ICAS 2021 SHANGHAI
International Council of the Aeronautical Sciences
32nd Congress of the International Council of the Aeronautical Sciences
September 6-10, 2021
Pudong Shangri-La, Shanghai, China

A DIGITAL EQUIVALENT DEVELOPMENT FOR FATIGUE AND DAMAGE TOLERANCE IN AIRCRAFT STRUCTURES

L. Zaplana Manzano[1], J. Romero Arana[1], J. Gómez-Escalonilla[1]

[1]Airbus Defence & Space, Madrid, Spain

Abstract

Data analytics and machine learning techniques are explored in this paper for the Fatigue and Damage Tolerance analysis of military aircraft structures, in order to establish the building blocks necessary for a Fatigue Digital Equivalent of the aircraft. The Fatigue Digital Equivalent is conceived as a digitalization tool which will enable certifying the aircraft for a complete usage region instead of for just one specific usage. This will provide the operator with an optimized maintenance program and great operational flexibility, maximizing the fatigue endurance life of the structure from the very beginning of the aircraft's operation.

Keywords: fatigue digital equivalent, aircraft usage, data analytics, machine learning

1. Introduction

The fatigue and damage tolerance (FDT) analysis process comprises the assessment of all those aircraft structural locations that are considered to be critical for FDT in accordance with the applicable airworthiness requirements (i.e., FAR/CS 25.571 for civil applications, and various national specifications for military environment). In a standard Airbus Defence & Space aircraft program, the number of locations analyzed ranges from several hundreds to several thousands, depending on the nature and size of the aircraft. When needed, a maintenance program is then established as a result of this analysis.

It is well known that structural fatigue is a very complex phenomenon, affected by a wide range of design factors (material properties, tolerances, surface finish, build stresses, etc). Once the fleet is being operated in service, additional aspects appear such as the usage of each individual serial number or the accidental damages (maintenance induced defects) and environmental damages (corrosion, temperature, etc) that can appear. Actually, from a fatigue perspective each individual aircraft is different not only due to its configuration (specific antennae, specific military/mission devices) but also due to its history and even its intended future usage.

Over the years, this phenomenon has been managed by engineers in a semi-manual fashion, with dedicated analyses based on the expected average condition of all the previous factors combined with a set of safety factor to cover possible variations, and with successive corrections once in-service experience is available in order to ensure the continued airworthiness of the platform. However, the current status of the technology enables a step forward, targeting a mostly automatic way of addressing this problem. This is the context of the creation of the Fatigue Digital Equivalent (FDE) concept by Airbus Defence & Space.

The FDE can be defined as a full-scale digital representation of the aircraft structure aimed to evaluate the impact of the different factors involved in the fatigue phenomenon in a probabilistic manner. The Fatigue Digital Equivalent is used for in-service aircraft, while its counterpart Fatigue Digital Prototype (FDP) is valid during all the development phases before the Entry into Service of the platform. Both the FDP and the FDE will be replaced in the long-term by the so-called Fatigue Digital Twin, which will be a 'true' representation of the whole structure associated response from a fatigue point of view. Table 1 shows the different features of Fatigue Digital Prototype, Fatigue Digital Equivalent and Fatigue Digital Twin.

	Fatigue Digital Prototype	Fatigue Digital Equivalent	Fatigue Digital Twin
Scope	Digital representation of the different design alternatives	Digital representation of the in-service aircraft	Digital representation of design alternatives and in-service aircraft
Period of application	Today and mid-term	Today and mid-term	Future (long-term)
Density of structural locations incorporated	Low	Medium	Very high
Fidelity in the geometry representation	Low	Medium/High	Very high/Ultra high
Accuracy	Medium	High (particularly after the end of full-scale test and the associated teardown)	Very high/Ultra high
Incorporation of test results	Development tests only	Development and certification tests	Development and certification tests
Incorporation of in-service experience	No	Environmental damages, accidental damages, usage patterns	Environmental damages, accidental damages, usage patterns

Table 1 - Comparison between Fatigue Digital Prototype, Fatigue Digital Equivalent and Fatigue Digital Twin

The models included in the Fatigue Digital Equivalent for the evaluation of the fatigue damage accrual are mainly physics-based, as the FDE is fed with the results of the standard analysis process used for qualification/certification, definition of individual and/or generic repairs, substantiation of modifications, etc. However, the FDE should not be understood as a mere automatization of all these analyses. On the contrary, the FDE incorporates an additional layer of data analytics and machine learning applications which behaves as the ‘brain’ of the system and enables not only the stochastic analysis of the required airframes in order to confirm that the real fleet operates within the design assumptions (or to provide our customers with complete and reliable data that allows them to make operational and/or sustainment decisions with impact on the structural integrity of the aircraft), but a critical assessment of the models used in the process.

The remaining sections of this paper are divided in three parts centered around the data analytics and machine learning applications of the FDE: In the first part (Section 2.2 through Section 2.4) an aircraft usage sensitivity analysis is performed in order to assess the impact of the different flight parameters of a typical tactical aircraft’s flight profile on the different critical structure locations for fatigue. Additionally, the fatigue damage distribution for all possible deviations from the aircraft’s certification mission mix is used to perform a different kind of sensitivity analysis where it is assessed whether if the fatigue life of a critical fatigue location is always above a certain threshold for all possible aircraft usages or if, on the contrary, part of the variations on the aircraft’s usage result in a fatigue life below this threshold. In the second part (Section 2.5), Machine Learning is used to assess the possibility of aircraft’s fatigue life prediction at different fatigue critical structural locations for all possible deviations from the mission mix for certification. In the third part (Section 3), a Machine Learning analysis is performed to assess the possibility of automatically allocating a typical strategic transport aircraft to a certain usage region, given a variation of the initially certified aircraft’s mission mix.

2. Sensitivity Analysis

Sensitivity analysis is used for diagnostic-modelling and in simulations, allowing system analysts to determine where to focus on system design to ensure robustness and accuracy across the range of inputs, [2]. This approach is adapted for this study to answer in a timelier manner one of the most recurring military aircraft operator's questions: how will the maintenance programme change (with regards to the fatigue life of the structure) due to deviations in the aircraft usage from that which has been certified initially?

Nowadays, there is no easy or fast way to assess this kind of impact and the analysis is performed on a case-by-case basis, generating a whole new fatigue analysis each time.

A sensitivity analysis is applied to fatigue spectra flight parameters and their influence on the obtained fatigue damage in different analysis locations of the aircraft structure. These analysis locations are those critical from a fatigue point of view of the structure and are referred to in this work as Fatigue Analysis Locations (FALs).

The Sensitivity Analysis for the FDE applications can be divided in:

- ***Fatigue lives distribution for assessment of the FALs usage sensitivity*** (Section 2.2).
- ***Determination of those FALs which fatigue life can be inferred from another FAL*** (Section 2.3)
- ***Flight Parameter and mission importance****:* Identification and prioritization of the flight parameters and missions that contribute most significantly to variability in the fatigue damage of an FAL (Section 2.4).

2.1 Dataset used in the analysis

The aircraft usage is characterized by the aircraft's mission mix, which is a representation of the appearance (in terms of percentage) of each one of the most occurring missions in the operator's usage. For this exercise, a typical Tactical aircraft is considered for which the most representative missions in the mission mix are seven: 2 different types of crew training (called for this study T1 and T2), one Ferry mission, and four different types of surveillance missions (called for this study S1, S2, S3 and S4). Each one of these missions is characterized by a specific flight profile, defined by the following flight parameters: Takeoff Weight (TOW), Fuel Weight (FW), Landing Weight (LW), Mission duration, Pressure Altitude, Dynamic Pressure, Differential Pressure and the number of landings.

To generate the dataset, these missions are combined by changing the percentage of appearance of those missions on the mission mix in a way that satisfies the condition that the sum of all the mission percentages is 100%. The combinatorics resource of "weak compositions" is used for this purpose [10] and a total of 8008 mission mixes are defined. Then, the flight parameter values associated to each one of these mission mixes are obtained by linear interpolation of the corresponding values of the seven missions that compose each new mission mix

The fatigue life for every new mission mix is then obtained for each selected Fatigue Analysis Location (FAL).

An extract of the structure of the dataset is shown in Table 2, where all values have been standardized:

Mission percentages in the Mission Mix	TOW (kg)	LW (kg)	Mission Duration (h)	Altitude (ft)	Flight Parameter n	Fatigue Life of FAL 1 (flight cycles)	Fatigue Life of FAL m (flight cycles)
85.0% 15.0% 0.0% 0.0% 0.0% 0.0% 0.0%	0.875	0.694	0.43	0.500	…	125	13
85.0% 10.0% 5.0% 0.0% 0.0% 0.0% 0.0%	0.869	0.694	0.55	1.000	…	65	7
…	…	…	…	…	…	…	…
0.0% 0.0% 0.0% 0.0% 50.0% 5.0% 45.0%	0.917	0.702	0.70	1.500	…	200	110
0.0% 0.0% 0.0% 0.0% 50.0% 0.0% 50.0%	0.914	0.701	0.65	0.750	…	230	115

Table 2 – Dataset structure for Tactical aircraft sensitivity and machine learning analysis

2.2 Aircraft Usage Sensitivity Assessment based on Fatigue Life Distribution

The fatigue life distribution of the dataset described in Section 2.1 is plotted in order to derive valuable insights regarding the aircraft's usage sensitivity at each FAL. Figure 1 through Figure 3 present the plots for these distributions for three selected FALs.

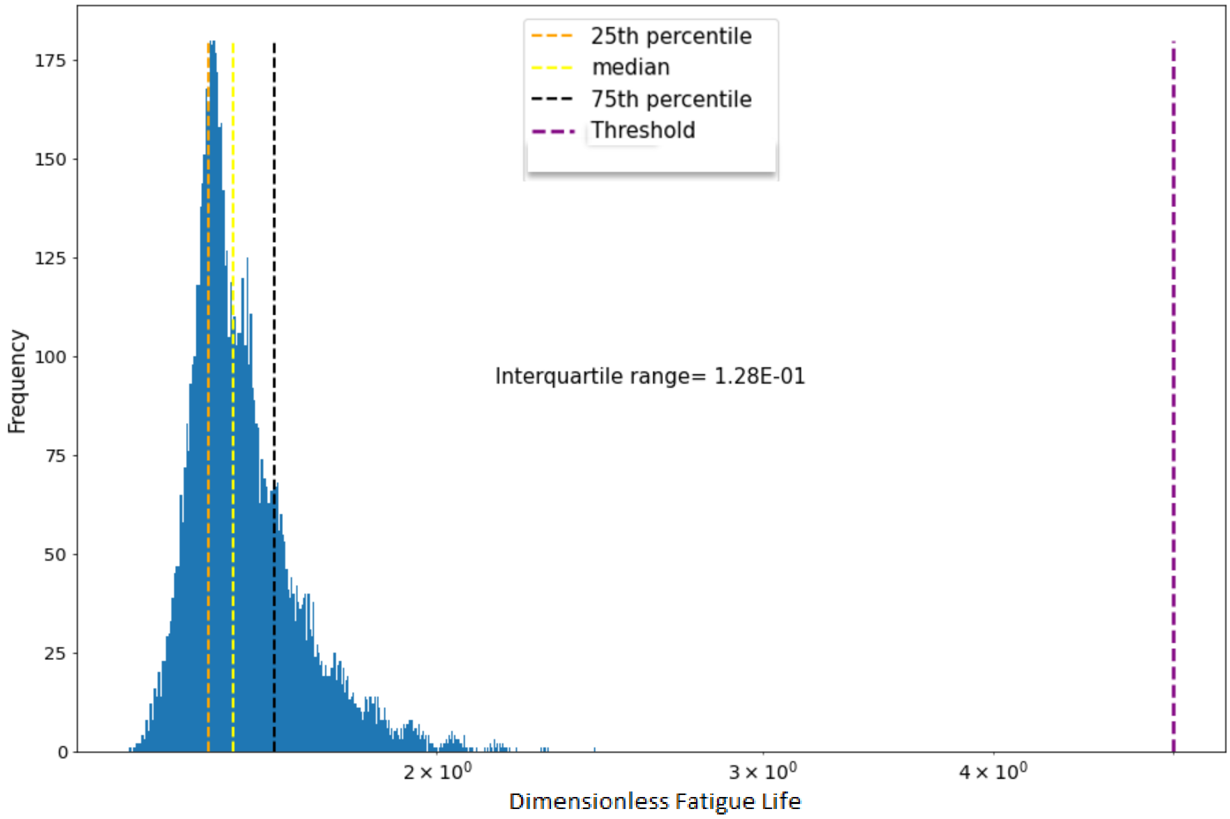

Figure 1 – Fatigue life distribution for FAL W001

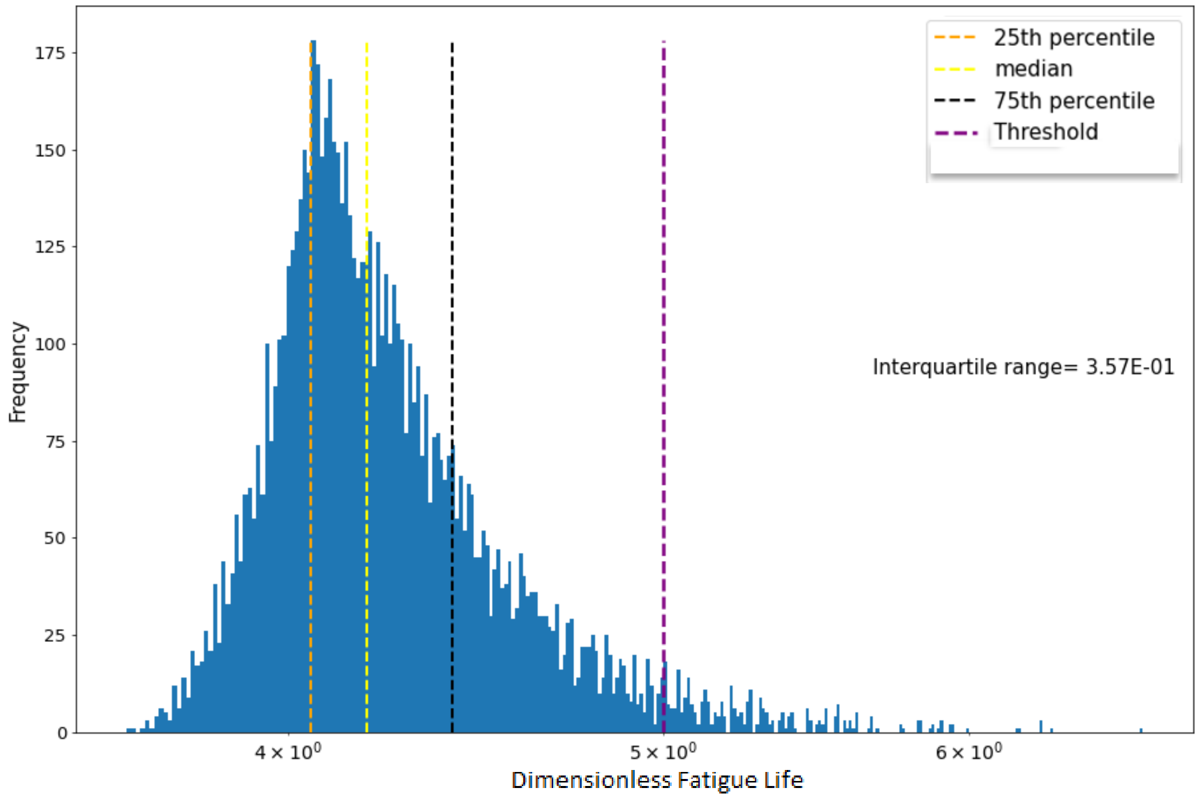

Figure 2 – Fatigue Life Distribution for FAL W002

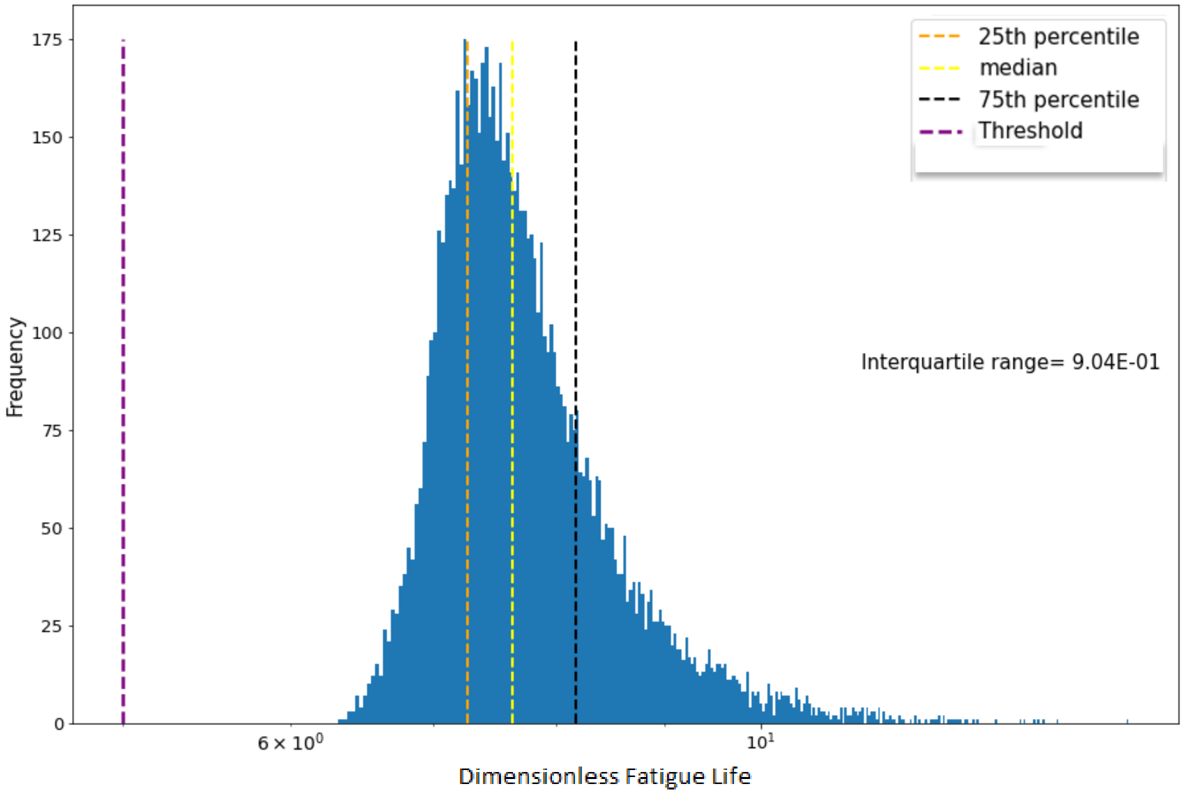

Figure 3 – Fatigue Life Distribution for FAL W003

It can be observed that the resulting fatigue life distribution is not a normal distribution but a right-skewed one. By calculating the interquartile range (which is the equivalent to the standard deviation for normally distributed data) as per (1), the usage sensitivity can be quantified.

Interquartile Range (IQR) = (Fatigue life at 75th percentile) - (Fatigue life at 25th percentile) (1),

Where the 75th percentile represents the middle value of the upper half of the data and the 25th percentile represents the middle value of the lower half of the data.

In this case, the higher the IQR range, the more sensitive to changes in the aircraft usage is a certain location. In the examples above, the most usage-sensitive location is W003, with an IQR of approximately 0.9.

However, while certain locations present a high IQR, it does not make them more critical as it could be the case that the lowest fatigue life value in the distribution is still higher than the criterion for validity selected.

From the examples above, and assuming that the threshold for validity is set at 5.0, location W001 would always have a fatigue life below the recommended threshold, no matter what the aircraft's usage is. On the contrary, location W003 will always have a fatigue life greater than the threshold regardless of the usage. The case of location W002 is an intermediate case of the previous ones, as for certain usages the fatigue life of the aircraft is always above the threshold but for some other usages (the majority) the fatigue life is below. In this case (and if the location requires it because of specific needs to optimize the maintenance programme at that location), the operator's efforts could be invested in trying to maximize those less severe usages in order to preserve the fatigue life at that location.

Therefore, these types of insights could provide the operator with an excellent flexibility to manage the structural maintenance programme in a more tailored manner, instead of having a fixed maintenance programme for each FAL from the beginning of the aircraft's life.

2.3 Predictive Power Score (PPS) to identify the most representative analysis locations

Another useful insight to obtain from this dataset would be to know which FALs fatigue lives are related to other FALs. This can be used to reduce the number of analysis locations needed in a certification for a change in aircraft usage, as the fatigue lives of some of these FALs could be inferred from the fatigue lives of other ones which are called pilot points for this study. This allows for a quicker assessment of the impact of the change in usage in the maintenance programmme.

In order to get the possible relationships between these FALs, the predictive power score (PPS) was used [6]. The PPS indicates the relationship between any two columns of the selected dataset, linear or nonlinear for categorical or numerical data. When obtaining the predictive power of column A over column B, the score can vary from 0 (no predictive power) to 1 (perfect predictive power). The score is also asymmetric as the predictive power of A over B does not necessarily need to be the same as of B over A.

The PPS was obtained for each possible pair of the selected FALs for this study (denoted as FR001 to FR008) and the results are presented in the matrix in Figure 4:

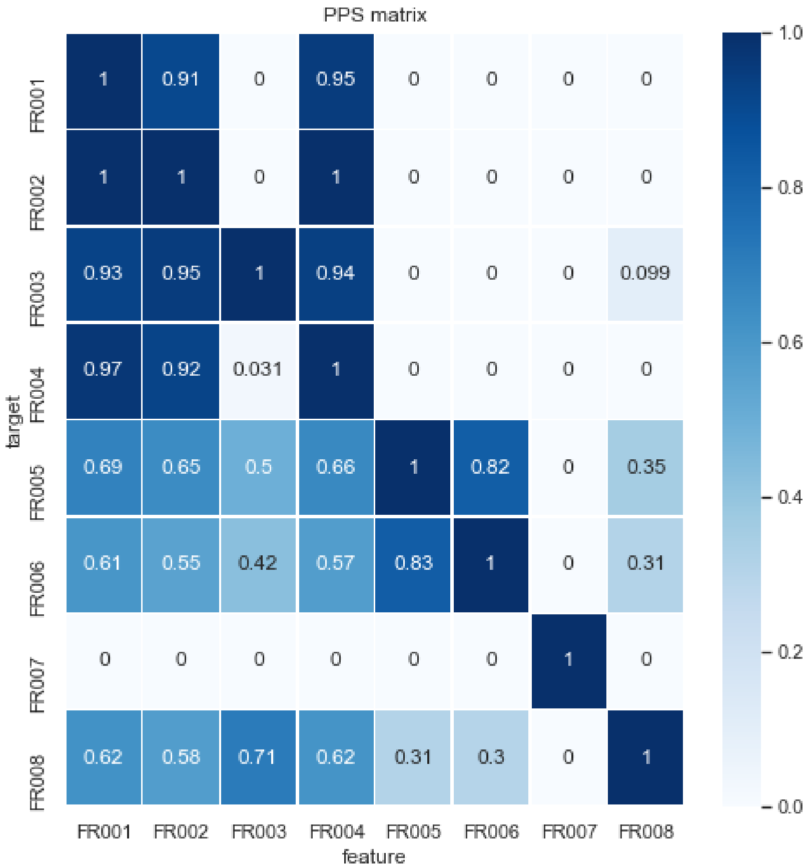

Figure 4 – Predictive Power Score matrix for several FALs

This matrix reflects all possible relationships between each pair of the FALs shown in terms of predictive power of each FAL over every other FAL in the sample. For example, there is no predictive power of FR001 over FR003 (the PPS is 0) but the predictive power of FR003 over FR001 is almost perfect with a PPS of 0.93. This is an example of an asymmetric prediction power score, where only one of the variables is able to predict the values of the other one.

There are also other cases where the predictive power is symmetric as in the case of FALs, FR001 and FR004, both with a PPS greater than 0.9 over each other.

FR007 has no predictive power over any FAL and no FAL has any predictive power over FR007, which would make this FAL an independent pilot point.

Different trials have shown that, when applied at full scale, this technique allows computational savings of approximately 35.0% in the evaluation of usages scenarios.

2.4 Feature Importance

A model-agnostic feature importance method called Permutation Feature Importance (PFI) [3] was used to determine the most important input features for each FAL. This exercise was performed not only as a sensitivity analysis but also as a check of the results obtained for the PPS in the section above: It was expected that if a certain pair of FALs have a mutually strong prediction power over each other, the same input features would be deemed as important by the PFI.

The feature importances for the FALs FR001-FR008 are shown in Table 3, where the top 5 most important features are presented. These results are divided into those FALs which have a mutually predictive power over each other and those which have no predictive power over any other FAL (as shown in the PPS matrix in Figure 4).

Chosen FALs with mutual predictive power					
FAL	**Parameter 1**	**Parameter 2**	**Parameter 3**	**Parameter 4**	**Parameter 5**
FR001	Altitude	% S3	% Ferry	Mission Duration	LW
FR002	Altitude	% Ferry	% S3	Mission Duration	TOW
FR004	Altitude	% S3	% Ferry	TOW	Mission Duration
Chosen FALs with no predictive power over any other FAL					
FAL	**Parameter 1**	**Parameter 2**	**Parameter 3**	**Parameter 4**	**Parameter 5**
FR007	No. Landings	% T1	% S4	% T2	Dynamic Pressure
FR008	Altitude	% S1	% S4	% Ferry	Dynamic Pressure

Table 3 - Feature Importances for FR001-FR008

For those locations with a mutually predictive power, the most important features are the altitude, the mission duration, the percentage of the Ferry mission on the mission mix (%Ferry), the percentage of the Surveillance type 3 mission on the mission mix (%S3) and the Mission's weight. These parameters are all related to each other and reflect the different loads that impact the fatigue damage on these locations, where pressure loads (given by the altitude) play a significant part along with those related to the aircraft's weight.

For those locations with no predictive power over any location, it can be seen that the relevant parameters such as the dynamic pressure and the altitude reflect the heterogeneous loading at these particular locations.

These results are in accordance with the PPS results in the previous section and with the domain-knowledge of an F&DT engineer. Additionally, they are obtained in a matter of a few seconds for all aircraft's FALs and provide the full list of flight parameters ranked in order of importance.

2.5 Machine Learning for Fatigue Damage Prediction

Machine learning and data analysis in the aeronautical industry are widely used in the areas of aircraft manufacturing monitoring and optimization, airline's operational efficiency increase and in the predictive maintenance of aircraft systems. All these applications rely on heavy amounts of data collected from sensors to derive the desired insights. In the aircraft structural analysis field, on the contrary, the amount of data available to perform a successful machine learning analysis is of a much lesser magnitude. This could demotivate the attempt to develop a machine learning application in this area, however, as shown in the following sections, these techniques can also offer valuable insights for the F&DT engineers.

Two different machine learning algorithms are used to assess their capability of prediction of the fatigue life of the FALs in the scenario of a change in the aircraft's intended usage. The dataset used for this study is that presented in Section 2.1.

The suitable machine learning algorithms that could be applied to the dataset in this study are part of two main families:

1) Supervised learning: these algorithms learn with labeled data consisting of a set of training instances, where each instance is a pair of an input and a desired output value. These models infer the best function that relates the input data to the output. Once this is done, the model is tested on unseen data (test data). These algorithms can be further split into classification ones: where the algorithm predicts which category the new data corresponds to, and regression ones: where a continuous numeric output is intended to be predicted. Some examples of supervised classification algorithms are support vector machines and nearest neighbors. Some examples of regression algorithms are the logistic regressor and the support vector regressor. Some algorithms can be used for both regression and classification problems like the XGBoost, the decision trees and the neural networks.

2) Unsupervised learning: The data in this case is not labeled, being the main objective for the algorithm to detect patterns in the data or grouping it into differentiated clusters. Some examples of algorithms in this family are k-means clustering and principal component analysis.

For this study, as the data is labeled (there is a set of input parameters: the flight parameters for each mission mix, which give the fatigue life of each FAL) and the value to be predicted (the fatigue life for each FAL) is continuous, the supervised learning regression algorithms eXtreme Gradient Boosting (XGBoost) and Support Vector Regressor are used.

XGBoost dominates structured or tabular datasets on classification and regression predictive modeling problems. Its main advantages are the computational efficiency and the overfitting control via the introduction of a regularization function [4].

Support Vector Machines (SVM) have been widely used for classification problems (where a class is assigned to a certain input or set of features) however, in 1996, Vladimir N. Vapnik et al. proposed a new version of SVM for regression [5] called Support Vector Regressor (SVR). Its main advantages for this study are its excellent generalization capability with also a low prediction error.

For the training and testing of both algorithms, the dataset is split into a training and a test set with an 80/20 ratio respectively. The model is trained and its hyperparameter selection and tuning are validated with the training set using a k-fold cross validation technique [11]. After the model is validated, it is evaluated on unseen data constituted by the test set to assess its predictive capability.

Machine Learning Prediction Results

To evaluate the quality of the predictions on the test set (blind test) of the selected model, the mean absolute error (MAE) of the prediction was obtained, which is defined as the mean of the absolute values of each prediction error on all instances of the test set. The prediction error is the difference between the actual value and the predicted value for that instance.

For both the XGBoost and the SVR algorithms, the MAE in terms of percentage of the fatigue life's mean value was below 5.0% for all locations evaluated, which demonstrates that fatigue life prediction using ML for a change in aircraft's usage is possible with a very small prediction error, which is within the typical scatter margins considered by aircraft manufacturers.

3. Usage Region Automatic Allocation with Machine Learning.

One of the most important applications of the FDE is a change of paradigm in the compliance with the airworthiness regulations by certifying an aircraft for a complete usage region (which is composed by a group of different mission mixes (Figure 5) instead of for a specific usage represented by only one mission mix. Being able to achieve this target, an optimized maintenance program could be offered to the operator from the beginning of the aircraft's operation, which would provide a great operational flexibility.

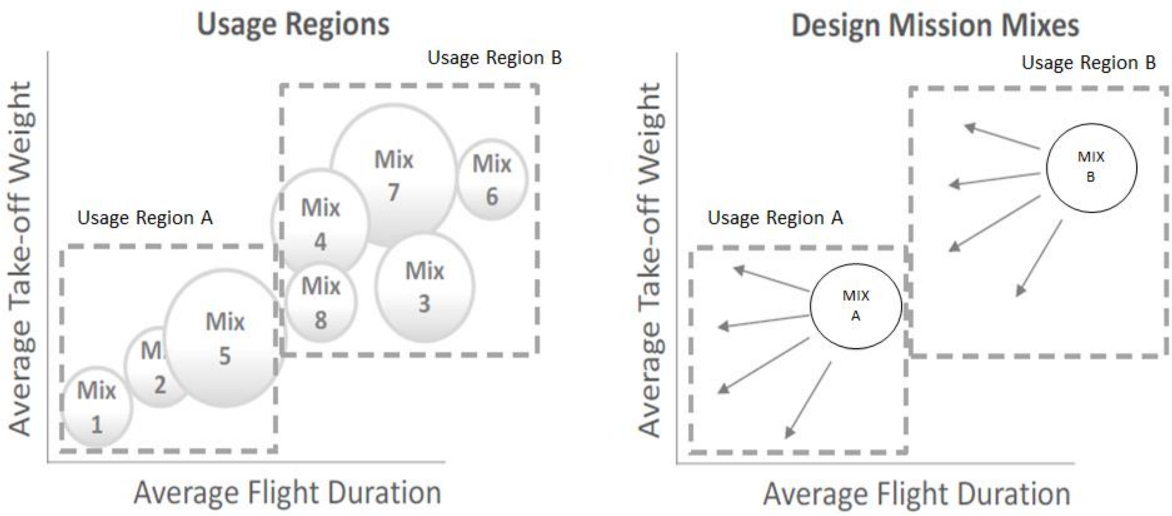

Figure 5 – Example of usage regions for a typical strategic transport aircraft, [1]. Each usage region (A or B) has a different maintenance program associated.

Another useful feature that is enabled for the engineers if the data they use is properly structured are the classification tools. In Section 2.5, machine learning has been used to solve a regression problem. Here, a simple example of a classifier is developed.

Typical classification problems in machine learning require predefined labeled data to be used as training data, and the new data is then assigned to one of those pre-established labels. The example developed here deals with the assignment of a typical strategic transport aircraft to a specific usage region attending to the cumulated usage (measured by flight parameters) in a certain period of time. The typical procedure of usage region allocation is accomplished in a non-automated manner, often involving the individual analysis of a high number of parameters with domain knowledge being also a key part of the loop. The proposal here is to generate and train a machine learning model that can automate this decision process.

For this purpose, data compressing fundamental flight parameters (such as Take-off Weight, Fuel Weights at Take-off and Landing) that is considered at initial stages of fatigue analysis is used as baseline here and employed in combination to its corresponding mission mix.

For the ongoing example, 6 missions and their associated flight parameters are considered representative of 2 different usage scenarios (3 missions for each usage region). By using a similar method as the one described in Section 2.1 and the resource of weak compositions [10], those missions are combined to obtain different additional combinations of missions and the labeled flight parameters. A total of 462 data points are obtained this way to serve as the initial dataset with the structure shown in Table 4, where all data have been standardized.

Mission percentages	TOW (kg)	LW (kg)	Parameter n	Target Label: Usage Scenario
85.0% 15.0% 0.0% 0.0% 0.0% 0.0%	0.8746	0.6944	…	0
85.0% 10.0% 5.0% 0.0% 0.0% 0.0%	0.8694	0.6936	…	0
0.0% 0.0% 0.0% 50.0% 5.0% 45.0%	0.9171	0.7023	…	1
0.0% 0.0% 0.0% 50.0% 0.0% 50.0%	0.9135	0.7010	…	1

Table 4 - Dataset structure for classifier

Features for the study are the flight parameters. The target to predict is the usage region label (the maintenance plan to which the aircraft will be assigned). The machine learning algorithm used for the prediction is a Decision Tree of the Python Scikit-Learn library with its default parameters.

Decision trees are one of the most common and widely spread classification algorithms. Evaluation of the results is accomplished with different metrics such as accuracy and the confusion matrix. The results obtained show that the classifier assigns all the test data correctly, obtaining an accuracy of 1.0 (the maximum value) and a diagonal confusion matrix.

The model classification capability is further tested with a series of real data aircraft usage records in Table 5. Those aircraft have been previously assigned to a usage region in a non-automated way and the goal is to check if the model is capable of obtaining the same region assignment.

Aircraft	TOW (kg)	LW (kg)	FATO (kg)	FAL (kg)	PL (kg)	AFT (h)	Parameter n	Usage Region assigned
Aircraft1	0.791	0.648	0.236	0.094	0.014	0.576	…	0
Aircraft2	0.797	0.645	0.243	0.091	0.013	0.594	…	0
…	…	…	…	…	…	…	…	…
Aircraft7	0.736	0.656	0.160	0.080	0.035	0.410	…	0
Aircraft8	0.953	0.666	0.391	0.104	0.021	0.911	…	1
Aircraft9	0.949	0.662	0.389	0.102	0.019	0.890	…	1
Aircraft10	0.959	0.664	0.398	0.103	0.020	1.000	…	1

Table 5 - Table of standardized real data usage records

The model achieves a perfect prediction (see Figure 6) for these real data usage records, allocating each one of them to the same usage region that the fatigue engineer would have decided to allocate it to. Obtaining a complete accuracy in this case can be a sign that this problem is a very easy one for the machine learning algorithm, which could be considered as an additional reason to incorporate this kind of techniques to assist the engineers in the decision-making process.

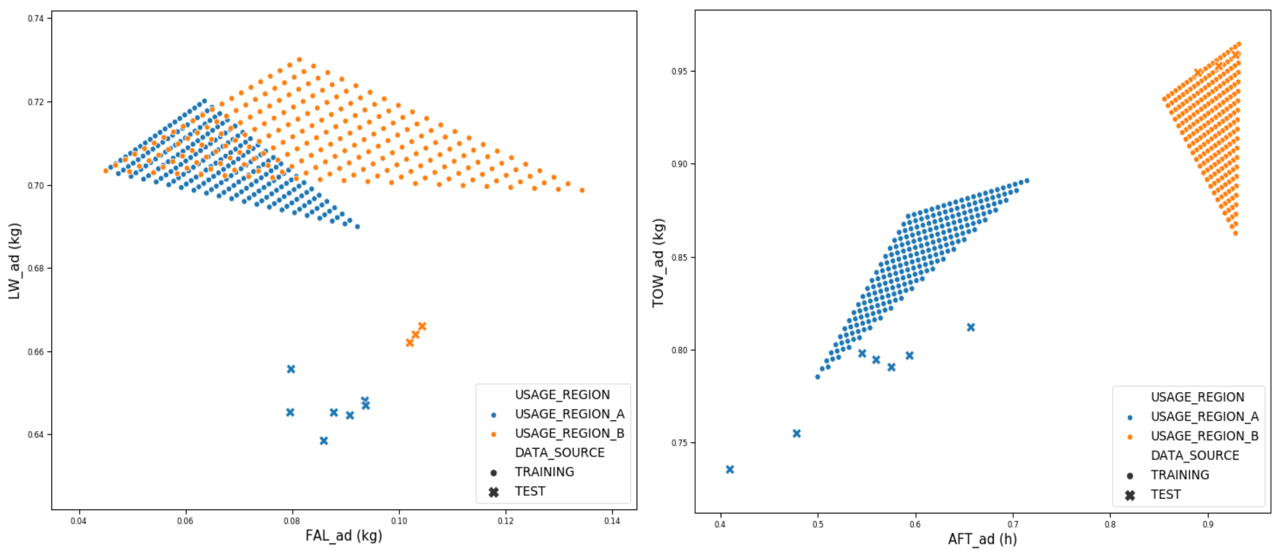

Figure 6 – Pairs of flight parameters and usage region allocation for training and actual data points

4. Conclusion and Way Forward

The concept of Fatigue Digital Equivalent (FDE) is being developed by Airbus Defence & Space in order to form the cornerstone of the qualification/certification processes of the near future, and to be the core of many of the continued airworthiness activities linked to structural fatigue. The symbiosis between physics-based models and data analytics can be considered as a new way to address some of the stochastic factors affecting the degradation of the airframe, thus providing FDT engineers with a novel set of tools to aid in a timelier assessment of the impact of a change in the aircraft's usage in the fatigue lives of the different structural significant items, among other factors.

This paper has presented some of the machine learning applications included under the FDE framework. It has been shown that, regardless of the usually small size of the available datasets in the military aircraft structural analysis field, the obtained results are promising. The next step will be the incorporation of these new techniques as part of the aircraft's certification process by achieving the airworthiness authorities' approval. For this purpose, a solid validation and verification framework is currently being consolidated. This is especially important as the machine learning models have an inherent prediction uncertainty when applied to new data due to their stochastic nature [7]. This uncertainty will be characterized by using probability theory, following the path that has already been explored for deep learning robotics applications [8] and in the medical diagnosis field [9].

5. Contact Author Email Address

Please contact Laura Zaplana, mailto: laura.zaplana-manzano@airbus.com

6. Copyright Statement

References

[1] E. Mirón et al. Continued airworthiness assessment of evolving usage patterns in military transport aircraft. *8th European Conference for Aeronautics and Space Sciences (EUCASS)*, Madrid, Spain, 2019

[2] Simske. S.. Meta-Analytics: Consensus approaches and system patterns for data analysis. 1st edition, Morgan Kaufman, 2019.

[3] Altmann, André; Tolosi, Laura; Sander, Oliver; Lengauer, Thomas. Permutation importance: a corrected feature importance measure, *Bioinformatics*, Volume 26, Issue 10, Pages 1340–1347, 2010

[4] Chen, Tianqi; Guestrin, Carlos. XGBoost: A scalable tree boosting system. *Cornell University*, arXiv:1603.02754, 2016

[5] Drucker, H. et al. Support vector regression machines. *Proceedings of the 9th International Conference on Neural Information Processing Systems*, Pages 155–161, 1996

[6] Wetschoreck, Florian et al., Predictive Power Score, DOI 10.5281/zenodo.4091345, 2020

[7] Hüllermeier, E., Waegeman, W. Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. *Machine Learning* 110, 457–506, 2021.

[8] A. Loquercio et al., A General Framework for Uncertainty Estimation in Deep Learning. *IEEE Robotics and Automation Letters,* 1907.06890v4, 2020

[9] Alizadehsani, R., Roshanzamir, M., Hussain, S. *et al.* Handling of uncertainty in medical data using machine learning and probability theory techniques: a review of 30 years (1991–2020). *Annals of Operations Research*, 2021

[10] Heubach, Silvia; Mansour, Toufik. Compositions of n with parts in a set. *Congressus Numerantium.* 168: 33–51. 2004

[11] Anguita, Davide et al. The ‘K’ in K-Fold cross validation. *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*. 2012