

# Analyse des biais dans la prédiction des déformations mécaniques d'un avion de combat par apprentissage automatique

C. Ferrassou<sup>1,2</sup>, P. Escande<sup>2</sup>, M. Duval<sup>1</sup>, L. Risser<sup>2</sup>, R. Bouclier<sup>3,2,4</sup>

<sup>1</sup> DGA-Techniques Aérospatiales, Balma, France (ferrassou@insa-toulouse.fr)

<sup>2</sup> Univ. Toulouse, INSA-Toulouse, CNRS UMR 5219, IMT, Toulouse, France ([paul.escande,laurent.risser]@math.univ-toulouse.fr)

<sup>3</sup> Univ. Toulouse, INSA-Toulouse, IMT-Albi, ISAE-SUPAERO, CNRS UMR 5312, ICA, Toulouse, France, (bouclier@insa-toulouse.fr)

<sup>4</sup> Institut Universitaire de France (IUF)

**Résumé** — La surveillance de la fatigue des structures aéronautiques est indispensable à la planification de la maintenance des avions militaires. Cette étude analyse les biais de prédiction dans les modèles d'apprentissage automatique appliqués à des données réelles de vol d'un avion de combat Rafale. L'approche combine outils d'explicabilité de l'apprentissage automatique et expertise métier pour isoler l'origine des biais. Nous proposons une décomposition du signal en composantes dynamique et environnementale, révélant que les biais proviennent exclusivement de la composante environnementale du signal de déformation.

**Mots clés** — jumeaux numériques, maintenance prédictive, approche basée sur les données, biais de prédiction, explicabilité, séries temporelles.

## 1 Introduction

La surveillance du vieillissement en fatigue est un enjeu crucial pour la gestion de l'intégrité structurale des aéronefs, avec des approches fondamentalement différentes pour les avions militaires et civils. Alors que les avions civils utilisent un suivi par heures de vol en raison de profils opérationnels prévisibles, les avions militaires présentent des défis uniques liés à leur utilisation opérationnelle variable, leurs configurations de charge diverses et leurs enveloppes de vol extrêmes [1]. Ces facteurs introduisent une dispersion significative dans l'accumulation de la fatigue, rendant les méthodes traditionnelles de suivi par heures de vol inadéquates.

L'approche historique concernant le suivi du vieillissement en fatigue des aéronefs militaires utilise un ensemble de paramètres de vol relatifs à l'environnement (pression statique et dynamique par exemple) et à l'attitude de l'avion (position des gouvernes, accélération normale, etc.). Ces données temporelles, enregistrées par chaque avion et à chaque vol, sont ensuite traitées par des modèles physiques, élaborés à partir de résultats d'essais et de simulations. Ces modèles permettent d'estimer les déformations mécaniques associées à ces données de vol en certains points critiques de la structure. Cependant, cette approche présente certaines limites : elle nécessite une connaissance exhaustive des propriétés structurales de l'avion, la résolution d'équations complexes, ainsi que des campagnes d'essai et de calibration onéreuses.

Une alternative consiste à mesurer directement les déformations de la cellule via des jauges d'extensométrie, mais cette solution est difficile à mettre en oeuvre à grande échelle sur une flotte entière en raison des difficultés et du coût opérationnel que représente la maintenance des jauges [2].

Les approches émergentes basées sur l'apprentissage automatique offrent un compromis prometteur [3]. Elles permettent d'équiper seulement quelques aéronefs de jauges, puis d'utiliser ces données pour entraîner des modèles généralisables à toute la flotte. Plusieurs travaux ont exploré ces méthodes, et concernent majoritairement des avions de ligne [4], des jets d'affaires [5] ou des drones [6] à enveloppe de vol limitée. Ils utilisent également parfois des données synthétiques obtenues à partir de simulations [7] plutôt que des déformations réelles mesurées en vol. Peu de travaux portent sur des avions de combat militaires en raison des coûts et de la complexité opérationnelle que cela implique [8].

Dans ce contexte, un avion de combat Rafale a été instrumenté par la *Direction Générale de l'Armement* (DGA) au moyen de jauges d'extensométrie. Ces valeurs de déformations sont mesurées en vol de

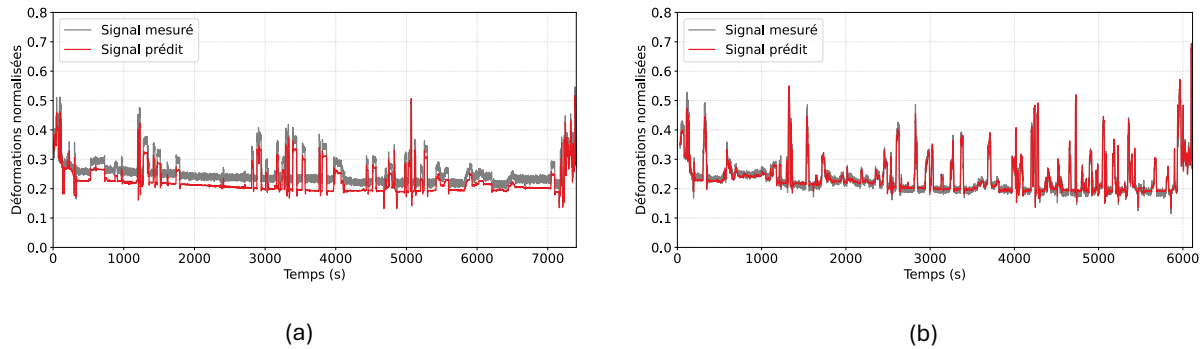


FIGURE 1 – Mise en évidence des biais dans la prédiction des déformations mécaniques. Les deux graphes comparent les déformations normalisées mesurées et prédites pour deux vols : **(a)** Un vol pour lequel le modèle reproduit bien la dynamique mais présente un décalage vertical persistant entre la prédiction et la mesure **(b)** Un vol correctement prédit au sens défini dans ce travail, *i.e.* ne présentant pas de biais de prédiction.

manière synchronisée avec l’ensemble des paramètres de vol correspondants. Le jeu de données obtenu permet ainsi de développer des modèles d’apprentissage capables de reproduire le comportement des jauges sous différentes conditions opérationnelles.

Le présent travail ne vise toutefois pas à optimiser les performances de tels modèles, mais plutôt à analyser les erreurs de prédiction qu’ils génèrent dans le cadre d’un avion de combat [9]. Plus précisément, nous mettons en évidence l’existence d’une source récurrente de biais de prédiction, définis comme des écarts systématiques persistants entre déformations prédites et mesurées, apparaissant occasionnellement sans raison apparente lors de certains vols. La Figure 1 illustre ce phénomène : *(a)* montre un exemple de prédiction biaisée où le modèle suit correctement la dynamique des déformations mais présente un décalage vertical marqué alors que *(b)* ne présente pas ce type d’écart. Ainsi, même des erreurs de prédiction isolées nécessitent une analyse approfondie compte tenu de la criticité de l’application.

De telles erreurs peuvent provenir de multiples sources, mentionnées dans la littérature. Il peut s’agir d’un problème de représentativité des données, d’une inadéquation de l’architecture des modèles, ou encore d’incertitudes liées aux capteurs physiques. Afin d’identifier l’origine de ces biais, nous proposons une démarche présentée comme suit. Dans un premier temps, nous décrivons le jeu de données, ainsi que le problème de régression étudié. Nous détaillons ensuite l’analyse comparative de plusieurs modèles de prédiction : eXtreme Gradient Boosting (XGBoost), Multilayer Perceptron (MLP) et Long Short-Term Memory (LSTM), afin de vérifier que les biais observés ne sont pas spécifiques à une architecture particulière. Puis, nous avons recours à des outils d’explicabilité (XAI) combinant méthodes d’analyse de gradients et expertise métier, afin d’identifier les origines physiques de ces écarts systématiques. Enfin, nous proposons une décomposition du signal de déformation en composantes basse et haute fréquence, ce qui permet d’isoler l’origine potentielle de ces erreurs et d’en proposer une explication physique.

## 2 Constitution du jeu de données de vol Rafale et formulation du problème de régression

Cette étude repose sur un jeu de données obtenu grâce à l’instrumentation d’un avion de combat Rafale par la DGA. L’analyse se concentre sur les mesures de la jauge située à la jonction voilure-fuselage. Les données de déformations sont enregistrées via une centrale d’acquisition MDR8, synchronisée avec l’ensemble des paramètres de vol essentiels (nombre de Mach, altitude, accélération normale, angle d’incidence, vitesses angulaires, positions des gouvernes, masses, etc.).

Le jeu de données initial comprend une centaine de vols, chacun d’une durée de plusieurs heures, acquis à une fréquence de 100 Hz puis rééchantillonnés à 1 Hz, fréquence suffisante pour l’étude des phénomènes liés à la fatigue. Les phases sol, décollage et atterrissage ont été retirées (seuils sur nombre de Mach  $< 0,3$  et altitude  $< 100$  m). Afin de réduire la variabilité liée aux configurations de charge, seuls les vols réalisés avec deux bidons externes de 2000 L ont été conservés, conduisant à un ensemble plus

homogène de 43 vols.

En raison du nombre élevé de variables présentant des redondances ou symétries, une sélection guidée par l’expertise métier a permis de retenir 14 paramètres pertinents pour la modélisation du comportement structural. Certaines grandeurs ont également été combinées pour obtenir des indicateurs plus directement reliés aux sollicitations mécaniques.

La division du dataset en ensembles d’apprentissage, validation et test a été réalisée avec un soin particulier pour préserver l’intégrité temporelle en effectuant la séparation au niveau du vol complet. Six partitionnements différents ont été générés avec à chaque fois : 8 vols dédiés à la validation des modèles, 5 vols jamais vus durant l’apprentissage et utilisés pour l’évaluation des modèles, et les 30 autres pour l’apprentissage. La sélection des vols de test a été optimisée en utilisant les deux premières composantes principales (captant 75% de la variance) pour garantir que les données de test restent dans l’enveloppe des données d’apprentissage, minimisant ainsi les risques d’extrapolation. Enfin, l’ensemble des variables (entrées et sortie) a été normalisé dans l’intervalle  $[0,1]$  à partir des statistiques de l’ensemble d’apprentissage. Pour des raisons de confidentialité, l’ensemble des visualisations et résultats présentés ont donc été normalisés dans cet intervalle.

Le problème d’apprentissage peut ensuite être formalisé comme suit : étant donné un ensemble de  $M$  vols, où chaque vol  $m$  comprend  $T_m$  pas de temps, avec à chaque pas de temps  $t$  : un vecteur d’entrée  $x_{m,t} = x_{m,t}^1, \dots, x_{m,t}^{14} \in \mathbb{R}^{14}$  représentant les 14 paramètres de vol et une valeur de sortie  $y_{m,t} \in \mathbb{R}$  représentant la mesure de déformation.

L’objectif est d’apprendre une fonction de régression  $f_{\Theta}$  qui minimise l’erreur quadratique moyenne sur l’ensemble du dataset :

$$\min_{\Theta} \frac{1}{\sum_{m=1}^M T_m} \sum_{m=1}^M \sum_{t=1}^{T_m} (y_{m,t} - f_{\Theta}(x_{m,t}))^2. \quad (1)$$

### 3 Mise en évidence de biais récurrents à travers l’analyse comparative de plusieurs modèles de prédiction

L’un des objectifs centraux de ce travail est de déterminer si les erreurs observées dans les prédictions de déformations proviennent d’un problème inhérent au jeu de données ou d’un effet lié au choix du modèle. Pour répondre à cette question, nous comparons plusieurs familles de modèles représentatives de degrés de complexité différents : XGBoost, un ensemble d’arbres boostés optimisé par recherche de grille sur les hyperparamètres clés (nombre d’arbres, profondeur, etc.); un réseau de neurones entièrement connecté de type MLP à une seule couche cachée (12 neurones) avec une fonction d’activation ReLU, choisi pour sa bonne performance tout en limitant le sur-apprentissage; et enfin un LSTM bi-directionnel de petite taille (8 unités par direction) traitant des fenêtres temporelles de 200 s afin de capturer explicitement la dépendance temporelle du signal. Les trois modèles ont été entraînés avec l’optimiseur Adam, une fonction de perte quadratique moyenne et un mécanisme d’early stopping basé sur l’ensemble de validation pour limiter le sur-apprentissage. Si des erreurs similaires apparaissent simultanément sur ces trois approches, cela suggère l’existence de limitations structurelles du problème ou du jeu de données, plutôt que d’une faiblesse liée à un modèle particulier.

TABLE 1 – Comparaison globale des performances des modèles XGBoost, MLP et LSTM sur l’ensemble des données de test.

Modèle	XGBoost	MLP	LSTM
MAE	0.01878	0.01807	0.01817
$R^2$	0.862	0.877	0.882
MBE (valeur absolue)	0.0010	0.0008	0.0007

Pour évaluer les performances globales des modèles, nous utilisons deux métriques classiques : l’erreur moyenne absolue (MAE) et le coefficient de détermination ( $R^2$  score). Nous introduisons également l’erreur de biais moyenne (MBE), une métrique complémentaire explicitement sensible aux offsets de prédiction et qui est utilisée ici en prenant sa valeur absolue. Leurs valeurs moyennes (Tableau 1)

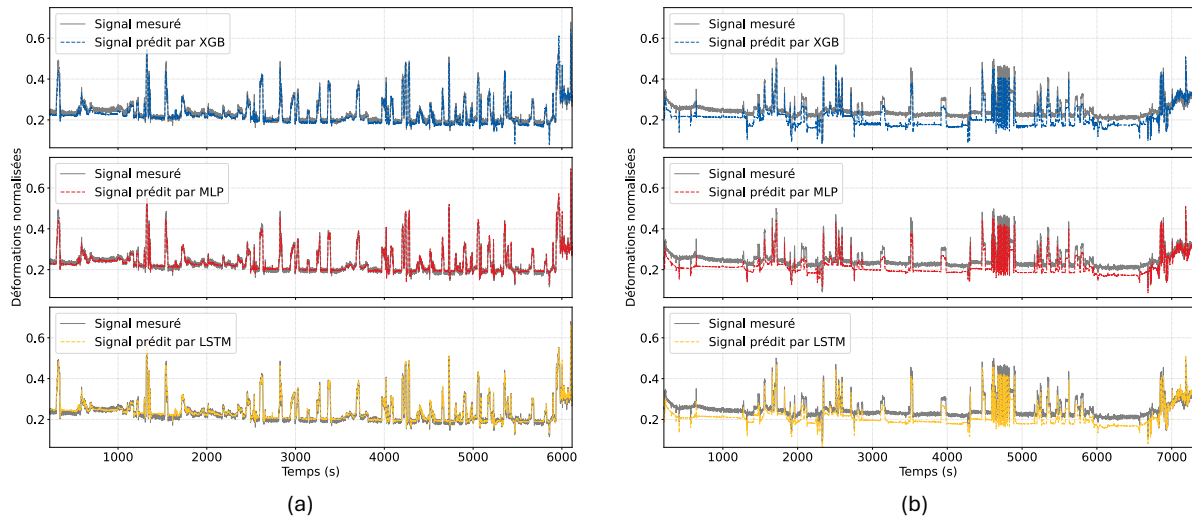


FIGURE 2 – Signaux de déformations normalisées mesurés et prédits par les trois modèles évalués dans cette étude, pour deux vols de test. **(a)** Les prédictions de ce vol sont précises, sans écart systématique : les trois modèles restituent correctement la tendance et la dynamique du signal. **(b)** Ce vol, en revanche, montre un décalage de prédiction clair et persistant : les trois modèles sous-estiment la déformation sur de longs segments du vol, malgré leurs architectures différentes.

montrent que les trois modèles obtiennent des performances similaires, qui sont très satisfaisantes au vu des valeurs de déformations qui sont comprises entre 0.0198 et 0.846. Cependant, cette évaluation moyenne masque des variations significatives entre les différents vols.

L’analyse détaillée des performances par vol révèle un phénomène intéressant : les trois modèles présentent systématiquement les mêmes comportements d’erreur, aux mêmes instants temporels et sur les mêmes vols. Les vols bien prédits par une architecture le sont également par les autres. Inversement, les vols présentant dans leurs prédictions des décalages persistants avec les déformations mesurées montrent ce biais simultanément pour les trois approches. Cette observation est illustrée avec la Figure 2.

Ces biais récurrents, qui représentent environ 20% des instants temporels des vols de test, apparaissent même sur des vols très proches de ceux vus durant l’apprentissage et se concentrent au niveau des tendances générales des signaux (basses fréquences), les pics de déformation (hautes fréquences) étant bien capturés. Ce résultat confirme que les erreurs observées ne proviennent pas d’un modèle particulier, mais d’une difficulté intrinsèque liée aux données ou à la physique sous-jacente du problème. Elles nécessitent donc une analyse explicative approfondie.

## 4 Analyse des biais de prédiction

### 4.1 Identification des sources de biais à l’aide d’outils d’explicabilité (XAI) couplés à la connaissance métier

La démarche repose sur deux piliers : des connaissances techniques en ingénierie dans le domaine d’application qui permettent de formuler des hypothèses sur les sources possibles du problème, et l’utilisation d’outils d’explicabilité (XAI) pour vérifier empiriquement ces hypothèses. Du point de vue physique, les déformations mesurées sur la cellule proviennent de deux types de contributions :

- Les effets aérodynamiques en régime permanent, liés à l’environnement et l’état de l’avion (modélisables à partir des variables d’altitude, pression, masses, vitesses).
- Les effets inertiels transitoires, liés aux manœuvres de l’aéronef (modélisations à partir des variables accélération normale, angle d’incidence, vitesses angulaires, position des gouvernes, etc.) et responsables des pics de déformation.

Dans les travaux de modélisation des charges structurelles aéronautiques, ces deux contributions sont traditionnellement traitées séparément car elles répondent à des phénomènes physiques distincts. Le fait que les modèles prédictifs reproduisent correctement les pics mais présentent un décalage sur la tendance

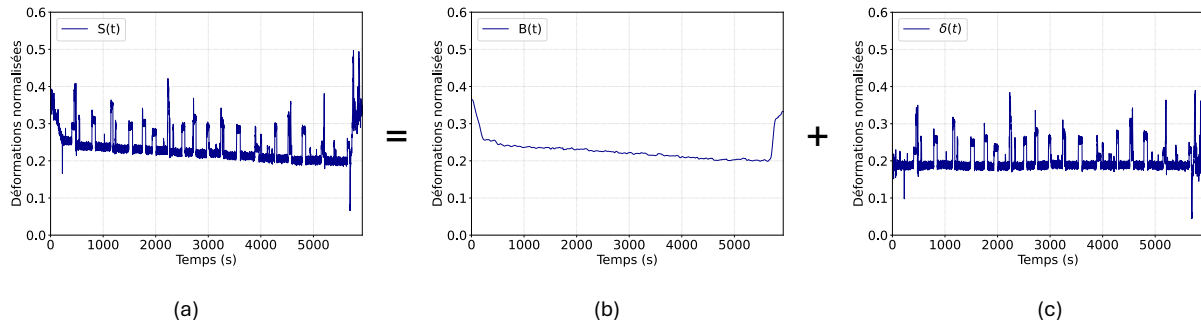


FIGURE 3 – Décomposition du signal de déformations  $S(t) = B(t) + \delta(t)$  avec (a)  $S(t)$  le signal original, (b)  $B(t)$  la tendance du signal et (c)  $\delta(t)$  la partie dynamique du signal.

suggère donc une surdépendance aux variables inertielles au détriment des variables environnementales.

Pour examiner ce point, nous utilisons une méthode d’explicitabilité, basée sur l’analyse de sensibilité des gradients, et appliquée au MLP (dont la structure simple facilite l’implémentation). Pour chaque point temporel et chaque variable d’entrée, nous évaluons la valeur absolue du gradient de la prédiction par rapport à cette variable. Cette stratégie, qui est une adaptation de SmoothGrad [10] à nos données de vol, n’est pas développée ici mais sera présentée à l’oral.

L’analyse révèle un phénomène déterminant : même dans les phases stabilisées sans manœuvre de vol, le modèle reste fortement sensible aux variables inertielles (notamment l’accélération normale) et très peu aux variables environnementales, alors même que ce sont ces dernières qui commandent la tendance du signal. En d’autres termes, le modèle apprend très bien à prédire les pics dynamiques, mais sous-représente la contribution aérodynamique lente, ce qui peut expliquer les écarts de prédiction observés. Ce constat conduit naturellement à l’hypothèse suivante : le biais provient de la partie "tendance" du signal de jauge. Cette observation motive la seconde étape de l’analyse : séparer explicitement les deux contributions du signal pour vérifier si c’est bien la tendance du signal qui en porte le biais.

## 4.2 Analyse des biais via la décomposition dynamique / environnementale du signal de déformations

Afin d’isoler la source du biais, le signal de jauge  $S(t)$  a été décomposé en :

- une composante environnementale  $B(t)$  (tendance) qui peut être modélisée avec les variables d’environnement de l’avion.
- une composante dynamique  $\delta(t)$  qui peut être modélisée avec les variables liées aux manœuvres de l’avion.

Pour extraire la composante environnementale du signal original, un processus en deux étapes est appliqué. Un filtrage par seuil est d’abord réalisé afin de sélectionner les segments où l’accélération normale est comprise entre  $-9$  et  $-11 \text{ m/s}^2$  (environ  $1g$ ). Puis, un filtre Gaussien est appliqué pour lisser le signal tout en conservant la tendance générale. La composante dynamique est ensuite obtenue par soustraction :  $\delta(t) = S(t) - B(t)$ . La Figure 3 illustre cette décomposition.

Chacune des deux composantes est ensuite prédite séparément par un MLP à partir de ses variables d’intérêt respectives. L’architecture et la stratégie d’entraînement utilisée est la même que celle utilisée pour la prédiction MLP du signal complet, décrit précédemment. Seul le nombre de neurones de la couche d’entrées varie en fonction des variables utilisées pour prédire chacun des deux signaux.

Les résultats de ces prédictions sont sans équivoque comme l’illustre la Figure 4 : la composante dynamique  $\delta(t)$  est prédite avec une très bonne précision, sans aucun biais visible alors que la composante environnementale  $B(t)$  présente exactement le même décalage systématique que le signal complet, démontrant que l’intégralité du biais provient de cette composante basse fréquence. Cette observation est confirmée quantitativement par l’analyse des valeurs absolues des MBE pour chaque vol de test, calculées séparément pour chaque composante du signal (Figure 5).

Ainsi, la décomposition du signal en partie dynamique et environnementale permet d’isoler la source du biais et constitue une stratégie de modélisation pertinente pour ce type de données. On peut donc en déduire que le biais observé est lié à une caractéristique spécifique de la composante environnementale.

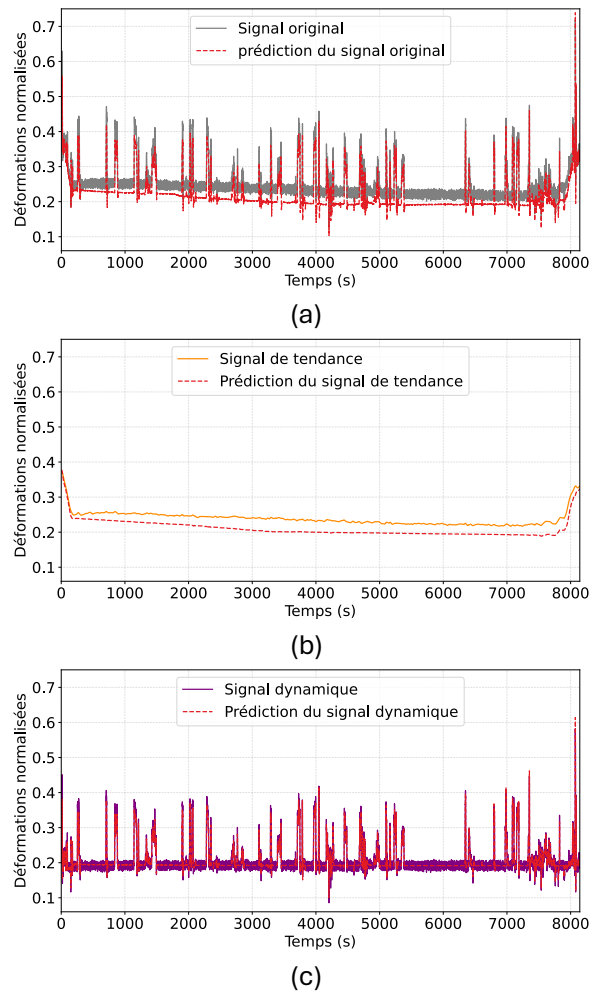


FIGURE 4 – (a) Prédiction du signal de déformation original par MLP avec toutes les variables retenues, montrant un décalage vertical au cours du vol entre valeurs mesurées et prédites. (b) Prédiction de la composante environnementale extraite, qui présente le même motif de biais que pour la prédiction du signal original. (c) Prédiction de la composante dynamique, qui montre une excellente précision de reconstruction sans biais observable.

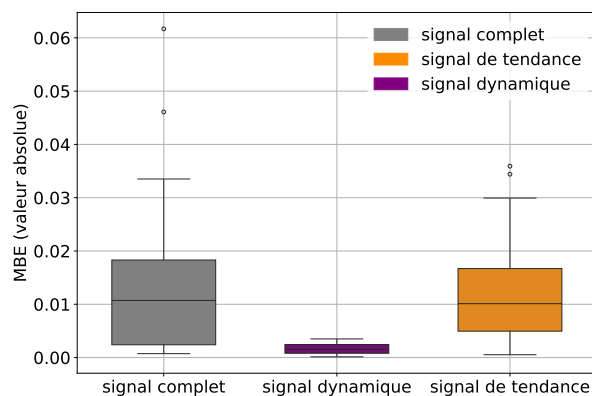


FIGURE 5 – Distribution du MBE (valeur absolue) pour les trois composantes du signal : signal complet, signal de tendance et signal dynamique, calculé sur l'ensemble des vols de test. La composante dynamique montre un MBE négligeable tandis que la composante environnementale reproduit le biais du signal complet.

## 5 Conclusion

Dans ce travail, nous avons étudié la prédiction des déformations mesurées sur un avion de combat Rafale à partir de paramètres de vol, en évaluant plusieurs familles de modèles (XGBoost, MLP et LSTM) afin d'analyser les biais récurrents observés dans leurs prédictions. Nous avons d'abord montré que, malgré de bonnes performances globales, les trois modèles présentent systématiquement les mêmes offsets sur des segments spécifiques de vol, indiquant un biais qui n'est pas lié à l'architecture des modèles.

Afin d'identifier l'origine de ce biais, nous avons combiné connaissances métier et outils d'explicabilité basés sur les gradients. Cette analyse a révélé un déséquilibre clair entre les contributions prises en compte par les modèles : les composantes dynamiques du signal, fortement corrélées aux variables inertielles, sont correctement apprises, tandis que la tendance lente pilotée par les variables environnementales est sous-représentée. Sur cette base, nous avons introduit une décomposition physique du signal en une composante dynamique et une composante environnementale. L'étude séparée de ces deux contributions a montré que le biais réside uniquement dans la composante lente, alors que la composante dynamique est prédite avec une grande précision. Cette décomposition a permis d'isoler le mécanisme responsable du biais et de confirmer que celui-ci provient des variables décrivant l'environnement du vol.

Ces observations suggèrent fortement que les offsets de prédiction ne découlent pas d'une limitation des modèles de régression eux-mêmes, mais plutôt d'incertitudes dans les mesures d'entrée associées à l'environnement. Parmi les facteurs plausibles, les phénomènes thermiques jouent un rôle central : ils peuvent induire des dérives significatives sur les capteurs, en particulier dans les contextes soumis à de fortes variations d'altitude et de température comme c'est le cas pour les avions de combat. Ce mécanisme a déjà été mentionné dans la littérature sur les avions de combat [11], et nos résultats en constituent un nouvel indice indirect. Les perspectives de ce travail incluent donc une investigation approfondie de ces effets thermiques. Cette étude ouvre la voie à des systèmes de surveillance plus robustes et interprétables pour les applications aéronautiques militaires critiques.

## Références

- [1] P.A. Lloyd. *Structural health monitoring systems – benefits and airworthiness issues*, The Aeronautical Journal, 285-289, 2008.
- [2] M. Wallace, H. Azzam, S. Newman. *Indirect approaches to individual aircraft structural monitoring*, Proceedings of the Institution of Mechanical Engineers, Part G : Journal of Aerospace Engineering, 329-346, 2004.
- [3] E. O'Higgins, K. Graham, D. Daverschot, J. Baris. *Machine Learning Application on Aircraft Fatigue Stress Predictions*, ICAF 2019 – Structural Integrity in the Age of Additive Manufacturing, 1031-1042, 2020.
- [4] J. Gómez-Escalonilla, J. García, M.M Andrés, J.I. Armijo. *Strain predictions using Artificial Neural Networks for a full-scale fatigue monitoring system*, AIAC-13 Thirteenth Australian International Aerospace Congress, 2009.
- [5] Pavao et al. *Aircraft Numerical “Twin” : A Time Series Regression Competition*, 20th IEEE International Conference on Machine Learning and Applications, 441-448, 2021.
- [6] X. Zhou, L. Dong, M. Dziendzikowski, K. Dragan. *Real-Time In-Service Load Tracking Toward Airframe Digital Twins*, AIAA Journal, 2025
- [7] A. Goichon, M. Ghienne, V. Tschannen, N. Peyret. *Temporal hierarchical modal neural network for structural strain prediction*, Computational Mechanics, 2025.
- [8] M.J. Scott, W.J.C. Verhagen, M.T. Bieber, P. Marzocca. *A Systematic Literature Review of Predictive Maintenance for Defence Fixed-Wing Aircraft Sustainment and Operations*, Sensors, 2022.
- [9] C. Ferrassou, P. Escande, M. Duval, R. Bouclier, L. Risser. *Understanding Biases in Machine Learning-based Strain Predictions for Fighter Aircraft*, submitted, 2025.
- [10] D. Smilkov, N. Thorat, B. Kim, F. Viégas, M. Wattenberg. *SmoothGrad : removing noise by adding noise*, 10.48550/arXiv.1706.03825, 2017.
- [11] J.A. Tikka. *Flight parameter based fatigue analysis approach for a fighter aircraft*, The Aeronautical Journal, 79-91, 2008.