Figure 75: Comparaison entre la durée de vie restante (RUL) Prédit par SVM et réelle pour Chaque Moteur .....	66
Figure 76:confusion matrix .....	68
Figure 77:Matrice de confusion .....	69
Figure 78:métriques du modèle pour chaque classe.....	69



## Introduction Générale

Dans un monde industriel en constante évolution, la maintenance des équipements critiques est devenue essentielle pour assurer la sécurité, la disponibilité et la rentabilité des opérations. Les moteurs turbo-fan, utilisés dans les systèmes de propulsion aéronautique, illustrent parfaitement cette réalité. Ces équipements sophistiqués, soumis à des contraintes mécaniques et environnementales intenses, nécessitent une gestion rigoureuse de leur cycle de vie afin d'éviter les pannes coûteuses et les arrêts imprévus.

La maintenance industrielle a connu une transformation majeure avec l'émergence de l'industrie 4.0 et des technologies modernes telles que l'intelligence artificielle (IA) et le machine learning. Ces outils permettent de passer d'une maintenance réactive, où l'intervention a lieu après la panne, à une maintenance prédictive, où les défaillances sont anticipées grâce à des modèles basés sur l'analyse des données. Cette transition offre aux entreprises un avantage stratégique en améliorant la disponibilité des équipements, en réduisant les coûts d'exploitation, et en augmentant la fiabilité globale des systèmes.

Ce projet s'inscrit dans cette dynamique et vise à développer une solution de maintenance prédictive pour les moteurs turbo-fan, en utilisant le jeu de données CMAPSS (Commercial Modular Aero-Propulsion System Simulation) fourni par la NASA. Ce jeu de données riche en informations réelles offre une base solide pour analyser les performances des moteurs et prédire leur durée de vie restante (Remaining Useful Life - RUL). L'objectif principal est d'exploiter les lois statistiques et les modèles d'apprentissage automatique pour anticiper les pannes et optimiser les stratégies de maintenance.

Pour assurer une compréhension complète et méthodique, ce document est organisé de manière à guider le lecteur à travers les différentes étapes du projet. Après avoir exploré les types de maintenance et présenté les données utilisées, nous détaillerons les **analyses statistiques** et les **lois sous-jacentes**. Par la suite, des modèles de **machine learning** seront développés pour **prédir le RUL**, avec une évaluation minutieuse de leur performance. Enfin, les **résultats** obtenus seront illustrés à l'aide de visualisations claires, accompagnées de recommandations pratiques pour intégrer ces solutions dans des stratégies de maintenance robustes et efficaces. Une attention particulière sera également accordée aux **contrats de maintenance**, en détaillant le **Plan de Continuité d'Activité (PCA)** et le **Plan de Reprise d'Activité (PRA)** en vue de garantir une gestion optimale et une continuité des opérations en cas d'imprévus.

En conclusion, ce projet vise à mettre en avant la valeur ajoutée des technologies modernes pour relever les défis de la maintenance prédictive. Il propose une approche structurée et des outils concrets pour anticiper les défaillances, optimiser les opérations industrielles et contribuer à la durabilité et à la performance des équipements critiques.

# Chapitre 1 : Concepts Fondamentaux de la Maintenance et Contexte du Projet

## Introduction

Ce chapitre explore les concepts fondamentaux de la maintenance, en abordant les différents types de maintenance (corrective, préventive et prédictive). Il met en lumière l'importance croissante de la maintenance prédictive, particulièrement dans le cadre de l'Industrie 4.0, où l'intégration des technologies avancées comme l'IoT et l'intelligence artificielle joue un rôle central. Enfin, il introduit le contexte de notre projet, qui utilise les données du système CMAPSS de la NASA pour prédire la durée de vie restante (RUL) des moteurs turbo-fan, un élément clé dans l'optimisation des stratégies de maintenance prédictive.

### I. Compréhension de la Maintenance : Concepts, Importance et Typologie

#### 1. Définition de la Maintenance

Selon la norme **AFNOR** (Association Française de Normalisation) **NF-X 60 000**, la **maintenance** est « *l'ensemble des activités ayant pour objectif de maintenir ou rétablir un bien dans un état spécifié de fonctionnement pour établir une fonction requise.* ».

La maintenance rassemble donc l'ensemble des actions visant à maintenir ou à restaurer un équipement dans un état opérationnel, afin qu'il puisse remplir la fonction pour laquelle il a été conçu. Cela inclut une série d'opérations techniques destinées à prévenir, détecter et réparer les pannes des équipements de production. Ces opérations peuvent concerner des machines, des systèmes, des infrastructures ou des installations complexes. La maintenance est donc essentielle pour garantir la performance, la fiabilité et la sécurité des équipements industriels. Selon l'**ISO 55000**, la maintenance comprend à la fois des actions de surveillance et des interventions telles que la réparation, l'ajustement, le nettoyage et le remplacement de pièces.

#### 2. Importance de la Maintenance

##### Impact sur la disponibilité des équipements :

La maintenance assure que les équipements sont disponibles au moment où ils sont requis, ce qui favorise une production plus efficace. Une maintenance préventive soigneusement planifiée aide à identifier et à résoudre les problèmes avant qu'ils ne provoquent des pannes inattendues, diminuant ainsi les interruptions et garantissant un fonctionnement optimal des machines.

##### Réduction des coûts liés aux pannes et aux arrêts de production :

Les coûts associés à une panne d'équipement peuvent être très élevés, non seulement en ce qui concerne les réparations, mais aussi en termes de perte de production et de temps d'arrêt. La maintenance préventive et prédictive s'avère particulièrement efficace pour réduire ces coûts en détectant les problèmes avant qu'ils ne se transforment en défaillances majeures. D'après une étude

de l'American Society of Mechanical Engineers (ASME), les coûts d'une panne imprévue peuvent atteindre jusqu'à huit fois ceux de la maintenance préventive. En anticipant et en planifiant les interventions, les entreprises peuvent ainsi éviter des dépenses considérables et maintenir un niveau de production stable.

### Amélioration de la sécurité et de la qualité des produits :

La maintenance joue un rôle crucial dans la sécurité des équipements et des travailleurs, en minimisant les risques liés à des pannes inattendues. En veillant à ce que les équipements opèrent selon les spécifications, elle assure également la qualité des produits fabriqués. Par conséquent, la maintenance contribue à maintenir des normes élevées de sécurité et de qualité tout au long du processus de production.

## 3. types de maintenance

La norme **NF-X 60 000** différencie deux grandes familles : la *maintenance corrective* et la *maintenance préventive*.

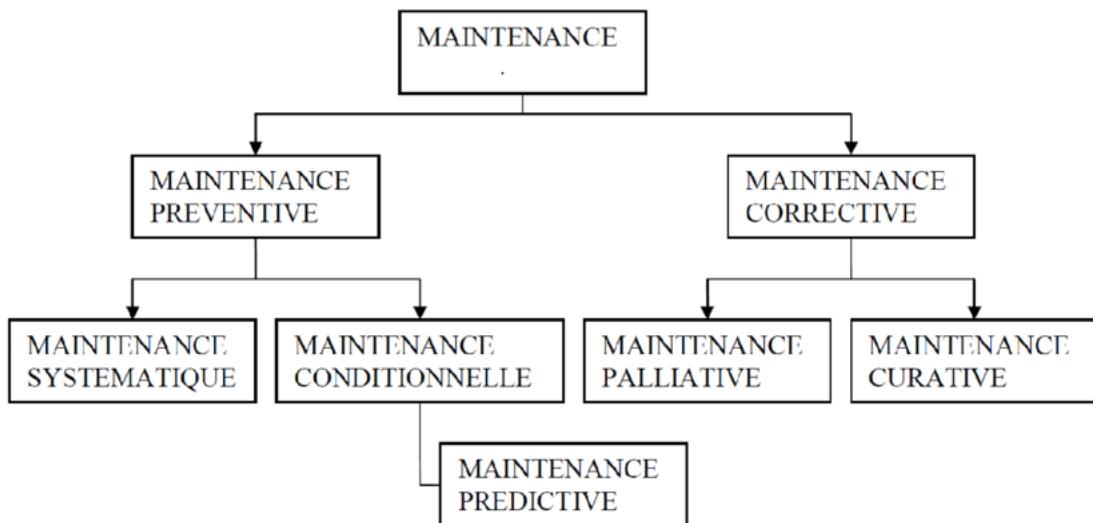


Figure 1:Les Types de Maintenance

### a) Maintenance Corrective

Définition **AFNOR** de la maintenance corrective : « *Maintenance exécutée après détection d'une panne et destinée à remettre un bien dans un état dans lequel il peut accomplir une fonction requise* ».

La maintenance corrective correspond à une opération engagée suite à une panne, un sinistre ou un aléa. Elle a pour objectif de remettre en fonctionnement les machines par un dépannage, une réparation ou le remplacement de pièces défectueuses. Elle se divise en deux sous-catégories :

#### Palliative :

La **maintenance palliative** correspond au **dépannage**. Elle consiste à remettre provisoirement une machine ou un outil en état, dans l'attente de sa réparation. La maintenance palliative permet d'éviter l'arrêt totale de production, on parle alors de reprise de production en mode « dégradé ».

**Dépannage** : actions physiques exécutées pour permettre à un bien d'accomplir sa fonction requise pendant une durée limitée jusqu'à ce que la réparation soit exécutée.

Exemple : Lorsqu'un tuyau d'arrosage est percé, une maintenance palliative consiste à réparer provisoirement la fuite avec du ruban adhésif.

### Curative :

La maintenance curative répare les causes et conséquences de la panne. Contrairement à la maintenance palliative, il s'agit d'une action en profondeur qui agit sur le long terme, souvent en remplaçant la pièce défectueuse par une neuve. L'équipement reprend alors une production normale.

**Réparation :** actions physiques exécutées pour rétablir la fonction requise d'un bien en panne.

Exemple : L'action de maintenance curative sera de changer le tuyau d'arrosage percé par un nouveau tuyau.

### b) Maintenance Préventive

Définition AFNOR de la maintenance préventive : « *Maintenance exécutée à des intervalles prédéterminés ou selon des critères prescrits et destinée à réduire la probabilité de défaillance ou la dégradation du fonctionnement d'un bien* ».

La **maintenance préventive** est donc, comme son nom l'indique, de la prévention par l'entretien régulier d'une machine ou d'un bien. Elle consiste à faire des contrôles selon des normes établies et/ou conformément aux instructions du fabricant, mais également en fonction de l'expérience humaine et des historiques d'interventions réalisés sur le bien.

La maintenance préventive a pour but de :

- garantir la disponibilité du bien,
- augmenter sa durée de vie,
- diminuer la probabilité des pannes et donc d'actions de maintenance corrective,
- contrôler les consommations d'énergie ou de pièces détachées en contribuant à la maîtrise des stocks.

On distingue plusieurs types de maintenance préventive :

#### Systématique :

La **maintenance préventive systématique** est effectuée selon un **échéancier**, en fonction du temps ou du nombre d'unités d'usage, mais **sans contrôle préalable** du bien.

La fréquence des interventions est établie à compter de la mise en fonctionnement du bien ou à la suite d'une révision partielle ou complète de celui-ci.

Exemple : Le changement d'huile sur une machine peut être défini au bout d'un certain nombre de semaines d'utilisation sans contrôle spécifique préalable.

#### Conditionnelle :

La **maintenance préventive conditionnelle** est basée sur une **surveillance** du bien en fonction d'un type d'évènement prédéfini (auto-diagnostique, information de capteurs...). Elle fait intervenir des informations recueillies en temps réel, au moyen de techniques comme la tribologie ou la thermographie par infra-rouge, avec pour objectif d'intervenir au moment où le bien a franchi le **seuil d'usure ou de dégradation** fixé préalablement. Il s'agit d'effectuer la bonne opération au bon moment, pour une meilleure maîtrise des coûts.

Exemple : L'analyse régulière des huiles d'une machine permet de détecter les pollutions, les usures de l'équipement et la qualité du lubrifiant, et de changer le lubrifiant lorsqu'il ne répond plus aux critères pour un bon fonctionnement.

## Prévisionnelle :

**La maintenance préventive prévisionnelle** (parfois appelée **prédictive**) est d'après la définition AFNOR « une maintenance conditionnelle exécutée en suivant les prévisions extrapolées de l'analyse et de l'évaluation de paramètres significatifs de la dégradation du bien ».

Elle consiste à **anticiper les besoins** de la maintenance en permettant de prévenir les pannes ou les défaillances. C'est la **maintenance 4.0**, avec une **surveillance en continu** des évolutions de certains paramètres, des seuils, des différentes data (**logiciel GMAO**, etc...), permettant une gestion précise pour déclencher la bonne intervention en fonction de l'ensemble des paramètres.

Un logiciel de **gestion de la maintenance** (GMAO) est une bonne solution informatique pour aider à un suivi efficace des interventions, récupérer des données fiables et élaborer son plan de maintenance ! Plus de prévisionnel et moins de correctif à la clé !

Dans notre projet, nous nous inscrivons pleinement dans cette démarche de maintenance préventive prévisionnelle, en utilisant des données issues de capteurs et des techniques avancées de modélisation pour anticiper les défaillances des moteurs turbo-fan et optimiser les interventions de maintenance

## II. Contexte du Projet et Données Utilisées

### 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). Bien que publié il y a plus d'une décennie, le dataset CMAPSS de la NASA, simulant la dégradation des moteurs turbo-fan, reste à la fois populaire et pertinent. À titre d'exemple, plus de 90 nouvelles publications de recherche ont vu le jour en 2020. Dans le cadre de notre projet, nous exploitons ce dataset pour développer et valider des modèles innovants capables de prédire la Durée de Vie Résiduelle (RUL - Remaining Useful Life) des moteurs turbo-fan, conformément aux objectifs définis : identifier les lois statistiques régissant l'apparition des pannes et anticiper les défaillances futures.

### 2. moteurs turbo-fan (moteur à turbines)

#### a) Qu'est-ce qu'un moteur à turbines ?

Un moteur turbo-fan est un type de moteur à réaction largement utilisé dans les avions commerciaux et militaires. Il appartient à la famille des moteurs à turbine à gaz et se distingue par le fait qu'il utilise un ventilateur pour accélérer l'air dans son mécanisme. Ce ventilateur, situé à l'avant du moteur, permet non seulement d'augmenter la poussée du moteur mais aussi de le rendre plus économique en carburant par rapport aux autres types de moteurs à réaction.



Figure 2:moteur à turbines

### b) composants d'un moteur à turbines

Les composants essentiels d'un moteur turbo-fan comprennent le ventilateur, le compresseur, la chambre de combustion, la turbine et la postcombustion. Chaque pièce joue un rôle essentiel dans le fonctionnement du moteur, et la conception globale est optimisée pour obtenir une efficacité et des performances maximales :

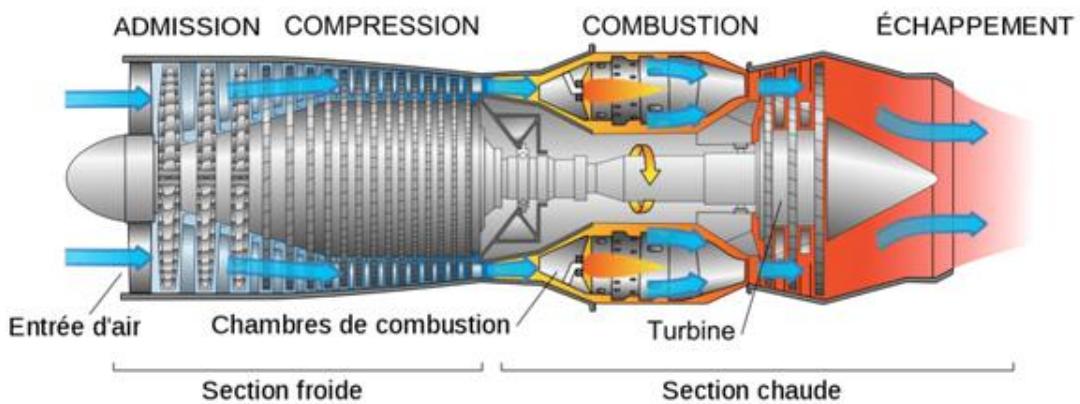


Figure 3:composants d'un turbo-fan

**Ventilateur** : situé à l'avant, le ventilateur initie le fonctionnement du moteur en aspirant l'air. Il propulse également une grande quantité d'air autour de l'extérieur du cœur du moteur, contribuant ainsi à la majeure partie de la poussée des moteurs à taux de dilution élevé.

**Le compresseur** : Il augmente la pression de l'air reçu du ventilateur avant qu'il ne pénètre dans la chambre de combustion. Le compresseur est constitué de plusieurs étages d'aubes rotatives et joue un rôle crucial dans l'efficacité du moteur.

Le compresseur d'un moteur turbo-fan peut être subdivisé en sections basse pression et haute pression

**Chambre de combustion** : c'est là que l'air comprimé se mélange au carburant et s'enflamme, créant ainsi des gaz à haute pression et à grande vitesse.

**Turbine** : Entraînée par les gaz en expansion de la chambre de combustion, la turbine génère de l'énergie pour entraîner le ventilateur et le compresseur.

**Postcombustion** : Composant optionnel qui permet d'augmenter encore la poussée en injectant et en allumant du carburant supplémentaire après la section de la turbine. Il est principalement utilisé dans les avions militaires pour améliorer les performances.

### c) Fonctionnement du moteur turbo-fan

Le moteur turbo-fan fonctionne en aspirant une grande quantité d'air à l'aide d'un ventilateur situé à l'avant, qui dirige une partie de cet air dans le cœur du moteur et une autre partie autour, générant une poussée importante. L'air entrant dans le cœur est comprimé par un compresseur, divisé en sections basse et haute pression, pour augmenter son efficacité avant d'atteindre la chambre de combustion. Dans cette dernière, l'air comprimé est mélangé au carburant et enflammé, produisant des gaz chauds à haute pression. Ces gaz traversent ensuite une turbine, qui utilise leur énergie pour entraîner le ventilateur et le compresseur. Enfin, certains moteurs disposent d'une postcombustion optionnelle pour injecter et allumer du carburant supplémentaire, augmentant la poussée, notamment pour les besoins des avions militaires.

Pour une démonstration visuelle approfondie du fonctionnement d'un moteur turbo-fan, je vous recommande de regarder la vidéo suivante : [\[Fonctionnement du moteur turbo-fan\]](#) Cette vidéo fournit une explication claire et illustrée qui complète les informations de ce rapport.

## 3. Description des Ensembles de Données

Les ensembles de données sont constitués de plusieurs séries temporelles multi-variées. Chaque ensemble est divisé en deux sous-ensembles : un ensemble d'entraînement et un ensemble de test. Chaque série temporelle provient d'un moteur différent, ce qui signifie que les données peuvent être considérées comme provenant d'une flotte de moteurs du même type.

Au début de chaque série temporelle, le moteur fonctionne normalement, et un défaut se développe progressivement jusqu'à une défaillance totale. Cette défaillance est atteinte dans l'ensemble d'entraînement, alors que dans l'ensemble de test, la série s'arrête avant que la défaillance ne survienne. L'objectif principal de cette étude est de prédire la durée de vie restante (RUL - **Remaining Useful Life**) des moteurs pour les données de test, exprimée en nombre de cycles d'opérations restants avant défaillance.

Les ensembles de données prennent en compte trois paramètres opérationnels qui influencent de manière significative les performances du moteur. Ces paramètres sont inclus dans les données, mais les mesures des capteurs sont affectées par un bruit de capteur, ajoutant une complexité supplémentaire à l'analyse. Chaque moteur démarre avec des degrés différents d'usure initiale et de variation de fabrication, considérés comme normaux et non comme des conditions de panne.

### Structure des Données

Les données sont fournies sous forme de fichiers texte compressés (ZIP) contenant 26 colonnes, avec des valeurs numériques séparées par des espaces. Chaque ligne représente une capture des données lors d'un cycle opérationnel, tandis que chaque colonne correspond à une variable spécifique. Voici les colonnes principales :

- **Colonne1** : Numéro d'unité (identifiant du moteur)
- **Colonne2** : Temps en cycles
- **Colonne 3** : Réglage opérationnel 1
- **Colonne 4**: Réglage opérationnel 2

- **Colonne 5:** Réglage opérationnel 3
- **Colonnes 6 à 26 :** Mesures des capteurs (de 1 à 26)

engine	time	op_setting_1	op_setting_2	op_setting_3	sm_1	sm_2	sm_3	sm_4	sm_5	...	sm_13	sm_14	sm_15	sm_16	sm_17	sm_18	sm_19	sm_20	sm_21
1	1	-0.0007	-0.0004	100.0	518.67	641.82	1589.70	1400.60	14.62	...	2388.02	8138.62	8.4195	0.03	392	2388	100.0	39.06	23.4190
1	2	0.0019	-0.0003	100.0	518.67	642.15	1591.82	1403.14	14.62	...	2388.07	8131.49	8.4318	0.03	392	2388	100.0	39.00	23.4236
1	3	-0.0043	0.0003	100.0	518.67	642.35	1587.99	1404.20	14.62	...	2388.03	8133.23	8.4178	0.03	390	2388	100.0	38.95	23.3442
1	4	0.0007	0.0000	100.0	518.67	642.35	1582.79	1401.87	14.62	...	2388.08	8133.83	8.3682	0.03	392	2388	100.0	38.88	23.3739
1	5	-0.0019	-0.0002	100.0	518.67	642.37	1582.85	1406.22	14.62	...	2388.04	8133.80	8.4294	0.03	393	2388	100.0	38.90	23.4044
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
100	196	-0.0004	-0.0003	100.0	518.67	643.49	1597.98	1428.63	14.62	...	2388.26	8137.60	8.4956	0.03	397	2388	100.0	38.49	22.9735
100	197	-0.0016	-0.0005	100.0	518.67	643.54	1604.50	1433.58	14.62	...	2388.22	8136.50	8.5139	0.03	395	2388	100.0	38.30	23.1594
100	198	0.0004	0.0000	100.0	518.67	643.42	1602.46	1428.18	14.62	...	2388.24	8141.05	8.5646	0.03	398	2388	100.0	38.44	22.9333
100	199	-0.0011	0.0003	100.0	518.67	643.23	1605.26	1426.53	14.62	...	2388.23	8139.29	8.5389	0.03	395	2388	100.0	38.29	23.0640
100	200	-0.0032	-0.0005	100.0	518.67	643.85	1600.38	1432.14	14.62	...	2388.26	8137.33	8.5036	0.03	396	2388	100.0	38.37	23.0522

Figure 4: captures d'écran des données

## Détails des Ensembles de Données

Les ensembles de données sont divisés en quatre configurations (FD001 à FD004) qui varient en termes de conditions opérationnelles et de modes de panne :

### FD001 :

- Trajectoires d'entraînement : 100
- Trajectoires de test : 100
- Conditions : Une seule (Niveau de la mer)
- Modes de panne : Un (Dégradation du compresseur haute pression - HPC)

### FD002 :

- Trajectoires d'entraînement : 260
- Trajectoires de test : 259
- Conditions : Six
- Modes de panne : Un (Dégradation du HPC)

### FD003 :

- Trajectoires d'entraînement : 100
- Trajectoires de test : 100
- Conditions : Une seule (Niveau de la mer)
- Modes de panne : Deux (Dégradation du HPC, Dégradation de l'hélice)

### FD004 :

- Trajectoires d'entraînement : 248
- Trajectoires de test : 249
- Conditions : Six
- Modes de panne : Deux (Dégradation du HPC, Dégradation de l'hélice)

Dans le cadre de notre projet, nous avons utilisé uniquement le premier ensemble de données, **FD001**

# Chapitre 2 : Lois Statistiques et Analyse de Temps de Fonctionnement

## Introduction

La fiabilité est une discipline essentielle pour analyser et gérer la maintenance des systèmes complexes. Elle permet d'évaluer la probabilité de bon fonctionnement d'un système sur une période donnée, tout en prévoyant les défaillances possibles. Ce chapitre propose une exploration approfondie des concepts fondamentaux de la fiabilité et des lois statistiques couramment utilisées pour modéliser les temps de fonctionnement.

Dans le cadre de notre projet, ces lois sont appliquées et ajustées aux données du dataset CMAPSS afin d'obtenir des indicateurs pertinents de performance et de durabilité.

### I. Concepts Généraux de la Fiabilité

#### 1. Définition

La fiabilité représente la probabilité qu'un produit accomplisse sa fonction requise sans défaillance dans des conditions d'utilisation spécifiques pendant une durée déterminée. Selon la norme **NF X50-500**, elle est définie comme une « *caractéristique d'un bien exprimée par la probabilité qu'il accomplisse une fonction requise dans des conditions données pendant un temps donné.* »

#### 2. Fonction de défaillance $F(t)$ - Fonction de fiabilité $R(t)$

Soit  $T$  une variable aléatoire représentant le temps de bon fonctionnement d'un équipement.

- On appelle **fonction de défaillance la fonction  $F$**  définie pour tout  $t \geq 0$

$$F(t) = P(T \leq t)$$

Le nombre  $F(t)$  représente la probabilité qu'un dispositif choisi au hasard ait une défaillance avant l'instant  $t$ .

- On distingue également la **fonction de densité de défaillance  $f(t)$** , qui décrit la densité de probabilité de défaillance à l'instant  $t$ , c'est-à-dire la probabilité qu'un appareil tombe en panne à un instant donné, par unité de temps.

La fonction de répartition  **$F(t)$**  et la fonction de densité  **$f(t)$**  sont liées par la relation suivante :

$$F(t) = \int_0^t f(x) dx$$

Cela signifie que la fonction de défaillance  **$F(t)$**  est l'intégrale de la fonction de densité  **$f(t)$**  sur l'intervalle **[0,t]**, représentant la probabilité cumulative de défaillance avant l'instant  $t$ .

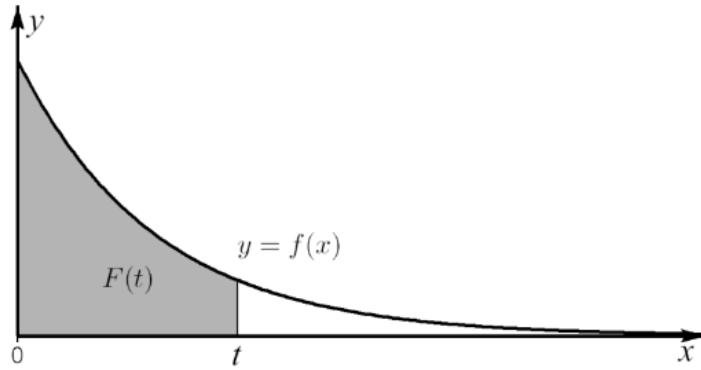


Figure 5:Fonction de défaillance  $F(t)$

Cette fonction nous amène naturellement une fonction associée : **la fonction de fiabilité  $R$**  définie pour tout  $t \geq 0$  par :

$$R(t) = 1 - F(t)$$

Le nombre  $R(t)$  représente la probabilité qu'un dispositif choisi au hasard dans la population n'ait pas de défaillance avant l'instant  $t$ .

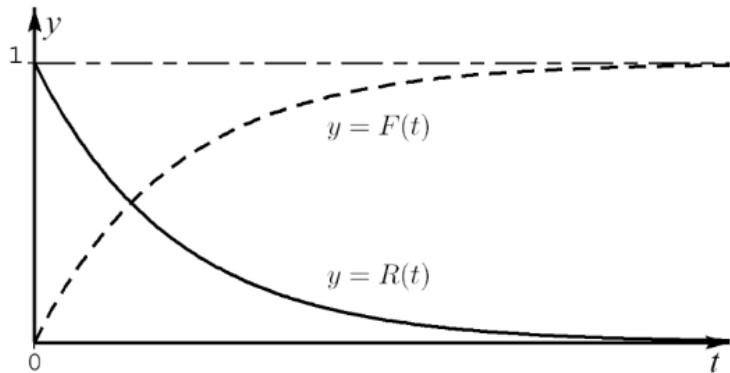


Figure 6:les fonction de fiabilité  $R(t)$  et de défaillance  $F(t)$

Enfin, le taux d'avarie moyen, c'est-à-dire la probabilité de défaillance dans un intervalle de temps  $[t, t+\Delta t]$  est donné par :

$$\frac{R(t) - R(t+\Delta t)}{R(t)} \times \frac{1}{\Delta t}$$

### 3. Taux de défaillance instantané

C'est la probabilité ( $0 \leq R \leq 1$ ) ; un produit doit accomplir de manière satisfaisante une fonction requise, sous des conditions données et pendant une période de temps donné.

L'écriture mathématique du taux de défaillance à l'instant  $t$ , noté  $\lambda(t)$ , défini sur  $R$  est la suivante :

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \left( \frac{1}{\Delta t} \cdot \frac{R(t) - R(t + \Delta t)}{R(t)} \right)$$

Physiquement le terme  $\lambda(t) \cdot \Delta t$ , mesure la probabilité qu'une défaillance d'un dispositif se produise dans l'intervalle de temps  $[t, t + \Delta t]$  sachant que ce dispositif a bien fonctionné jusqu'à l'instant  $t$ .

$$\begin{aligned}\lambda(t) &= - \frac{dR(t)}{dt} \cdot \frac{1}{R(t)} \\ &= \frac{dF(t)}{dt} \cdot \frac{1}{R(t)} \\ &= \frac{f(t)}{R(t)}\end{aligned}$$

où  $R$  est la fonction de fiabilité de ce matériel.

## II. lois de Fiabilité

Il est toujours possible d'associer à une variable aléatoire une probabilité et définir ainsi une loi de probabilité. Lorsque le nombre d'épreuves augmente indéfiniment, les fréquences observées pour le phénomène étudié tendent vers les probabilités et les distributions observées vers les distributions de probabilité ou loi de probabilité. Une loi de probabilité est un modèle représentant "au mieux", une distribution de fréquences d'une variable aléatoire

On distingue deux types

- ✓ Lois discrètes
- ✓ Lois continues

### 1. Lois discrètes

Une loi est dite discrète si elle prend ses valeur dans  $N$  c'est à dire des valeurs entières comme par exemple celle qui compte le nombre de pannes. Parmi les lois discrètes on peut citer :

- Loi Uniforme
- Loi de Bernoulli
- Loi Binomiale
- Loi Binomiale négative
- Loi Géométrique
- Loi Hypergéométrique
- Loi de Poisson

Cependant, dans notre cas, nous allons nous concentrer sur les lois continues, qui sont plus adaptées pour modéliser le temps de fonctionnement d'un moteur jusqu'à sa défaillance.

### 2. Les lois continues

Dans le cadre de l'étude de la fiabilité, les lois continues sont couramment utilisées pour modéliser le temps de fonctionnement jusqu'à la défaillance d'un système

Voici quelques-unes des lois continues les plus utilisées dans l'analyse de fiabilité :

- La loi exponentielle
- La loi Gamma
- Loi Inverse Gamma
- La loi Bêta
- La Loi normal
- La loi Log normale
- La loi de Weibull

#### a) La loi exponentielle:

La fonction de densité de probabilité (PDF) de la loi exponentielle est donnée par :

$$f(t; \lambda) = \lambda e^{-\lambda t}, \quad t \geq 0, \lambda > 0$$

#### b) La loi Gamma :

la fonction de densité de probabilité de la loi Gamma est :

$$f(t; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t}, \quad t > 0,$$

- $\alpha$  : Paramètre de forme .
- $\beta$  : Paramètre d'échelle .
- $\Gamma(\alpha)$  : Fonction Gamma, qui est une généralisation de la factorielle.

#### c) Loi Inverse Gamma :

la fonction de densité de probabilité de la loi Inverse Gamma est :

$$f(t; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} t^{-(\alpha+1)} e^{-\frac{\beta}{t}}, \quad t > 0, \alpha > 0, \beta > 0$$

- $\alpha > 0$  : paramètre de forme .
- $\beta > 0$  : paramètre d'échelle.

#### d) La loi Bêta :

la fonction de densité de probabilité de la loi Bêta est :

$$f(t; \alpha, \beta) = \frac{t^{\alpha-1} (1-t)^{\beta-1}}{B(\alpha, \beta)}, \quad 0 \leq t \leq 1, \alpha > 0, \beta > 0$$

- $\alpha$  et  $\beta$  : Paramètres de forme.

- **B( $\alpha, \beta$ )** : Fonction Beta, qui est une généralisation de la factorielle pour des valeurs continues

#### e) Loi Normale (ou Loi Gaussienne) :

la fonction de densité de probabilité de la loi Normale est :

$$f(t; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}, \quad -\infty < t < \infty, \sigma > 0$$

- $\mu$  : Moyenne
- $\sigma$  : Écart-type

#### f) Loi Log-Normale :

la fonction de densité de probabilité de la loi Log-Normale est :

$$f(t; \mu, \sigma) = \frac{1}{t\sigma\sqrt{2\pi}} e^{-\frac{(\ln t - \mu)^2}{2\sigma^2}}, \quad t > 0, \sigma > 0$$

- $\mu$  : Moyenne du logarithme
- $\sigma$  : Écart-type du logarithme

#### g) Loi de Weibull :

la fonction de densité de probabilité de la loi Weibull est :

$$f(t; \alpha, \beta) = \frac{\alpha}{\beta} \left( \frac{t}{\beta} \right)^{\alpha-1} e^{-(t/\beta)^\alpha}, \quad t \geq 0, \alpha > 0, \beta > 0$$

- $\alpha$  : Paramètre de forme .
- $\beta$  : Paramètre d'échelle .

### III. Analyse du Temps de Fonctionnement et Sélection des Lois de Distribution

Dans cette section, nous allons réaliser une analyse approfondie des temps de fonctionnement des moteurs turbo-fan afin de choisir la loi de distribution la plus adaptée. L'objectif est d'identifier la loi de distribution qui reflète le mieux ces phénomènes de défaillance , et de modéliser avec précision le comportement des défaillances des moteurs.

#### 1. Analyse du Temps de Fonctionnement

##### a) Préparation des Données

Dans cette étape, nous nettoyons les données ( éliminant les valeurs manquantes ainsi que les doublons)

Nous nous concentrons ensuite sur les temps de fonctionnement des moteurs turbo-fan, que nous calculons en déterminant le nombre maximal de cycles effectués avant chaque défaillance.

```
failure_time = data_FD001.groupby('engine')['time'].max()
```

### b) la répartition des temps de défaillance

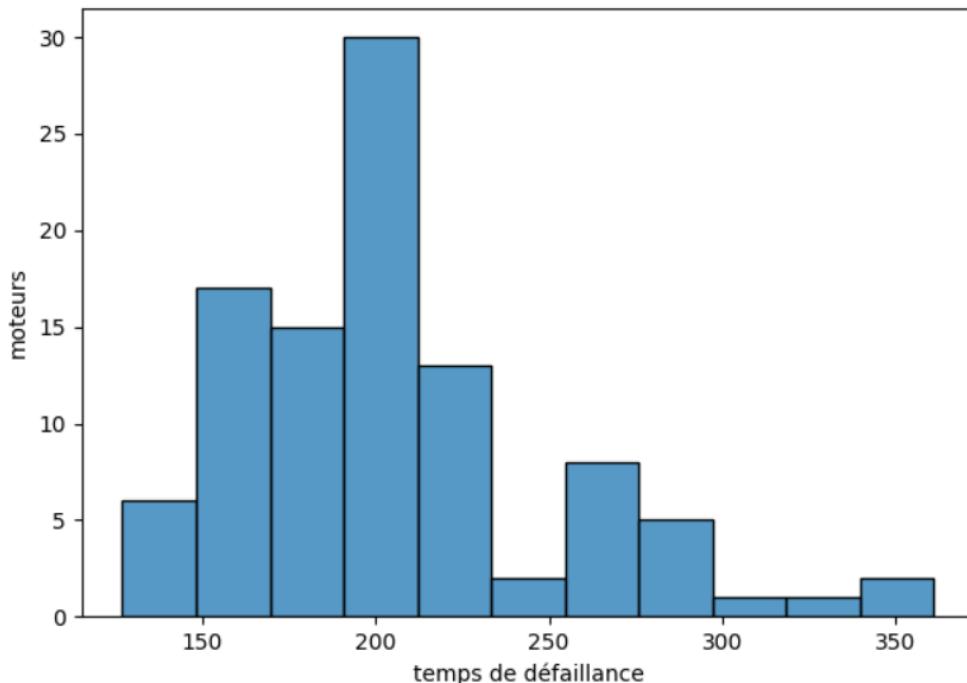


Figure 7:Distribution des temps de défaillance (failure time) par moteur

D'après la distribution des temps de défaillance, on constate une asymétrie marquée avec une concentration des valeurs entre **150 et 200 cycles** et une queue allongée, indiquant des comportements typiques de phénomènes de défaillance. En conséquence, les lois candidates qui semblent les plus adaptées pour modéliser ces données sont :

- Log-Normale
- Gamma
- Weibull
- Beta
- Normale

Ces lois seront testées pour identifier celle qui modélise le mieux les données des temps de fonctionnement des moteurs turbo-fan.

## 2. Ajustement des Lois sur les Données

Pour cette étape, nous estimons les paramètres des différentes lois de distribution candidates en utilisant des méthodes d'ajustement statistique. Les paramètres sont ajustés pour chaque loi en fonction des données de temps de fonctionnement, dans le but de réduire l'écart entre les données

observées et les distributions théoriques. Voici un extrait de code utilisé pour ajuster les paramètres des lois :

```
from scipy.stats import weibull_min, norm, lognorm, gamma, beta

# Define the distributions you want to fit
distributions = {
    "Beta": beta,
    "Gamma": gamma,
    "Normal": norm,
    "Log-Normal": lognorm,
    "Weibull": weibull_min
}

# Store the results of fitting the distributions
fitted_params = {}

# Fit each distribution to the data and store the parameters
for name, dist in distributions.items():
    params = dist.fit(failure_time) # Fit the distribution to the data
    fitted_params[name] = params # Store the fitted parameters
```

Après l'ajustement des différentes lois de distribution aux données des temps de défaillance, nous comparons leurs fonctions de densité (PDF) à la distribution observée des données. Ce graphique superpose l'histogramme des temps de défaillance aux courbes des densités ajustées pour chaque loi candidate, permettant d'évaluer visuellement laquelle correspond le mieux aux données.

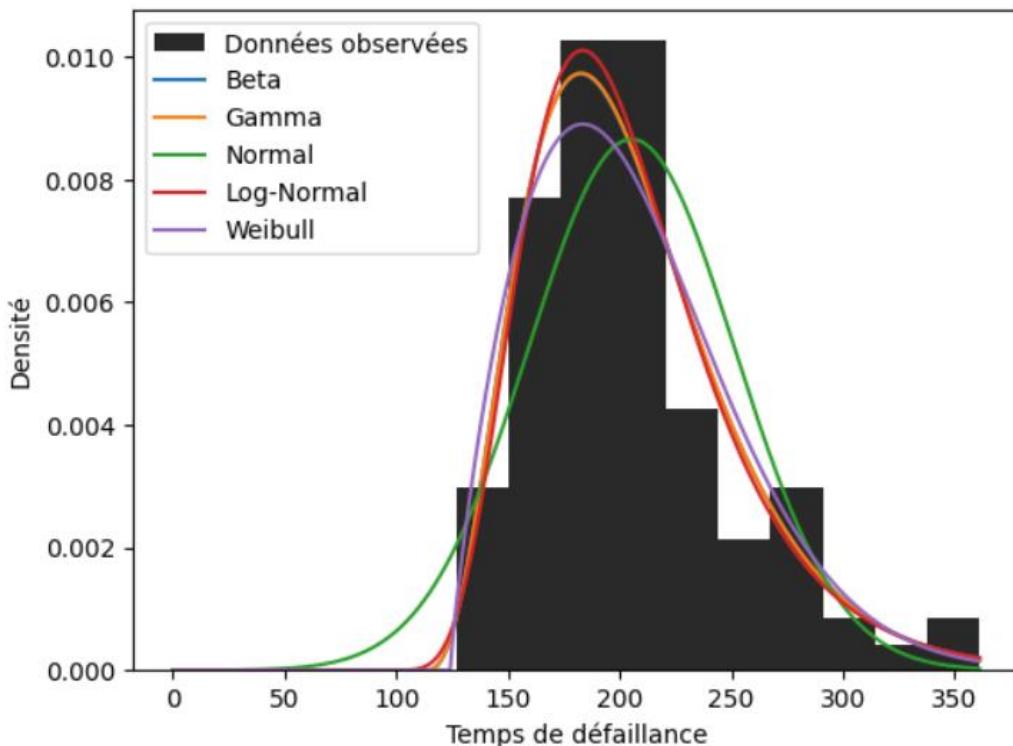


Figure 8: les fonctions de densités (pdf) ajustées pour chaque distribution

D'après ce graphe, on peut constater que, comme mentionné précédemment, ces lois sont des candidates plausibles pour modéliser les données, notamment la **Log-Normale**, **Gamma**, **Bêta** (avec Gamma et Bêta apparaissant confondues sur le graphique), et **Weibull**. Cependant, pour déterminer avec précision la loi la plus adaptée, nous appliquerons plusieurs tests statistiques pour valider cette analyse.

### 3. Validation de l'Ajustement

Dans cette étape, nous allons appliquer des tests statistiques pour évaluer quantitativement l'adéquation des lois ajustées aux données observées et identifier la loi qui offre le meilleur compromis entre précision et complexité.

#### a) Test de Kolmogorov-Smirnov (KS)

Le **test de Kolmogorov-Smirnov** est défini comme un test statistique non paramétrique qui quantifie la différence entre la fonction de distribution empirique d'un échantillon et la fonction de distribution cumulative d'une distribution de référence, ou les fonctions de distribution empiriques de deux échantillons.

En termes plus simples, le test de Kolmogorov-Smirnov t'aide à comprendre si deux ensembles de données proviennent de la même distribution. Le test calcule la distance maximale K entre les fonctions de distribution cumulative (FDC) de deux échantillons ou d'un échantillon et d'une distribution de référence.

$$K = \sup |F(x) - F_0(x)|$$

où,  $F(x)$  et  $F_0(x)$  sont les fonctions de distribution cumulées empiriques des deux échantillons. Pour un échantillon et une distribution de référence,  $F_0(x)$  serait remplacé par la fonction de distribution cumulative de la référence

- Plus la valeur K est petite, plus il est probable que les deux ensembles suivent la même distribution.
- Ce test repose sur une hypothèse ( $H_0$ ) : les données suivent la distribution de référence. (Rejet de  $H_0$  si K est trop grand)

Les résultats du test de Kolmogorov-Smirnov (KS) pour chaque distribution candidate sont présentés dans le tableau ci-dessous. Ces résultats incluent la statistique **KS**, qui mesure la distance maximale entre les fonctions de distribution, ainsi que la **p-value**, qui indique si l'hypothèse nulle  $H_0$  peut être acceptée ou rejetée.

	KS	p_value
Log-Normal	0.071910	0.652432
Gamma	0.078931	0.535430
Beta	0.079032	0.533805
Weibull	0.094090	0.318623
Normal	0.138333	0.039376

Figure 9:Les résultats du test de Kolmogorov-Smirnov (KS)

D'après ces résultats, les lois **log-normale**, **gamma**, et **beta** se distinguent comme les meilleures candidates pour modéliser les données de temps de défaillance

## b) Le test du Chi-deux (ou test de $\chi^2$ )

le **test du  $\chi^2$  d'ajustement**, ou d'**adéquation**, qui compare globalement la distribution observée dans un échantillon statistique à une distribution théorique.

Le test du  $\chi^2$  est basé sur la somme des carrés des écarts normalisés entre les fréquences observées et les fréquences attendues. La formule est la suivante :

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

où :

- $O_i$  est la fréquence observée dans la i-ème classe (ou intervalle)
- $E_i$  est la fréquence attendue dans la i-ème classe, calculée selon la loi théorique,
- k est le nombre total de classes dans l'histogramme

Une fois que le  $\chi^2$  a été calculé, il est comparé à une distribution de  $\chi^2$  avec  $k-1$  degrés de liberté

La **valeur p** est la probabilité d'obtenir une valeur de  $\chi^2$  aussi extrême, ou plus extrême, que celle calculée à partir de l'échantillon. Elle est déterminée à l'aide de la fonction de distribution cumulée de  $\chi^2$ .

le test du  **$\chi^2$  d'ajustement** a été appliqué pour chaque loi de distribution candidate (Weibull, Log-Normal, Gamma, Beta, Normal) afin de vérifier si l'ajustement des données observées suit une distribution théorique.

La statistique  $\chi^2$  et la valeur p correspondante ont été calculées pour chaque loi de distribution.

Voici les résultats obtenus :

	$\chi^2$	p_value
Log-Normal	7.036038	0.633367
Beta	7.658930	0.568849
Gamma	7.717748	0.562822
Weibull	10.394762	0.319482
Normal	29.389340	0.000557

Figure 10:résultats de test chi-deux

Même si les lois **Log-Normale, Beta, et Gamma** se distinguent comme les meilleures candidates pour modéliser les données de temps de défaillance.

## c) AIC (Akaike Information Criterion) et BIC (Bayesian Information Criterion)

Les critères AIC et BIC sont utilisés pour évaluer la qualité d'un modèle statistique tout en tenant compte de sa complexité. Ces deux critères permettent de comparer plusieurs modèles et de choisir

celui qui offre le meilleur compromis entre ajustement aux données et simplicité (complexité du modèle).

**L'AIC** est défini par la formule suivante :

$$AIC = 2k - 2 \ln(\hat{L})$$

Avec :

- $k$  : Le nombre de paramètres estimés dans le modèle.
- $\hat{L}$  : La valeur de la fonction de vraisemblance maximisée du modèle.

Le modèle avec l'AIC le plus bas est considéré comme le meilleur parmi les candidats, car il pénalise les modèles plus complexes (avec plus de paramètres)

Le **BIC** est similaire à l'AIC, mais il pénalise davantage les modèles complexes, surtout pour des tailles d'échantillons plus grandes. La formule est la suivante :

$$BIC = \ln(n)k - 2 \ln(\hat{L})$$

Avec :

- $n$  : Le nombre d'observations (taille de l'échantillon).
- $k$  : Le nombre de paramètres estimés dans le modèle.
- $\hat{L}$  : La fonction de vraisemblance maximisée du modèle.

Comme pour l'AIC, un BIC plus faible indique un modèle préféré

En appliquant ces critères à nos données de temps de défaillance, nous avons obtenu les valeurs suivantes pour chaque distribution :

	<b>Beta</b>	<b>Gamma</b>	<b>Normal</b>	<b>Log-Normal</b>	<b>Weibull</b>
AIC	1039.859333	1037.857388	1053.995642	1037.872591	1039.032455
BIC	1050.280014	1045.672899	1059.205982	1045.688101	1046.847966

Figure 11:résultats AIC et BIC pour chaque loi ajusté

#### d) Conclusion

Les tests de Kolmogorov-Smirnov, Khi-deux, AIC et BIC ont révélé que les distributions **Log-Normale**, **Gamma** et **Beta** sont les meilleures options pour modéliser les données de temps de défaillance. Parmi celles-ci, les distributions **Log-Normale** et **Gamma** se sont particulièrement démarquées par leur adéquation aux données, surtout lors des évaluations AIC et BIC.

En conclusion, nous avons choisi de continuer notre étude avec les distributions **Log-Normale** et **Gamma**, en approfondissant leur application pour l'analyse des temps de défaillance.

## IV. La Loi Log-Normale

La loi Log-Normale est une distribution de probabilité utilisée pour modéliser des données strictement positives, où la variable d'intérêt suit une distribution normale après une transformation logarithmique.

En d'autres termes, si une variable aléatoire X suit une loi Log-Normale, alors  $Y = \ln(X)$  suit une loi Normale.

### 1. Relations Mathématiques Associées

Pour une loi Log-Normale avec les paramètres  $\mu$  (moyenne logarithmique) et  $\sigma$  (écart-type logarithmique) :

#### Fonction de densité de probabilité (PDF) :

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} e^{-\frac{(\ln(t)-\mu)^2}{2\sigma^2}}, \quad t > 0$$

- Elle représente la fréquence des défaillances à différents moments. Un pic dans  $f(t)$  indique le moment où les défaillances sont les plus probables.

#### La fonction de distribution cumulative $F(t)$ (ou fonction de défaillance) :

$$F(t) = \int_0^t f(u) du = \Phi\left(\frac{\ln(t) - \mu}{\sigma}\right), \quad t > 0$$

où :

- $\Phi$  est la fonction de répartition de la loi normale standard,
- $\mu$  et  $\sigma$  sont les paramètres de la distribution log-normale
- $F(t)$  donne la probabilité qu'une défaillance se produise avant  $t$ . Dans un contexte de maintenance

#### La fonction de fiabilité $R(t)$ (ou Fonction de survie):

$$R(t) = 1 - F(t) = 1 - \Phi\left(\frac{\ln(t) - \mu}{\sigma}\right)$$

- Elle représente la probabilité qu'une unité (moteur) fonctionne correctement sans défaillance jusqu'à un instant  $t$ .

#### La fonction de taux de défaillance instantané $\lambda(t)$ (ou Taux de défaillance) :

$$\lambda(x) = \frac{f(x)}{R(x)} = \frac{\frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}}{1 - \Phi\left(\frac{\ln(x)-\mu}{\sigma}\right)}, \quad x > 0$$

- Elle quantifie la probabilité qu'une défaillance se produise dans un intervalle infinitésimal autour de  $t$ , étant donné que l'unité a survécu jusqu'à  $t$ .

### Fonction de risque cumulatif (H) :

$$H(x) = -\ln(R(x)) = -\ln\left(1 - \Phi\left(\frac{\ln(x) - \mu}{\sigma}\right)\right), \quad x > 0$$

- **H(t)** est utile pour comprendre l'évolution du risque de défaillance sur le temps.
- Une croissance rapide de **H(t)** signale un système en phase de vieillissement critique.

## 2. Estimation des Paramètre et Visualisation des Fonctions :

Après avoir ajusté les données à une distribution log-normale, les paramètres ajustés jouent un rôle essentiel dans la compréhension de la forme de la loi et dans l'évaluation de la pertinence de l'ajustement. Ces paramètres incluent :

- **Moyenne logarithmique ( $\mu$ )** :  $\mu=5.3012$
- **Écart-type logarithmique ( $\sigma$ )** :  $\sigma=0.2132$

Ces valeurs représentent respectivement la moyenne et la dispersion logarithmiques des données de défaillance observées.

Pour évaluer l'adéquation de l'ajustement à la loi Log-Normale, nous avons généré un ensemble d'échantillons simulés à partir des paramètres ajustés ( $\mu$  et  $\sigma$ ). Ces échantillons simulés représentent les temps de défaillance théoriques selon la distribution Log-Normale.

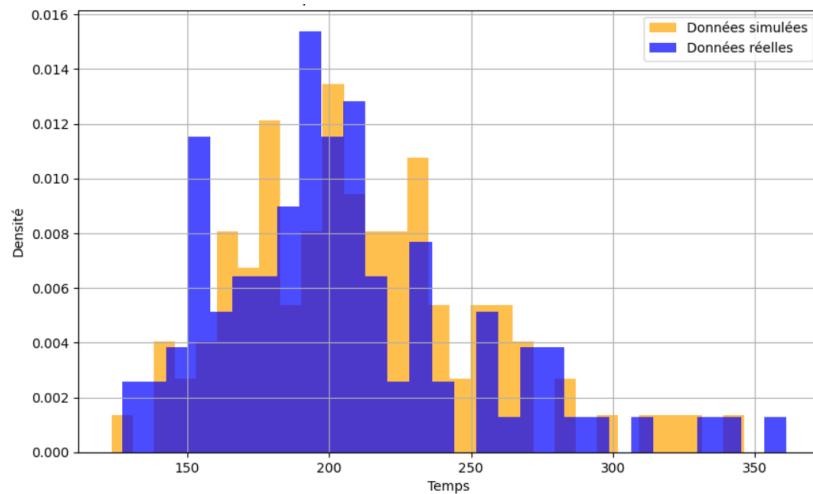


Figure 12:Comparaison des données simulées et réelles

Pour mieux comprendre le comportement de la distribution log-normale ajustée, les principales fonctions mathématiques associées sont représentées ci-dessous.

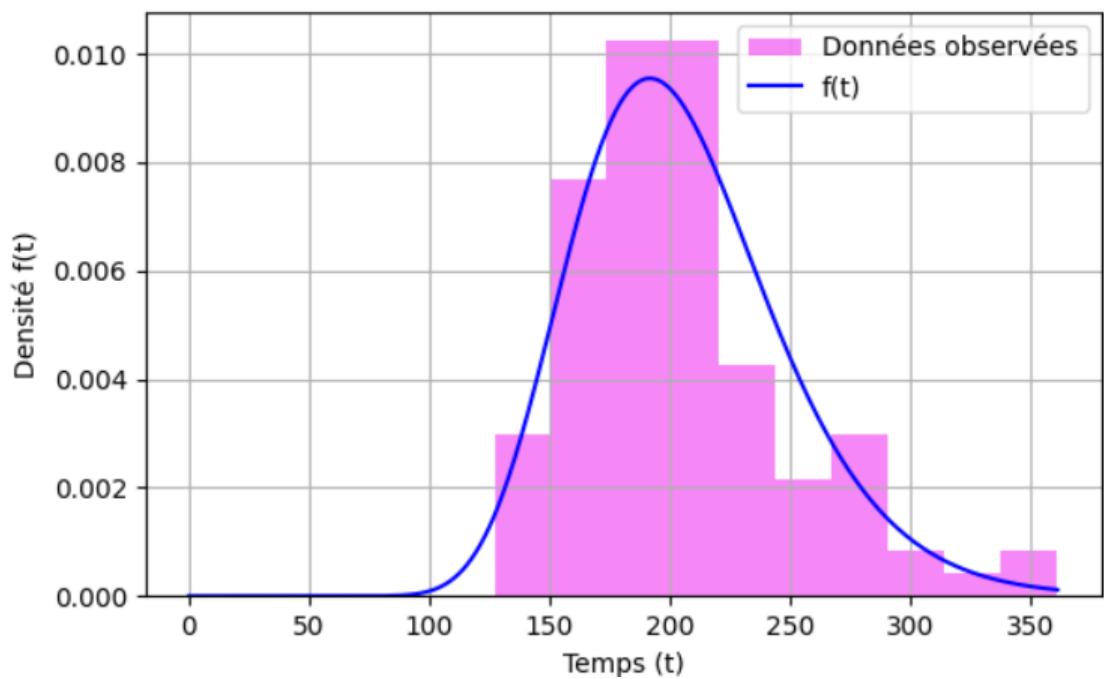


Figure 13:densité de probabilité  $f(t)$  - Log Normal

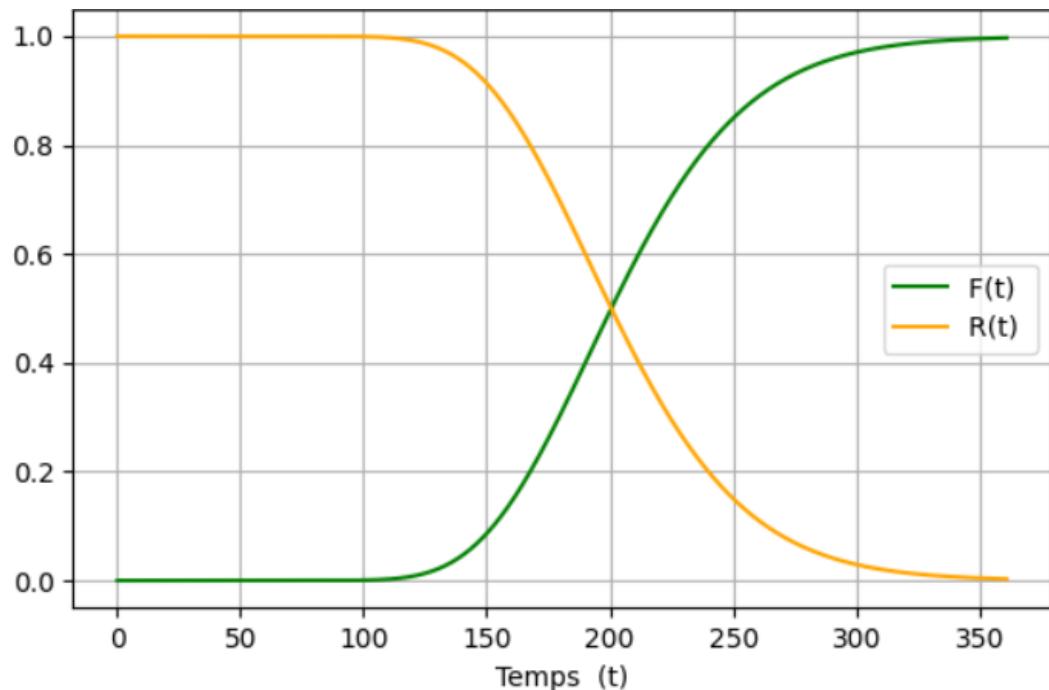


Figure 14:Fonction de défaillance  $F(t)$  et de fiabilité  $R(t)$  - Log Normal

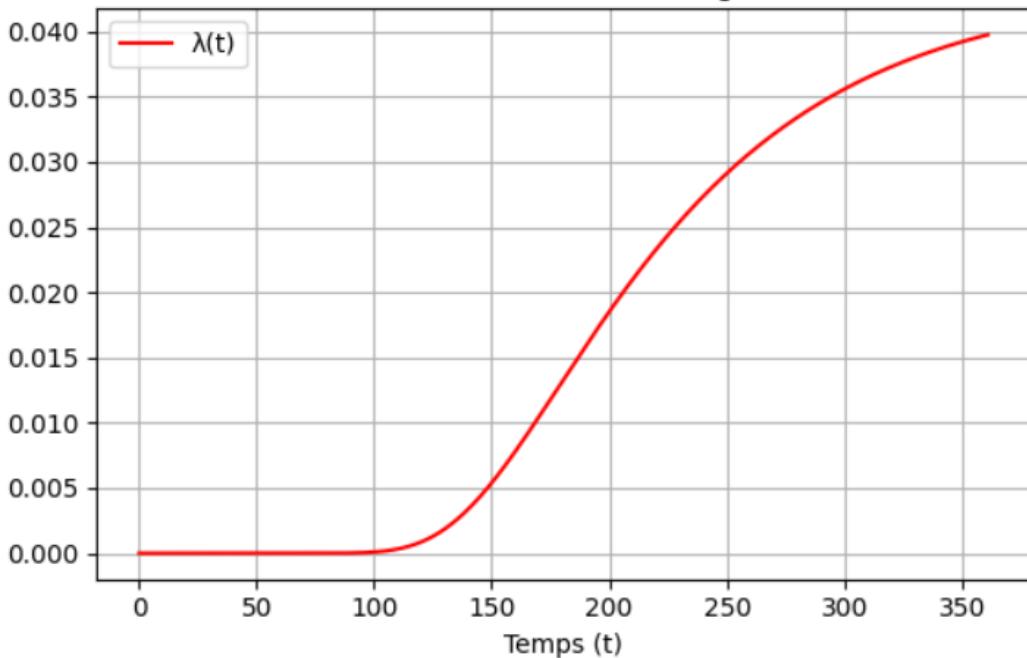


Figure 15:Taux de défaillance  $\lambda(t)$  - Log Normal

### 3. Calcul de MTTF (Mean Time To Failure)

Le **MTTF (Mean Time To Failure)** représente la durée de vie moyenne avant défaillance , il correspond à l'espérance mathématique du temps avant une défaillance. Dans le cas de la loi log-normale, il est donné par la formule :

$$MTTF = e^{\mu + \frac{\sigma^2}{2}}$$

Où :

- $\mu$  est la moyenne logarithmique des données
- $\sigma$  est l'écart-type logarithmique des données

Résultat calculé :

Le calcul du **MTTF (Mean Time To Failure)**, basé sur les paramètres ajustés de la distribution log-normale ( $\mu=5.3012$ ,  $\sigma=0.2132$ ), montre que la durée de vie moyenne des moteurs avant défaillance est estimée à **205 cycles** .

### 4. Prédiction de la Durée de Vie Restante (RUL) Basée sur la Loi Log-Normale

#### a) Estimation du temps avant défaillance à une fiabilité donnée

Pour une fiabilité souhaitée **R**, la durée avant défaillance est obtenue en inversant la fonction de répartition **F(t)**, La relation mathématique est donnée par :

$$F(t) = 1 - R(t)$$

Ainsi, pour une fiabilité donnée **R**, la durée avant défaillance estimée est :

$$t = F^{-1}(1 - R)$$

### Exemple :

Dans notre cas, pour une fiabilité de 95 % (**R=0.95**), nous obtenons :

Le temps estimé avant défaillance est de **145 cycles**

### b) Estimation Conditionnelle après un Nombre de Cycles Réalisés

L'estimation conditionnelle de la durée de vie restante  $t_{sup}$  après un certain nombre de cycles  $t_0$  réalisés repose sur la probabilité conditionnelle. En d'autres termes, une fois que le moteur a fonctionné pendant  $t_0$  cycles, nous pouvons estimer combien de cycles supplémentaires il peut encore effectuer pour atteindre une fiabilité **R** donnée.

#### Probabilité Conditionnelle de Survie :

La probabilité conditionnelle que le moteur continue à fonctionner pendant un temps supplémentaire  $t$  après avoir déjà fonctionné pendant  $t_0$  cycles est donnée par :

$$P(T > t_0 + t \mid T > t_0) = \frac{R(t_0 + t)}{R(t_0)}$$

L'objectif est de déterminer la durée restante  $t_{sup}$ , c'est-à-dire, combien de cycles supplémentaires le moteur peut fonctionner pour une fiabilité **R** donnée. Pour cela, nous fixons une fiabilité **R**, puis nous résolvons l'équation suivante :

$$R = \frac{R(t_0 + t_{sup})}{R(t_0)}$$

$$1 - F(t_0 + t_{sup}) = R \cdot R(t_0)$$

$$t_{sup} = F^{-1}(1 - R \cdot R(t_0)) - t_0$$

### Exemple :

Considérons un moteur ayant réalisé  $t_0=50$  cycles, et souhaitons estimer les cycles supplémentaires pour une fiabilité de 90% (**R=0.90**) :

- **$t_{sup} = 104$  cycles supplémentaires**

### c) Estimation de la Durée Restante(RUL) pour l'Ensemble de Test

Ici, nous allons appliquer la méthode d'estimation conditionnelle décrite précédemment pour évaluer la durée de vie restante des équipements dans un ensemble de tests, en tenant compte des cycles déjà effectués et de la fiabilité souhaitée.

#### Choix de la Fiabilité :

Le choix de la fiabilité **R** est un élément essentiel dans cette estimation. En général, une fiabilité élevée est souhaitable, car elle assure une forte probabilité que l'équipement continue à fonctionner sans

défaillance durant la période d'observation restante. Cependant, dans la pratique, nous optons pour un seuil de fiabilité qui correspond aux contraintes et aux exigences de maintenance.

Pour illustrer, nous commencerons avec une fiabilité de 90% ( $R=0.90$ ), qui représente souvent un bon compromis entre la durée de vie estimée et la précision de la prévision. Une fiabilité plus élevée peut être envisagée pour des scénarios plus prudents.

### Estimation de la Durée Restante :

Étape 1 : Préparation du jeu de données de test et calcul du nombre maximum de cycles réalisés.

```
# DataFrame des cycles réalisés par moteur
test_cycles_realises = test_FD001.groupby('engine')['time'].max()
```

Étape 2 : Calcul de la durée restante estimée, sachant que la fiabilité est de 90%.

```
# Calculer la durée de vie restante pour chaque moteur
duree_vie_restante = test_cycles_realises.apply(
    lambda x: estimer_duree_supplementaire(x, fiabilite_souhaitee=0.9)
)

# Afficher les résultats sous forme de DataFrame
resultats = pd.DataFrame({
    'engine': test_cycles_realises.index,
    'cycles_realises': test_cycles_realises.values,
    'duree_vie_restante_estime': duree_vie_restante.values,
    'duree_vie_restante_reel': rul_true.RUL
})
```

	engine	cycles_realises	duree_vie_restante_estime	duree_vie_restante_reel
0	1	31	159	112
1	2	49	141	98
2	3	126	64	69
3	4	106	84	82
4	5	98	92	91

### Évaluation du Modèle :

Pour évaluer la qualité de l'estimation de la durée restante, nous utiliserons deux critères de performance courants :

- **RMSE (Root Mean Square Error)** : La racine carrée de l'erreur quadratique moyenne est une mesure de l'écart moyen entre les valeurs estimées et les valeurs réelles.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{\text{réel},i} - y_{\text{prédict},i})^2}$$

- **MAE (Mean Absolute Error)** : L'erreur absolue moyenne mesure la différence absolue moyenne entre les valeurs estimées et les valeurs réelles.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_{\text{réel},i} - y_{\text{prédict},i}|$$

- **R<sup>2</sup> (Coefficient de Détermination)** : Ce coefficient mesure la proportion de la variance des données expliquée par le modèle d'estimation.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_{\text{réel},i} - y_{\text{prédict},i})^2}{\sum_{i=1}^n (y_{\text{réel},i} - \bar{y}_{\text{réel}})^2}$$

Donc, maintenant, nous allons évaluer la performance des estimations en fonction de la fiabilité souhaitée **R**. Nous allons calculer les principales métriques de performance, à savoir **R<sup>2</sup>**, l'**Erreur Absolue Moyenne** (MAE) et l'**Erreur Quadratique Moyenne** (RMSE). Les graphiques ci-dessous montrent comment ces métriques évoluent en fonction de la fiabilité choisie.

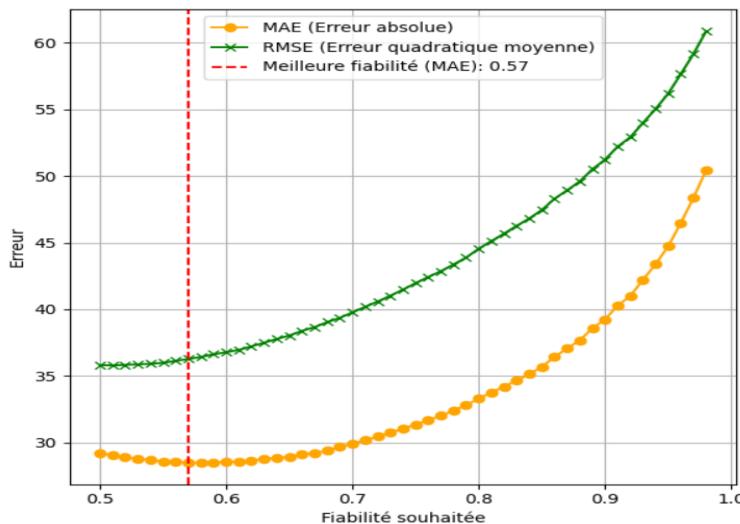


Figure 16: Évolution des erreurs en fonction de la fiabilité souhaitée

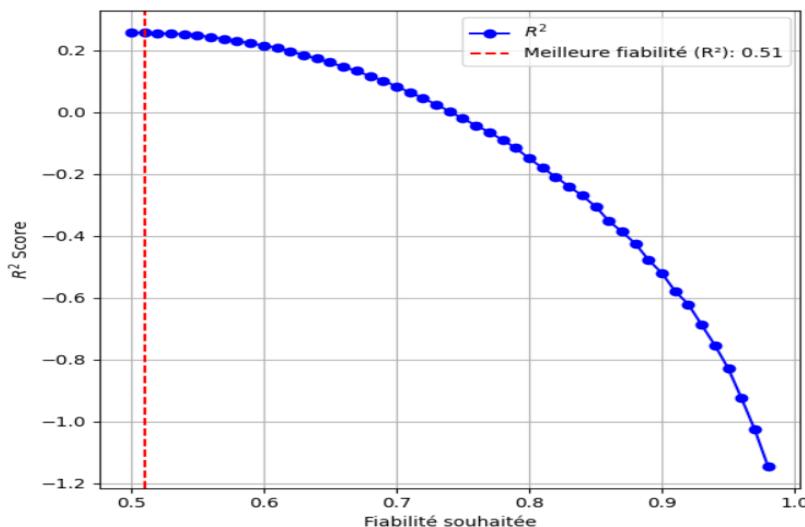


Figure 17: Évolution du R<sup>2</sup> en fonction de la fiabilité souhaitée

Pour une fiabilité de 57 %, on constate des erreurs minimales .

## V. La Loi Gamma

Nous allons suivre la même approche que pour la loi log-normale, mais cette fois-ci avec la loi gamma. Nous procéderons à l'ajustement des données, à l'estimation des paramètres, et à l'évaluation des métriques de performance pour les prédictions basées sur cette nouvelle distribution.

Une distribution Gamma ou loi Gamma est un type de loi de probabilité des variables aléatoires réelles positives. Une distribution Gamma est caractérisée par deux paramètres :

- $\alpha$  (paramètre de forme) : détermine la forme de la fonction de densité.
- $\beta$  (paramètre d'échelle) : contrôle l'étirement ou la compression de la distribution.

Ces deux paramètres influencent la représentation graphique de sa fonction de densité

### 1. Relations Mathématiques Associées

#### Fonction de densité de probabilité :

$$f(t; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t}, \quad t > 0,$$

où :

- $\alpha > 0$  est le paramètre de forme,
- $\beta > 0$  est le paramètre d'échelle,
- $\Gamma(\alpha)$  est la fonction Gamma donnée par.  $\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx$

#### Fonction de défaillance :

$$F(t; \alpha, \beta) = \frac{\gamma(\alpha, \beta t)}{\Gamma(\alpha)}$$

où :

- $\gamma(\alpha, x) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx$  est la fonction Gamma incomplète.

#### Fonction de fiabilité :

$$R(t; \alpha, \beta) = 1 - F(t) = \frac{\Gamma(\alpha, \beta t)}{\Gamma(\alpha)},$$

#### Fonction de taux de défaillance instantané :

$$\lambda(t; \alpha, \beta) = \frac{f(t; \alpha, \beta)}{R(t; \alpha, \beta)} = \frac{\beta^\alpha t^{\alpha-1} e^{-\beta t}}{\Gamma(\alpha, \beta t)}$$

## 2. Estimation des Paramètre et Visualisation des Fonctions :

Après avoir ajusté les données à une distribution Gamma, les paramètres ajustés jouent un rôle essentiel dans la compréhension de la forme de la loi et dans l'évaluation de la pertinence de l'ajustement. Ces paramètres incluent :

- **paramètre de forme ( $\alpha$ )** :  $\alpha = 4.11$
- **paramètre d'échelle ( $\beta$ )** :  $\beta = 0.04$

Générer les temps de défaillance théoriques selon la distribution Gamma.

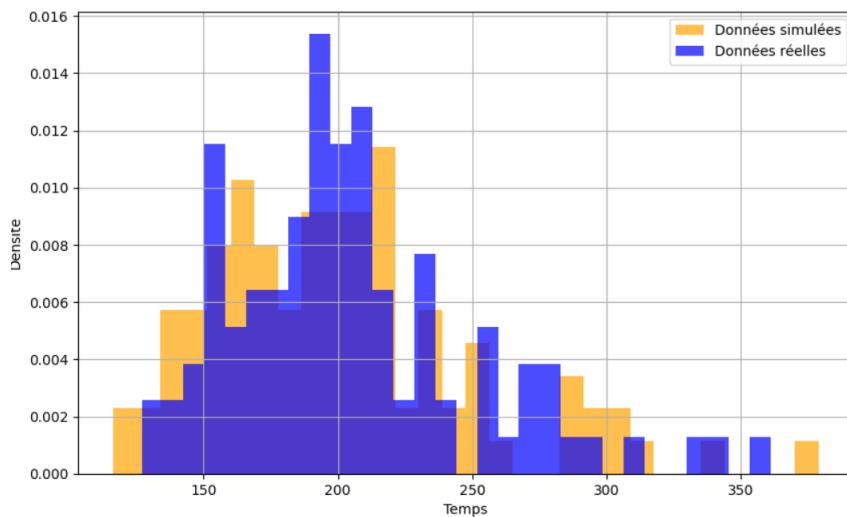


Figure 18:Comparaison des données simulées (par distribution Gamma) et réelles

les principales fonctions mathématiques associées sont représentées ci-dessous.

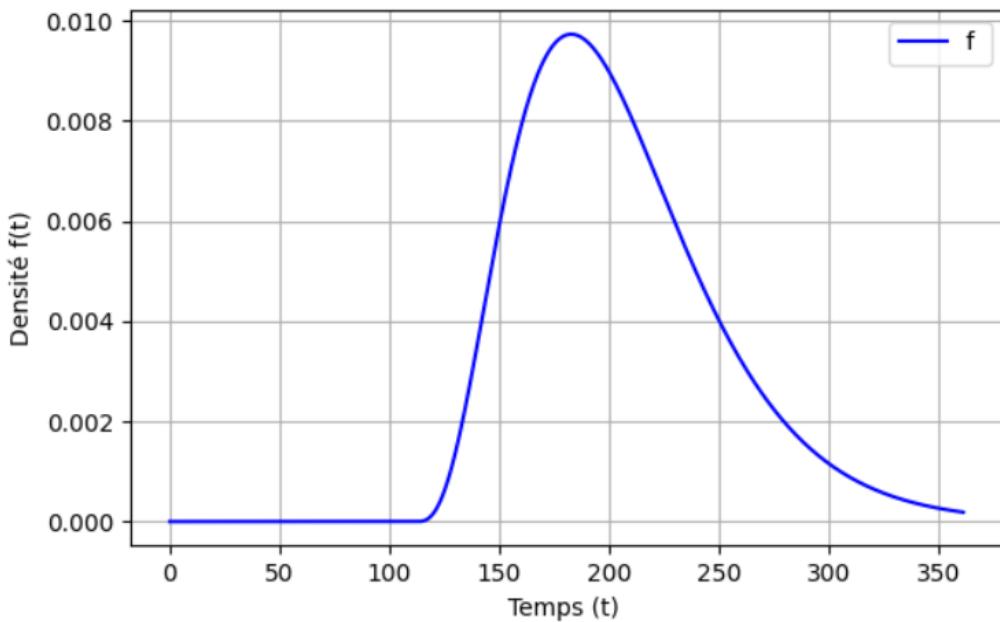


Figure 19:Fonction de densité de probabilité  $f(t)$  (pour la loi Gamma )

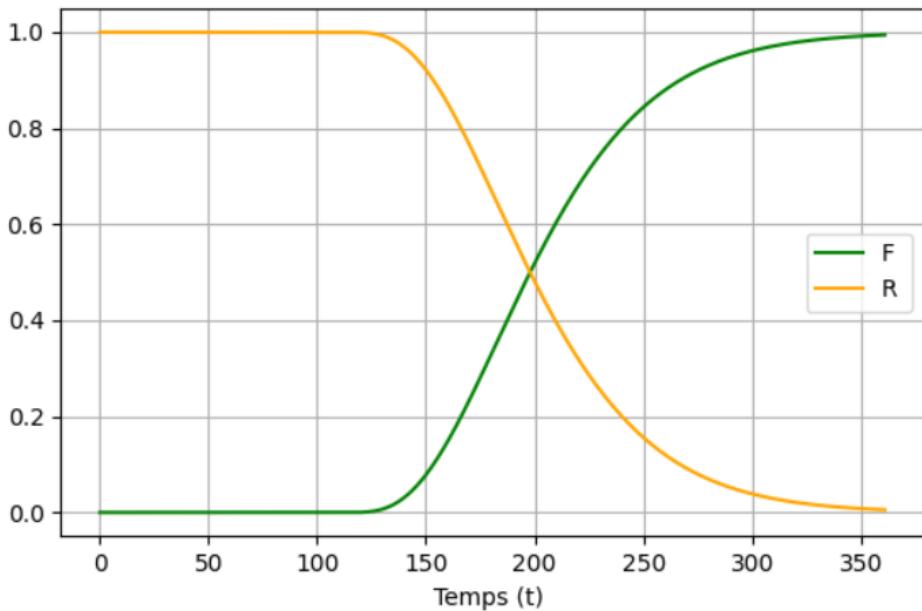


Figure 20:Fonction de défaillance  $F(t)$  et de fiabilité  $R(t)$  (pour la loi Gamma )

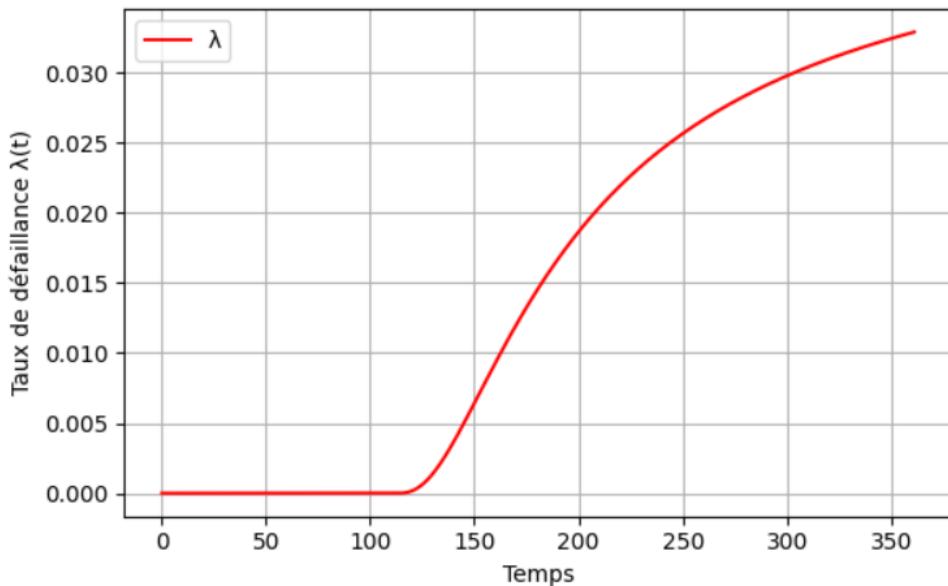


Figure 21:Taux de défaillance ( $\lambda(t)$ ) ( pour la loi Taux de défaillance ( $\lambda(t)$ ) ( pour la loi Gamma)

### 3. Estimation de la Durée Restante(RUL) pour l'Ensemble de Test :

Nous allons suivre la même approche ici pour estimer la durée de vie restante (RUL) des équipements en utilisant la distribution Gamma. Comme pour la loi Log-Normale, nous appliquerons la méthode d'estimation conditionnelle pour évaluer la durée de vie restante en fonction des cycles déjà effectués et de la fiabilité souhaitée. Nous choisirons également un seuil de fiabilité approprié pour cette estimation, et nous évaluerons la performance du modèle en utilisant des métriques telles que l'Erreur Quadratique Moyenne (RMSE), l'Erreur Absolue Moyenne (MAE) et le Coefficient de Détermination ( $R^2$ ). Les graphiques suivants montreront l'évolution de ces métriques en fonction de la fiabilité choisie pour la distribution Gamma.

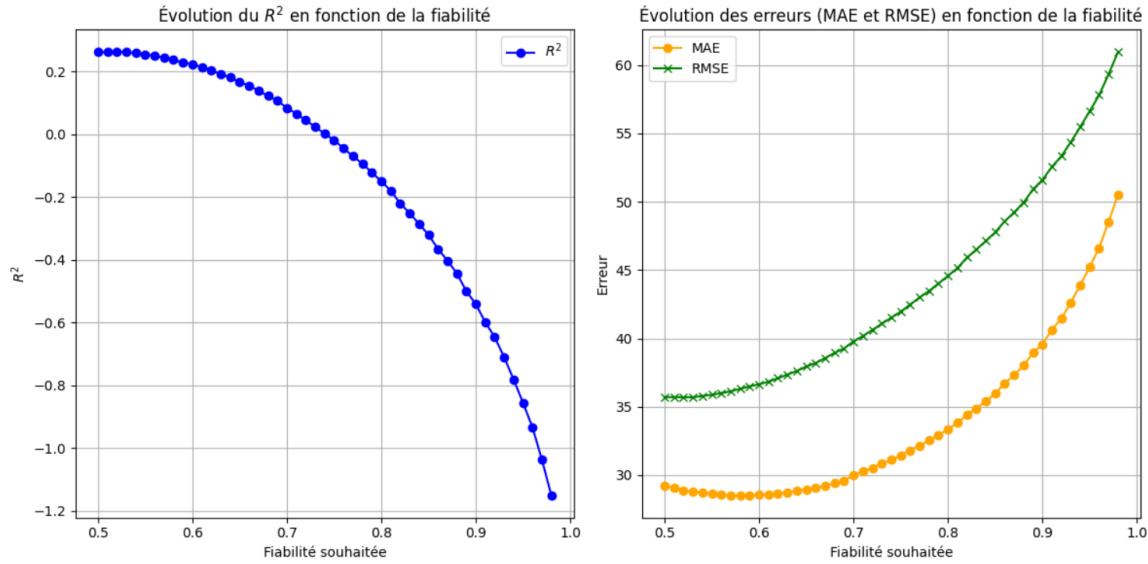


Figure 22:l'évolution de ces métriques en fonction de la fiabilité choisie pour la distribution Gamma

Pour une fiabilité de 57 %, on constate des erreurs minimales.

## VI. Comparaison des Résultats et Synthèse

### 1. Résultats

Dans cette section, nous comparons les résultats des estimations de la durée de vie restante (RUL) obtenues à partir des deux lois de probabilité : Log-Normale et Gamma, en utilisant une fiabilité fixée à 57%. Cette analyse vise uniquement à comparer les performances des deux lois dans le contexte des temps de fonctionnement, sans tenir compte des valeurs des capteurs. **Il est important** de noter que cette fiabilité est relativement faible, et normalement, une fiabilité plus élevée serait souhaitée pour assurer des prévisions plus fiables. Cependant, cette étude sert avant tout de référence pour les modélisations futures en Machine Learning (ML), qui intégreront les données de capteurs et permettront de construire des modèles plus précis.

Les résultats des deux lois sont comparés à travers les graphiques suivants, qui montrent l'évolution du RUL réel par rapport au RUL prédict par chaque loi. De plus, un tableau récapitule les erreurs de chaque modèle, incluant les valeurs de l'Erreure Absolue Moyenne (MAE), de l'Erreure Quadratique Moyenne (RMSE) et du Coefficient de Détermination ( $R^2$ ).

#### a) l'évolution du RUL réel par rapport au RUL prédict

**loi Log-Normal :**

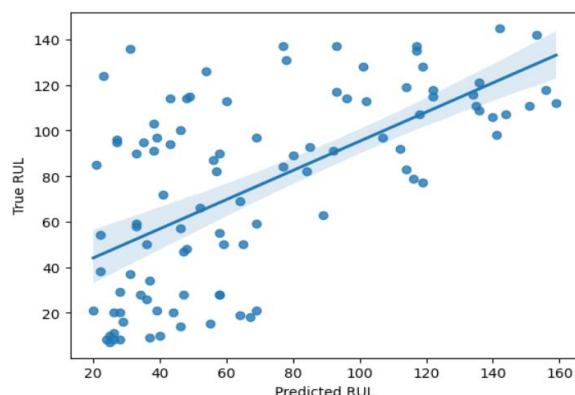


Figure 23:Comparaison entre la durée de vie restante estimée et réelle (RUL) - loi Log-Normal

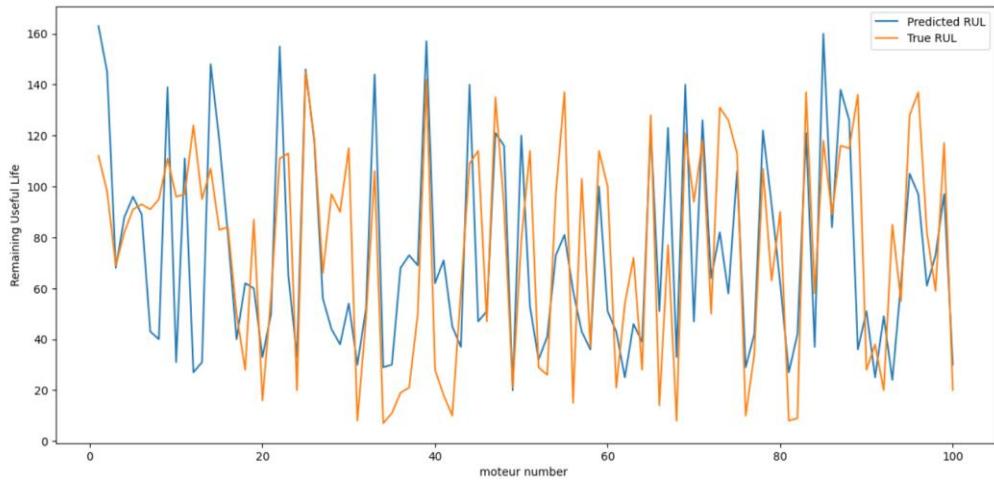


Figure 24: Évolution de la durée de vie restante estimée et réelle par moteur - loi Log-Normal

### Loi Gamma :

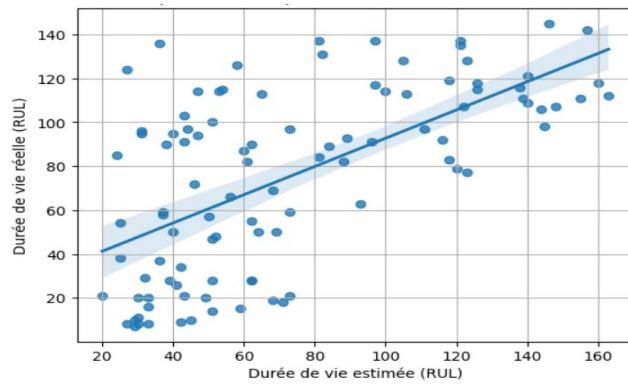


Figure 25: Comparaison entre la durée de vie restante estimée et réelle (RUL) - loi Gamma

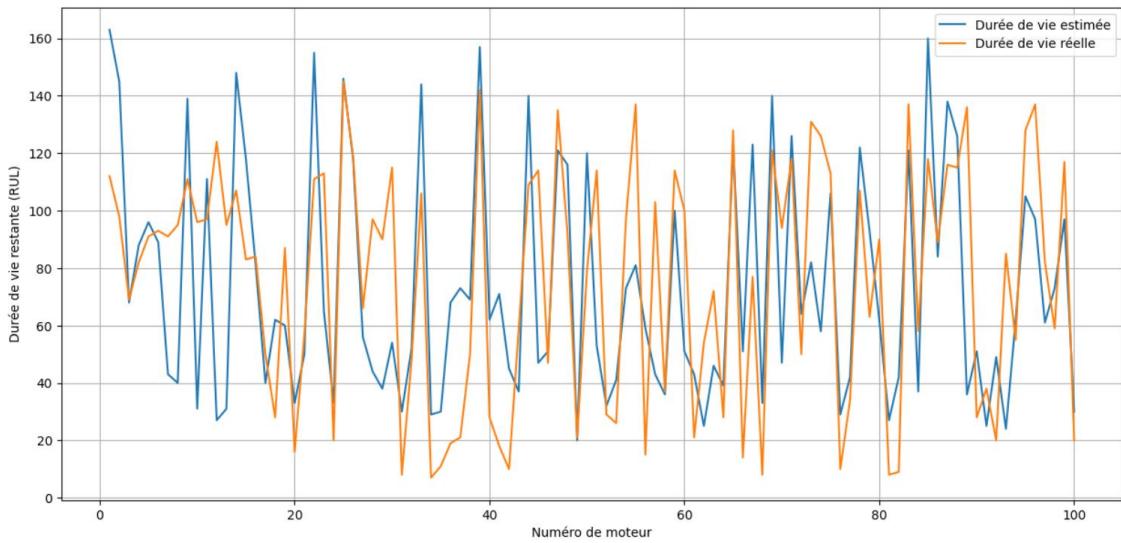


Figure 26: Figure 24: Évolution de la durée de vie restante estimée et réelle par moteur - loi Gamma

### b) Comparaison des métriques d'erreur

loi	erreur absolue moyenne (MAE)	erreur quadratique moyenne (RMSE)	coefficient de détermination ( $R^2$ )
<b>Log-Normal</b>	28.51	36.13	0.24
<b>Gamma</b>	28.78	35.69	0.26

Figure 27:métriques d'erreur pour les lois Gamma et Log-Normale à une fiabilité de 0.57

## Conclusion

Cette étude se concentre sur l'analyse des temps de fonctionnement en utilisant les lois Log-Normale et Gamma pour estimer la durée de vie restante (RUL). Bien que les résultats fournissent une première base de comparaison entre les deux lois, cette étude ne prend en compte que les temps de fonctionnement et ne tient pas compte des valeurs des capteurs.

En conclusion, cette analyse sert de **référence** pour les futures modélisations en Machine Learning, qui intégreront les données de capteurs et permettront d'améliorer les prévisions. Dans les chapitres suivants, nous approfondirons l'utilisation des technologies d'analyse de données et de Machine Learning pour la construction de modèles de maintenance plus avancés. Ces modèles permettront une prédiction plus précise de la durée de vie restante en prenant en compte des variables supplémentaires issues des capteurs.

# Chapitre 3 : Prétraitement et Exploration des Données

## Introduction

L'analyse prédictive fondée sur le machine learning requiert des données de haute qualité et des caractéristiques significatives pour assurer des résultats pertinents. Ce chapitre se concentre sur la préparation des données provenant des capteurs et leur exploration en vue de modéliser la durée de vie restante (RUL). Nous commencerons par les étapes essentielles de nettoyage et de prétraitement des données pour éliminer les incohérences, normaliser les variables et garantir leur adéquation pour l'analyse. Ensuite, nous procéderons à une sélection rigoureuse des caractéristiques, en nous basant sur des critères tels que la monotonie, la tendabilité et la prognosabilité, afin d'identifier les variables les plus informatives. Enfin, une analyse exploratoire des données (EDA) sera réalisée pour visualiser les relations entre les caractéristiques sélectionnées et le RUL, tout en identifiant les schémas et tendances utiles. Ce processus servira de fondation pour le développement des modèles de machine learning, qui seront abordés dans les chapitres suivants. L'objectif ultime de ce chapitre est de s'assurer que les données brutes sont transformées en une forme exploitable et que les caractéristiques retenues maximisent les performances des modèles prédictifs.

## I. Préparation des Données Brutes

### 1. Structure des Données et Définition des Colonnes

Voici la structure des colonnes principales :

- **Colonnes principales et paramètres d'opération :**
  - **engine** : Identifiant unique du moteur.
  - **time** : Cycle opérationnel (nombre de cycles depuis le début de la simulation).
  - **op\_setting\_1, op\_setting\_2, op\_setting\_3** : Paramètres d'opération simulés.
- **Mesures des capteurs :**
  - sm\_1 à sm\_21 : Mesures issues des capteurs.

Voici une description des capteurs :

- sm\_1 : Température à l'entrée du ventilateur.
- sm\_2 : Température à la sortie du compresseur basse pression (LPC).
- sm\_3 : Température à la sortie du compresseur haute pression (HPC).
- sm\_4 : Température à la sortie de la turbine basse pression (LPT).
- sm\_5 : Pression à l'entrée du ventilateur.
- sm\_6 : Pression dans le conduit de dérivation.
- sm\_7 : Pression à la sortie du compresseur haute pression (HPC).
- sm\_8 : Vitesse physique du ventilateur.
- sm\_9 : Vitesse physique du noyau.
- sm\_10 : Rapport de pression du moteur (P50/P2).
- sm\_11 : Pression statique à la sortie du compresseur haute pression (HPC).
- sm\_12 : Ratio du débit de carburant par rapport à Ps30.
- sm\_13 : Vitesse corrigée du ventilateur.
- sm\_14 : Vitesse corrigée du noyau.

- sm\_15 : Ratio de dérivation.
- sm\_16 : Ratio carburant-air dans le brûleur.
- sm\_17 : Enthalpie du prélèvement d'air.
- sm\_18 : Vitesse requise du ventilateur.
- sm\_19 : Vitesse de conversion requise du ventilateur.
- sm\_20 : Flux d'air de refroidissement des turbines haute pression.
- sm\_21 : Flux d'air de refroidissement des turbines basse pression.

## 2. Vérification et Nettoyage des Données

Avant d'utiliser les données pour l'analyse ou l'entraînement des modèles, il est crucial d'évaluer leur qualité. Cela inclut la vérification des valeurs manquantes, des colonnes constantes et des anomalies potentielles.

Le tableau récapitulatif ci-dessous résume les vérifications effectuées :

	missing value	number unique value	dtype
engine	0	100	int64
time	0	362	int64
op_setting_1	0	158	float64
op_setting_2	0	13	float64
op_setting_3	0	1	float64
sm_1	0	1	float64
sm_2	0	310	float64
sm_3	0	3012	float64
sm_4	0	4051	float64
sm_5	0	1	float64
sm_6	0	2	float64
sm_7	0	513	float64
sm_8	0	53	float64
sm_9	0	6403	float64
sm_10	0	1	float64
sm_11	0	159	float64
sm_12	0	427	float64
sm_13	0	56	float64
sm_14	0	6078	float64
sm_15	0	1918	float64
sm_16	0	1	float64
sm_17	0	13	int64
sm_18	0	1	int64
sm_19	0	1	float64
sm_20	0	120	float64

Figure 28: Valeurs Manquantes, Unicité et Types de données

**Constantes** : Les colonnes suivantes ont des valeurs constantes et ne contribuent pas à l'analyse :

- **op\_setting\_3, sm\_1, sm\_5, sm\_10, sm\_16, sm\_18, sm\_19:** Ces colonnes seront supprimées.

Après nettoyage, les données sont prêtes pour les prochaines étapes. Voici un aperçu des colonnes restantes :

engine	time	op_setting_1	op_setting_2	sm_2	sm_3	sm_4	sm_6	sm_7	sm_8	sm_9	sm_11	sm_12	sm_13	sm_14	sm_15	sm_17	sm_20	sm_21	
0	1	1	-0.0007	-0.0004	641.82	1589.70	1400.60	21.61	554.36	2388.06	9046.19	47.47	521.66	2388.02	8138.62	8.4195	392	39.06	23.4190
1	1	2	0.0019	-0.0003	642.15	1591.82	1403.14	21.61	553.75	2388.04	9044.07	47.49	522.28	2388.07	8131.49	8.4318	392	39.00	23.4236
2	1	3	-0.0043	0.0003	642.35	1587.99	1404.20	21.61	554.26	2388.08	9052.94	47.27	522.42	2388.03	8133.23	8.4178	390	38.95	23.3442
3	1	4	0.0007	0.0000	642.35	1582.79	1401.87	21.61	554.45	2388.11	9049.48	47.13	522.86	2388.08	8133.83	8.3682	392	38.88	23.3739
4	1	5	-0.0019	-0.0002	642.37	1582.85	1406.22	21.61	554.00	2388.06	9055.15	47.28	522.19	2388.04	8133.80	8.4294	393	38.90	23.4044
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	

Figure 29: aperçu des colonnes restantes

## II. Sélection des Caractéristiques

La sélection des caractéristiques est une étape essentielle pour identifier les variables les plus pertinentes dans la prédiction de la durée de vie restante (Remaining Useful Life, RUL). Cette démarche repose sur l'analyse des données par le biais de trois critères principaux : la **monotonie**, la **tendabilité (trendability)** et la **pronosticabilité (prognosability)**. Avant cela, nous introduisons la méthode de la **moyenne roulante** pour lisser les fluctuations aléatoires dans les données.

### 1. Importance de la Moyenne Roulante

La moyenne roulante (ou glissante) est une technique de prétraitement des données qui permet de lisser les variations aléatoires dans une série temporelle tout en conservant les tendances sous-jacentes.

$$MA(t) = \frac{1}{n} \sum_{i=t-n+1}^t x_i$$

où :

- **MA(t)** est la moyenne roulante au temps **t**,
- **n** est la fenêtre de la moyenne roulante,
- **xi** représente les valeurs de la série temporelle.

**Importance :**

- **Réduction du bruit** : En éliminant les variations aléatoires, la moyenne roulante aide à mieux visualiser les tendances générales des données.
- **Stabilisation des caractéristiques** : Elle permet d'identifier des relations cohérentes et fiables entre les variables et le temps.

### 2. Critères de Sélection des Caractéristiques

#### a) Monotonie :

La **monotonie** mesure dans quelle mesure une caractéristique suit une évolution régulière (croissante ou décroissante) au fil du temps.

$$Monotonicity(x) = \frac{1}{m} \sum_{j=1}^m \frac{|\text{Nombre de différences positives} - \text{Nombre de différences négatives}|}{n - 1}$$

Où :

- **m** : nombre d'unités,
- **n** : nombre total de points de données

Une forte monotonie indique que la caractéristique suit une tendance constante et peut être utile pour prédire la défaillance d'un système.

**Résultats de Monotonie :**

(Avant de calculer la monotone, nous avons utilisé la moyenne glissante pour lisser les données et réduire le bruit)

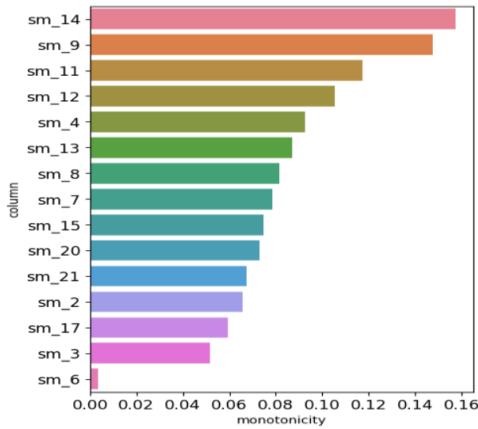


Figure 30: Scores de Monotonie par Colonne

- Les colonnes **sm\_6**, **sm\_17**, et **sm\_3** présentent une faible monotonie, ce qui indique qu'elles ne suivent pas une tendance claire (croissante ou décroissante) et pourraient être moins utiles pour la modélisation prédictive. Ces colonnes peuvent être éliminées ou examinées de manière approfondie.

### b) Tendance(Trendability):

La **tendabilité** mesure la capacité d'une caractéristique à suivre une tendance significative avec le temps.

$$Trendability(x) = \frac{1}{M} \sum_{j=1}^M |\text{corr}(x_j, t_j)|$$

Où :

- **corr( $x_j, t_j$ )** : corrélation entre la caractéristique  $x_j$  et le temps  $t_j$ .

Une forte tendabilité indique que la caractéristique est capable de capturer une tendance prévisible.

### Résultats de Tendabilité :

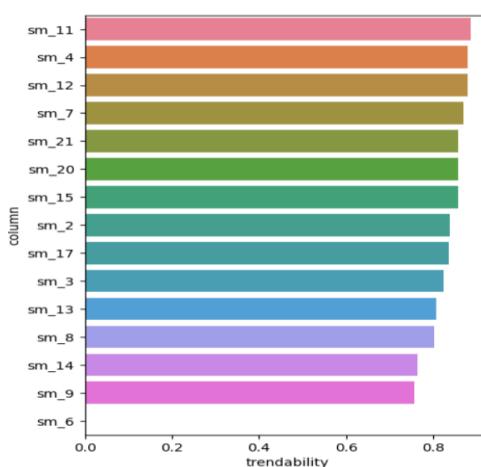


Figure 31: Scores de Tendabilité par Colonne

- La colonne **sm\_6** a un score de tendabilité nul, ce qui justifie son élimination.

### c) pronosticabilité:

La **pronosticabilité** mesure la variance d'un prédicteur en fin de vie (End of Life, EoL) pour l'ensemble des unités. Formellement, la pronosticabilité d'un prédicteur (caractéristique) est définie comme

$$\text{Prognosabilité} = \exp \left( -\frac{\text{std}_j(x_j(N_j))}{\text{mean}_j|x_j(1) - x_j(N_j)|} \right)$$

Où :

- **std<sub>j</sub>(x<sub>j</sub> (N<sub>j</sub>))** : écart-type des dernières valeurs d'une caractéristique pour l'unité j,
- **mean<sub>j</sub> |x<sub>j</sub>(1) – x<sub>j</sub>(N<sub>j</sub>)|** moyenne des différences entre les premières et dernières valeurs.

Une forte pronosticabilité indique une meilleure capacité à prévoir les défaillances futures.

### Résultats de pronosticabilité:

Ici, nous avons calculé la pronosticabilité pour les 50 derniers cycles avant la défaillance.

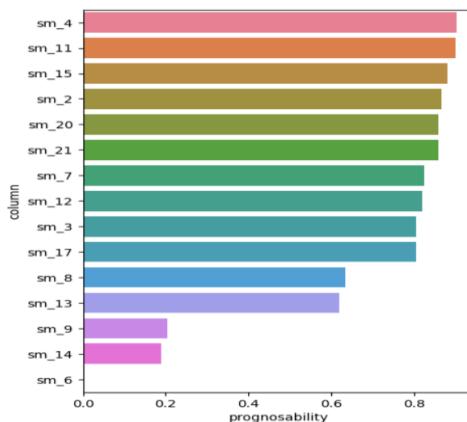


Figure 32: Scores de Prognosabilité par Colonne

- Les colonnes **sm\_6**, **sm\_9**, et **sm\_14** présentent une pronosticabilité faible, ce qui reflète une variance réduite des valeurs à la fin de vie (EoL). Cela réduit leur utilité pour prévoir les comportements futurs des équipements surveillés.

## 3. Conclusion : Sélection Finale des Caractéristiques

### Colonnes éliminées :

- sm\_6 en raison de ses faibles scores dans les trois critères (monotonie, tendabilité, pronosticabilité).

## III. Analyse Exploratoire des Données (EDA)

L'objectif principal de cette étape est de mieux comprendre les relations entre les caractéristiques sélectionnées et la durée de vie restante (RUL) des moteurs. Cela inclut l'analyse de corrélation et la visualisation des comportements des capteurs.

## 1. Analyse de Corrélation

L'analyse de corrélation permet d'identifier les relations linéaires ou non linéaires entre les capteurs sélectionnés et la durée de vie restante (RUL). Nous avons calculé la matrice de corrélation pour toutes les caractéristiques.

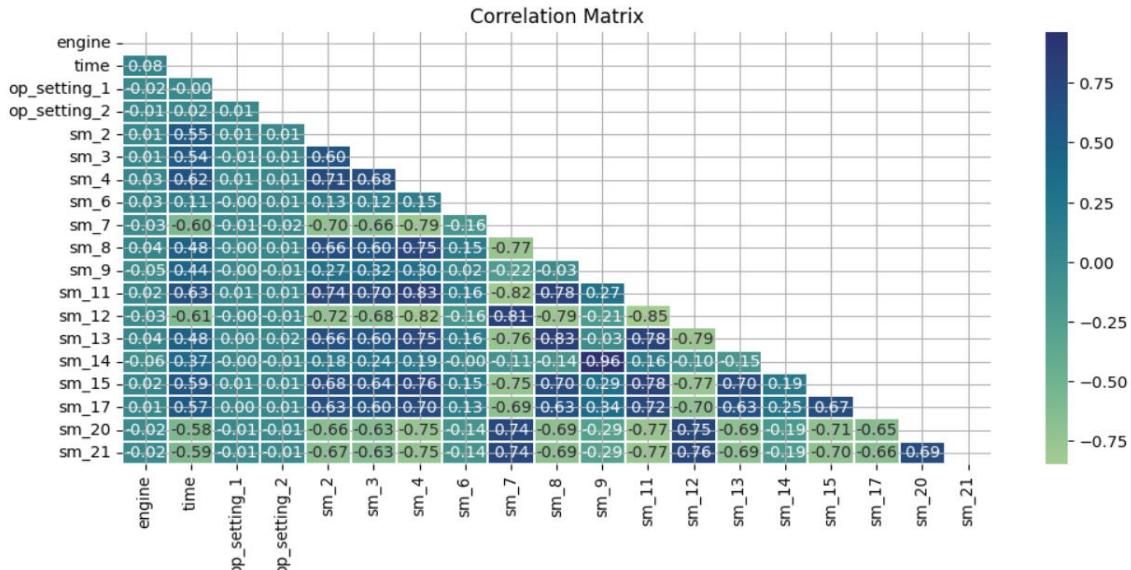


Figure 33:la la matrice de corrélation entre les colonnes

Le graphique ci-dessous montre les caractéristiques avec une corrélation élevée (> 0,9)

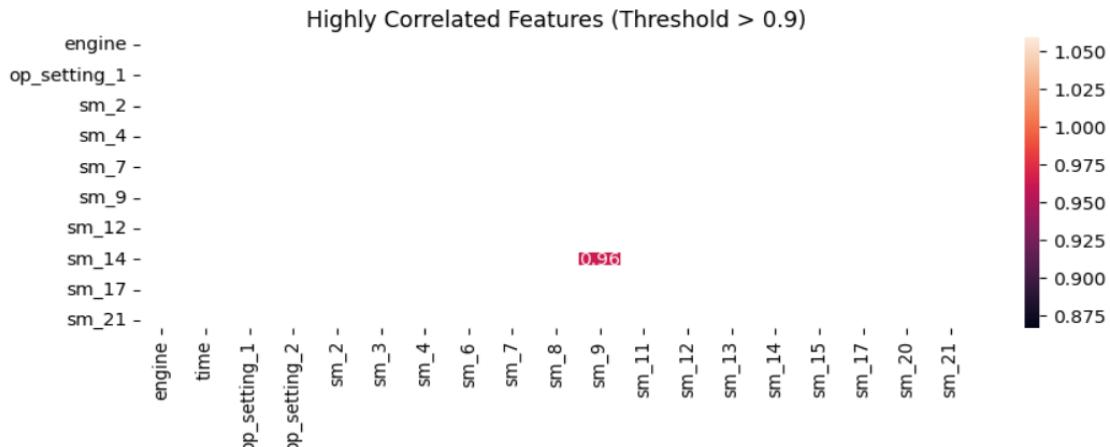


Figure 34:corrélations élevées (seuil > 0.9)

- Les colonnes **sm\_9** et **sm\_14** sont fortement corrélées avec une valeur de corrélation de 0.96

Afin de réduire la redondance dans les données, nous avons décidé d'éliminer la colonne **sm\_14**, qui est redondante par rapport à **sm\_9**.

## 2. Visualisation des Comportements des Signaux des Capteurs

Pour approfondir la compréhension de la relation entre les capteurs sélectionnés et la durée de vie restante (RUL), nous avons analysé l'évolution des signaux des capteurs lors des derniers cycles avant la défaillance. Une moyenne glissante a été utilisée pour lisser les fluctuations et révéler les tendances sous-jacentes.

Pour illustrer cette analyse, nous avons choisi d'afficher les signaux pour 10 moteurs sélectionnés aléatoirement, offrant ainsi une vue représentative des variations observées.

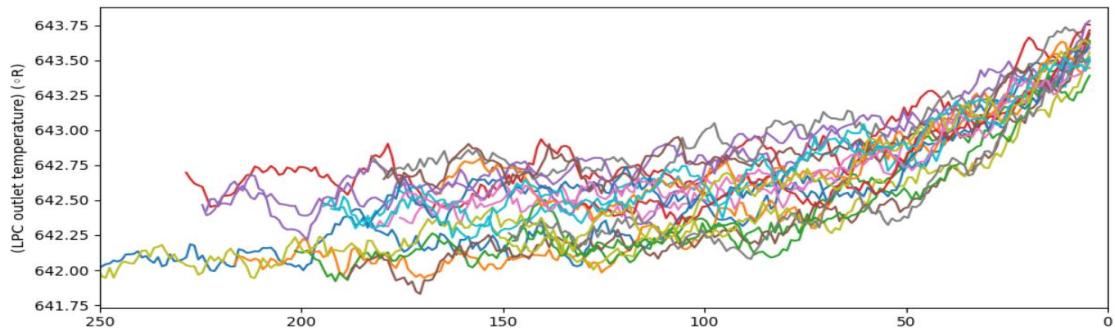


Figure 35:Température à la sortie du compresseur basse pression (LPC)

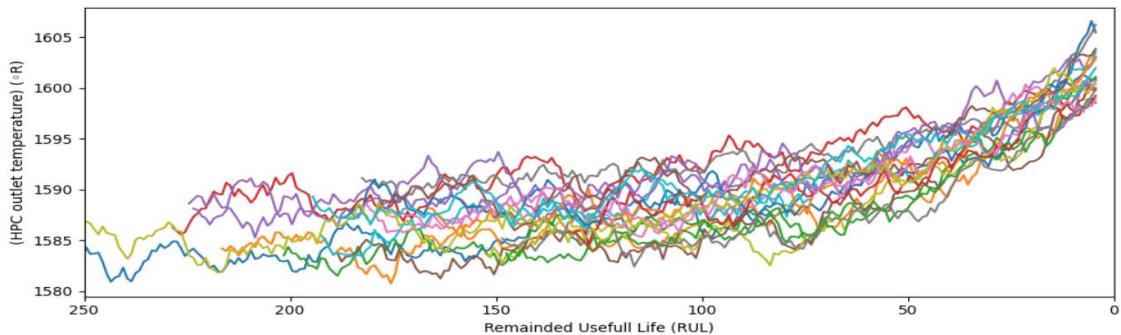


Figure 36:Température à la sortie du compresseur haute pression (HPC)

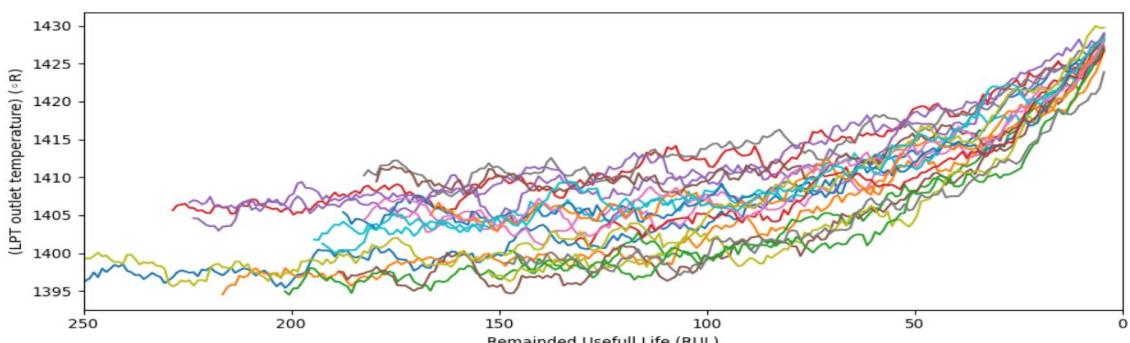


Figure 37:Température à la sortie de la turbine basse pression (LPT)

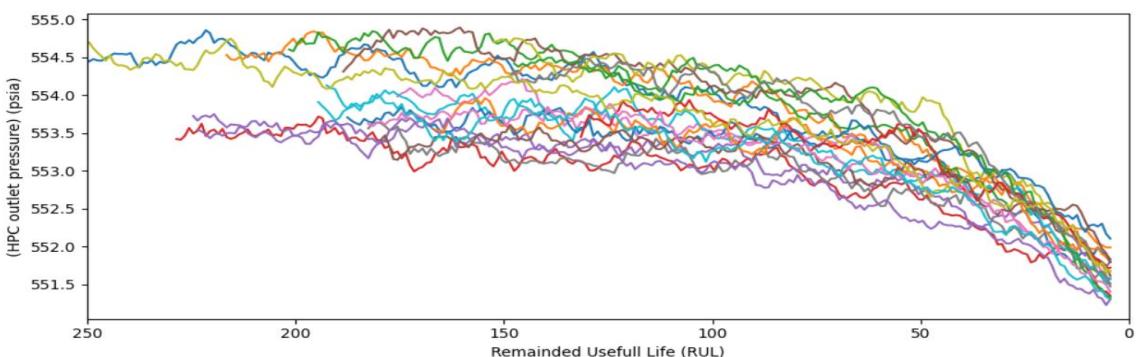


Figure 38:Pression à la sortie du compresseur haute pression (HPC)

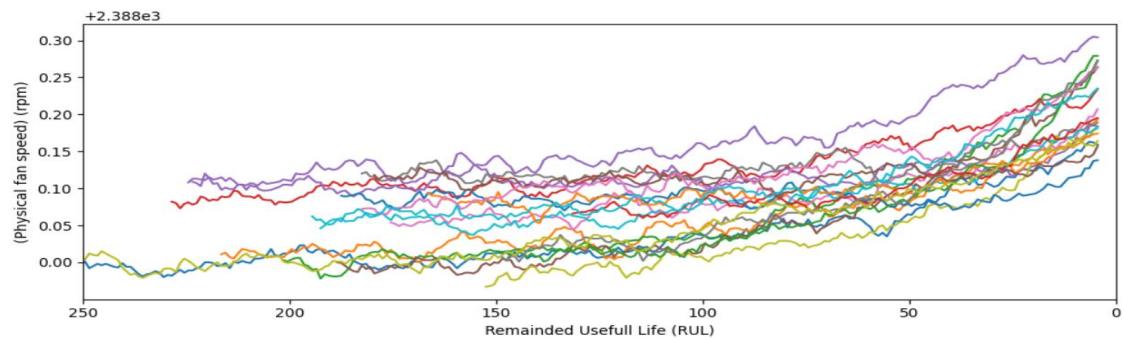


Figure 39: Vitesse physique du ventilateur

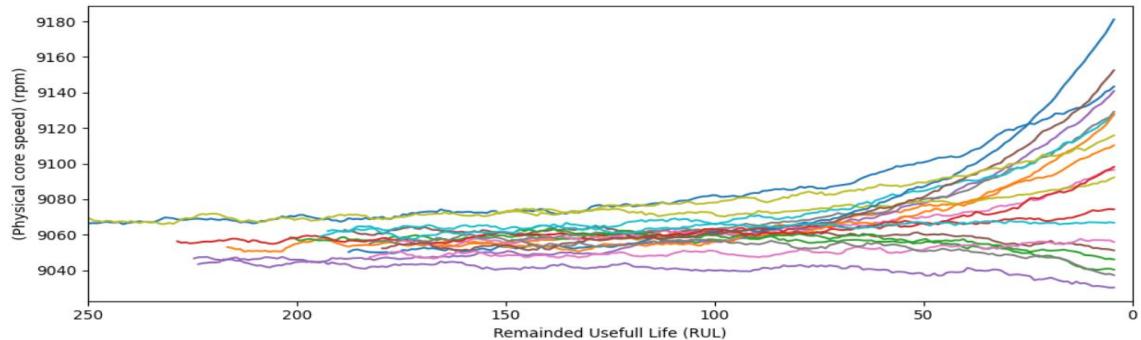


Figure 40: Vitesse physique du noyau

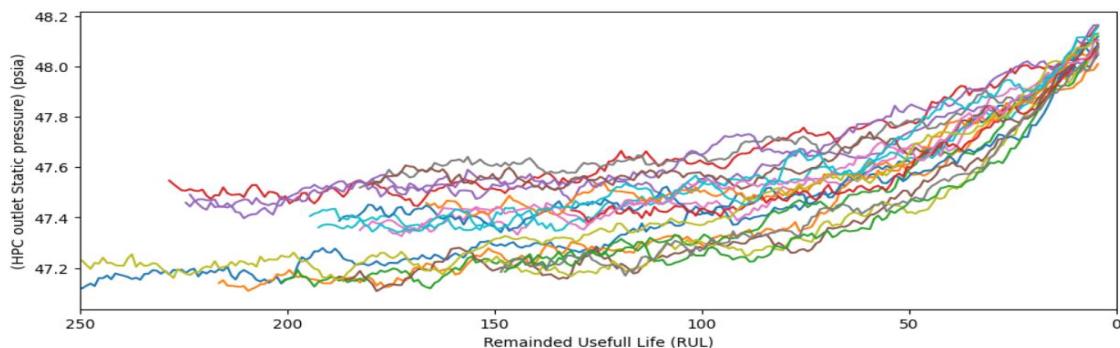


Figure 41: Pression statique à la sortie du compresseur haute pression (HPC)

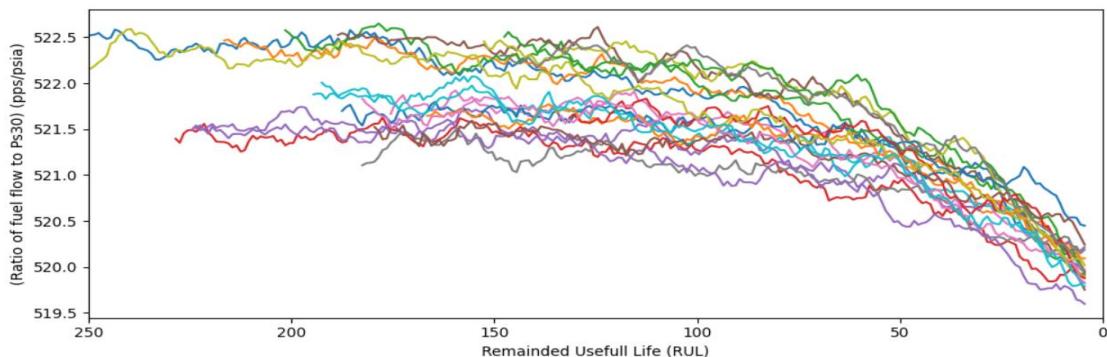


Figure 42: Rapport du débit de carburant à Ps30

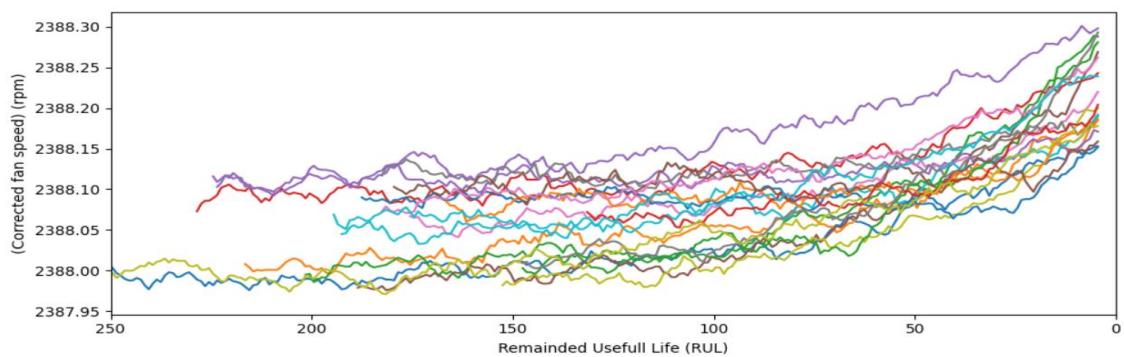


Figure 43: Vitesse corrigée du ventilateur

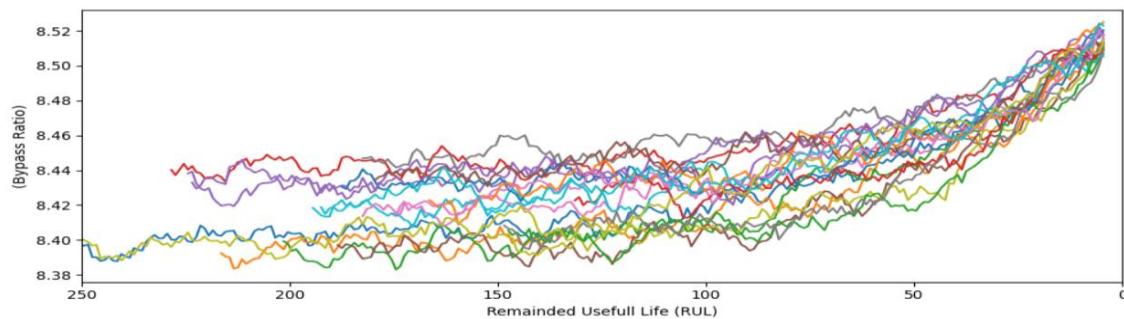


Figure 44: Rapport de dérivation

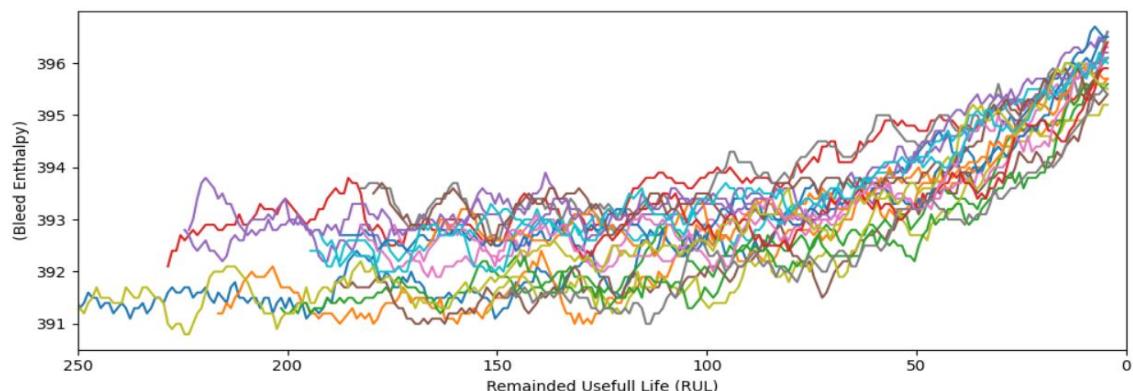


Figure 45: Enthalpie du prélèvement d'air

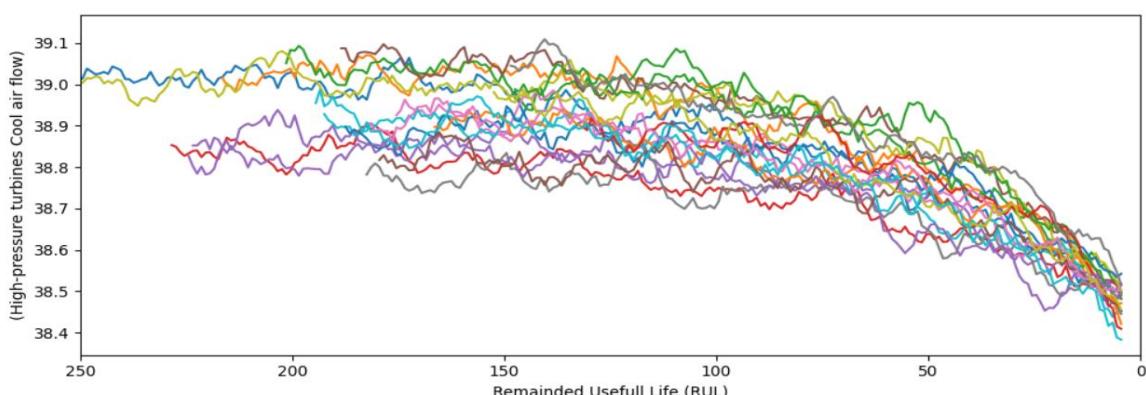


Figure 46: Flux d'air de refroidissement des turbines haute pression

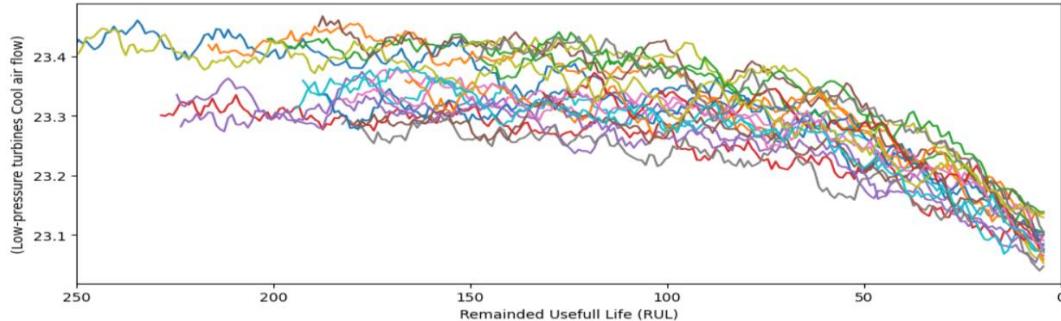


Figure 47:Flux d'air de refroidissement des turbines basse pression

### Observations sur les signaux des moteurs et des capteurs :

Nous observons un signal fort dans les derniers cycles avant la défaillance d'un moteur, ce qui peut être quantifié à l'aide de différentes techniques.

De plus, le capteur **sm\_9 (Physical core speed) (rpm)** (figure 40) semble avoir des valeurs différentes pour chaque moteur, ce qui pourrait ne pas être utile pour notre analyse. Cependant, ne le supprimons pas pour l'instant.

## Conclusion

D'après les méthodes appliquées (trendability, monotonocity, prognosability) et l'analyse des corrélations, voici les colonnes sélectionnées qui peuvent être considérées comme informatives :

- op\_setting\_1
- op\_setting\_2
- sm\_2 (Température à la sortie du compresseur basse pression)
- sm\_3 (Température à la sortie du compresseur haute pression)
- sm\_4 (Température à la sortie de la turbine basse pression)
- sm\_7 (Pression à la sortie du compresseur haute pression)
- sm\_8 (Vitesse physique du ventilateur)
- sm\_9 (Vitesse physique du noyau)
- sm\_11 (Pression statique à la sortie du compresseur haute pression)
- sm\_12 (Rapport du débit de carburant à Ps30)
- sm\_13 (Vitesse corrigée du ventilateur)
- sm\_15 (Rapport de dérivation)
- sm\_17 (Enthalpie du prélèvement d'air)
- sm\_20 (Flux d'air de refroidissement des turbines haute pression)
- sm\_21 (Flux d'air de refroidissement des turbines basse pression)

Ces caractéristiques seront essentielles pour la construction et l'entraînement des modèles de machine learning dans les chapitres à venir. Dans les sections suivantes, nous examinerons deux approches principales :

Une **approche de régression** pour estimer la durée de vie restante (RUL) des moteurs, en utilisant diverses méthodologies :

- Modélisation de la dégradation exponentielle
- RUL basé sur la similarité
- Modèles classiques tels que la régression linéaire, SVM et Random Forest

Une **approche de classification** pour évaluer l'état des moteurs, qui se décline en trois classes : Bon, Avertissement , et Danger.

# Chapitre 4 : Modélisation de la Dégradation Exponentielle pour la Prédiction de la RUL

## Introduction

Dans ce chapitre, nous abordons la modélisation de la dégradation exponentielle pour estimer la durée de vie utile restante (**RUL**) des moteurs. L'objectif est de capturer la tendance exponentielle de la dégradation en utilisant des indicateurs de santé dérivés des données.

Cette approche repose sur plusieurs étapes, depuis la préparation des données jusqu'à la prédiction de la RUL. Nous nous concentrons également sur l'identification des moteurs présentant des comportements de dégradation similaires et la définition d'un seuil de panne.

## I. Identification de l'Indicateur de Santé

Le but de cette section est de développer un indicateur de santé fiable pour surveiller l'état des équipements en fonction des données collectées.

### 1. Préparation des Données

- Pour garantir une analyse robuste, nous avons appliqué plusieurs techniques de prétraitement

**Moyenne Glissante :** La méthode de la moyenne glissante a été utilisée pour réduire le bruit dans les données brutes. Cette étape permet d'obtenir des séries temporelles lissées, facilitant l'analyse des tendances.

**Normalisation Min-Max :** Une normalisation Min-Max a ensuite été appliquée pour mettre les données à l'échelle, dans l'intervalle [0, 1]. Cette approche permet de comparer les variables sur une base commune tout en préservant les relations relatives entre les valeurs.

La relation utilisée pour la normalisation est la suivante :

$$X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

Ce choix est justifié par la simplicité d'interprétation des résultats et la stabilité des calculs.

### Sélection des Caractéristiques :

Les colonnes présentant une forte monotonie ont été sélectionnées pour refléter la dynamique des données et leur corrélation avec l'évolution du système.

Nous avons défini un seuil de monotonie **mon>0.055**, et retenu uniquement les colonnes répondant à ce critère

```
feats = mon_df.columns[mon_df.monotonicity > 0.055]
feats
✓ 0.0s
['sm_9', 'sm_11', 'sm_12', 'sm_4', 'sm_13', 'sm_8', 'sm_7', 'sm_15', 'sm_20', 'sm_21', 'sm_2', 'sm_17']
```

Cette démarche vise à ne conserver que les caractéristiques les plus pertinentes dans la construction de l'indicateur de santé.

## 2. Réduction de Dimensionnalité avec PCA :

Pour fusionner les caractéristiques sélectionnées, une Analyse en Composantes Principales (PCA) a été appliquée. Nous avons choisi de retenir les trois premières composantes principales (PC1, PC2, PC3), représentant la majorité de la variance des données.

Pour explorer les composantes principales, nous avons visualisé PC1, PC2, et PC3 entre eux

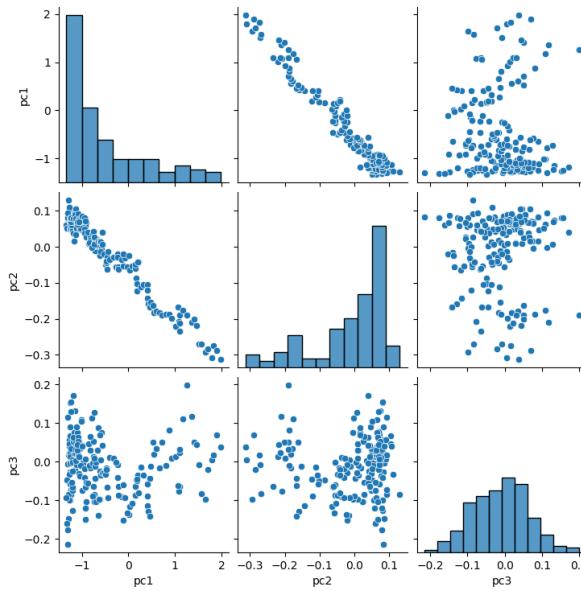


Figure 48:composantes principales

Cette visualisation a révélé une relation significative entre les deux premières composantes principales (PC1 et PC2)

### Visualisation des Composantes Principales PC1 et PC2 :

Une visualisation des relations entre les deux premières composantes principales (PC1 et PC2) a été réalisée en fonction de la durée de vie restante (RUL). L'exemple ci-dessous illustre cette relation pour un moteur particulier :

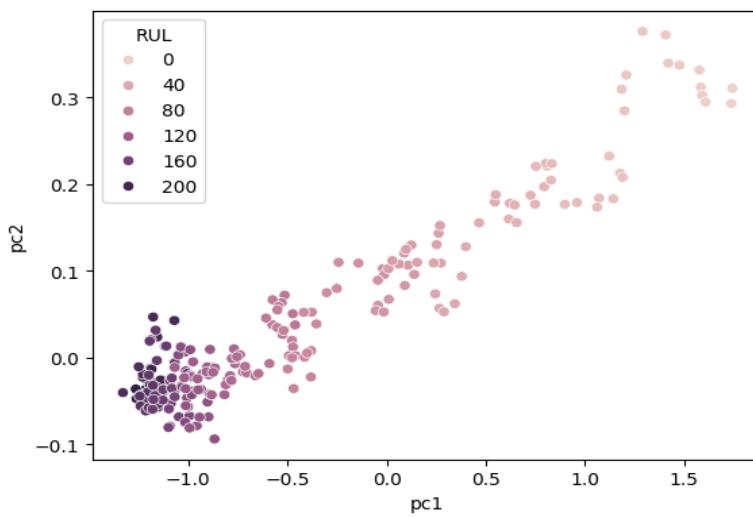


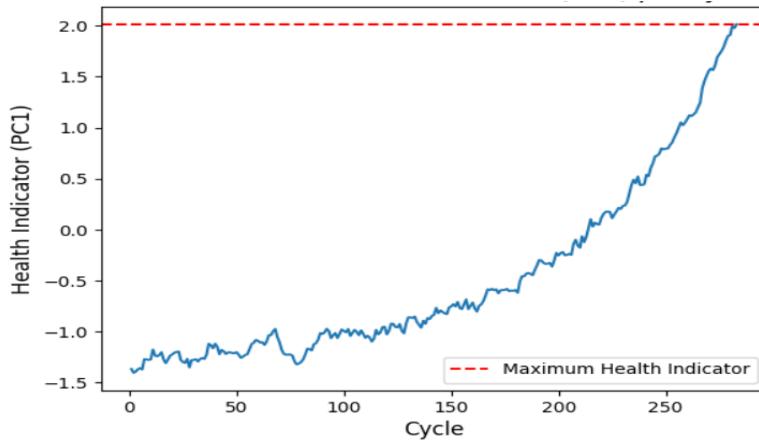
Figure 49:les composants pC1 et PC2 en fonction de RUL

Les résultats montrent une corrélation significative entre PC1, PC2 et RUL, suggérant que **PC1** peut servir d'indicateur de santé fiable.

## Utilisation de PC1 comme Indicateur de Santé :

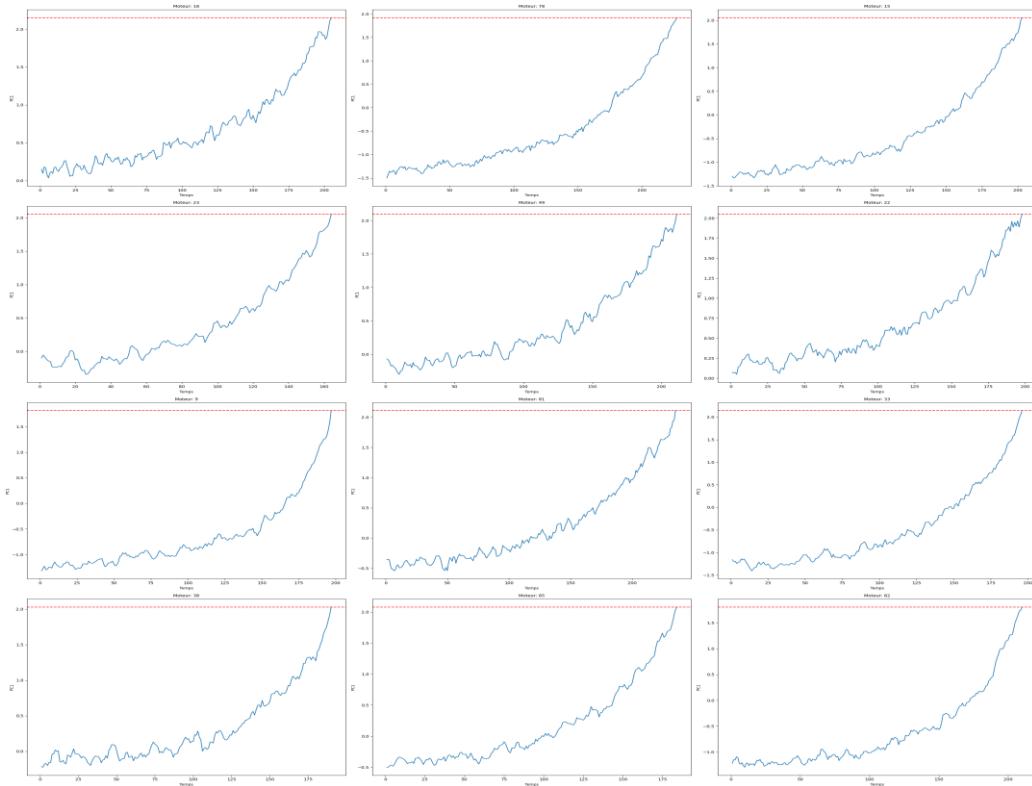
Nous avons choisi d'adopter la première composante principale (PC1) comme indicateur principal de santé.

La figure suivante illustre l'évolution de PC1 au fil des cycles pour un moteur donné :



## Visualisation des Tendances des Moteurs :

Pour mieux comprendre l'évolution des indicateurs de santé (PC1) pour différents moteurs, nous avons sélectionné 12 moteurs au hasard et avons tracé l'évolution du PC1 pour chaque moteur au fil du temps. Ce processus a permis de visualiser la tendance générale du PC1 avant et au moment de défaillance du moteur.



En observant les courbes générées pour les différents moteurs, nous constatons que le PC1 suit généralement une croissance progressive, semblable à une courbe exponentielle, indiquant une amélioration ou une stabilité de l'état du moteur. Cependant, à l'approche de la défaillance, les valeurs

du PC1 se concentrent autour de 2, ce qui marque une réduction notable de l'indicateur de santé. Ce phénomène peut être observé juste avant que le moteur ne tombe en panne, suggérant que PC1 est un indicateur pertinent pour détecter l'usure des moteurs et prédire leur défaillance.

### 3. Détermination d'un Seuil de Défaillance

#### Visualisation des Valeurs des Indicateurs de Santé selon Différentes Règles :

Afin d'explorer la distribution des valeurs de l'indicateur de santé (PC1) en fonction du Duré de vie Restante (RUL), nous avons tracé les distributions des valeurs de PC1 pour différentes valeurs de RUL. Ces visualisations permettent de mieux comprendre comment les valeurs de l'indicateur de santé évoluent à mesure que le RUL diminue.

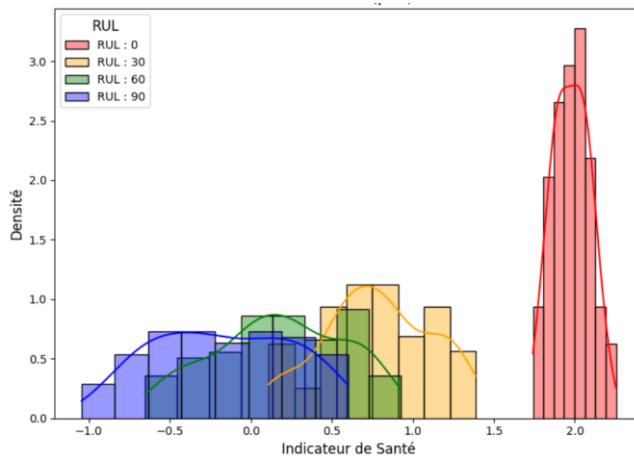


Figure 52: Distribution de l'Indicateur de Santé (pc1) selon les Niveaux de RUL

Comme nous pouvons le constater dans les graphiques ci-dessus, la dispersion des valeurs de l'indicateur de santé (PC1) diminue à mesure que le RUL approche de zéro. Cela montre que l'indicateur devient plus concentré au fur et à mesure que le moteur approche de sa défaillance, ce qui est un signe de dégradation du système.

#### Seuil de Défaillance :

Sur la base des observations précédentes, le seuil de défaillance a été déterminé à partir de la moyenne des valeurs de PC1 lorsque le RUL est égal à zéro.

Ce seuil correspond à une valeur de  $\mu = 1.97$ , avec un écart-type associé de  $\sigma = 0.11$ .

## II. Modèle de l'Indicateur de Santé (Exponentiel)

Dans cette section, nous appliquons un modèle exponentiel pour modéliser l'évolution de l'indicateur de santé (Health Indicator, HI) au fil du temps pour chaque moteur. Ce modèle a pour but de capturer la dégradation de l'indicateur de santé en fonction du temps, en s'appuyant sur la forme exponentielle de la courbe de dégradation des moteurs.

### 1. modélisation

Le modèle exponentiel que nous utilisons est défini par l'expression suivante :

$$HI(t) = A + B \cdot e^{\lambda t}$$

Où  $A$ ,  $B$ , et  $\lambda$  sont des constantes qui caractérisent respectivement le niveau de base, l'amplitude, et le taux de dégradation exponentielle.

L'objectif principal de cette modélisation est d'ajuster les paramètres  $A$ ,  $B$ , et  $\lambda$  pour chaque moteur à partir des données de son cycle et de l'**indicateur de santé pc1**. Nous cherchons à minimiser l'erreur entre les valeurs observées de l'indicateur de santé et celles prédictes par le modèle.

## 2. Ajustement du Modèle pour Chaque Moteur

- Pour chaque moteur de notre jeu de données, nous avons utilisé la **méthode des moindres carrés** pour ajuster le modèle exponentiel.

- L'optimisation permet de déterminer les paramètres  $A$ ,  $B$ , et  $\lambda$  qui minimisent la différence entre les données observées et les prédictions du modèle. Cette procédure a été appliquée à tous les moteurs, de manière à obtenir les paramètres de dégradation pour chaque moteur.

engine		A	B	lambda_
0	1	-0.880568	0.075135	0.019556
1	2	-1.337524	0.068601	0.013763
2	3	-0.937691	0.024805	0.026614
3	4	-0.516098	0.010315	0.029437
4	5	-1.283207	0.062999	0.014831
...	...	...	...	...
95	96	-0.859759	0.077333	0.010943
96	97	-0.747404	0.004845	0.031526
97	98	-1.213762	0.122798	0.021208
98	99	-0.876814	0.056936	0.021865
99	100	-0.608983	0.093579	0.017105

Figure 53:les paramètres A, B, et  $\lambda$  pour chaque moteur

Le graphique ci-dessous présente la comparaison entre l'indicateur de santé observé et l'indicateur de santé prédit pour un moteur donné, ainsi que la ligne de seuil de défaillance

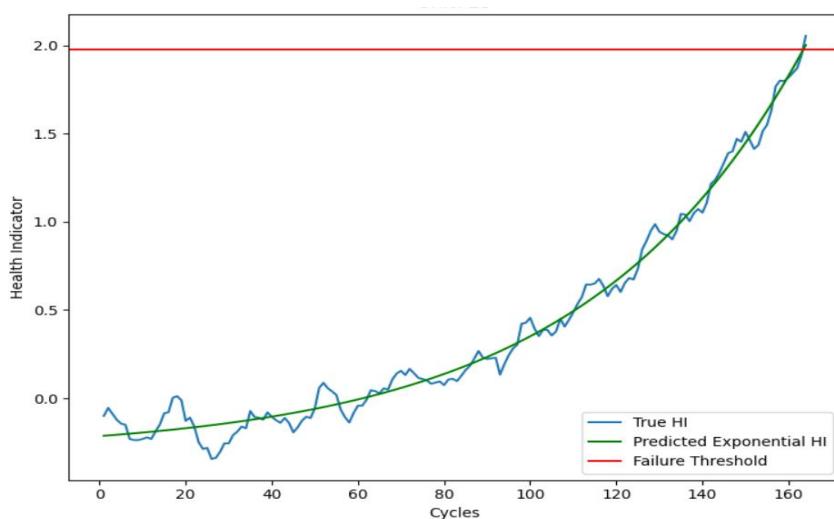


Figure 54:l'indicateur de santé observé et prédit pour un moteur donné (moteur 23)

- L'application de ce modèle exponentiel permet de simuler la dégradation des moteurs au fil du temps et de prédire l'évolution de leur indicateur de santé.

- On peut utiliser cette modélisation pour anticiper les pannes et améliorer la gestion de la maintenance des moteurs.

### III. Prédiction de la Durée de Vie Restante (RUL)

Cette section se concentre sur la prédiction de la **Durée de Vie Restante (RUL)** pour les unités de test en utilisant un modèle de dégradation exponentielle.

#### 1. Méthodologie :

##### Étape 1 - Préparation des données de test :

- Nous avons commencé par préparer les données de test en effectuant l'Analyse en Composantes Principales (PCA) et en extrayant la première composante principale (**PC1**).
- Les PCA ont été calculés en utilisant les mêmes colonnes que celles utilisées pour l'entraînement (les colonnes sélectionnées présentant une monotonie supérieure à **0.055**), afin d'assurer une cohérence.
- Les données ont ensuite été normalisées à l'aide de la méthode Min-Max, suivant exactement le même procédé appliqué aux données d'entraînement.

##### Étape 2 - Ajustement du modèle :

Pour chaque moteur de **test**, nous avons ajusté un modèle exponentiel en utilisant les paramètres (**A**, **B**,  $\lambda$ ) obtenus à partir des **données d'entraînement** comme référence.

Les paramètres sont ajustés à l'aide des données d'indicateur de santé (**pc1**) et des cycles , avec des bornes définies par les valeurs des **25e** et **75e** percentiles des modèles exponentiels obtenus à partir des **données d'entraînement**.

**Remarque :** L'utilisation des paramètres obtenus à partir de l'entraînement comme référence permet de tirer parti des connaissances déjà acquises sur les comportements typiques des moteurs. Cela garantit une certaine cohérence entre les modèles appliqués aux données d'entraînement et de test.

De plus, en définissant les bornes des paramètres selon les 25e et 75e percentiles, nous avons cherché à capturer la variabilité naturelle des données tout en évitant les valeurs extrêmes ou aberrantes. Cette approche assure que l'ajustement reste réaliste et représentatif, tout en évitant un sur-ajustement ou une sous-estimation des dynamiques réelles des moteurs

##### Étape 3 - Calcul du nombre total de cycles jusqu'à la défaillance:

Nous avons calculé le nombre total de cycles nécessaires pour que l'indicateur de santé atteigne le **seuil de défaillance** à l'aide de la formule :

$$\text{total\_cycles} = \frac{\log\left(\frac{\text{threshold}-A}{B}\right)}{\lambda}$$

##### Étape 4 - Estimation de la RUL :

- La Durée de Vie Restante (**RUL**) est calculée comme suit :

$$RUL = \text{total\_cycles} - \text{cycle actuel}$$

Répéter les **étapes 2 à 4** pour toutes les moteurs de test.

Le tableau ci-dessous montre un aperçu des résultats obtenus pour quelques moteurs de Test.

engine	A	B	lambda_	Pred_RUL	True_RUL
0	-0.858803	0.079980	0.023414	125.422608	112
1	-0.303772	0.079980	0.023414	98.124962	98
2	-0.249964	0.079980	0.019196	51.332748	69
3	-0.364340	0.079980	0.019606	70.261633	82
4	-0.282076	0.054107	0.023406	65.458355	91

Figure 55: Comparaison des Prédictions de RUL pour Quelques Moteurs de Test

Le graphique suivant présente les résultats pour un moteur spécifique sélectionné aléatoirement. Il compare l'évolution réelle de l'indicateur de santé (**HI**) avec la dégradation prédictive par un modèle exponentiel, et inclut le seuil de défaillance ainsi que le RUL réel pour le moteur en question.

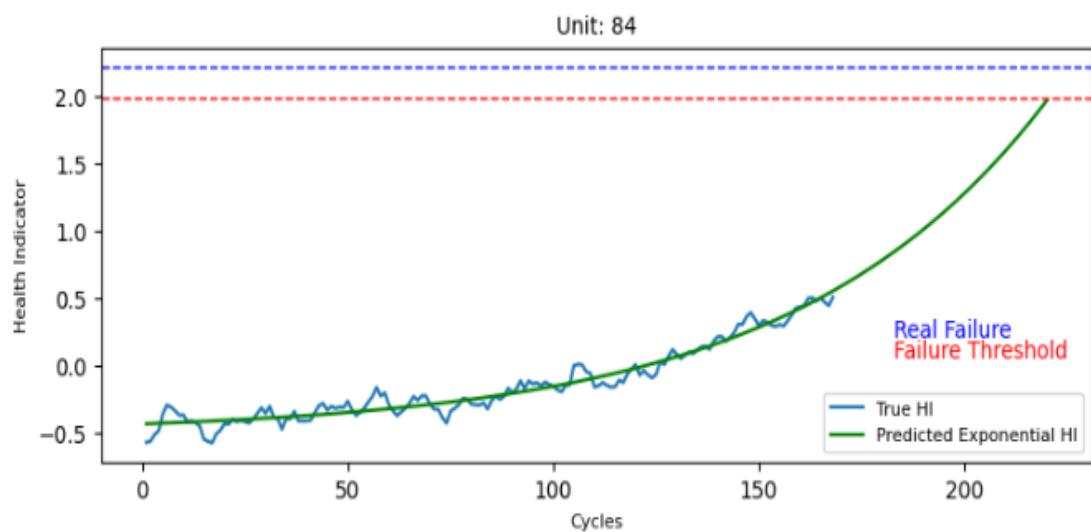


Figure 56:Comparaison entre l'Indicateur de Santé Réel et la Dégradation Prédite pour le Moteur 84

## 2. Résultats et Analyse des Performances

### 1. Comparaison entre les RUL prédict et réel :

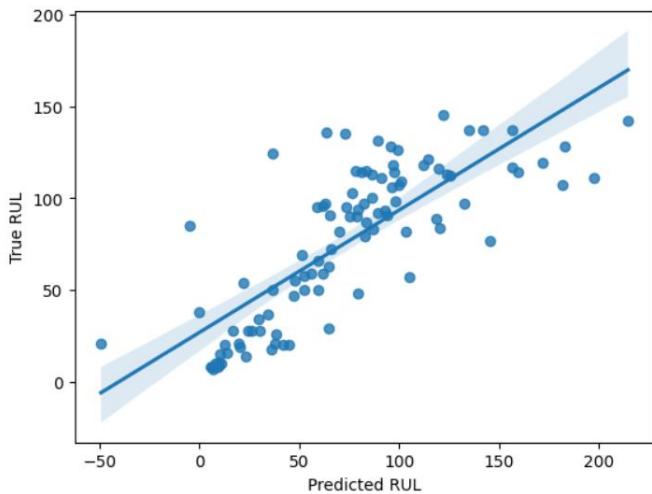


Figure 57: Comparaison entre la durée de vie restante ((RUL) Prédit et réelle

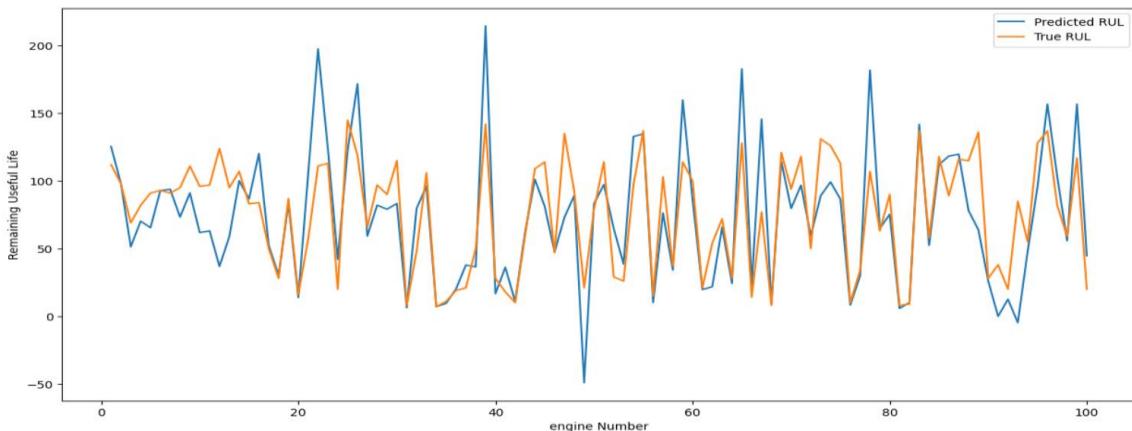


Figure 58: Comparaison entre la durée de vie restante (RUL) Prédit et réelle pour Chaque Moteur

## 2. métriques d'erreur

métrique	erreur absolue moyenne (MAE)	erreur quadratique moyenne (RMSE)	coefficient de détermination ( $R^2$ )
valeur	20.95	30.51	0.46

Figure 59: Évaluation des Performances du Modèle : RMSE, RAE et  $R^2$

## 3. Analyse des résultats

- Amélioration par rapport à une approche basée uniquement sur le temps de fonctionnement (loi gamma et log-normal) :

Cette méthode utilise l'historique des indicateurs de santé pour ajuster un modèle exponentiel, ce qui permet de mieux capturer les tendances de dégradation. Cela se traduit par des prévisions de RUL plus fiables, surtout pour les moteurs ayant fonctionné un nombre significatif de cycles.

- Limites pour les moteurs en début de vie :

Lorsque l'historique est limité (par exemple, un moteur ayant fonctionné seulement 10 cycles), le modèle ne dispose pas d'une tendance de dégradation claire, rendant les prévisions moins précises. En revanche, pour un moteur ayant fonctionné 100 cycles ou plus, l'indicateur de santé montre une tendance croissante bien définie, ce qui permet au modèle d'estimer le RUL de manière plus précise.

Les graphiques suivants comparent le RUL prédict et le RUL réel d'un moteur dans deux scénarios différents : le premier graphique présente un moteur ayant effectué plus de 200 cycles, tandis que le deuxième graphique concerne un moteur ayant réalisé moins de 30 cycles. Ces deux graphiques montrent comment le RUL prédict évolue par rapport au RUL réel dans chaque cas.

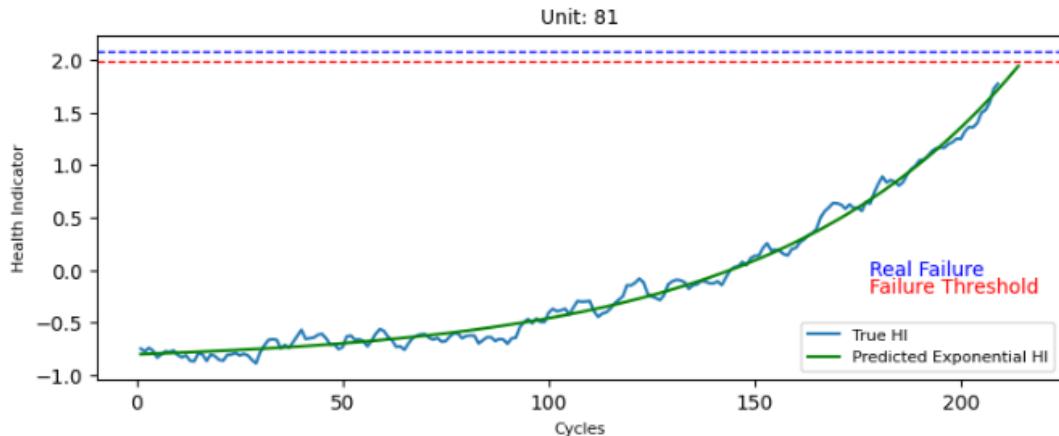


Figure 60:RUL Prédit et Réel pour un Moteur avec Plus de 200 Cycles

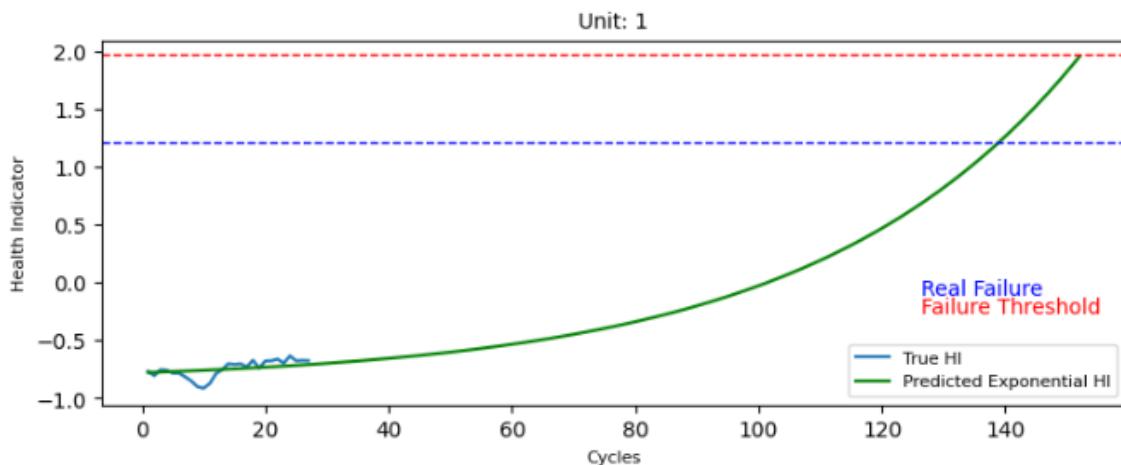


Figure 61:Prédit et Réel pour un Moteur avec moins de 30 Cycles

## Conclusion :

En conclusion, l'approche basée sur l'historique des indicateurs de santé et l'ajustement d'un modèle exponentiel présente des avantages considérables par rapport à une méthode simple reposant uniquement sur le temps de fonctionnement, notamment pour les moteurs ayant effectué un nombre conséquent de cycles.

Cette approche, bien que moins précise pour les moteurs moins âgés, bénéficie d'une grande flexibilité qui permet d'ajuster les prévisions en fonction des caractéristiques spécifiques des moteurs à différents stades de leur cycle de vie. Elle démontre ainsi une forte valeur prédictive et constitue une méthode robuste et prometteuse pour les prévisions de RUL.

# Chapitre 5 : Prédition du RUL Basée sur l'Analyse de Similarités Multi-Facteurs

## Introduction :

Après avoir abordé l'approche de la modélisation de la dégradation exponentielle, ce chapitre se concentre sur une méthodologie avancée : l'**Analyse de Similarités Multi-Facteurs** (ASMF). Cette approche permet de prédire le RUL de manière plus précise en intégrant plusieurs paramètres mesurés à partir de capteurs, notamment ceux liés aux **Conditions d'Activité** (CA) : l'ensemble des paramètres de la machine contrôlables par les opérateurs (vitesse, pression, alimentation) et aux **États du Système** (ES) : les grandeurs physiques intrinsèques à l'équipement non contrôlable par l'opérateur (vibration, qualité d'huile, température résiduelle).

L'ASMF repose sur l'idée que la similarité entre différentes moteurs ou différents états d'une même moteur, observée à travers des séries temporelles, peut fournir des informations cruciales pour prédire la durée de vie restante. Cette méthodologie s'appuie sur plusieurs étapes clés :

**Identification des Régimes de Fonctionnement** : Dans un premier temps, les moteurs sont regroupées en fonction de leurs conditions de fonctionnement, à savoir les **CA**, afin de normaliser les données selon chaque régime de fonctionnement.

**Identification des Paramètres Prépondérants** : Une fois les régimes de fonctionnement identifiés, il est crucial de sélectionner les capteurs les plus pertinents pour prédire la fin de vie des moteurs. Cette sélection se base sur les **corrélations** et les **tendances** des données des capteurs, en isolant les paramètres ayant le plus grand impact sur la durée de vie restante.

**Fusion des Données** : L'étape suivante consiste à agréger les données provenant de différents capteurs pour générer un **indicateur de survie** global pour chaque moteur. Cet indicateur est essentiel pour mesurer l'état de santé de la moteur au fil du temps et pour prédire la probabilité de défaillance.

**Prédiction du RUL** : Enfin, la prédition du RUL est réalisée en utilisant une approche basée sur les **k voisins les plus proches** (k-NN). En comparant les courbes de survie des différentes moteurs, cette méthode permet d'identifier celles dont les comportements sont les plus similaires à celui de la machine en cours d'analyse, fournissant ainsi une estimation fiable du RUL.

## I. Analyse de Similarités Multi-Facteurs (ASMF)

### 1. Identification des Régimes de Fonctionnement

Dans notre cas, les **Conditions d'Activité (CA)** sont représentées par les paramètres **op\_setting\_1** et **op\_setting\_2**. Dans notre cas il existe un seul régime de fonctionnement. Par conséquent, une normalisation des données a été réalisée à l'aide de la méthode de **normalisation en Z** afin de garantir l'homogénéité des valeurs des différents paramètres.

La **normalisation en Z**, également connue sous le nom de normalisation standard, est une technique utilisée pour transformer les données afin qu'elles aient une moyenne de 0 et un écart-type de 1. Cela permet de standardiser différentes caractéristiques en éliminant les écarts dus à leurs différentes échelles.

Pour une donnée  $x_i$  d'une caractéristique donnée  $\mathbf{X}$ , la normalisation en **Z** est définie par la formule suivante :

$$z_i = \frac{x_i - \mu}{\sigma}$$

Où :

$z_i$ : valeur normalisée (ou score Z) pour  $x_i$ .

$x_i$  : valeur originale de la donnée.

$\mu$  : moyenne des valeurs de la caractéristique X.

$\sigma$  : écart-type des valeurs de la caractéristique X.

## 2. Sélection des Paramètres Prépondérants

La sélection des paramètres a été réalisée lors de la sélection des **caractéristiques informatives** dans le **chapitre 4**. Les paramètres retenus sont considérés comme les plus pertinents pour la prédiction de la durée de vie restante. Ils servent d'entrée pour les étapes de fusion et de modélisation

## 3. Fusion des Données

Dans un premier temps, nous avons élaboré un modèle de dégradation linéaire pour intégrer les différentes caractéristiques identifiées.

- Supposons que l'état de santé d'une unité se dégrade de manière linéaire. Pour modéliser la dégradation d'une unité de manière linéaire, on peut définir l'**Indicateur de Santé (HI)** comme suit :

$$HI = 1 - \frac{\text{Cycle Actuel}}{\text{Cycle Maximal}}$$

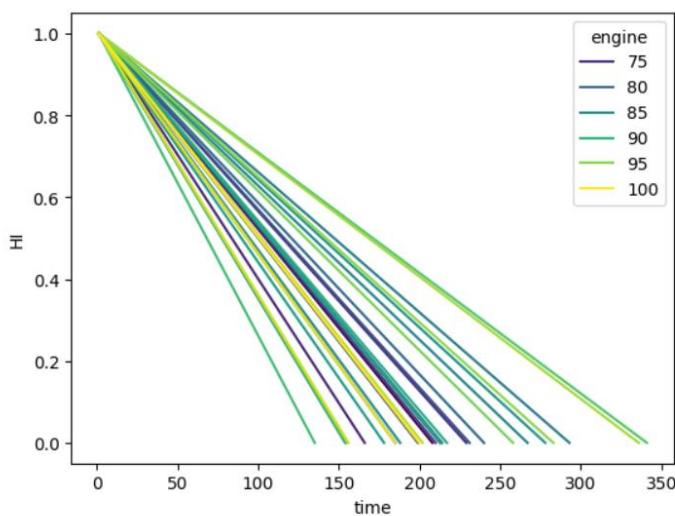


Figure 62: Indicateur de Santé Initial pour les moteur 70 à 100

### Construction d'un Modèle de Régression Linéaire pour l'Indicateur de Santé :

Un modèle de régression linéaire est utilisé pour modéliser la relation entre les caractéristiques sélectionnées et l'indicateur de santé (HI).

$$\text{New HI} = \sum_{i=1}^n (\text{coef}_i \times \text{feature}_i)$$

Les coefficients de la régression linéaire, obtenus lors de l'entraînement du modèle, servent à pondérer les caractéristiques.

Ces coefficients illustrent l'impact relatif de chaque caractéristique sur l'indicateur de santé et montrent leur contribution à la dégradation de l'unité. Une caractéristique fortement corrélée à la dégradation recevra un coefficient plus élevé, tandis qu'une caractéristique moins influente aura un coefficient plus faible.

Cette méthode assure que le **nouvel indicateur de santé** reflète fidèlement l'état de l'unité en prenant en compte les informations des caractéristiques les plus pertinentes.

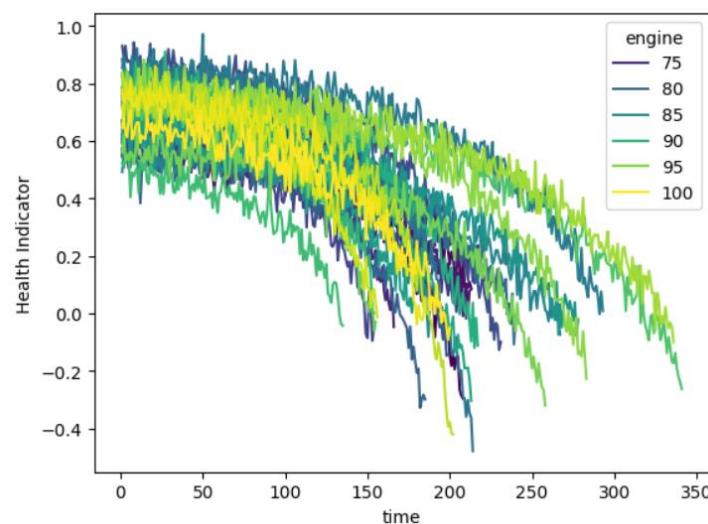


Figure 63: Indicateur de Santé Final pour les Unités 71 à 100

### Lissage de l'Indicateur de Santé Final :

Un lissage a été appliqué en calculant une moyenne mobile de l'indicateur de santé final

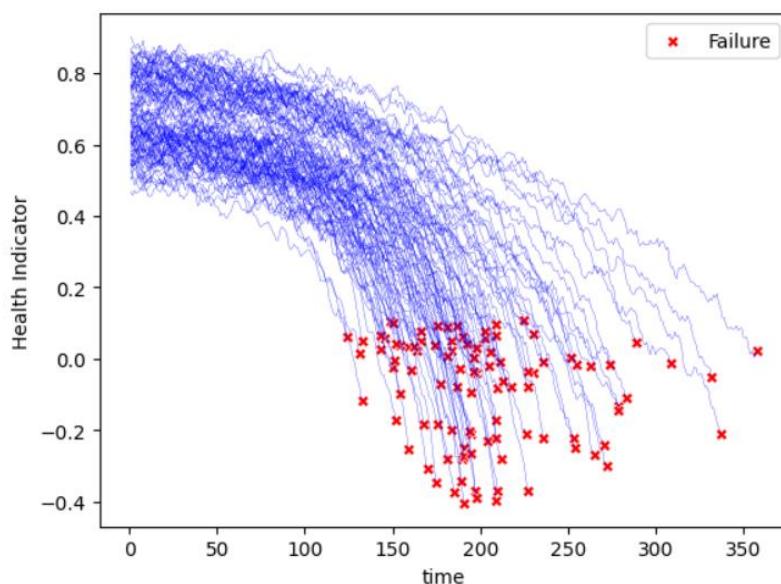


Figure 64: Indicateur de Santé Final et Points de Défaillance

## Modélisation de l'Évolution de l'Indicateur de Santé

Dans cette étape, une **fondction polynomiale de second ordre** est ajustée sur les indicateurs de santé (HI) afin de modéliser leur évolution au fil des cycles de fonctionnement. La fonction est définie par l'équation suivante :

$$y = \theta_0 + \theta_1 x + \theta_2 x^2$$

où :

- **y** représente l'indicateur de santé (**HI**).
- **x** correspond au cycle de fonctionnement.
- $\theta_0, \theta_1, \theta_2$  sont les coefficients du polynôme, déterminés par un ajustement sur les données réelles

### Estimation des Coefficients :

Les coefficients du polynôme sont calculés à l'aide de la méthode des **moindres carrés**, qui consiste à minimiser la somme des erreurs quadratiques entre les valeurs observées et celles prédites par le modèle. Cette méthode permet de déterminer les coefficients optimaux ( $\theta_0, \theta_1, \theta_2$ ) qui représentent l'ajustement le plus précis de la courbe

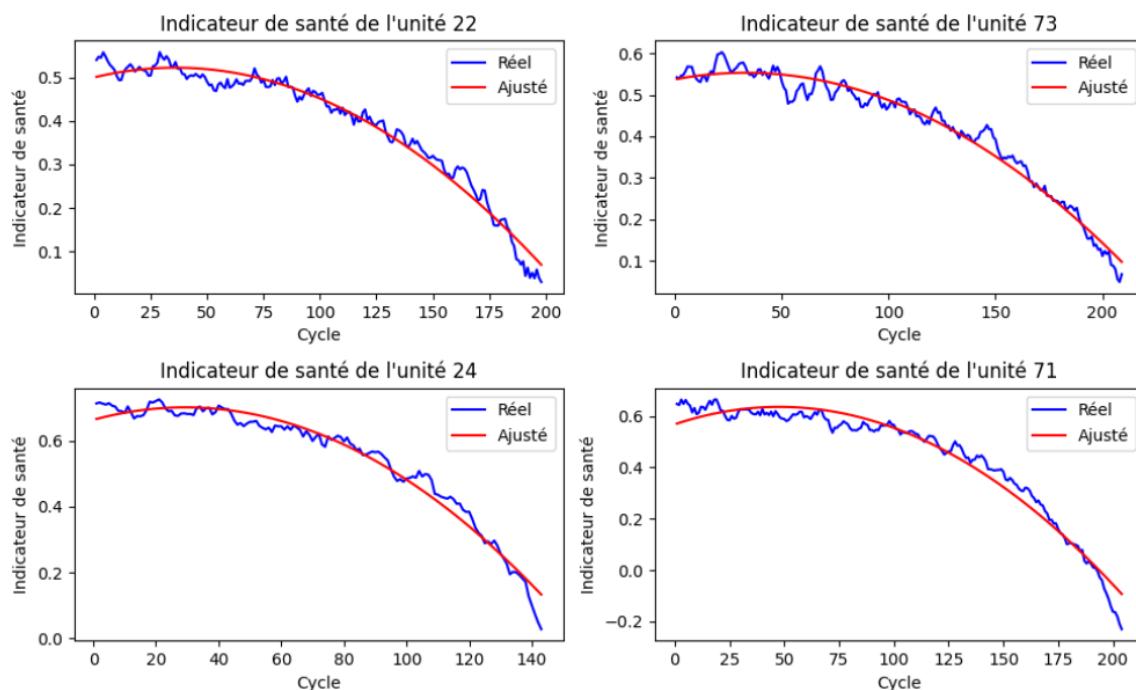


Figure 65:Comparaison de l'Indicateur de Santé Réel et Ajusté pour différentes moteurs

## 4. Prédiction du RUL : Construction d'un modèle RUL basé sur la similarité

L'objectif de cette étape est de prédire la durée de vie restante (RUL) des moteurs en se basant sur la similarité des indicateurs de santé entre les moteurs du jeu de données d'entraînement et de test. Cette approche repose sur la comparaison des modèles polynomiaux ajustés aux indicateurs de santé, et sur l'utilisation de la méthode des **voisins les plus similaires** pour prédire le RUL.

## Étapes Méthodologiques

- **Générer un indicateur de santé :** en multipliant les poids des caractéristiques sélectionnées dans les données de test.

$$HI = \sum_{i=1}^n (\text{coef}_i \times \text{feature}_i)$$

- **Application de la moyenne roulante sur l'indicateur de santé :** Afin de réduire le bruit et d'obtenir une courbe plus lissée de l'indicateur de santé, nous appliquons une moyenne roulante sur l'indicateur de santé calculé pour chaque moteur

- **Comparer les modèles polynomiaux ajustés avec les indicateurs de santé des données de test :** et trouver les résidus en norme 1 en utilisant l'équation ci-dessous

$$d(i, j) = \|y_j - \hat{y}_{j,i}\|_1$$

Où :

$\hat{Y}_{j,i}$ : l'indicateur de santé ajusté prédit par le modèle polynomial pour l'unité **j** et le moteur **i**

- **Calculer le score de similarité** en utilisant la formule suivante :

$$\text{score}(i, j) = \exp(-d(i, j)^2)$$

- **Sélection des modèles les plus similaires :**

Une fois que nous avons calculé les scores de similarité, nous sélectionnons les **K** modèles les plus similaires pour chaque moteur de test

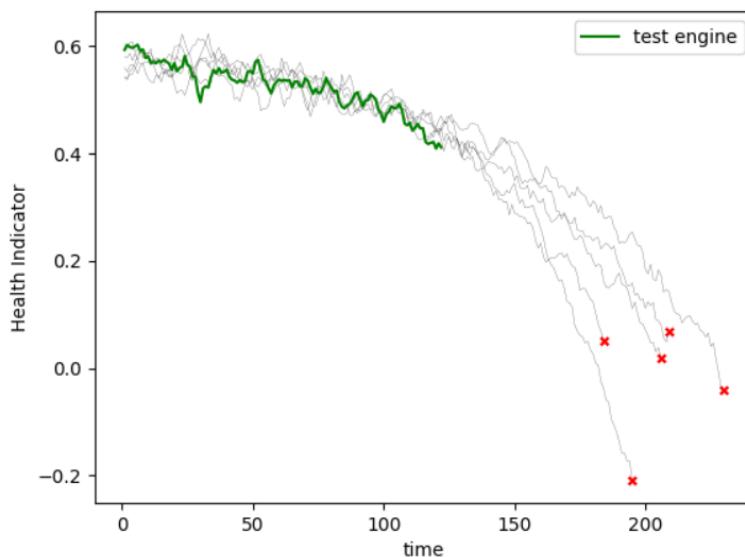


Figure 66 :5 Modèles les Plus Similaires pour un moteur de test

- **Calcul de la prédition du RUL :**

Enfin, la prédition du RUL est calculée en soustrayant le cycle actuel du nombre total de cycles des unités les plus similaires. Nous utilisons la médiane des cycles de vie totaux des moteurs similaires (médiane des K modèles les plus proches) pour déterminer le RUL estimé, car cela donne des résultats plus robustes par rapport à la moyenne.

Pour le calcul du RUL, il s'agit de la différence entre le total (médiane de **K**) et le cycle actuel.

## Choix du meilleur K pour la prédiction du RUL :

Dans cette étape, nous utilisons les valeurs de **RMSE** (Root Mean Squared Error) et de  $R^2$  (coefficient de détermination) pour évaluer la performance du modèle de prédiction du RUL en fonction de la valeur de **K**. Nous testons différentes valeurs de **K** allant de 1 à 50, et pour chaque valeur, nous calculons le  $R^2$  et le **RMSE**.

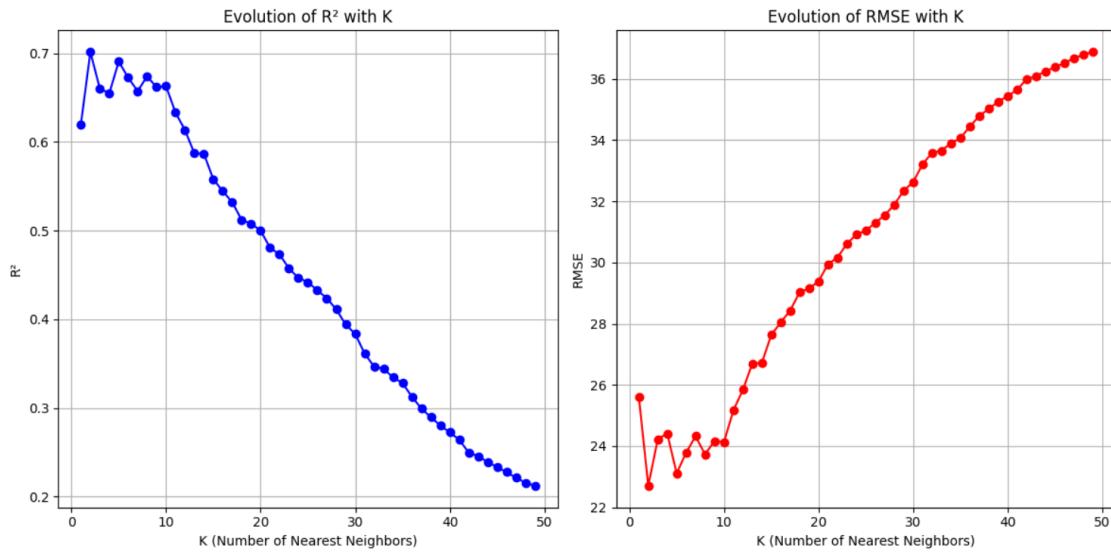


Figure 67: évolution de  $R^2$  et RMSE en fonction de K

Dans ce cas, la meilleure valeur de K selon le **RMSE** et le  $R^2$  est K=2, ce qui nous permet d'obtenir la meilleure précision dans la prédiction du RUL.

## II. Résultats et Analyse des Performances

### 1. Comparaison entre les RUL prédict et réel :

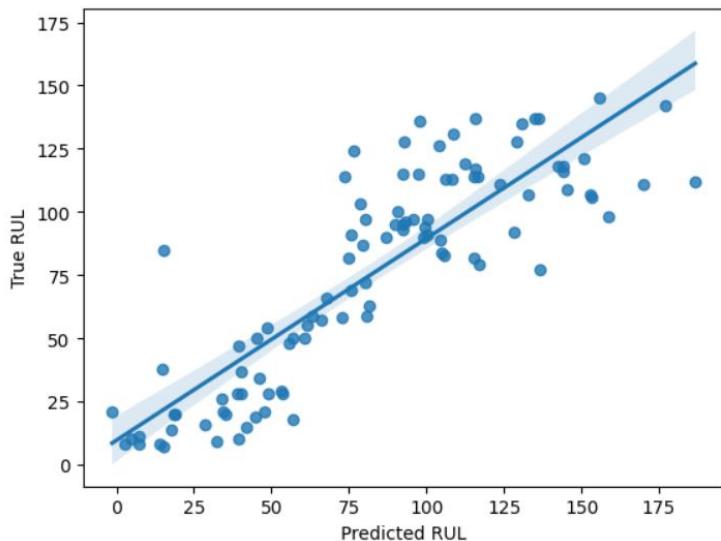


Figure 68: Comparaison entre la durée de vie restante((RUL) Prédit et réelle

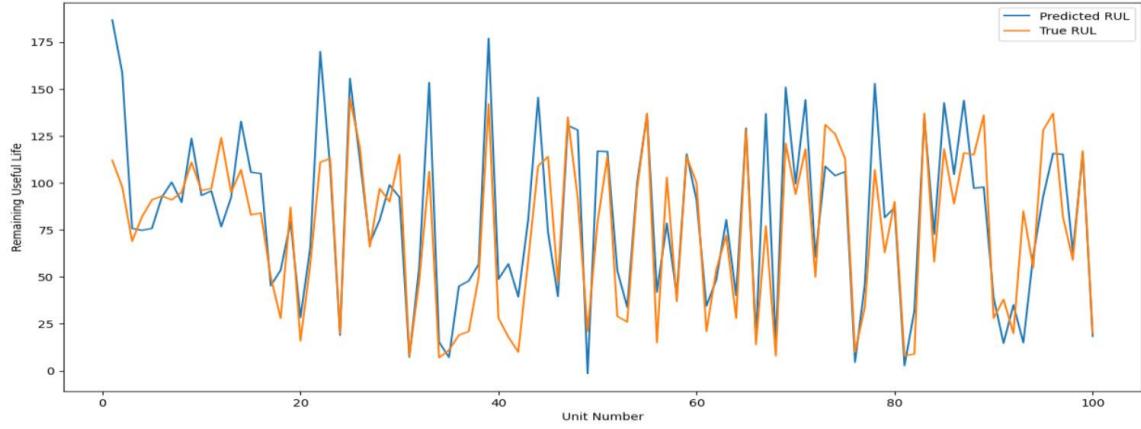


Figure 69: Comparaison entre la durée de vie restante (RUL) Prédit et réelle pour Chaque Moteur

## 2. métriques d'erreur

métrique	erreur absolue moyenne (MAE)	erreur quadratique moyenne (RMSE)	coefficient de détermination ( $R^2$ )
valeur	17.93	22.70	0.70

Figure 70: Évaluation des Performances du Modèle : RMSE, RAE et R2

## 3. Analyse des résultats

Les résultats obtenus pour la prédiction du RUL avec cette méthode de similarité sont satisfaisants, surtout en comparaison avec les approches précédentes utilisées pour la prédiction de la durée de vie restante des moteurs . Ces résultats montrent une bonne précision du modèle, avec un faible niveau d'erreur et un coefficient de détermination assez élevé. L'utilisation de la méthode de similarité pour trouver les unités les plus proches permet d'obtenir des prédictions plus robustes et adaptées aux spécificités des moteurs.

## Conclusion :

Dans ce chapitre, nous avons examiné la méthode de prédiction du RUL fondée sur la similarité, qui a donné des résultats prometteurs, surtout pour les moteurs ayant effectué plusieurs cycles. Cette approche s'avère particulièrement efficace lorsque l'historique des moteurs est suffisamment détaillé pour permettre une comparaison pertinente avec d'autres moteurs similaires.

Cependant, tout comme la méthode de dégradation exponentielle, elle présente certaines limites lorsqu'il s'agit de moteur avec peu ou pas d'historique. Dans ces situations, l'ajustement du modèle peut être moins précis, car il n'existe pas assez de données pour réaliser des comparaisons fiables.

L'approche de similarité dépend également du choix du nombre de voisins K, et un mauvais réglage de ce paramètre peut affecter la performance du modèle, surtout pour les moteur qui n'ont pas un historique de cycles assez étendu pour permettre une estimation correcte de leur RUL.

Dans les chapitres suivants, nous nous pencherons sur l'application de modèles tels que les machines à vecteurs de support (SVM), la régression linéaire et les forêts aléatoires (Random Forest). Nous examinerons également des méthodes de classification pour prédire les états des moteurs, notamment les catégories "Bon", "Avertissement" et "Danger", afin d'améliorer la détection et la gestion des pannes des moteurs.

# Chapitre 6 : Modèles Classiques pour la Prédiction et Classification de l'État des Moteurs

## Introduction

Après avoir examiné des méthodes basées sur la similarité et la dégradation exponentielle pour estimer la durée de vie restante (RUL) des moteurs, nous nous tournons maintenant vers l'utilisation de modèles classiques de machine learning. Parmi ces modèles, nous trouvons la **régression linéaire**, **les machines à vecteurs de support** (SVM) et **les forêts aléatoires** (Random Forest), qui sont couramment employés dans les domaines de l'analyse prédictive et de la classification.

Dans ce chapitre, nous appliquerons ces techniques à deux axes principaux :

- **Prédiction du RUL** : Utilisation de modèles supervisés pour estimer la durée de vie restante des moteurs avec une précision améliorée.
- **Classification de l'état des moteurs** : Adoption d'une approche de classification pour diviser les moteurs en trois états distincts – "Bon", "Avertissement" et "Danger". Cette méthode permet d'évaluer le niveau de dégradation des moteurs et de prévoir les interventions de maintenance nécessaires.

## I. Modèles supervisés : Régression Linéaire, SVM et Random Forest

### 1. préparation de données

Avant d'entamer l'entraînement des modèles supervisés, les données ont été soigneusement préparées en respectant les étapes suivantes :

- **Sélection des caractéristiques pertinentes** : Les variables les plus significatives ont été retenues à partir de l'analyse réalisée lors de l'étape de sélection des caractéristiques (Chapitre 4).
- **Normalisation Min-Max** : Une mise à l'échelle des caractéristiques a été effectuée pour les ramener dans l'intervalle [0, 1], favorisant ainsi une convergence optimale des modèles.
- **Calcul du RUL (Remaining Useful Life)** : Pour chaque ligne de données, la durée de vie restante a été calculée en appliquant la formule suivante :

$$\text{RUL} = \text{Max de cycle du moteur} - \text{Cycle actuel.}$$

```
data['rul'] = data.groupby('engine')['time'].transform('max') - data['time']
```

### 2. Régression Linéaire :

La régression linéaire est une méthode de modélisation permettant d'établir une relation linéaire entre une variable continue dite "variable expliquée" ou dépendante et un ensemble d'autres variables continues dites "variables explicatives" ou indépendantes. Plus spécifiquement elle propose un modèle explicatif qui permet de prédire la variable dépendante en fonction des variables indépendantes

**Entraînement:** Le modèle a été entraîné sur les données d'entraînement normalisées. Les performances ont ensuite été évaluées en termes de **RMSE (Root Mean Squared Error)** et de **R<sup>2</sup> (coefficient de détermination)**.

## Résultats:

- **ensemble d'entraînement:** RMSE = 40.43 et R<sup>2</sup> = 65.54%
- **ensemble de test:** RMSE = 31.36 et R<sup>2</sup> = 43.05%

## 3. Support Vector Machine (SVM)

Les SVMs sont une famille d'algorithmes d'apprentissage automatique qui permettent de résoudre des problèmes tant de classification que de régression ou de détection d'anomalie.

Comme le montre la figure ci-dessous, leur principe est simple : il ont pour but de séparer les données en classes à l'aide d'une frontière aussi « simple » que possible, de telle façon que la distance entre les différents groupes de données et la frontière qui les sépare soit maximale. Cette distance est aussi appelée « marge » et les SVMs sont ainsi qualifiés de « séparateurs à vaste marge », les « vecteurs de support » étant les données les plus proches de la frontière

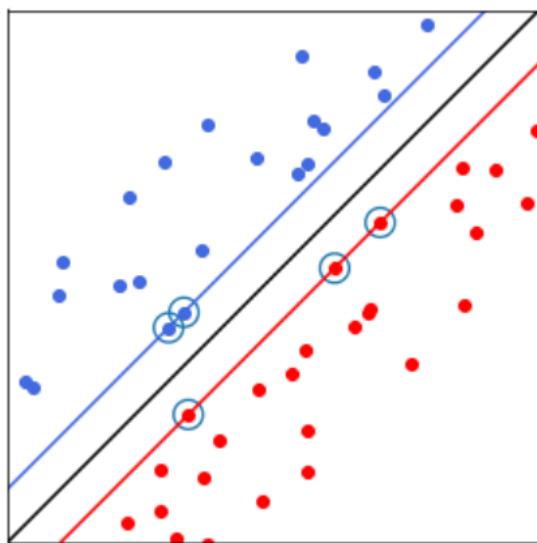


Figure 71: Support Vector Machine (SVM)

Pour l'entraînement, le modèle a été appliqué sur l'ensemble d'entraînement. Les performances initiales sur les ensembles d'entraînement et de test sont les suivantes :

- **Ensemble d'entraînement :** RMSE = 42.91 et R<sup>2</sup> = 61.54%
- **Ensemble de test :** RMSE = 26.15 et R<sup>2</sup> = 60.41%

## Ajustement des hyperparamètres :

Le réglage des hyperparamètres consiste à identifier et à choisir les meilleurs hyperparamètres pour entraîner un modèle. Un bon réglage des hyper paramètres permet de réduire la fonction de perte des modèles, ce qui signifie que ceux-ci sont entraînés pour atteindre la meilleure précision possible.

## Hyperparamètres du SVM ajustés :

Pour le modèle SVM, les principaux hyperparamètres ajustés sont :

- **C** : Ce paramètre contrôle la régularisation du modèle, équilibrant la complexité du modèle (variance) et la minimisation des erreurs (biais).
- **Gamma** : Définit l'impact d'un seul point de données sur la décision. Une valeur élevée de gamma signifie que chaque point influence fortement la décision, alors qu'une valeur basse le fait moins.
- **Kernel** : Fonction noyau utilisée pour transformer les données. Les options incluent des noyaux linéaires, polynomiaux et radiaux (RBF).

### Résultats après ajustement :

- **Ensemble d'entraînement** : RMSE = 37.85 et  $R^2 = 69.80\%$
- **Ensemble de test** : RMSE = 24.91 et  $R^2 = 64.06\%$

## 4. Modèle Random Forest

Le **Random Forest** (La forêt d'arbres décisionnels) est un algorithme de machine learning couramment utilisé, qui combine les résultats de plusieurs arbres de décision pour obtenir un résultat unique. Sa facilité d'utilisation et sa flexibilité ont favorisé son adoption, car il gère à la fois les problèmes de classification et de régression.

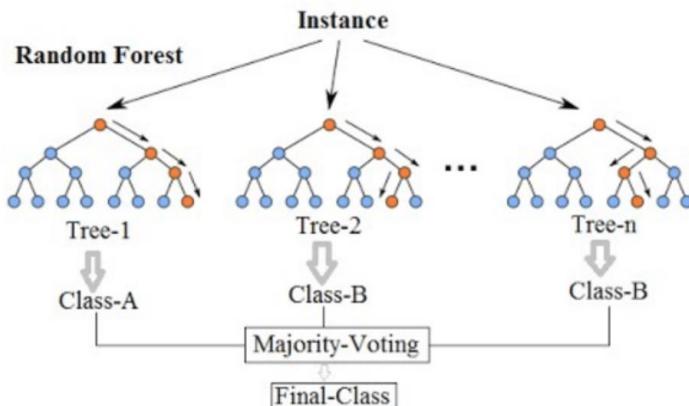


Figure 72: Illustration de l'algorithme Random forest

### Hyperparamètres ajustés :

Les principaux hyperparamètres ajustés sont :

- **n\_estimators** : Le nombre d'arbres dans la forêt. Plus ce nombre est élevé, plus le modèle est stable, mais cela augmente également le temps de calcul.
- **max\_depth** : La profondeur maximale des arbres, qui limite leur complexité et aide à éviter le surapprentissage.
- **min\_samples\_split** : Le nombre minimum d'échantillons requis pour diviser un nœud interne.
- **min\_samples\_leaf** : Le nombre minimum d'échantillons dans une feuille pour éviter des branches trop spécifiques.
- **max\_features** : Le nombre maximum de caractéristiques considérées lors de la division d'un nœud.

### Résultats après ajustement :

- **Ensemble d'entraînement** : RMSE = 30.05 et  $R^2 = 80.97\%$
- **Ensemble de test** : RMSE = 25.03 et  $R^2 = 63.72\%$

## 5. Résultats et Analyse

## Comparaison des métriques d'erreur :

Modèle	erreur quadratique moyenne (RMSE)	coefficient de détermination ( $R^2$ )
Régression linéaire	31.36	0.43
random forest	25.03	0.63
SVM	24.91	0.64

Figure 73: comparaison des Performances des trois Modèles

nous observons que le modèle SVM offre les meilleurs résultats en termes de performance sur l'ensemble de test. En particulier, le SVM présente une meilleure précision avec des métriques de RMSE et de  $R^2$  qui surpassent celles des autres modèles, confirmant ainsi son efficacité pour cette tâche de prédiction.

## Visualisation de la relation entre le RUL prédit par SVM et le RUL réel :

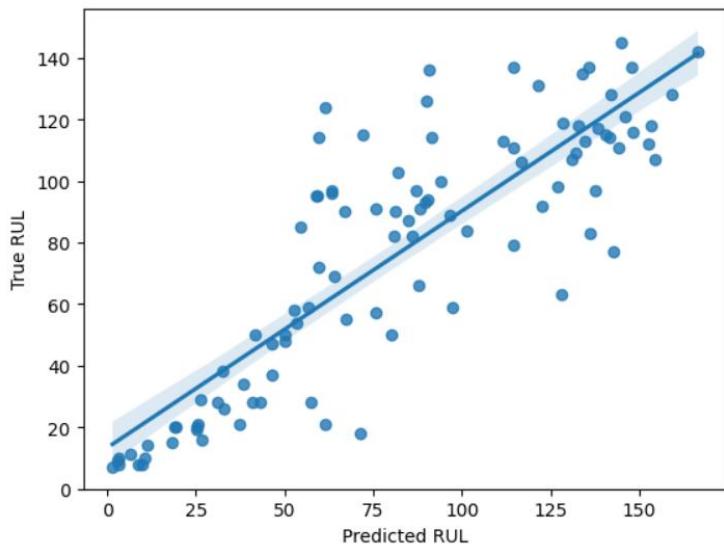


Figure 74: Comparaison entre la durée de vie restante(RUL) Prédit par SVM et réelle

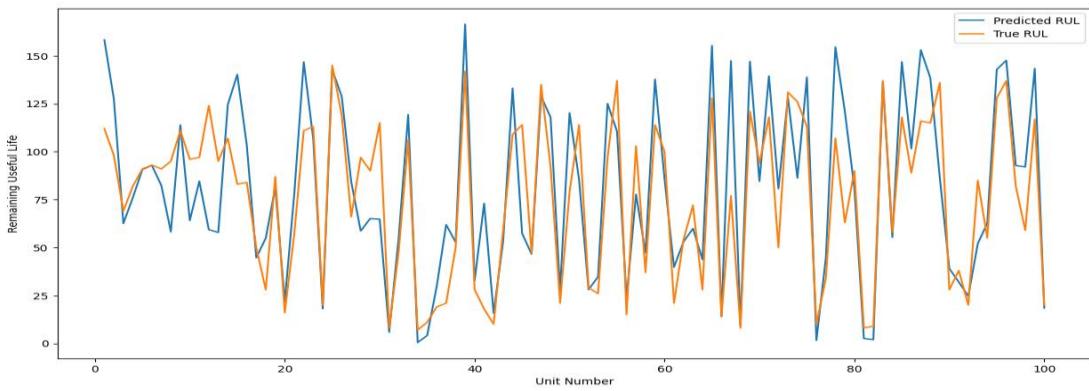


Figure 75: Comparaison entre la durée de vie restante (RUL) Prédit par SVM et réelle pour Chaque Moteur

## II. Approche de Classification pour la Prédiction de l'État des Moteurs

Dans cette approche, au lieu de prédire directement la durée de vie restante (RUL), nous utilisons une méthode de classification pour évaluer l'état des moteurs. Cette méthode permet de classer les moteurs en trois états distincts:

- **Bon** : lorsque le RUL est supérieur à 40 cycles, indiquant un moteur en bonne santé.
- **Avertissement** : lorsque le RUL se situe entre 10 et 40 cycles, signalant une éventuelle intervention à prévoir.
- **Danger** : lorsque le RUL est inférieur à 10 cycles, nécessitant une intervention immédiate.

Les valeurs seuils (<10, entre 10 et 40, et >40) ont été fixés de manière empirique en l'absence de recommandations ou d'informations provenant d'experts dans le domaine.

Cette catégorisation facilite l'interprétation des résultats et permet une prise de décision rapide en matière de maintenance.

## 1. Préparation de données :

### Génération de la variable d'état :

une nouvelle variable, appelée **État**, est créée selon les seuils définis précédemment.

- **État = 2** si RUL > 40 (Bon).
- **État = 1** si  $10 \leq \text{RUL} \leq 40$  (Avertissement).
- **État = 0** si RUL < 10 (Danger).

```
conditions = [
    (test_rul['RUL'] > 40),
    (test_rul['RUL'] >= 10) & (test_rul['RUL'] <= 40),
    (test_rul['RUL'] < 10)
]
choices = [2, 1, 0] # Bon, Avertissement, Danger
test_rul['État'] = np.select(conditions, choices)
```

## 2. Entraînement d'un modèle SVM pour la classification :

Une fois la variable **État** générée. Un modèle **SVM** (Support Vector Machine) est ensuite entraîné pour prédire la classe correspondante à l'état du moteur. ( Les données sont normalisées (Min-Max Scaling) avant l'entraînement, tout comme pour les modèles précédents.)

### Évaluation des performances :

Les performances du modèle SVM sont évaluées à l'aide de métriques classiques en classification :

- ✓ **Matrice de confusion** : La matrice de confusion (ou matrice d'erreur) est une méthode qui permet de visualiser les résultats d'un algorithme de classification. Plus précisément, il s'agit d'une table qui décompose le nombre d'instances de vérité terrain d'une classe donnée par rapport au nombre d'instances de classes prédictes.

	Positive	TP	FN
Actual value			
Negative		FP	TN
Positive		Negative	
Predicted value			

Figure 76:confusion matrix

**TP (Vrais Positifs)** : prédictions correctes pour une classe donnée.

**FP (Faux Positifs)** : prédictions incorrectes, où la classe prédite diffère de la classe réelle.

**FN (Faux Négatifs)** : classe réelle qui a été manquée.

**TN (Vrais Négatifs)** : prédictions correctes pour toutes les autres classes.

- ✓ **Accuracy (Précision globale)** : L'accuracy mesure le pourcentage de prédictions correctes par rapport à l'ensemble des prédictions effectuées.

$$\text{Accuracy} = \frac{\text{Nombre de prédictions correctes}}{\text{Total des prédictions}}$$

- ✓ **Precision (Précision)** : La précision mesure la proportion de prédictions positives correctes pour une classe donnée.

$$\text{Precision} = \frac{\text{Vrais positifs}}{\text{Vrais positifs} + \text{Faux positifs}}$$

Dans notre cas, cela signifie combien de fois le modèle a correctement prédit un moteur en "Danger" ou "Avertissement" parmi toutes ses prédictions pour ces états.

- ✓ **Recall (Rappel ou Sensibilité)** : Le rappel mesure la capacité du modèle à identifier tous les exemples pertinents d'une classe donnée.

$$\text{Recall} = \frac{\text{Vrais positifs}}{\text{Vrais positifs} + \text{Faux négatifs}}$$

Cela correspond à la proportion de moteurs effectivement en "Danger" ou "Avertissement" que le modèle a bien classifiés.

- ✓ **F1-Score** : L'équilibre entre précision et rappel est donné par le F1-score, une métrique harmonique.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

## Quelle métrique privilégier dans le contexte des moteurs turbo-fan ?

Dans le cadre de la prédiction des états des moteurs turbo-fan, les métriques à considérer varient selon les objectifs :

- **Danger (Critique)** : Le rappel (Recall) est primordial pour cette catégorie, car il est essentiel de ne pas négliger un moteur en danger afin d'éviter des pannes catastrophiques. Un faible rappel dans ce cas signifierait que des moteurs critiques pourraient passer inaperçus.
- **Avertissement** : La précision pourrait être cruciale pour réduire les fausses alertes et éviter des interventions inutiles.
- **Bon** : Un équilibre entre précision et rappel (via l'accuracy ou le F1-score) est suffisant, car cette catégorie est généralement moins critique.

### 3. Résultats et Analyse des Performances

**Précision globale :** Le modèle SVM atteint une précision globale (accuracy) de **85%**, ce qui indique que 85% des prédictions faites par le modèle sont correctes sur l'ensemble de test.

**Matrice de confusion :**

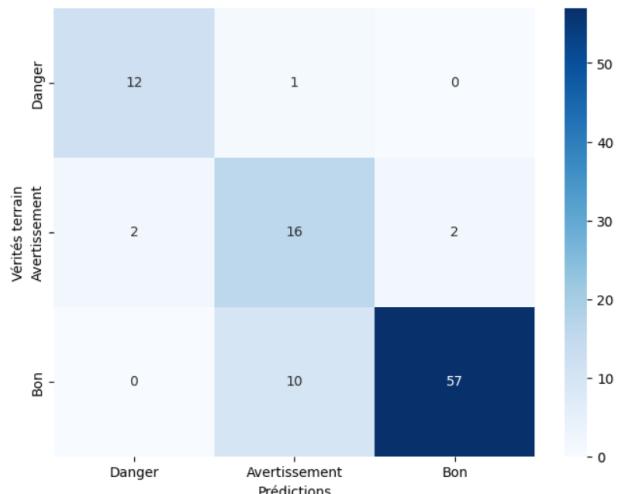


Figure 77: Matrice de confusion

**Métriques par classe :**

Classe	Précision	Rappel	F1-Score
<b>Danger</b>	0.86	0.92	0.89
<b>Avertissement</b>	0.59	0.80	0.68
<b>Bon</b>	0.97	0.85	0.90

Figure 78: métriques du modèle pour chaque classe

**Interprétation des métriques :**

Classe « **Danger** » :

- Avec une précision de **86%** et un rappel de **92%**, le modèle se distingue par son efficacité à identifier les moteurs en état critique.
- Le F1-score élevé (**89%**) atteste d'un bon équilibre entre la précision et le rappel pour cette catégorie, ce qui est essentiel pour ne pas négliger les moteurs en situation dangereuse.

#### Classe « Avertissement » :

- La précision pour cette classe est plutôt faible (**59%**), ce qui signifie que le modèle a parfois du mal à distinguer les états "Avertissement" des classes "Danger" ou "Bon".
- Cependant, un rappel de **80%** indique que le modèle réussit à détecter la plupart des moteurs réellement en état d'avertissement, même si certaines de ses prédictions ne sont pas tout à fait correctes.

#### Classe « Bon » :

- Le modèle est très efficace pour détecter les moteurs en bon état, affichant une précision élevée de **97 %**.
- Cependant, un rappel légèrement inférieur de **85 %** indique qu'il y a quelques erreurs où des moteurs jugés "bons" sont incorrectement classés en "avertissement".

#### Analyse globale :

- Le modèle est très fiable pour prédire les moteurs dans des états extrêmes ("Danger" et "Bon").
- Les performances sur la classe intermédiaire "Avertissement" sont moins précises, ce qui pourrait être attribué à une ambiguïté dans les données ou à un chevauchement dans les caractéristiques.
- L'équilibre global des métriques (Précision globale : **85%**) indique que le modèle peut être utilisé efficacement comme un outil d'aide à la décision pour évaluer l'état des moteurs.

# Chapitre 7 : Développement d'un Tableau de Bord interactif pour le Suivi et la Visualisation des États des Moteurs

## Introduction

Pour rendre les résultats des analyses précédentes exploitables, ce chapitre se concentre sur la création d'un tableau de bord interactif. L'objectif principal est de fournir une interface intuitive qui permettra de :

- Visualiser les prédictions du RUL (Remaining Useful Life) et l'état des moteurs.
- Suivre les indicateurs de santé pour évaluer l'état actuel des moteurs.
- Analyser la dégradation exponentielle des moteurs afin de détecter les tendances de dégradation.
- Afficher des moteurs similaires grâce à des approches basées sur la similarité.
- Suivre plusieurs indicateurs clés de performance (KPI), tels que la fiabilité et les probabilités de défaillance.
- Gérer efficacement les cycles de maintenance en fonction des prédictions fournies.

Le tableau de bord intégrera les différents modèles développés, combinant les résultats du chapitre précédent avec l'analyse statistique des temps de fonctionnement, les approches de dégradation exponentielle, basées sur la similarité, ainsi que les prédictions de régression et de classification. Il permettra également d'afficher des visualisations dynamiques.