

Élimination des conditions extérieures fondée sur la physique : normalisation adaptative de données pour la surveillance santé

Maxime LEIBER^{1,2} Yosra MARNISSI² Sébastien RAZAKARIVONY² Dong Quan VU²
Dohy HONG² Laurent MASSOULIÉ¹

¹INRIA, DI/ENS, Université PSL, 2 Rue Simone IFF, 75012 Paris, France

²Safran Tech, Digital Sciences & Technologies Department, Rue des Jeunes Bois-Châteaufort, 78114 Magny-les-Hameaux, France

Résumé – Cet article propose une nouvelle méthode de prétraitement pour normaliser les grandeurs physiques mesurées et indicateurs d’un système physique, en fonction des conditions externes (opérationnelles et environnementales). Notre approche transforme ces variables en variables corrigées qui capturent le comportement interne du système tout en éliminant l’impact des variables externes sur ce comportement. Nous démontrons l’efficacité de notre approche à travers deux expériences sur des données aéronautiques. Notre normalisation comme prétraitement permet l’utilisation de données cohérentes pour l’analyse et la prédiction dans différentes conditions, améliorant ainsi l’exactitude et la fiabilité des systèmes de surveillance santé en mécanique.

Abstract – This article proposes a new preprocessing method for normalizing the physical quantities measured and indicators of a physical system, based on external (operational and ambient) conditions. Our approach transforms these variables into corrected variables that capture the internal behavior of the system while eliminating the impact of external variables on this behavior. We demonstrate the effectiveness of our approach through two experiments on aeronautical data. Our normalization as a preprocessing step enables the use of consistent data for analysis and prediction under different conditions, thus improving the accuracy and reliability of health monitoring systems in mechanics.

1 Introduction

La surveillance de la santé est cruciale dans de nombreuses industries pour assurer un fonctionnement sûr et efficace des systèmes. Elle permet de détecter et d’analyser les données acquises à partir de machines, d’équipements et de structures afin d’identifier les problèmes potentiels, permettant ainsi d’effectuer une maintenance préventive avant qu’une défaillance ne se produise. La surveillance des systèmes mécaniques se fait généralement à partir de quantités physiques mesurées telles que la température ou la vitesse, ou à partir d’indicateurs de santé calculés sur des quantités physiques tels que les moments statistiques. Cette surveillance peut être manuelle ou automatisée en utilisant des algorithmes d’apprentissage automatique. L’automatisation permet une surveillance continue des systèmes mécaniques et l’analyse de grandes quantités de données. Cependant, la fiabilité de ces méthodes de surveillance dépend fortement des méthodes statistiques qui supposent que les données utilisées lors de l’apprentissage et de l’analyse sont mesurées dans les mêmes conditions de fonctionnement et environnementales, ce qui n’est pas toujours le cas dans la réalité. En effet, ces variables (quantités physiques et indicateurs) dépendent souvent non seulement du comportement interne du système mais également des conditions externes [10], telles que la pression extérieure ou la vitesse du moteur dans un exemple aéronautique. De plus, ces conditions environnementales et opérationnelles changent constamment [1, 3]. Pour relever ce défi, il est nécessaire d’ajuster les variables pour éliminer leur impact de ces facteurs externes et ne capturer que les variations internes du système. Par exemple, si l’on souhaite surveiller la température d’un moteur de voiture, l’analyse de la température ne sera pas la même en fonction de la température extérieure. Il est donc nécessaire de corriger

la température mesurée sur le moteur par rapport à la température extérieure afin d’analyser les mesures collectées dans différentes conditions. Il en va de même pour les vibrations d’un système mécanique, qui dépendent de variables externes telles que la vitesse de rotation ou la charge.

Dans le domaine de l’apprentissage automatique, il existe des techniques de prétraitement classiques qui abordent des problèmes similaires en transformant les données en un format normalisé qui peut être plus facilement analysé [2], tels que la normalisation dans une distribution standard ou la mise à l’échelle des données dans une plage donnée. Dans le même esprit, la surveillance de l’état de santé des systèmes mécaniques, en particulier des moteurs, nécessite la normalisation des données pour obtenir des mesures fiables et non-biaisées par les conditions externes.

Plusieurs articles ont été publiés sur ce sujet. [4] propose une méthode de normalisation statistique en corrigeant la moyenne et l’écart-type des données d’entrée à des valeurs standards, usuelles ou historiques. [8] utilise un classificateur SVM (Support Vector Machine) à classe unique pour normaliser les données. Ils entraînent ce classificateur uniquement sur des signaux sains dans des conditions extérieures variables et projettent les signaux de test dans un sous-espace invariant aux conditions externes. Cette méthode est basée sur l’apprentissage automatique. [7] poursuivent les travaux précédents en recherchant le sous-espace optimal avec une Analyse en Composantes Principales et en mesurant la distance entre les signaux de test projetés et les signaux sains dans les mêmes conditions projetés. Ces deux dernières méthodes supposent que les conditions extérieures pendant la phase de test auront déjà été apprises pendant l’entraînement et que donc toutes les conditions auront été fournies au modèle pendant l’entraînement. De plus, l’ensemble des techniques de normalisation

utilisent les statistiques et l'apprentissage automatique pour normaliser les données sans prendre en compte la physique sous-jacente. [6] utilise également différents modèles non linéaires pour normaliser. [11] corrige les données d'un moteur aéronautique à partir des équations de la thermodynamique. Nous étendons cette approche à l'ensemble des données mesurées et pas seulement au cas des mesures thermodynamiques. De plus notre méthode s'adapte aux données et ne nécessite pas de connaître a priori les équations.

Dans cet article, nous proposons une méthode de normalisation basée sur la physique et qui s'adapte aux données pour corriger les quantités physiques et indicateurs de santé et de performance par rapport aux variables externes. Notre objectif est de transformer ces grandeurs en variables normalisées qui capturent le comportement interne du système tout en éliminant l'impact des variables opérationnelles et environnementales sur ce comportement. Nous présentons notre approche de normalisation dans la section 2, nous illustrons l'efficacité de notre approche sur deux bases de données en aéronautique en partie 3 et nous concluons dans la partie 4.

2 Méthode proposée

2.1 Objectif

Considérons un ensemble d'indicateurs $j \in 1, \dots, J$ calculés à partir de quantités physiques mesurées sur un système mécanique. Ces indicateurs sont souvent utilisés dans des tâches liées à la surveillance de la santé du système telles que le diagnostic et le pronostic. Notre méthode vise à corriger les effets des variations des conditions externes sur les indicateurs en les plaçant dans des conditions externes de référence. Pour cela, nous proposons une transformation appliquée aux indicateurs physiques avant d'apprendre une méthode pour la tâche cible.

2.2 Modèle

Soit $I^{(j)}$ la valeur d'un indicateur j dans une certaine condition externe \mathcal{E} . Nous cherchons à estimer la valeur $I_{\text{ref}}^{(j)}$ du même indicateur j comme si le système fonctionnait sous une autre condition externe, désignée par \mathcal{E}_{ref} , que nous utilisons comme environnement de référence. Les conditions de externes sont définies par un ensemble de K variables e_1, \dots, e_K . Il s'agit d'apprendre, pour chaque indicateur j , la fonction $f_j : \mathbb{R} \times \mathbb{R}^K \times \mathbb{R}^K \rightarrow \mathbb{R}$, qui produit une estimation $\hat{I}_{\text{ref}}^{(j)} = f_j(I^{(j)}, \mathcal{E}, \mathcal{E}_{\text{ref}})$ de $I_{\text{ref}}^{(j)}$. En particulier, nous posons les hypothèses suivantes : (i) *Le taux de variation (des valeurs) des indicateurs lors du changement des conditions externes dépend des ratios respectifs des variables de conditions externes ;* (ii) *Les différentes variables des conditions externes agissent de manière indépendante sur les indicateurs.* En d'autres termes, nous cherchons à trouver les meilleurs modèles $\psi_k^{(j)}$ pour chaque indicateur j tel que

$$\Delta I^{(j)} = \prod_{k=1}^K \psi_k^{(j)}(\Delta e_k) \quad (1)$$

où $\Delta I^{(j)} = I^{(j)}/I_{\text{ref}}^{(j)}$, $\Delta e_k = e_k/e_{\text{ref}k}$. Ces hypothèses sont cohérentes avec de nombreux principes et lois qui régissent les systèmes physiques (loi des gaz parfaits, deuxième loi de

Newton, loi de la conservation de l'énergie mécanique, etc.). En suivant ces hypothèses, nous pouvons montrer que (1) a une forme simple de produit de puissance comme le montre la proposition suivante.

Proposition 2.1. *Sous les hypothèses (i) and (ii), la fonction qui lit les ratios des indicateurs aux ratios des conditions d'essais est donnée par :*

$$\Delta I^{(j)} = \prod_{k=1}^K \Delta e_k^{a_k^{(j)}} \quad (2)$$

où $a_k^{(j)}$ sont des paramètres réels inconnus assignés à chaque indicateur j .

Démonstration. La fonction $\psi_k^{(j)}$ satisfait les propriétés suivantes : (i) $\psi_k^{(j)}(1) = 1$ (ii) $\psi_k^{(j)}(\Delta e_k) \times \psi_k^{(j)}(\Delta e'_k) = \psi_k^{(j)}(\Delta e_k \times \Delta e'_k)$. On en déduit que $\psi_k^{(j)}$ est un automorphisme du groupe \mathbb{R}_+^* munis de l'opérateur \times . Or, on peut montrer¹ que chaque automorphisme continu de (\mathbb{R}_+^*, \times) s'écrit comme une fonction puissance $x \rightarrow x^r$, ce qui implique (2). \square

Les exposants $a_k^{(j)}$ dans (2) représentent l'élasticité des indicateurs par rapport aux conditions externes, c'est-à-dire la variation de $I^{(j)}$ pour une variation d'une unité de e_k . En pratique, étant donné que toutes les quantités physiques sont positives, la recherche de a_k équivaut à trouver les coefficients d'un modèle log-log linéaire pour chaque indicateur j , c'est-à-dire $\log(\Delta I) = \sum_{k=1}^K a_k^{(j)} \log(\Delta e_k)$.

2.3 Mise en oeuvre

2.3.1 Apprentissage

Dans la phase d'apprentissage, notre approche repose sur un ensemble de données constitué de N points de données $(I_i^{(j)}, \mathcal{E}_i)$, $1 \leq i \leq N$, où $I_i^{(j)}$ représente l'indicateur physique j mesuré sur le même système lorsqu'il fonctionne dans l'environnement \mathcal{E}_i . Pour construire notre ensemble de données d'apprentissage pour le prétraitement, nous calculons les ratios entre toutes les paires de points $(I_i^{(j)}, \mathcal{E}_i)$. Cela nous permet d'apprendre les paramètres $a_k^{(j)}$ de notre modèle qui estime les ratios d'indicateurs à partir des ratios d'environnements tels que dans (2). Ce prétraitement doit être entraîné sur des données provenant d'un système dont l'état interne demeure constant. Ainsi, toute variabilité peut être attribuée uniquement aux changements des conditions externes. De plus, Pour garantir un pré-traitement efficace et non biaisé, il est essentiel que les différentes variables externes dans les données d'entraînement soient suffisamment hétérogènes et varient de manière indépendante les unes des autres.

¹D'une part, le groupe (\mathbb{R}_+^*, \times) est isomorphe au groupe $(\mathbb{R}, +)$ à travers l'opérateur \log . D'autre part, il est connu que tout automorphisme continu de $(\mathbb{R}, +)$ s'écrit sous la forme $t \rightarrow rt$ avec $r \in \mathbb{R}$. Soit ξ un automorphisme continu de (\mathbb{R}_+^*, \times) . Ainsi, $\log \circ \xi \circ \exp$ est un automorphisme continu de $(\mathbb{R}, +)$. Il existe alors $r \in \mathbb{R}$ tel que $\forall t \in (\mathbb{R}, +)$, $\log \circ \xi \circ \exp(t) = rt$. En faisant le changement de variable $x = \log(t) \in (\mathbb{R}_+^*, \times)$, on en déduit que, $\xi(x) = \exp(r \log(x)) = x^r$.

2.3.2 Test

Dans la phase de test, pour garantir la cohérence pour la tâche cible, les valeurs d'indicateurs dans les données de test sont ajustées pour correspondre à une condition externe unique, désignée par \mathcal{E}_{ref} , en utilisant les coefficients appris $a_k^{(j)}$. L'environnement de référence \mathcal{E}_{ref} est défini arbitrairement. Pour les conditions opérationnelles, les valeurs de références peuvent par exemple correspondre au fonctionnement en mode nominal du système. Pour les variables environnementales, les valeurs usuelles telles que les conditions normales de température et pression peuvent être utilisées. D'autres approches plus statistiques consistent à considérer l'environnement moyen, médian, le plus représenté dans les données ou le barycentre des environnements.

La figure 1 résume les étapes d'apprentissage et de test.

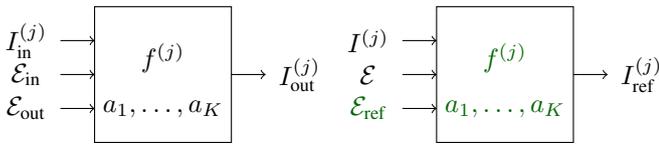


FIGURE 1 : Description de la méthode en apprentissage (à gauche) et en test (à droite), en vert les valeurs sont fixes.

Remarque 2.1. Si l'indicateur j est une fonction g_j non-triviale d'une grandeur physique plutôt que de la grandeur physique elle-même, il est possible que l'hypothèse (i) ne soit pas satisfaite. Dans ce cas, des corrections doivent être apportées plutôt à la grandeur physique. Une fois corrigée, la fonction g_j peut être utilisée pour obtenir l'indicateur dans des conditions de référence. Le cas où g_j est inconnue fera l'objet de nos travaux futurs.

Remarque 2.2. Une approche alternative à la correction de variables pourrait être d'inclure les conditions externes en tant qu'entrées du modèle pour la tâche cible. Une telle approche présente cependant un risque de surapprentissage, car ces variables peuvent biaiser le modèle appris. Prenons l'exemple des expériences d'endurance jusqu'à défaillance, où un système est soumis à des tests répétitifs jusqu'à ce qu'il tombe en panne. Il est courant de soumettre le système à une grande charge au début pour accélérer l'apparition du défaut. Si les données ne sont pas corrigées par rapport à la charge ou si la charge est incluse comme variable explicative, le modèle apprendra qu'une forte charge est synonyme d'absence de défaut, et inversement, ce qui n'est pas le comportement souhaité. Ainsi, inclure les conditions externes en entrées peut affecter la précision et la fiabilité des modèles obtenus.

3 Expériences

3.1 Données NASA C-MAPSS

Les données utilisées dans cet exemple sont fournies par la NASA [9]. Elles comprennent des mesures de performances telles que la température d'huile, la pression du carburant, la vitesse de rotation, la température de sortie de la turbine etc. Ces données sont collectées à partir de moteurs d'avion opérant dans des conditions de vol simulées et contiennent également des informations sur les défaillances mécaniques des moteurs. Nous avons utilisé ces données pour construire un

modèle de survie permettant de prédire la durée de vie restante des moteurs, en utilisant le modèle proposé dans [5]. Nous avons évalué les performances de ce modèle en utilisant différentes méthodes de normalisation. En particulier, nous avons comparé les résultats obtenus avec quatre approches : (A) Les données brutes non-normalisées ; (B) Les données normalisées selon une distribution normale ; (C) Les données normalisées selon une distribution normale par condition d'essai [5] ; (D) Les données normalisées par notre méthode puis normalisées selon une distribution normale. Le tableau 1 montre les résultats d'estimation. Les meilleurs résultats sont obtenus par la normalisation (C), notre méthode (D) donne des résultats proches de la méthode (C) tandis que les méthodes (A) et (B) donnent de moins bons résultats en ignorant la variabilité des conditions externes. Cependant, il est important de souligner que la normalisation (C), contrairement à notre méthode, est possible uniquement parce que les conditions d'essais varient de manière discrète, qu'il y a peu d'environnements différents (six au total) et qu'il y a assez de mesures par environnement différent pour faire la normalisation. Cette méthode de normalisation par condition externe est donc restreinte aux simulations. Dans les cas réels, les conditions externes varient de manière continue, ce qui rend plus difficile l'identification d'environnements distincts. Au contraire, notre méthode est plus générique et est capable d'apprendre sur des conditions externes plus variées.

TABLE 1 : Résultats de régression du temps de reste à vivre sur la base de données de test FD002 (racine carrée de l'erreur quadratique moyenne (Err) et coefficient de détermination (R2)).

	A	B	C	D
Err	53.8	35.6	29.4	30.3
R2	-2.97	0.56	0.70	0.68

Remarque 3.1. La relation physique entre deux indicateurs, lorsqu'elle existe, est souvent perturbée par des conditions externes changeantes. L'apport de notre méthode de correction est qu'elle permet de rendre ces relations plus stables et plus fiables.

3.2 Données vibratoires aéronautiques

Nous avons mené des expériences sur certains indicateurs de santé calculés à partir de données vibratoires mesurées par 4 accéléromètres et montés sur 4 roulements dans l'industrie aéronautique. Les enregistrements ont été effectués à la fois en état sain et défectueux. Notre objectif est de corriger les indicateurs de santé par rapport à la vitesse (désignée par V) et à la charge (désignée par C). Contrairement à l'expérience 3.1, ces conditions varient de manière continue. Il est connu que la charge et la vitesse affectent les vibrations des roulements. Pour construire notre ensemble de données pour le prétraitement, nous avons construit tous les croisements entre les triplets (I, V, C) en état sain de chaque roulement et de chaque accéléromètre pour chaque indicateur. Dans cette étude, nous avons utilisé un processus de validation de type "leave-one-out" : nous avons entraîné notre modèle sur tous les roulements sauf un utilisé pour le test, puis nous avons répété le processus pour tous les roulements. La Figure 2 illustre les valeurs

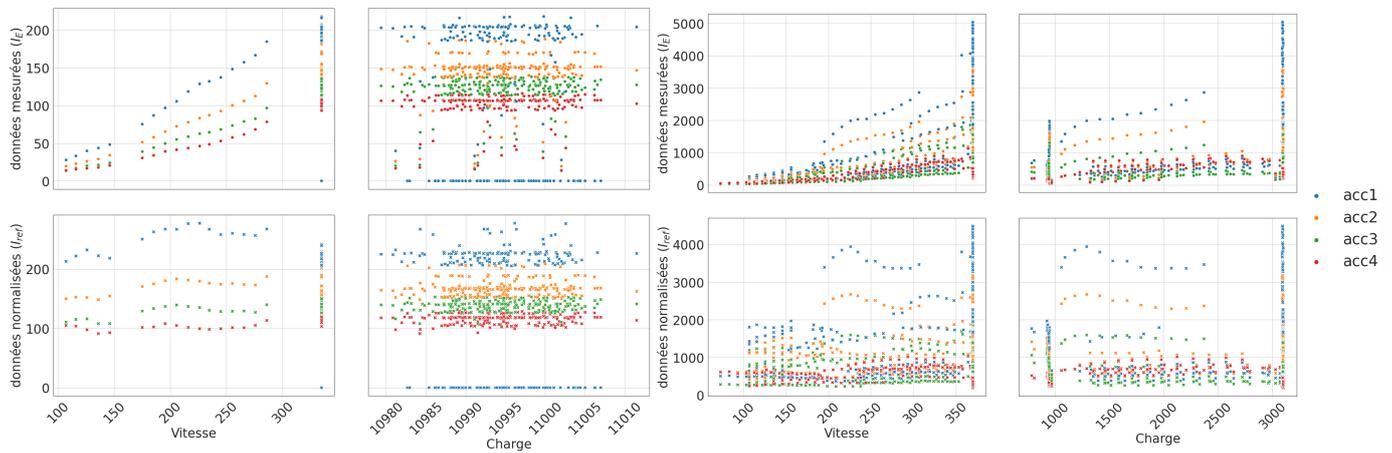


FIGURE 2 : Un indicateur de santé sur les états sains (gauche) et défectueux (droite) d'un roulement.

brutes et corrigées d'un indicateur de santé pour un roulement particulier. Le Tableau 2 présente les résultats d'une tâche de régression visant à prédire la taille de fissure sur 4 roulements. Nos résultats d'expérience montrent que notre approche de normalisation améliore significativement la fiabilité des indicateurs de santé dans la détection des états défectueux des roulements. En effet, notre approche corrige les effets de la charge et de la vitesse, ce qui se traduit par des indicateurs de santé plus précis et propres à l'état interne du système.

TABLE 2 : Résultats de régression en termes d'Erreurs Absolues Moyennes.

capteurs	indicateurs bruts	indicateurs corrigés
acc1	0.85 ± 0.45	0.74 ± 0.10
acc2	0.99 ± 0.27	0.62 ± 0.21
acc3	0.75 ± 0.13	0.52 ± 0.20
acc4	1.02 ± 0.59	0.87 ± 0.53

4 Conclusion

Nous avons présenté une méthode de normalisation basée sur la physique pour corriger les variables en fonction des conditions externes, qu'elles soient opérationnelles ou environnementales. Cette méthode a été testée sur des systèmes mécaniques avec des conditions externes variables. Les résultats montrent que l'approche proposée est efficace pour normaliser les valeurs des indicateurs dans différentes conditions. En résumé, cette méthode a le potentiel d'améliorer la précision des tâches de surveillance de la santé des systèmes mécaniques.

Références

- [1] D. ABOUD, J. GRIFFATON, Y. MARNISSI, J. L. GOMEZ CHIRINOS et M. EL BADAoui : Vibration-based diagnostic of aircraft bearings operating under strong speed variations. *In Turbo Expo : Power for Land, Sea, and Air*, 2022.
- [2] P. J. M. ALI, R. H FARAJ et E. KOYA : Data normalization and standardization : a technical report. *Mach Learn Tech Rep*, 2014.
- [3] W. BARTELMUS et R. ZIMROZ : Vibration condition monitoring of planetary gearbox under varying external load. *Mechanical systems and signal processing*, 2009.
- [4] R. J. BRONSON, H. R. DEPOLD, R. RAJAMANI, S. DEB, W. H. MORRISON et K. R. PATTIPATI : Data normalization for engine health monitoring. *In Turbo Expo : Power for Land, Sea, and Air*, 2005.
- [5] N. COSTA et L. SÁNCHEZ : Variational encoding approach for interpretable assessment of remaining useful life estimation. *Reliability Engineering & System Safety*, 2022.
- [6] C. FLEXA, W. GOMES et C. SALES : Data normalization in structural health monitoring by means of nonlinear filtering. *In 2019 8th Brazilian Conference on Intelligent Systems (BRACIS)*. IEEE, 2019.
- [7] K. KUMAR et S. AGARWAL : Optimal residual subspace model for structural damage diagnosis : an approach independent of operational and environmental variations. *International Journal of Structural Engineering*, 2022.
- [8] K. KUMAR, P. K. BISWAS et N. DHANG : Data normalization technique in shm under environmental variations using one-class svm. *In 23rd International Congress on Sound and Vibration*, 2016.
- [9] A. SAXENA, K. GOEBEL, D. SIMON et N. EKLUND : Damage propagation modeling for aircraft engine run-to-failure simulation. *In 2008 international conference on prognostics and health management*. IEEE, 2008.
- [10] H. SOHN : Effects of environmental and operational variability on structural health monitoring. *Philosophical Transactions of the Royal Society A : Mathematical, Physical and Engineering Sciences*, 2007.
- [11] A. J. VOLPONI : Gas turbine parameter corrections. *In Turbo Expo : Power for Land, Sea, and Air*, 1999.