Data-Driven Model to Predict Aircraft Vibration Environment

Stéphane Février, Stéphane Nachar, Lionel Mathelin, Frédéric Giordano, Bérengère Podvin

A data-driven model to predict aircraft vibration environment

Stéphane Février* and Lionel Mathelin†
Université Paris-Saclay, CNRS, Laboratoire interdisciplinaire des sciences du numérique, 91405, Orsay, France.

Stéphane Nachar‡ and Frédéric Giordano§
Dassault Aviation, France.

Bérengère Podvin¶
Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire EM2C, 91190, Gif-sur-Yvette, France.

Vibration levels that onboard equipment must be able to withstand throughout their life for correct operation are mainly determined experimentally, as predicting the dynamic behavior of a complete aircraft requires computational means and methods that are currently difficult to implement. We present a data-driven methodology that leverages flight test accelerometer data to produce a predictive model. This model, based on an ensemble of artificial neural networks, performs a multi-output multivariate regression to estimate vibration spectra from a set of aircraft general parameters without having to characterize excitation sources. The model is compared with baseline models over two protocols: (i) standard training and testing and (ii) extrapolation to high dynamic pressures, in order to assess physical consistency. While the first protocol shows that all models can produce results accurate enough for this context, the second protocol shows that only the ensemble model is able to correctly extrapolate the energy. Using the SHAP method, we show that these results can be explained by the ability of our model to identify the dynamic pressure as the core feature used in the extrapolation protocol. The proposed model can be used in multiple applications, such as anomaly detection and vibration flight envelope opening.

Nomenclature

$\mathbf{A}$ = matrix

$\mathbf{a}$ = vector

a = scalar

B = dimension of $\mathcal{Y}$, number of frequency bands

*CNRS PhD student, stephane.fevrier@lisn.upsaclay.fr
†CNRS Researcher.
‡Data scientist.
§Vibration flight test specialist.
¶CNRS Research director.

$\mathcal{D}$ = dataset

N = number of datapoints in $\mathcal{D}$

P = dimension of $\mathcal{X}$, number of parameters

$\mathbf{s}$ = PSD (Power Spectral Density) estimation of $\mathbf{u}$

$\tilde{\mathbf{s}}$ = Smoothed-out PSD estimation of $\mathbf{u}$, broadband only

$\mathbf{u}$ = accelerometer temporal signal (g)

$\mathcal{X}$ = input space

$\tilde{\mathbf{x}}$ = unnormalized input vector, set of parameters

$\mathbf{x}$ = input vector, set of parameters $\tilde{\mathbf{x}}$ after normalization

$\mathcal{Y}$ = output space

$\tilde{\mathbf{y}}$ = reduced smoothed-out PSD estimation, RMS (Root Mean Square) level per frequency band

$\mathbf{y}$ = output vector, log-transformed and normalized $\tilde{\mathbf{y}}$

$\hat{\mathbf{y}}$ = output vector prediction

Θ = parameter space

θ = model parameters

$\mathbf{\Phi}$ = ensemble of models ϕ

ϕ = model

Subscripts

a = accelerometer

b = frequency band

i = datapoint

I. Introduction

Specifying the vibration environment of an aircraft is crucial to ensure mechanical strength of equipment [1]. It must be done at an early stage of the design, for civil and military aircraft. The vibration environment corresponds to the $10 - 2000$ Hz dynamic loads undergone by the structure and the onboard equipment during the use of the aircraft. These excitations are only statistically stationary during certain flight phases, such as the stabilized flight phases. They can originate from external forces, such as with aerodynamic field interactions, but can also originate from the aircraft itself, such as with any rotating machinery – engines, pumps and turbines – or on-board system. As reported in [2], examples of internal excitations include imbalance of rotating engine parts, misalignment in hardware assembly, additional magnetic, aerodynamic or hydrodynamic forces within the unit, etc.

The vibration environment is monitored for each equipment via uniaxial or triaxial accelerometers, generally located

at their root. The signal measured by each accelerometer is usually analyzed in terms of its local frequency spectrum or periodogram to estimate its PSD (Power Spectral Density). Each of the many sources of vibration contributes differently to a given sensor frequency spectrum, depending on the structural path and the distance from the source. Typically, interactions with the aerodynamic field will result in a broadband response, while rotating machinery will be identifiable by the resulting kinematic lines – energy concentration at certain frequency that depends on the rotating speed.

Specifying equipment consists of providing the robustness and fatigue levels that equipment should withstand. Robustness levels represent the maximum levels equipment might encounter, while fatigue levels represent the average vibration that equipment will encounter during their life. To complete the specification process, the robustness and fatigue levels are applied on equipment through random vibration test using shakers. Any equipment not complying with its specifications would have to be specified again or redesigned. Determining the vibration environment levels is done in the early design phase of the aircraft but can only be verified during flight tests. The validation process is cumbersome, costly, and occurs late in the development program.

There is currently no reliable and comprehensive approach to evaluate the dynamic behavior of a complete aircraft over the full range of frequencies considered, which extends up to 2000 Hz. As a result, specifications are based on existing standards and experience of previous aircraft, but they may not be precise enough. This may lead to equipment over-dimensioning, which has a negative impact on the overall aircraft mass.

One solution to produce more precise specifications would be to overcome current computational limitations and to develop reliable and comprehensive numerical models of the full aircraft. Various models of the dynamic environment of an aircraft have been developed for different ranges of frequencies.

In the low-frequency domain, up to 100 Hz, approaches focus on a modal behavior to extract localized responses - which is possible as the modal overlap is low. Deterministic techniques such as FEM (Finite Element Method) [3] or BEM (Boundary Element Method) [4] are used. In practice, these methods are limited to low frequencies, as they require a very fine discretization and a high degree of interpolation to remove pollution and dispersion effects [5].

In the high-frequency domain, from 800 to 2000 Hz, the interest is in describing a global behavior in terms of energy since the modal overlap is high. SEA (Statistical Energy Analysis) [6], [7], [8] is well suited for this frequency domain as it models the vibration energy transfers across partitions – subsystems – of a structure. The energy flow between them are described using a set of linear equations for the input, storage, transmission and dissipation of energy with a set of coefficients that needs to be determined. The problem is that the response is averaged over frequency and space, and is hence not suitable for local characterization.

In the mid-frequency range, from 100 to 800 Hz, neither low- nor high-frequency methods are suitable: low frequency methods become too costly due to the required high mesh refinement to reduce pollution effect, and the latter does not give the local response. Mid-frequency methods can be divided into hybrid approaches [9], [10], [11], [12] – combining low and high frequency methods – and Trefftz approaches – approximating the solution as a linear

combination of shape functions that satisfy *a priori* the equilibrium equations. An overview of these methods is provided in [13]. Some of the hybrid approaches try to leverage experimental data [14]. However they still need to be adapted for implementation at the industrial scale.

Aside numerical models, data-driven techniques could be used. Among machine learning models, ANNs (Artifical Neural Networks) have been shown to be successful in predicting vibrations in a wide variety of applications such as aerodynamics, mining or civil engineering. In the temporal domain, LSTMs (Long Short-Term Memory) have been shown to be very effective in order to predict the response of complex mechanical systems to a broadband excitation [15], [16]. They were also able to predict excess vibration events for engine aircraft [17]. LSTMs have also been developed to predict the vibration environment of electric cars and were found to be superior to classical ARMAX models [18]. In the frequency domain, ANNs have been shown to successfully predict structural stress [19]. They have also been used in aerodynamics to predict noise levels from limited pressure measurements in jets [20] and at the trailing-edge of an airfoil [21]. Other applications of ANNs include the prediction of ground vibration due to blasting for mining applications [22], where their performance was found to be superior to that of conventional regression [23]. ANNs have also been found to determine successfully ground vibration levels due to the passage of oncoming trains, [24].

In this paper, we develop an ANN-based model to predict the vibration level at any measurement point of an aircraft using flight test data which, to the best of our knowledge, has not been reported in the literature. Given the complexity of the data, consisting of sparse and noisy measurements, we adopt a robust viewpoint and use an ensemble-learning strategy which also provides uncertainty quantification. To characterize the performance of our method to be used in a real-world context, we define two evaluation protocols and compare the ensemble model prediction with two classical models: (i) standard linear regression, and (ii) XGBoost [25], a popular algorithm based on gradient-boosted decision trees. The paper is organized as follows: section 2 describes how data is preprocessed in order to provide a dataset for the models, while section 3 details the different models. Results are given in section 4, and section 5 presents a conclusion as well as perspectives.

II. Data preprocessing

In this section, we present the available data and the preprocessing steps implemented to generate a dataset used to fit the models.

A. Available data

In what follows, data from the flight test campaign of a civil jet aircraft are considered. 23 flight tests are used, which amounts to 65 hours of flight. Each flight test produces a large number of time series, each of which is associated with a sensor. Among the considered sensors, vibration environment accelerometers are sampled at 5000 Hz and the dataset to be handled is therefore very large.

The aircraft does not spend equal time at all flight points, making the measurement sparse along the flight domain. Moreover, measurements are contaminated with noise and possibly faulty sensors. The first step of data processing therefore consists in identifying usable sequences and discarding unreliable data.

We focus on the measurements along the vertical axis of $A = 5$ triaxial accelerometers, represented in Figure 1. This set was chosen to be representative of the whole aircraft as these sensors are known to be influenced by different vibration sources: V4Z and V5Z are heavily influenced by the engines, while V1Z and V2Z are mostly influenced by aerodynamic field interaction, V3Z being a middle ground.

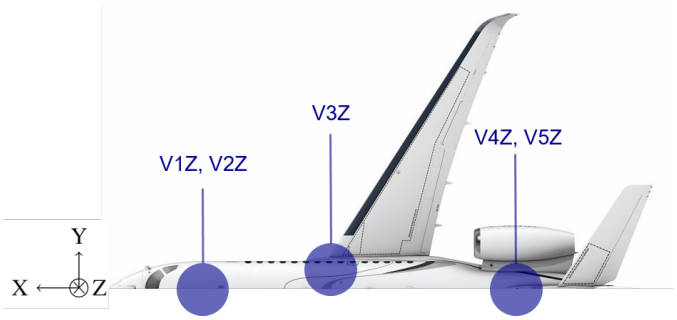


Fig. 1 Position of the 5 accelerometer channels considered in this study.

Along with vibration measurements, a large quantity of other variables are measured, giving information about the state and attitude of the aircraft: altitude, control surfaces position, engines rpm, etc. All of these parameters that help to describe the general state of the aircraft are what we call *general parameters*. These parameters are acquired through various sensors at different sample rates.

B. Dataset generation

The goal is to generate a dataset usable to train a predictive model. It must contain aircraft general parameters – our inputs –, and vibration levels at specific points – our outputs. Considering the large size of the temporal data, a choice has been made to consider the vibration levels through a compressed frequency representation, and to associate them with scalar general parameters. The preprocessing pipeline is presented in Figure 2, and allows to produce a sufficiently diverse yet compact dataset to train a regression model.

The first step consists in identifying flight phases, a difficult problem [26] since the time series are unlabelled. We will focus our approach on stabilized flight phases, where the aircraft conditions can be considered steady, which eases flight phase identification and avoids non stationary phases. A total of 342 sequences, accounting for 7.5 flight hours, are identified based on a set of logical rules – such as invariance of speed or altitude. This allows us to perform PSD estimations, a standard way of representing the frequency domain of random processes that requires the original signal to be stationary and ergodic.

For each sequence, sliding PSDs of 20 seconds are computed with a step of 1 second. For each spectrum $i = 1 \dots N$,

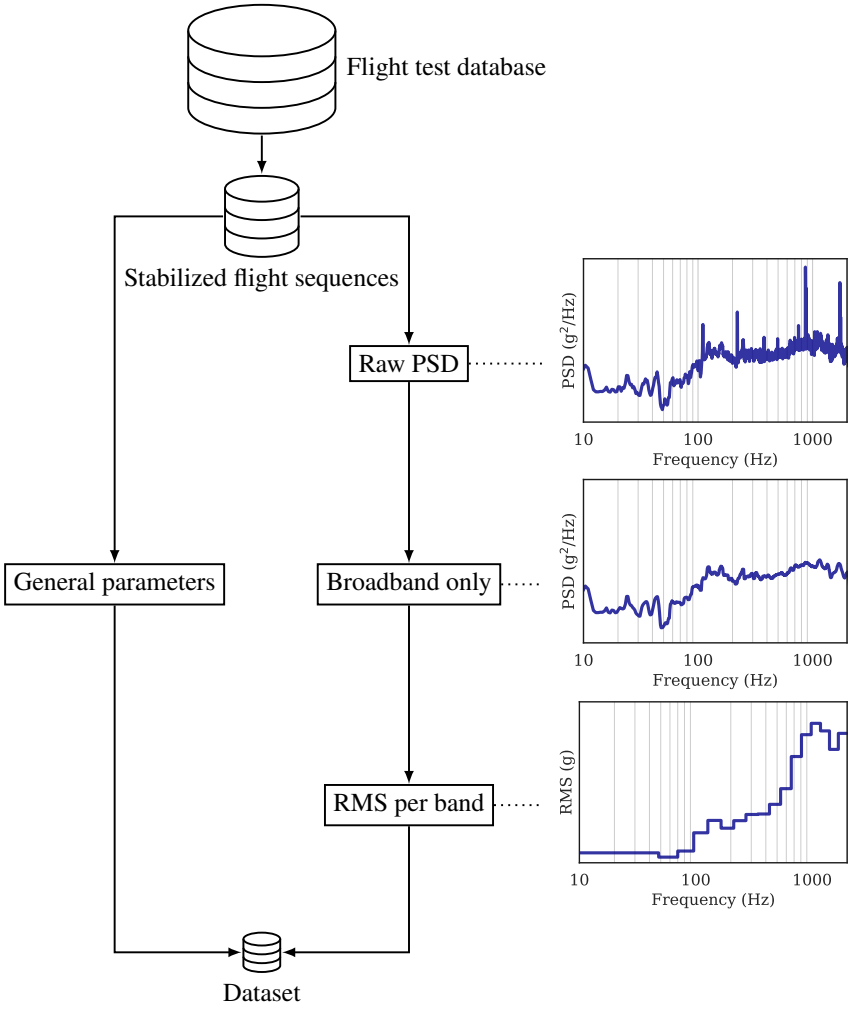


Fig. 2 Diagram of the preprocessing pipeline, generating a dataset from the flight test database.

the median of a set of general parameters is also computed to be used as an input in the regression model. Using a median instead of a time series is justified since the signal is stationary. The set of general parameters is detailed in the next subsection. A PSD represents the vibratory energy of the temporal signal over a frequency range. It can be expressed as:

$$\mathbf{s} = \mathrm{PSD}(\mathbf{u}) = \lim_{T \to +\infty} \frac{2}{T} \left| \mathrm{FFT}(\mathbf{u}) \right|^2 , \tag{1}$$

with $\mathrm{FFT}(\mathbf{u})$ the Fast Fourier Transform of the temporal signal $\mathbf{u}(t)$ and T the length of the signal. The result is expressed in g^2/Hz as the original signal is in g. Its spectral resolution β is given by:

$$\beta = \frac{\text{sample rate}}{\text{FFT window size}}. \tag{2}$$

In practice, a PSD can be estimated using Welch's method [27]. To compress the information, further transformations are applied on spectra. First, only the broadband energy is considered and the kinematic lines are removed. Kinematic lines are generated by rotating machinery and correspond to concentration of energy at a frequency varying with the rotating speed. Prediction of kinematic lines is left for another study as our interest is in estimating a spectra resulting from a large quantity of non separable sources.

To consider only the broadband signal, kinematic lines are removed by applying a moving median with a Δf window whose width grows linearly with the frequency. This allows to keep more modal information in the low frequency range, and less in the high frequency range. The value of a smoothed out signal $\tilde{\mathbf{s}}_i$ is given at any frequency by:

$$\tilde{s}_i(f) = \text{median}\left(\mathbf{s}_i\left[f - \frac{\Delta f}{2}, f + \frac{\Delta f}{2}\right]\right), \tag{3}$$

where $\mathbf{s}_i\left[f - \frac{\Delta f}{2}, f + \frac{\Delta f}{2}\right]$ contains the values of the original signal $\mathbf{s}_i$ from $f - \frac{\Delta f}{2}$ to $f + \frac{\Delta f}{2}$. To reduce information size, smoothed out spectra are compressed by expressing them as RMS (Root Mean Square) levels per frequency band, which is consistent with how vibration specifications are generally given. The RMS value of a reduced spectrum $\tilde{\mathbf{y}}_i$ in a given band $b = [f_1, f_2]$, is defined as:

$$\tilde{y}_{ib} = \sqrt{\int_{f_1}^{f_2} \tilde{s}_i(f)\mathrm{d}f}. \tag{4}$$

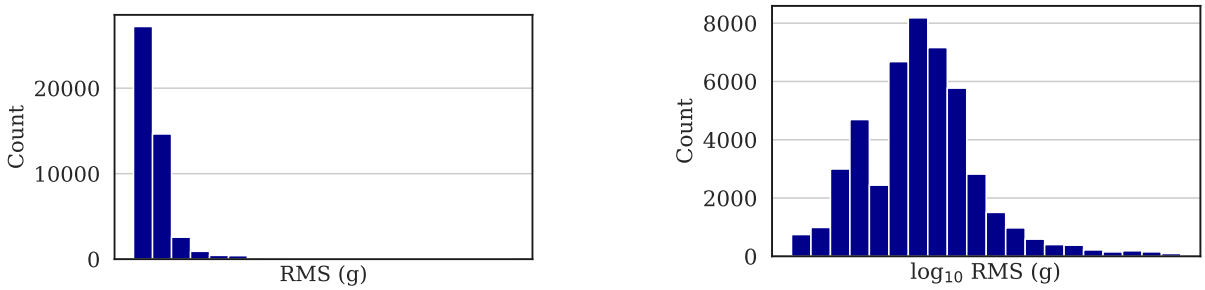
The choice of frequency bands is specific to domain knowledge and their width grows in an almost logarithmic way, although the first band is wider to incorporate known structural modes. A total of $B = 17$ frequency bands are chosen. Their width influences their variance: the wider they are, the higher the number of points used to compute the RMS value, thus the lower the noise. A sensitivity study of models error to the width of RMS bands is given in Appendix A.

Finally, as spectra can take values spreading over several orders of magnitudes, they are log transformed, see Figure 3. A min-max scaler is applied which constrains the data to lie in the $[0, 1]$ range using the min and max values across all $i = 1 \ldots N$ samples. The overall transformation from a compressed spectrum $\tilde{\mathbf{y}}_i$ to its final form $\mathbf{y}_i$ is given by:

$$\mathbf{y}_i = \frac{\log_{10}\tilde{\mathbf{y}}_i - \min_i(\log_{10}\tilde{\mathbf{y}})}{\max_i(\log_{10}\tilde{\mathbf{y}}) - \min_i(\log_{10}\tilde{\mathbf{y}})}, \qquad i = 1 \ldots N. \tag{5}$$

The final dataset $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1\ldots N}$ is composed of a set of N spectra $\mathbf{y}_i$ of dimension B associated with a vector of general parameters $\mathbf{x}_i$ of dimension P, each datapoint representing a *flight point*. The choice of parameters $\mathbf{x}_i$ is detailed in the next section. We also note that the measurement noise is high. Giving a precise estimate is hard, but taking into account the whole data acquisition system leads to a rough estimate of 5-10%.

The compression is such that 20 seconds of temporal data sampled at 5000 Hz are compressed into a frequency representation of $B = 17$ scalars, which amounts to a compression rate of 0.02%.



(a) Raw data **(b) Log-transformed data**

Fig. 3 RMS-levels log-transformation in the B_1 band.

C. Features selection process

From the generated dataset, the goal is to train a predictive model to estimate spectra $\mathbf{y}_i$ from an input $\mathbf{x}_i$, containing a set of P parameters or *features*. A constrain is that the model should be agnostic to equipment position in the aircraft, meaning that the same set P must be used to make prediction whether or not the same set of sources are influential. This allows to quickly train a new model without having to change model's architecture.

Among all the monitored parameters, roughly 50 are pre-selected via domain knowledge, since most parameters are not related to the vibration environment. From this set of 50 parameters, only $P = 13$ of them are kept by removing low variance, redundant and non influential parameters. There are presented below and express the state of the aircraft relative to several categories: attitude P_{A_1}, engines $P_{\mathrm{E}_{1\ldots3}}$, flight point $P_{\mathrm{FP}_{1\ldots3}}$, configuration $P_{\mathrm{C}_{1\ldots6}}$.

The input-output correlation matrix of each accelerometer is one tool used to analyze the relationship between features and frequency bands energy. As non-linear relationships are expected, Spearman's rank correlation coefficient r_s, see (7), is considered to reveal monotonic relationships instead of linear ones. It corresponds to Pearson's correlation coefficient ρ, see (6), applied on rank variables, i.e. variables with values ordered.

$$\rho(\mathbf{x}, \mathbf{y}) = \frac{\mathrm{cov}(\mathbf{x}, \mathbf{y})}{\sigma(\mathbf{x})\sigma(\mathbf{y})} \tag{6}$$

$$r_s(\mathbf{x}, \mathbf{y}) = \rho(\mathrm{rg}_{\mathbf{x}}, \mathrm{rg}_{\mathbf{y}}) \tag{7}$$

It must be noted that correlations are only a partial indication of features influence, as they only consider pairwise relationships. Moreover, some flight points are under-represented in the dataset, and several variables are almost categorical, which are difficult to explain using correlations. For the continuous variables, flight test constraints make features distributions highly multi modal as the flight domain is not covered in a continuous manner, see Figure 4 for an example. Because of this, the correlation value can be different from the importance of a feature in a specific model. Though, computing the correlations between general parameters and energy bands is still informative and can be used

easily when performing flight test campaigns.

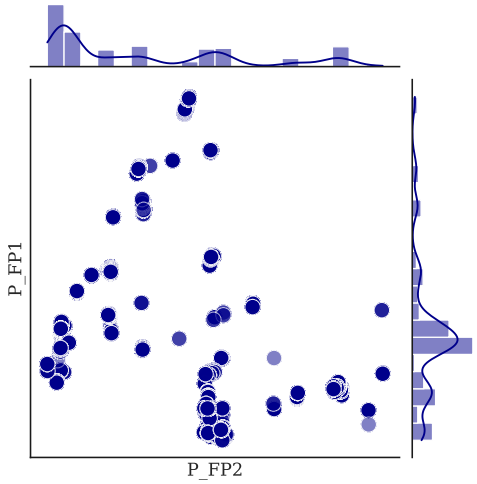


Fig. 4 Scatter plot with marginal distribution of flight point parameters P_{FP_1} and P_{FP_2}.

As an example, the input-output correlation matrix of the accelerometer V3Z is presented in Figure 5, where the inputs are on the y-axis and the outputs on the x-axis, from low to high frequency. This matrix contains only the $P = 13$ selected features.

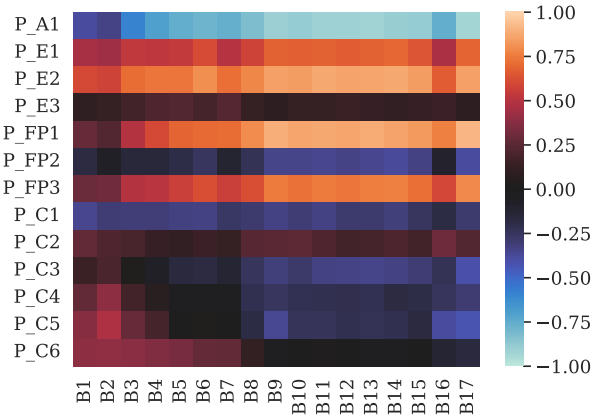


Fig. 5 Correlation matrix for accelerometer channel V3Z between inputs and outputs.

Low frequencies are mostly influenced by parameters related to the aircraft configuration. The parameters related to the engines also have an influence, but mainly because they are linked to the flight point which in turn is linked to the configuration. Mid and high frequencies are mostly influenced by parameters related to aircraft attitude, engines and flight point.

Even though kinematic lines are removed from spectra, engines are highly correlated with the RMS levels across all bands. This is, in addition to engines being linked to other parameters defining the flight point, because high %rpm

values generates broadband levels.

It can also be noted that several features show weak correlations with the outputs, but are yet retained. As noted before, this is because they were deemed to have a certain predictive power not encapsulated by the correlation value.

Before using the selected set of features, each feature vector $\tilde{\mathbf{x}}_i$ is normalized using a standard scaler with the empirical mean μ and standard deviation σ across all N datapoints:

$$\mathbf{x}_i = \frac{\tilde{\mathbf{x}}_i - \mu(\tilde{\mathbf{x}})}{\sigma(\tilde{\mathbf{x}})}, \qquad i = 1 \dots N, \tag{8}$$

which allows to ease the training process by forcing all variables to be in the same range of values.

III. Data-driven models

We now turn to the definition of the models used to predict spectrum from the general flight parameters selected above. After introducing the general problem formalism we present the baseline models implemented for comparison, then the ensemble model.

A. Problem formalism

Our goal is to predict a spectra $\hat{\mathbf{y}}_i \in \mathcal{Y} \subset \mathbb{R}_+^B$ from general parameters $\mathbf{x}_i \in \mathcal{X} \subset \mathbb{R}^P$, using a model $\phi \in \Phi$ parameterized by $\theta \in \Theta$ such that $\phi : \mathcal{X} \times \Theta \to \mathcal{Y}$. To do so, the parameters of the model are learned using a training dataset $\mathcal{D}_{\text{train}} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1\dots N_{\text{train}}}$ and a cost function $C(\theta)$ to minimize, quantifying the difference between the predictions $\hat{\mathbf{y}}_i$ and true values $\mathbf{y}_i$. It can also include a regularization term so as to promote sparsity and improve model generalization. Thus, we need to find the best parameters such that:

$$\theta^\star \in \arg\min_{\theta \in \Theta} C(\theta). \tag{9}$$

This corresponds to *supervised learning*, where a function is learned using a dataset of labeled data. Moreover, here we want to achieve a *multi-output multivariate regression*. To select the *hyperparameters*, i.e. parameters not adjusted during the learning, a process called hyperparameters optimization has to be performed where the best combination is sought for. There are various approaches to explore the hyperparameter space. A *grid search* involves exploring every possible combination of hyperparameter values within predefined intervals, while a Bayesian optimization can help to sample the space more efficiently and is convenient to use when the objective function is expensive to compute. Here, Bayesian optimisation is first performed with the Python implementation from Optuna [28] to identify a region of interest in the hyperparameters space, where a grid search is then conducted.

To use the available data in an efficient way and limit information leakage, k-fold *cross validation* is used which consists in splitting the training set into k subsets, training on $k-1$ subsets, and evaluating on the remaining one. Each

subset is called a *fold*, and a value of k=5 folds is used here. To have a non biased estimate of model performance, a final test is performed on a separate test set, yet unseen and never used for the optimization.

B. Linear Regression

The simplest regression model we can implement is a linear regression, and it is the first baseline model to evaluate. It is defined by:

$$\phi(\mathbf{x}_i) = \hat{\mathbf{y}}_i = \mathbf{A}\mathbf{x}_i + \mathbf{b}. \tag{10}$$

In our case, the ordinary least squares linear regression is used, which minimizes the residual sum of squares between target values and prediction:

$$C(\theta) = \sum_{i=1}^{N} (\mathbf{y}_i - \hat{\mathbf{y}}_i)^2. \tag{11}$$

The simplicity of this model allows interpretability, and makes training very fast. We also consider second-order and third-order polynomial bases, but no significant gain is obtained as reported in Appendix B.

C. XGBoost

To handle non-linearity between inputs and outputs, a more expressive model should be used. XGBoost [25] is a library implementing tree-based gradient boosting techniques with parallel computing. It is used in many applications such as ground vibration prediction [29] and vehicle driving risk prediction [30]. No vector notations is used here as multi-output regression is not supported, thus one model is learned for each frequency band b, and we denote the i-th output for a band b as y_{ib}. The idea is to perform gradient boosting, where an ensemble of weak predictors are combined in an additive way such that:

$$\hat{y}_{ib}^{(t)} = \hat{y}_{ib}^{(t-1)} + f_t(x_{ib}), \tag{12}$$

where f_t is a a classification and regression tree (CART). It can be defined as:

$$f_t(x_{ib}) = w_{q(x_{ib})}, \quad w \in \mathbb{R}^T, \quad q : \mathcal{X} \to \{1, 2, \ldots, T\}, \tag{13}$$

where w is the vector on leaves and $q(x_{ib})$ is a function that attributes each data point x_{ib} to a specific leaf on the current tree t. We can define the complexity of a tree as:

$$\omega(f_t) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2, \tag{14}$$

where $\gamma > 0$ and $\lambda > 0$ are hyperparameters. Finally, the cost function is:

$$C(\theta) = \sum_{i=1}^{N} (y_{ib} - \hat{y}_{ib})^2 + \sum_{t=1}^{T} \omega(f_t), \tag{15}$$

which aims to minimize the MSE (Mean Squared Error) as a loss function, while a regularization term on trees complexity $\omega(f_t)$ is included. One major problem with XGBoost is that predictions are only based on a sum of values attached to tree leaves: no transformation is applied. As a result, it is not able to perform out-of-sample extrapolation where it can only predict a constant.

D. Artificial Neural Network

An ANN (Artificial Neural Network) is a model comprised of different layers: an input layer, one or more hidden layers, and an output layer. First, let us consider the case of Fully Connected Neural Networks (FCNN) where each neuron in a layer is connected to all neurons in the next layer, see Figure 6. The value of all the neurons in a layer l is given by:

$$\mathbf{a}^{(l)} = \mathbf{g}^{(l)}(\mathbf{W}^{(l)}\mathbf{a}^{(l-1)} + \mathbf{b}^{(l)}) = \mathbf{f}^{(l)}(\mathbf{a}^{(l-1)}), \tag{16}$$

with $\mathbf{a}^{(l)}$ the neurons values at layer l, $\mathbf{W}^{(l)}$ the connection weights, $\mathbf{b}^{(l)}$ the bias, and $\mathbf{g}^{(l)}$ the *activation function* which purpose is to introduce non-linearity. Several activation functions exists, such as sigmoid, tanh or ReLU (Rectified Linear Unit). Denoting $\mathbf{f}^{(l)}$ the transformation from layer $l-1$ to layer l, the general equation of a L-layer FCNN can be written as:

$$\phi(\mathbf{x}_i) = (\mathbf{f}^{(L)} \circ \ldots \circ \mathbf{f}^{(3)} \circ \mathbf{f}^{(2)})(\mathbf{x}_i) = \hat{\mathbf{y}}_i. \tag{17}$$

Each new layer is built on the previous one and adds a level of complexity. The number of neurons in a given layer defines its dimension.

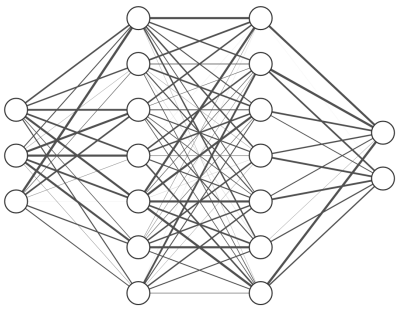


Fig. 6 Illustration of a neural network with 2 hidden layers.

Training is based on the same cost function as for the linear regression, MSE, see Eq. (11). To ease the hyperparameters optimization process, each hidden layer is chosen to have the same structure. Model stability is however fostered by using batch normalization, i.e. centering and rescaling of the layers. Each layer performs the following transformations:

$$\text{Dense} \rightarrow \text{BatchNormalization} \rightarrow \text{Activation}$$

In order to select the optimal configuration, tests were carried out for architectures of 2 to 6 layers. The final architecture is made of 3 hidden layers, each of them having 30 neurons and using a ReLU (Rectified Linear Unit) as activation function.

E. ANNs ensemble

When using neural networks, a problem that can arise is overfitting, where the model is unable to generalize. Another issue is that no indicator tells us how confident one can be in the predictions of the model.

A deep ensemble [31] is an ensemble of neural networks allowing to quantify model uncertainties. It was recently shown that this method produces better results than other uncertainty quantification methods [32]. In our case, we use an ensemble of ANNs: 8 models, with identical hyperparameters but different initial conditions, are trained on the same dataset for 5000 epochs each with a batch size of 32, using an Adam optimizer. Contrary to the deep ensemble presented in [31], the individual models are outputting scalar values and trained by minimizing the MSE, instead of outputting the mean μ and variance σ^2 of a distribution and training by minimizing the negative log-likelihood criterion.

The prediction of the ANNs ensemble is considered to be a Gaussian distribution parameterized by a mean μ and standard deviation σ over all predictions $\hat{\mathbf{y}}_i$ of the 8 networks:

$$\phi(\mathbf{x}_i) \sim \mathcal{N}(\mu(\hat{\mathbf{y}}_i), \sigma(\hat{\mathbf{y}}_i)). \tag{18}$$

Prediction disagreement between individual models of the ensemble produce large standard deviation σ and indicates that prediction should be taken carefully.

F. Evaluation methodology

For each datapoint i and energy band b, model predictive performances are assessed using the Absolute Log Error:

$$\text{ALE}(y_{ib}, \hat{y}_{ib}) = |(\log_{10} y_{ib} - \log_{10} \hat{y}_{ib}) \times 100|, \tag{19}$$

where a 100% error amounts to a 1 order of magnitude error. Only results for the unseen test data are presented, which are datapoints that were not used for model training. As mentioned in the introduction, we use two protocols which

correspond to different ways of splitting the dataset. For each protocol, all models are trained and evaluated on the same data.

The first protocol corresponds to a standard random splitting of the dataset (constituted of $N = 9430$ datapoints), where 90% of the data is used for training ($N_{\text{train}} = 8487$ datapoints) and 10% for test ($N_{\text{test}} = 943$ datapoints). The goal is assess whether the model is expressive enough to represent the data.

The second protocol aims to evaluate the physical consistency of the model, through its ability to model the dynamic pressure. The dynamic pressure, here labelled P_{FP_1}, is directly linked with the broadband aerodynamic loads on the aircraft; the RMS-levels are expected to increase linearly with it, all other things being equal. In order to check this, the training dataset in the second protocol contains only samples with a dynamic pressure value lower than a given threshold, and the model is tested on the remaining datapoints, corresponding to a higher dynamic pressure. We chose a threshold of $\text{p}_{\text{dyn}} = 100$ hPa, which yielded respective values for the training and the test dataset of $N_{\text{train}} = 7232$ and $N_{\text{test}} = 2198$. Figure 7 shows the train-test split for the dynamic pressure and another flight point-related parameter.

This second protocol may seem curious at first, as no i.i.d. (independent and identically distributed) hypothesis holds for the training and test sets. However, as mentioned before, we expect that the output will have a linear dependence on this variable. Thus, such a dataset split allows us to specifically evaluate the ability of the model to learn this relationship, as the training set outputs corresponding to a low dynamic pressure – are influenced by many other variables, and the test set outputs mostly depend on the dynamic pressure.

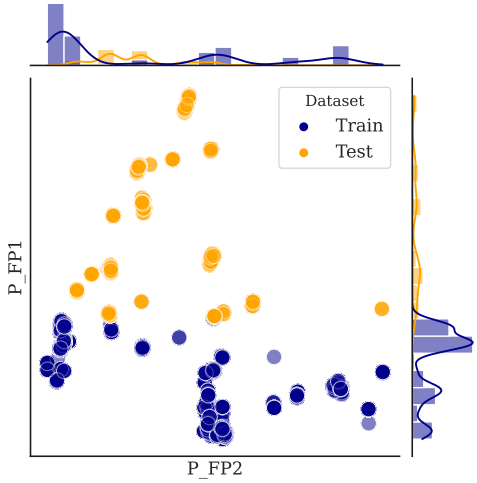


Fig. 7 Scatter plot with marginal distribution of flight point parameters P_{FP_1} and P_{FP_2} using the extrapolation protocol.

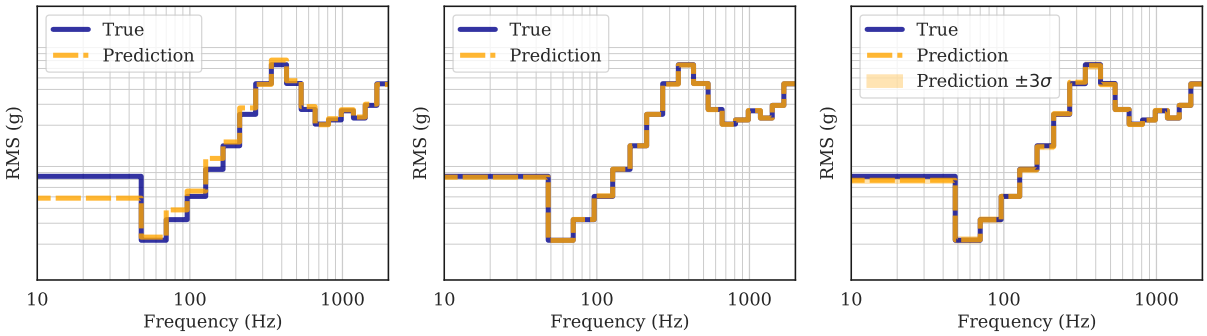
IV. Results

This section details the predictive performances of each model for both evaluation protocols. In each case, we first present an example of prediction for a specific datapoint, then the distribution of error over frequency bands for a specific accelerometer, then the global results for all accelerometers. In addition to that, for the extrapolation protocol we show the projection of the prediction over the dynamic pressure variable. Finally, after presenting the prediction performances, models are compared through their SHAP values.

A. Classical evaluation

1. Prediction example for one flight point

First, we consider one flight point randomly taken from the test set and compare the performance of each model. Figure 8 shows a typical output i.e. one spectrum prediction for an input datapoint, associated with the sensor V3Z which is in the middle of the aircraft, thus representative of equipment subjected to various sources. In the case of the ensemble model, a confidence interval corresponding to 3 standard deviations is also provided. All predictions fit almost perfectly the measurements, with some little variability. In this specific example, the linear regression underestimates the first frequency band, which is highly influenced by aircraft configuration and thus hard to predict through a linear regression.



(a) Linear regression **(b) XGBoost** **(c) ANNs ensemble**

Fig. 8 Predicted spectrum of a given datapoint for the accelerometer V3Z, classical evaluation.

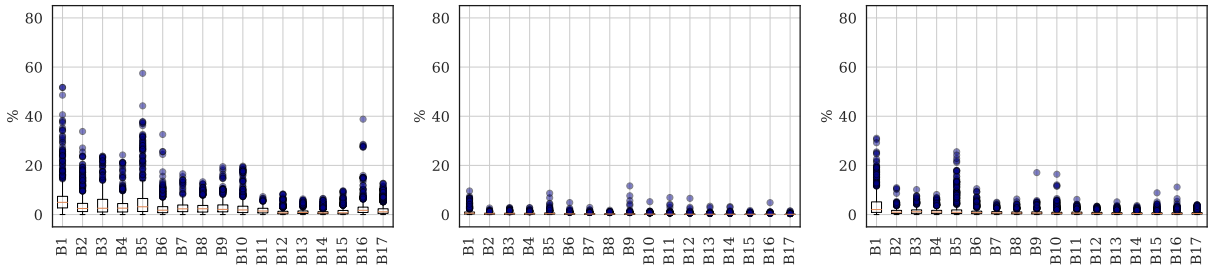
2. Error distribution over frequency bands for an accelerometer

To better characterize models errors, we analyze the mean error over each frequency band b over all N datapoints i:

$$E_b = \frac{1}{N}\sum_{i=1}^{N} \mathrm{ALE}(y_{ib}, \hat{y}_{ib}). \tag{20}$$

Figure 9 shows the boxplot of this error for each frequency band on the accelerometer V3Z across all three models. We first notice that they all have good performances, the end of distributions being consistently under 20%, and even less

most of the time. However, a large amount of outliers can be seen, which correspond to under-represented flight points. Even though the outliers are far from the distribution, these errors are acceptable for the XGBoost and ANNs ensemble models. Finally, The bands B_1 and B_5 are consistently harder to predict, which is notable for the linear regression and ANNs ensemble.



(a) Linear regression **(b) XGBoost** **(c) ANNs ensemble**

Fig. 9 Boxplot of the error per band for the accelerometer V3Z, classical evaluation.

3. Global error

Finally, the overall predictive performance of each model is summarized in Table 1. For each accelerometer, the average of the ALE is computed across every test datapoints i and frequency band b, see (21), and the last column provides the average across all $A = 5$ accelerometers.

$$E = \frac{1}{B}\frac{1}{N}\sum_{b=1}^{B}\sum_{i=1}^{N} \mathrm{ALE}(y_{ib}, \hat{y}_{ib}). \tag{21}$$

	Accelerometer					
	V1Z	V2Z	V3Z	V4Z	V5Z	Mean
Linear Regression	3.9%	5.1%	2.8%	2.6%	3.1%	3.5%
XGBoost	0.3%	0.4%	0.3%	0.3%	0.4%	0.3%
ANNs ensemble	**1.3%**	**1.4%**	**1.1%**	**0.9%**	**1.3%**	**1.2%**

Table 1 Mean error for each accelerometer depending on the model, classical evaluation.

XGBoost offers the best performances with this test protocol, followed by the ANNs ensemble, then the linear regression. The error is relatively similar across all the accelerometers, even though V2Z at the front of the aircraft seem to be slightly harder to predict. Since engines are located at the rear of the aircraft, this sensor is mostly influenced by the aerodynamic field and the flight configuration.

The performance of each ANN of the ensemble is given in Figure 10. The variability between ANNs depends on the sensor, but is relatively low in this case. As expected, the ensemble error is lower than the average error of all the ANNs by approximately 30% on average. Here, the ensemble error is even lower than any single ANN at all times. By

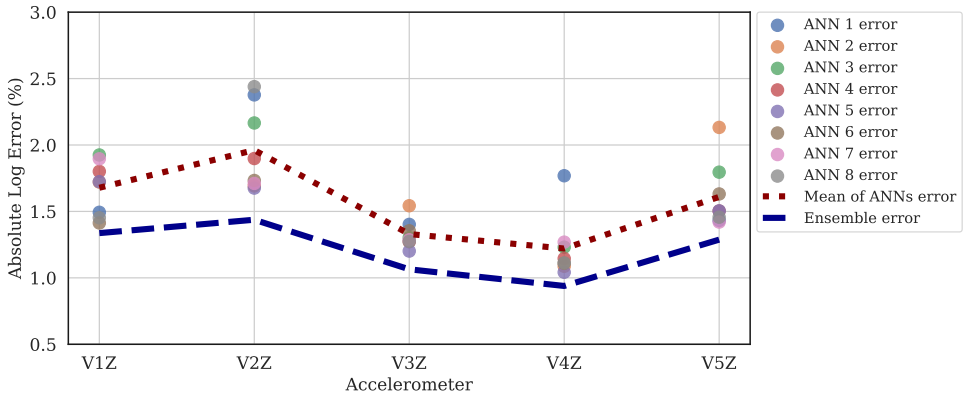


Fig. 10 Mean error for each accelerometer depending on the neural network, classical evaluation.

studying the confidence interval of predictions, we find that 93.2% of predictions are within a $\pm 3\sigma$ interval, and 96.9% of predictions are below $\mu + 3\sigma$, which is deemed to be conservative. A large variability between individual predictors indicate that the flight point is difficult to predict for these models and that the predicted values should be taken carefully. We note that some uncertainty quantification can be obtained with a single gradient boosting model [33], but was not used here.

The good performances of all models show that the set of inputs is sufficient to predict the outputs. In fact, as the measurement error can be roughly estimated to be around 5-10%, taking into account the whole data acquisition system, it is likely that overfitting is present. We note that learning a dataset without generalization ability can be useful to generate spectra instead of querying a database. This can be used for anomaly detection, where the validity of a new measurement can be determined by comparing it with its predicted value in the model. In any case, overfitting is not possible in the second evaluation protocol, as the training set and the test sets are disjoint in parameter space.

As a side note, learning a dataset without generalization ability can be useful to generate spectra instead of querying a database. This can be used for anomaly detection, where a measurement is compared against such synthetic data.

B. Extrapolation evaluation

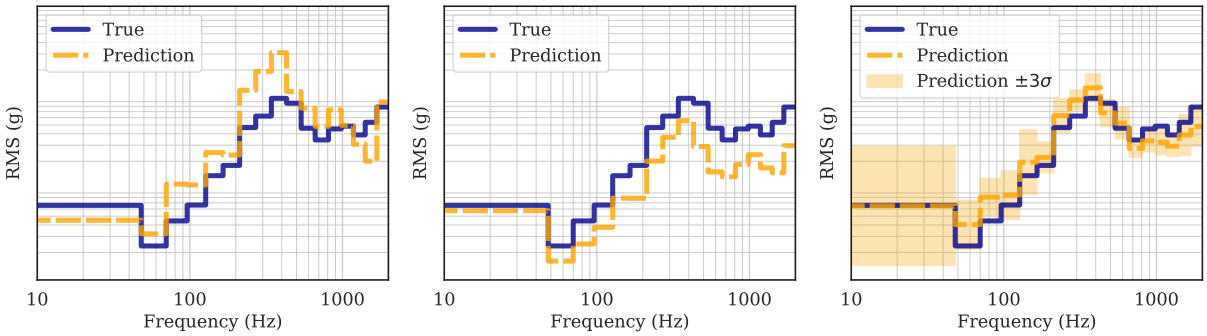
The classical evaluation protocol described above relies on the hypothesis that variables are independent and identically distributed (i.i.d.), and gives a measure of how well a model can interpolate. Here however, we possess knowledge about the physics that generated the data, namely there is a linear dependence between the energy spectra and the dynamic pressure. The goal of the new evaluation protocol is to determine whether the model is able to extrapolate, which is generally not possible for neural networks, but should be possible in some specific cases [34]. To this end, all data where the dynamic pressure is below 100 hPa is used for training, and all the remaining data is used for testing. The value of 100 hPa corresponds to the mid-range value of the dynamic pressure.

1. Prediction example for one flight point

Again, we consider one flight point randomly taken from the test set and compare the performance of each model, see Figure 11. One spectrum prediction for an input datapoint is shown, associated with the sensor V3Z. The datapoint corresponds to a dynamic pressure of 210 hPa, which is far outside the range of the training set.

For linear regression, under-prediction is observed in the lower frequency band, but substantial over-prediction is present at almost all frequencies except in the highest range. In particular, the peak of the spectrum over the frequency range is overestimated by a factor of three.

In contrast, above the first frequency band, XGBoost consistently under-predicts the spectrum by a nearly constant factor of 2. Best results are observed for the ensemble model, which slightly over-predicts the energy for frequencies lower than the peak frequency, and slightly under-predicts it for higher frequencies. We note that the true spectrum lies within the confidence interval of $\pm 3\sigma$.



(a) Linear regression **(b) XGBoost** **(c) ANNs ensemble**

Fig. 11 Predicted spectrum of a given datapoint for the accelerometer V3Z, extrapolation evaluation.

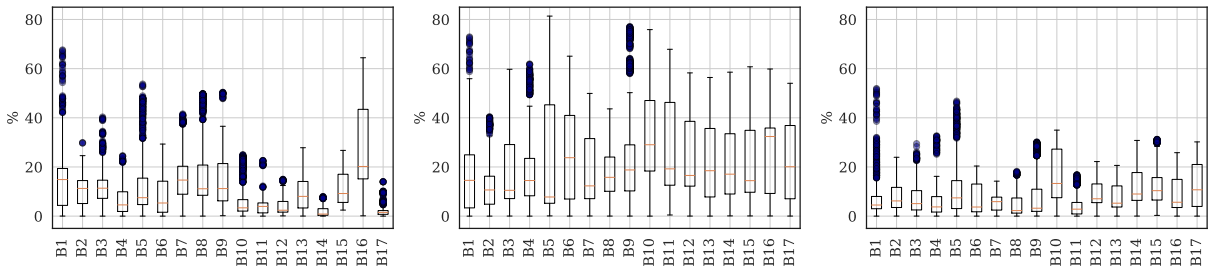
2. Error distribution over frequency bands for an accelerometer

We then represent the boxplot of the error averaged over each frequency band b. Figure 12 shows this for all three models applied on the accelerometer V3Z.

The ensemble model shows the best performances and has its outliers closer to its distributions than with the other models. Distribution spread itself is much higher than for the classical evaluation protocol, as it is a much harder problem. We note that the frequency band B_1 is constantly a high error band. This is explainable by its higher variability and link to many variables affecting the aerodynamic field.

3. Dynamic pressure projection

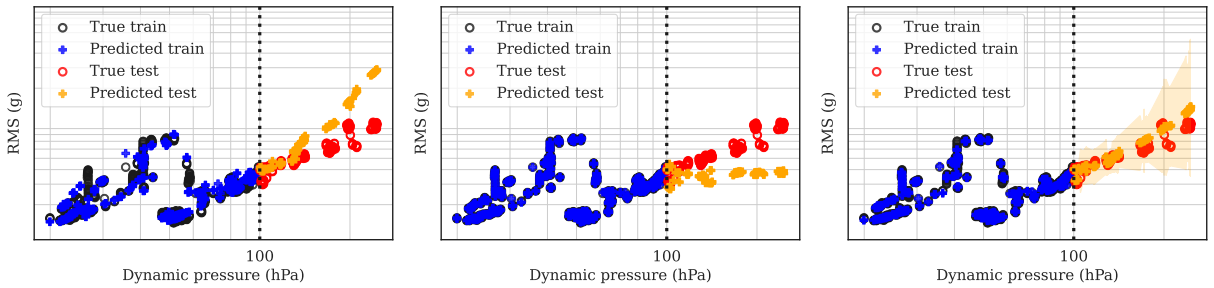
To better understand the behavior of the models, Figure 13 compares how predicted and real RMS levels in a specific frequency band vary with the dynamic pressure for both training and test sets. We selected the frequency band B_8, which provides a stringent test of the models, as it is highly influenced by aircraft configuration for dynamic pressure



(a) Linear regression **(b) XGBoost** **(c) ANNs ensemble**

Fig. 12 Boxplot of the error per band for the accelerometer V3Z, extrapolation evaluation.

values under 100 hPa and mostly by the dynamic pressure when it is over 100 hPa.



(a) Linear regression **(b) XGBoost** **(c) ANNs ensemble**

Fig. 13 RMS-levels on the B_8 frequency band depending on the dynamic pressure for the accelerometer V3Z. The black dotted line separates the training set (left) and the test (set). The shaded region corresponds to a value of $\pm 3\sigma$ for the ensemble.

Linear regression is not able to fit accurately the training set which is likely to originate from a lack of expressivity of the model. In the test set, a positive trend is predicted although its slope is overestimated in this case. We emphasize that since this is a one-dimensional projection from a much higher-dimensional space, we do not observe a strictly linear dependence as other variables come into play in the linear representation.

XGBoost almost perfectly reproduces the training set but is unable to extrapolate to the test set, as it predicts an almost constant RMS with the dynamic pressure. This is consistent with the fact that it is based on decision trees. It has likely overfitted the data.

In contrast, the ensemble model fits well the training set while extrapolating correctly the increasing trend of the test set. One can see that the uncertainty associated with the model also increases with the dynamic pressure, which is consistent with the increasing distance between the test datapoints and the training set.

4. Global error

Finally, we conclude this evaluation by giving the overall predictive performance of each model in Table 2. For each accelerometer, average of the ALE error is computed across every test datapoints and frequency band, while the last

column gives the average across all $A = 5$ accelerometers.

	Accelerometer					
	V1Z	V2Z	V3Z	V4Z	V5Z	Mean
Linear Regression	15.1%	16.2%	10.2%	16.9%	15.9%	14.9%
XGBoost	21.5%	22.3%	22.8%	22.9%	20.0%	21.9%
ANNs ensemble	**9.1**%	**12.8**%	**8.6**%	**10.2**%	**13.3**%	**10.8**%

Table 2 Mean error for each accelerometer depending on the model, extrapolation evaluation.

The worst results are obtained with XGBoost, with a global mean error of 21.9%. As stated before, it is based on an ensemble of trees that do not apply any transformation to the data. As a result, its extrapolation abilities are limited to a constant function which explains its poor performance. The error obtained with linear regression is smaller (about 15%), but still inferior to the ensemble model, which provides an error of only 10.8%. The difference can be attributed to the limited expressivity of the linear regression. In addition to the lower error, the ensemble model provides a confidence interval that captures 98.3% of the predictions into a $\pm 3\sigma$ interval, while 98.7% of the predictions are below $\mu + 3\sigma$.

The error of each neural network composing the ANNs ensemble is given in Figure 14. The ensemble error is significantly lower than the average error of ANNs: for V2Z, the mean of ANNs error is two times higher than the ensemble error.

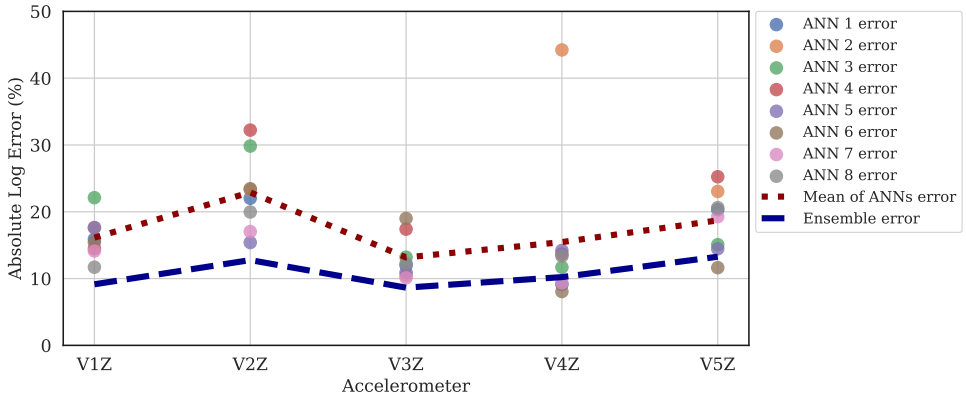


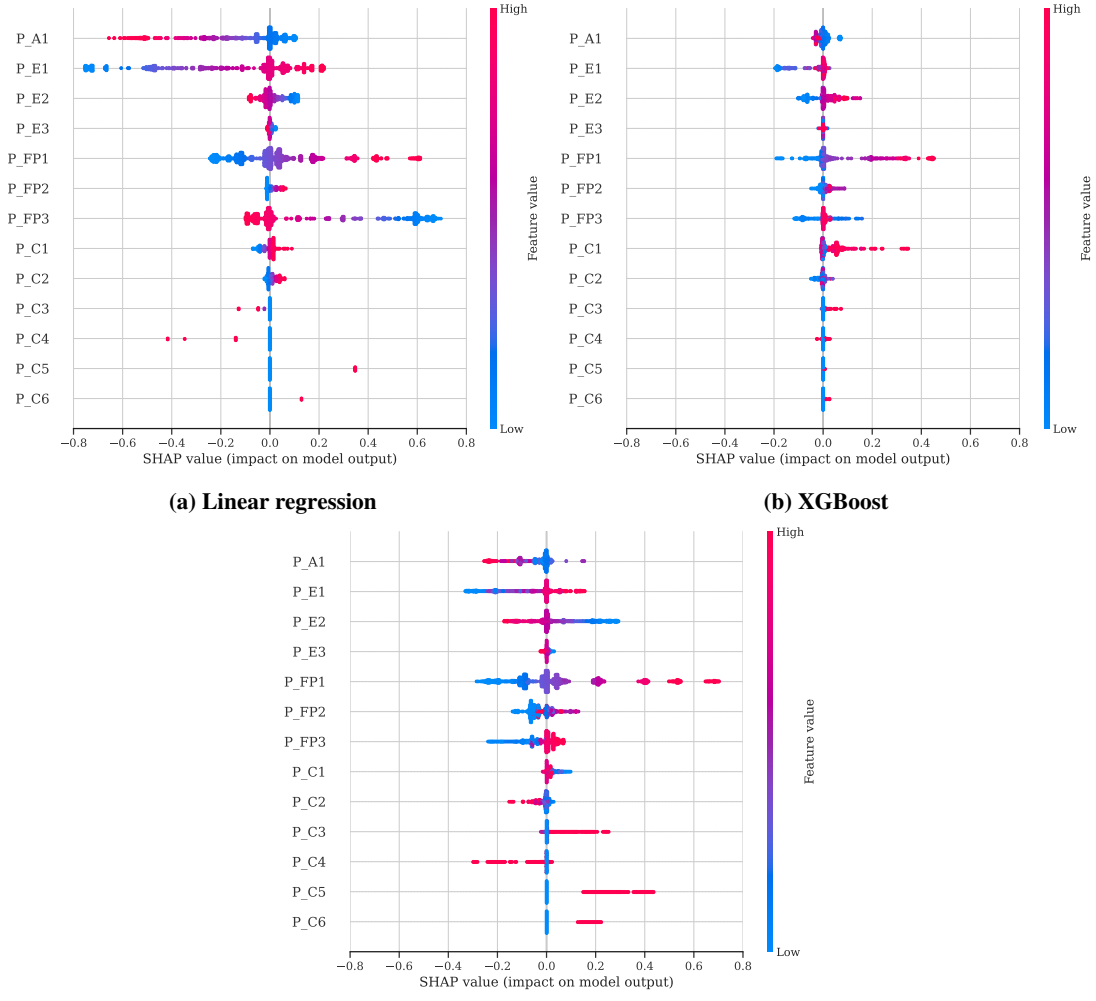
Fig. 14 Mean error for each accelerometer depending on the neural network, extrapolation evaluation.

These results show that the ANN-based ensemble model is best suited to extrapolate the energy along the dynamic pressure variable. This is consistent with results indicating that ANNs using a ReLU activation function are able to extrapolate linearly [34]. Extrapolation properties for the model may be desirable in order to enforce good prediction outside the range of the training set, which is typically non-convex.

C. SHAP values

While comparing two linear regression model can be a simple task, comparing widely different models – including black box ones such as ANNs – is much harder and necessitates a dedicated method. To do so, we use a model agnostic approach: SHAP (SHapley Additive exPlanations) [35]. It allows to efficiently quantify the importance of each feature for any given prediction.

The method is based on Shapley values [36], which represent the mean marginal contribution of each feature value across all possible values in the feature space. As Shapley values are hard to compute, SHAP values were introduced. They are the Shapley values of a conditional expectation function of the model. We advise the interested reader to refer to the original SHAP paper [35] for more details.



(a) Linear regression **(b) XGBoost**

(c) ANN 1

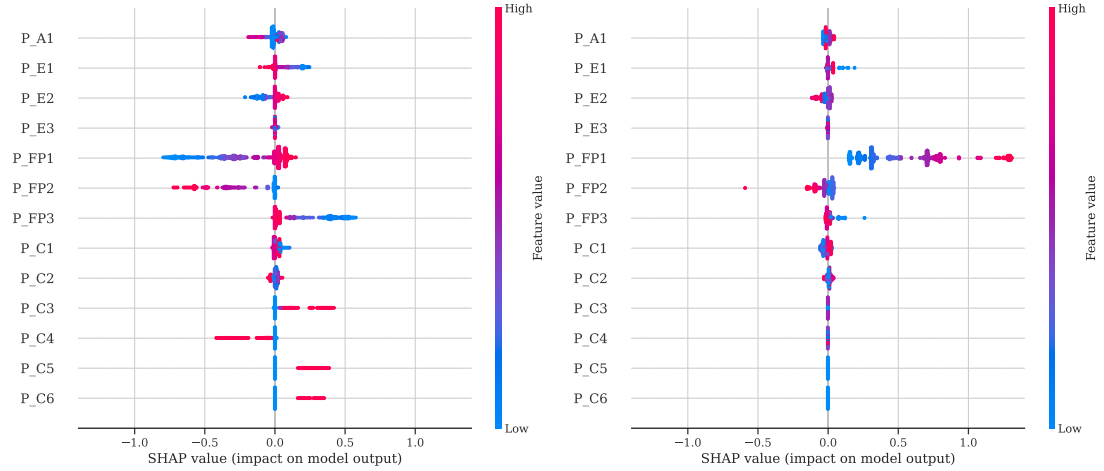
Fig. 15 SHAP values on the training set of the classic protocol, for different models. Accelerometer V3Z, frequency band B_8.

In our case, we use the implementation of the Kernel SHAP which has the advantage of being model agnostic.

For consistency, we study the frequency band B_8 and the sensor V3Z. Figure 15 shows the SHAP values of the entire training set of the classic evaluation protocol, for each model. For the ensemble, we consider the SHAP values of only one ANN, which is acceptable as the different ANNs have similar SHAP values.

SHAP values are computed relatively to a background datapoint, quantifying how much of a impact on the output a feature has. For this Figure, the background data consist of the median datapoint of the training set. It corresponds to an often encountered flight point with a clean configuration – when all hypersustentation surfaces and air brakes are retracted to minimize drag. We see that the dynamic pressure P_{FP_1} share similar importance between models. Engines related parameters P_{E_1} and P_{E_2} are important for each models, but at different levels. Configuration-related variables – with the exception of P_{C_1} – have surprisingly low importance for the XGBoost model.

It is interesting to compare the results of Figure 15 with the input-output correlation matrix given in Figure 5. First, it should be noted that while the sign of a correlation value relates to the sign of the relationship, SHAP values can be shifted as they are relative to the background datapoint. Considering only the B_8 column of the correlation matrix for comparison purposes, several differences can be observed. Multiple low-correlated variables can have significant SHAP values, such as P_{C_1} and other configuration-related variables. Conversely, strongly correlated variables can have small SHAP values, for instance if they are linearly related with a small proportionality coefficient. For the linear regression model, the features with the highest SHAP values are E1, FP3, FP1 and A1, which are also the most weighted features.



(a) Training set **(b) Test set**

Fig. 16 SHAP values on the training and test sets of the extrapolation protocol, for one ANN of the ensemble. Accelerometer V3Z, frequency band B_8.

Finally, Figure 16 shows the SHAP values of both the training and test set, but for the extrapolation protocol. Only one ANN is presented, and the band / accelerometer are the same as in Figure 15. Again, the background data is the median datapoint of the training set, for both the training and test set evaluation. It corresponds to a common flight point with a clean configuration. The differences of feature importance between the training and test set are significant.

While the dynamic pressure P_{FP_1} is almost the only important variable for the model in the test set, the SHAP values of the training set are much more consistent with what is computed with the classical protocol, which is an expected result. For brevity, the values for the linear regression and XGBoost models are not given for the extrapolation protocol. In the test set, linear regression predictions are highly impacted by P_{FP_1}, P_{FP_3} and P_{E_1}, while XGBoost has extremely low SHAP values for all variables.

V. Conclusion

We propose a general methodology to predict the vibration environment generated from a variety of unknown sources, at different locations of an aircraft, from general flight parameters. Time series acquired at a high sample rate during flight tests are fed through a preprocessing pipeline which generates a reduced representation of broadband energy spectra. This representation is based on discrete frequency bands, allowing for an easy manipulation and interpretation of the data.

A model is trained to predict the RMS levels of the frequency bands from a limited set of general flight parameters. The model architecture does not depend on the sensor, which makes it easy to train the model for a new accelerometer. To carry out this multi-output multivariate regression task, the model is built from an ensemble of artificial neural networks and its prediction are compared with that of two common baseline models: linear regression and XGBoost. Two different evaluation protocols are considered. The first protocol corresponds to standard evaluation and is based on a random split of the database into a training and a test sets. In this case, XGBoost and the ensemble model produce very good performances with average errors of less than 2%. In the second protocol, the models are trained and evaluated for extrapolation along the dynamic pressure dimension in order to assess their physical consistency and ability to generalize. It is found that linear regression generally does not capture the correct trend, while XGBoost is unable to extrapolate. The ANNs ensemble model provides acceptable predictions, with a correct trend prediction and a mean error of 10% across all accelerometers, compared to 15% for the linear regression and 20% for XGBoost. In addition to the good performances, the ensemble model provides a confidence interval by considering the output as a Gaussian distribution.

The ability to predict a spectrum without having to characterize the vibration sources and the excitation propagation allows for many applications such as anomaly detection – by generating synthetic spectra on the fly and comparing them to measurement –, and vibration flight envelope opening – by using extrapolation abilities.

However, it is also possible to use the model to study excitation propagation through the aircraft structure. A key issue for this is to provide a relevant representation of the structure and to characterize links between sensors. To this end, we have recently developed a method to learn a graph representation using graph signal processing methods [37].

Appendix A

Reducing PSD spectra into RMS levels per frequency provides results that are closer to the actual engineering needs,

increases robustness to noise, decreases training time, and allows for easier interpretability of the results.

However, the choice of frequency bands can induce a bias in the model: under the assumption of uncorrelated noise, the noise is expected to increase as the band width decreases, since it includes less points.

Here, a sensitivity analysis of models to the width of RMS bands is conducted. The 3 models are compared using the classical evaluation protocol, for 4 different discretization levels: 8, 16, 32 bands, and 3897 points –corresponding to maximal resolution and no reduction–, represented in Figure 17.

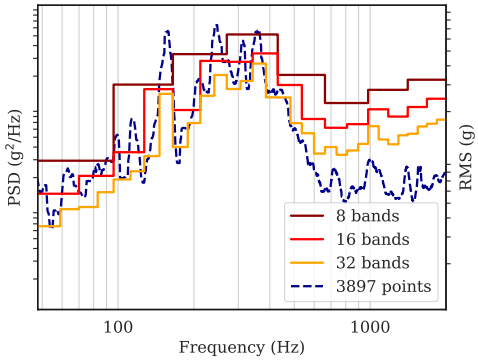


Fig. 17 Illustration on the same spectrum of the different representations compared in this study.

Results show that the error tends to grow with the number of outputs, see Table 3, which is not surprising since the complexity of the problem increases. However the differences can be deemed non-significant when comparing different RMS band level discretization of 8, 16 and 32 bands.

	Discretization level (# points)			
	8	16	32	3897
Linear Regression	2.5%	2.6%	2.7%	5.7%
XGBoost	0.2%	0.3%	0.3%	1.0%
ANNs ensemble	**0.8**%	**0.9**%	**1.1**%	**2.8**%

Table 3 Error on the classical evaluation protocol depending on spectra discretization level.

Appendix B

As a further point of comparison between models, results for second-order and third-order polynomial regression are presented. Compared to a linear regression, adding higher polynomial orders helps the model to better fit the data and lower the error for the classical evaluation protocol, see Table 4. However, the error is still higher than both ANNs and XGBoost.

For the extrapolation protocol, see Table 5, results are significantly worse than for all other models, indicating a lack

	Accelerometer					
	V1Z	V2Z	V3Z	V4Z	V5Z	Mean
Linear Regression	3.9%	5.1%	2.8%	2.6%	3.1%	3.5%
Second-order regression	3.0%	4.0%	2.1%	2.0%	2.3%	2.7%
Third-order regression	2.6%	3.3%	1.8%	1.5%	1.9%	2.2%
XGBoost	0.3%	0.4%	0.3%	0.3%	0.4%	0.3%
ANNs ensemble	**1.3**%	**1.4**%	**1.1**%	**0.9**%	**1.3**%	**1.2%**

Table 4 Mean error for each accelerometer depending on the model, considering polynomial regression, classical evaluation.

	Accelerometer					
	V1Z	V2Z	V3Z	V4Z	V5Z	Mean
Linear Regression	15.1%	16.2%	10.2%	16.9%	15.9%	14.9%
Second-order regression	71.6%	78.0%	31.7%	39.3%	45.4%	53.2%
Third-order regression	248.4%	362.0%	291.1%	196.1%	215.8%	262.7%
XGBoost	21.5%	22.3%	22.8%	22.9%	20.0%	21.9%
ANNs ensemble	**9.1**%	**12.8**%	**8.6**%	**10.2**%	**13.3**%	**10.8%**

Table 5 Mean error for each accelerometer depending on the model, considering polynomial regression, extrapolation evaluation.

of robustness.

Funding Sources

This work was supported by Dassault Aviation.

References

[1] Woods, A., "Design for Acceptable Aircraft Vibration," *Journal of Sound and Vibration*, Vol. 20, No. 3, 1972, pp. 353–358. https://doi.org/10.1016/0022-460X(72)90616-5.

[2] Himelblau, H., Kern, D. L., Manning, J. E., Piersol, A. G., and Rubin, S., *NASA Handbook 7005: Dynamics Environmental Criteria*, Vol. NASA-HDBK-7005, 2001.

[3] Zienkiewicz, O. C., *The Finite Element Method*, McGraw-Hill, 1977.

[4] Banerjee, P. K., and Banerjee, P. K., *The Boundary Element Methods in Engineering*, 2nd ed., McGraw-Hill, London ; New York, 1994.

[5] Deraemaeker, A., Babuska, I., and Bouillard, P., "Dispersion and Pollution of the FEM Solution for the Helmholtz Equation in One, Two and Three Dimensions," *International Journal for Numerical Methods in Engineering*, Vol. 46, No. 4, 1999, pp. 471–499. https://doi.org/10.1002/(SICI)1097-0207(19991010)46:4<471::AID-NME684>3.0.CO;2-6.

[6] Lyon, R. H., and Maidanik, G., "Power Flow between Linearly Coupled Oscillators," *The Journal of the Acoustical Society of America*, Vol. 34, No. 5, 1962, pp. 623–639. https://doi.org/10.1121/1.1918177.

[7] Langley, R., and Brown, A., "The Ensemble Statistics of the Energy of a Random System Subjected to Harmonic Excitation," *Journal of Sound and Vibration*, Vol. 275, No. 3-5, 2004, pp. 823–846. https://doi.org/10.1016/S0022-460X(03)00780-6.

[8] Cotoni, V., Langley, R., and Kidner, M., "Numerical and Experimental Validation of Variance Prediction in the Statistical Energy Analysis of Built-up Systems," *Journal of Sound and Vibration*, Vol. 288, No. 3, 2005, pp. 701–728. https://doi.org/10.1016/j.jsv.2005.07.012.

[9] van Hal, B., Desmet, W., and Vandepitte, D., "Hybrid Finite Element—Wave-Based Method for Steady-State Interior Structural–Acoustic Problems," *Computers & Structures*, Vol. 83, No. 2-3, 2005, pp. 167–180. https://doi.org/10.1016/j.compstruc.2004.04.019.

[10] Deckers, E., Van Genechten, B., Vandepitte, D., and Desmet, W., "Efficient Treatment of Stress Singularities in Poroelastic Wave Based Models Using Special Purpose Enrichment Functions," *Computers & Structures*, Vol. 89, No. 11-12, 2011, pp. 1117–1130. https://doi.org/10.1016/j.compstruc.2010.11.012.

[11] Cicirello, A., and Langley, R. S., "The Vibro-Acoustic Analysis of Built-up Systems Using a Hybrid Method with Parametric and Non-Parametric Uncertainties," *Journal of Sound and Vibration*, Vol. 332, No. 9, 2013, pp. 2165–2178. https://doi.org/10.1016/j.jsv.2012.05.040.

[12] Atak, O., Bergen, B., Huybrechs, D., Pluymers, B., and Desmet, W., "Coupling of Boundary Element and Wave Based Methods for the Efficient Solution of Complex Multiple Scattering Problems," *Journal of Computational Physics*, Vol. 258, 2014, pp. 165–184. https://doi.org/10.1016/j.jcp.2013.10.034.

[13] Atak, O., Pluymers, B., Desmet, W., and Katholieke Universiteit te Leuven (1970-), *MID-Frequency- CAE Methodologies for MID-Frequency Analysis in Vibration and Acoustics*, 2012.

[14] Clot, A., Meggitt, J., Langley, R., Elliott, A., and Moorhouse, A., "Development of a Hybrid FE-SEA-experimental Model," *Journal of Sound and Vibration*, Vol. 452, 2019, pp. 112–131. https://doi.org/10.1016/j.jsv.2019.03.027.

[15] Singh, I., Conrad, P., Chowdhury, P., Bakos, J. D., and Downey, A., "Real-Time Forecasting of Vibrations with Non-stationarities," *Data Science in Engineering, Volume 9*, edited by R. Madarshahian and F. Hemez, Springer International Publishing, Cham, 2022, pp. 21–29. https://doi.org/10.1007/978-3-030-76004-5_4.

[16] Najera, D., and Brink, A., "Efficient Random Vibration Analysis of Nonlinear Systems with Long Short-Term Memory Networks for Uncertainty Quantification," *ISMA 2018 International Conference on Noise and Vibration Engineering and USD2018 International Conference on Uncertainty in Structural Dynamics*, Vol. SAND2018-9830C, 2018.

[17] ElSaid, A., Wild, B., Higgins, J., and Desell, T., "Using LSTM Recurrent Neural Networks to Predict Excess Vibration Events in Aircraft Engines," *2016 IEEE 12th International Conference on E-Science (e-Science)*, IEEE, Baltimore, MD, USA, 2016, pp. 260–269. https://doi.org/10.1109/eScience.2016.7870907.

[18] Dostal, L., Grossert, H., Duecker, D. A., Grube, M., Kreuter, D. C., Sandmann, K., Zillmann, B., and Seifried, R., "Predictability of Vibration Loads From Experimental Data by Means of Reduced Vehicle Models and Machine Learning," *IEEE Access*, Vol. 8, 2020, pp. 177180–177194. https://doi.org/10.1109/ACCESS.2020.3027499.

[19] Wilmes, L., Olympio, R., de Payrebrune, K. M., and Schatz, M., "Structural Vibration Tests: Use of Artificial Neural Networks for Live Prediction of Structural Stress," *Applied Sciences*, Vol. 10, No. 23, 2020, p. 8542. https://doi.org/10.3390/app10238542.

[20] Tenney, A. S., Glauser, M. N., Ruscher, C. J., and Berger, Z. P., "Application of Artificial Neural Networks to Stochastic Estimation and Jet Noise Modeling," *AIAA Journal*, Vol. 58, No. 2, 2020, pp. 647–658. https://doi.org/10.2514/1.J058638.

[21] Arina, R., and Ferrero, A., "Data-Driven Aeroacoustic Modelling: Trailing-Edge Noise," *AIAA AVIATION 2021 FORUM*, 2021. https://doi.org/10.2514/6.2021-2237.

[22] Singh, T., "Artificial Neural Network Approach for Prediction and Control of Ground Vibrations in Mines," *Mining Technology*, Vol. 113, No. 4, 2004, pp. 251–256. https://doi.org/10.1179/037178404225006137.

[23] Bisoyi, S. K., and Pal, B. K., "Prediction of Ground Vibration Using Various Regression Analysis," *Journal of Mining Science*, Vol. 56, No. 3, 2020, pp. 378–387. https://doi.org/10.1134/S1062739120036665.

[24] Paneiro, G., Durão, F. O., Costa e Silva, M., and Falcão Neves, P., "Artificial Neural Network Model for Ground Vibration Amplitudes Prediction Due to Light Railway Traffic in Urban Areas," *Neural Computing and Applications*, Vol. 29, No. 11, 2018, pp. 1045–1057. https://doi.org/10.1007/s00521-016-2625-9.

[25] Chen, T., and Guestrin, C., "XGBoost: A Scalable Tree Boosting System," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794. https://doi.org/10.1145/2939672.2939785.

[26] Leoni, J., Zinnari, F., Villa, E., Tanelli, M., and Baldi, A., "Flight Regimes Recognition in Actual Operating Conditions: A Functional Data Analysis Approach," *Engineering Applications of Artificial Intelligence*, Vol. 114, 2022, p. 105016. https://doi.org/10.1016/j.engappai.2022.105016.

[27] Welch, P., "The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging over Short, Modified Periodograms," *IEEE Transactions on Audio and Electroacoustics*, Vol. 15, No. 2, 1967, pp. 70–73. https://doi.org/10.1109/TAU.1967.1161901.

[28] Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M., "Optuna: A Next-generation Hyperparameter Optimization Framework," *Proceedings of the 25th {ACM} {SIGKDD} International Conference on Knowledge Discovery and Data Mining*, 2019. https://doi.org/10.1145/3292500.3330701.

[29] Nguyen, H., Bui, X.-N., Bui, H.-B., and Cuong, D. T., "Developing an XGBoost Model to Predict Blast-Induced Peak Particle Velocity in an Open-Pit Mine: A Case Study," *Acta Geophysica*, Vol. 67, No. 2, 2019, pp. 477–490. https://doi.org/10.1007/s11600-019-00268-4.

[30] Shi, X., Wong, Y. D., Li, M. Z.-F., Palanisamy, C., and Chai, C., "A Feature Learning Approach Based on XGBoost for Driving Assessment and Risk Prediction," *Accident Analysis & Prevention*, Vol. 129, 2019, pp. 170–179. https://doi.org/10.1016/j.aap.2019.05.005.

[31] Lakshminarayanan, B., Pritzel, A., and Blundell, C., "Simple and Scalable Predictive Uncertainty Estimation Using Deep Ensembles," *Advances in Neural Information Processing Systems*, Vol. 30, edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Curran Associates, Inc., 2017. https://doi.org/10.48550/ARXIV.1612.01474.

[32] Fort, S., Hu, H., and Lakshminarayanan, B., "Deep Ensembles: A Loss Landscape Perspective," *arXiv:1912.02757 [cs, stat]*, 2020. https://doi.org/10.48550/arXiv.1912.02757.

[33] Malinin, A., Prokhorenkova, L., and Ustimenko, A., "Uncertainty in Gradient Boosting via Ensembles," 2021. https://doi.org/10.48550/arXiv.2006.10562.

[34] Xu, K., Zhang, M., Li, J., Du, S. S., Kawarabayashi, K.-i., and Jegelka, S., "How Neural Networks Extrapolate: From Feedforward to Graph Neural Networks," *International Conference on Learning Representations*, 2021. https://doi.org/10.48550/ARXIV.2009.11848.

[35] Lundberg, S., and Lee, S.-I., "A Unified Approach to Interpreting Model Predictions," *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017, pp. 4768–4777. https://doi.org/10.48550/ARXIV.1705.07874.

[36] Shapley, L. S., "Notes on the N-Person Game — II: The Value of an N-Person Game," Tech. rep., RAND Corporation, Aug. 1951.

[37] Février, S., Nachar, S., Mathelin, L., Giordano, F., and Podvin, B., "Apprentissage de Graphe Pour La Reconstruction de l'Environnement Vibratoire," *15ème Colloque National En Calcul Des Structures*, Université Polytechnique Hauts-de-France [UPHF], 83400 Hyères-les-Palmiers, France, 2022.