



HAL
open science

Prediction of Lightning Currents on Fastening Assemblies of an Aircraft Fuel Tank With Machine Learning Methods

Paul Monferran, Charles Guille-Escuret, Christophe Guiffaut, Alain Reineix

► **To cite this version:**

Paul Monferran, Charles Guille-Escuret, Christophe Guiffaut, Alain Reineix. Prediction of Lightning Currents on Fastening Assemblies of an Aircraft Fuel Tank With Machine Learning Methods. IEEE Transactions on Electromagnetic Compatibility, 2023, 65 (3), pp.812 - 822. 10.1109/temc.2023.3267280 . hal-04209349

HAL Id: hal-04209349

<https://unilim.hal.science/hal-04209349v1>

Submitted on 17 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Prediction of Lightning Currents on Fastening Assemblies of an Aircraft Fuel Tank With Machine Learning Methods

Paul Monferran , Charles Guille-Escuret, Christophe Guiffaut , *Member, IEEE*,
and Alain Reineix , *Member, IEEE*

Abstract—An important challenge for the aircraft industry consists to predict the currents on the fastening assemblies in order to avoid sparking, which can lead to accident, especially for fuel tank fasteners. In the literature, it has been demonstrated that the contact resistance plays a major role in the current path on fasteners. Nevertheless, these contact resistances cannot be well determined and vary greatly. As a result, the prediction of current must be done in a statistical way. Usually, it requires several aircraft simulations with several set of contact resistances, which represents a significant computational cost. This article proposes a machine learning model, which allows us to predict the currents in the fastening assemblies of an aircraft fuel tank in a few seconds. This model is built from a database of FDTD simulations of the aircraft fuel tank in the lightning frequency range 100 Hz to 1 MHz. The FDTD modeling is depicted in detail in this article based on previous work. From this database, several machine learning approaches are explored (*k*-nearest neighbors, support vector regression, XGBoost, and a neural network). As a result of this study, XGBoost presents the best performances. Further investigations using XGBoost highlights the ability of the model to predict well the current for most fasteners and frequencies, even with a small amount of simulations as training data. Moreover, the proposed model allows us to perform a parametric analysis, which underline the ability of the model to provide results in agreement with the physical effects of the issue (current paths, resistive effects, inductive effects, etc.). The results presented are promising for the use of the proposed methodology in the aeronautical industry.

Index Terms—Aircraft fuel tank, fastening assemblies, lightning currents, machine learning methods, parametric analysis.

I. INTRODUCTION

THE prediction of the lightning currents on fastening assemblies of an aircraft fuel tank is a huge challenge for the aeronautic industry. Indeed, an accurate fastening lightning current estimation on an aircraft fuel tank could be useful

in order to prevent aircraft accident. Undoubtedly, when the lightning current is flowing through fasteners a sparking can occur and can lead to some accident in particular in the fuel tank area [1], [2]. Nevertheless, only few fastener models are proposed in the literature [3], [4], [5], [6], [7], [8], [9]. The proposed models represent fastener as a circuit model where all the contact resistances are taken into account [3], [5], [8] or with a simplified circuit with an equivalent resistor [4], [6], [7], [8], [9]. The shared issue of these models is that the main parameter cannot be accurately determined. Hence, the models found in the literature usually use nominal values around few milliohms for these contact resistances, which lead to an inaccurate estimation of the current flowing through fasteners. The variability of the resistive parameter has been proven to be high [5], [7], [10]. In particular, in [7], the results of the proposed model defined from a measurement database highlights that the variation of the contact resistance can lead to several tens of decibels variation of the fastening lightning currents. Thus, a statistical approach has been chosen in order to solve this issue in previous works [7], [8], [9]. Merging a well-chosen uncertainties model with an FDTD fastener model, lightning current measurement are surrounded by simulations [9]. This approach presents good results in terms of modeling. Nevertheless, this kind of approach has some drawbacks. In particular, it requires an important number of simulations in order to propose an accurate surrounding. Hence, the process is a probabilistic estimation, which can have an important cost in terms of computing resources. Furthermore, in practice, the aeronautic authorities mainly deal with the worst case and most likely occurring case. These cases are difficult to identify due to the huge variability of the contact resistance values and the complex relationships between the different elements. Moreover, in such a scenario, the FDTD approach is not appropriate to do a parametric study, such as evaluating the effect of the fastener mapping or the materials conductivity. In this work, we consider machine learning methods to tackle these challenges. Indeed, machine learning approaches may provide a fast and computationally efficient way to predict the lightning currents while also facilitating parametric studies through tools depending of the chosen model. To the best of authors' knowledge, such methods have never been applied to this setting, thus, we propose a thorough preliminary study of the potential of well-known machine learning algorithms. We start by creating a dataset made of 2000 FDTD simulations,

Manuscript received 29 October 2022; revised 19 January 2023 and 24 March 2023; accepted 1 April 2023. This work was supported by the XLIM Laboratory. (Paul Monferran and Charles Guille-Escuret are co-first authors.) (Corresponding author: Paul Monferran.)

Paul Monferran, Christophe Guiffaut, and Alain Reineix are with the XLIM Laboratory, 87000 Limoges, France (e-mail: p.monferran75@gmail.com; christophe.guiffaut@xlim.fr; alain.reineix@xlim.fr).

Charles Guille-Escuret is with the Montreal Institute of Learning Algorithms, Montreal, QC H3T 1N8, Canada (e-mail: cge705@gmail.com).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TEMC.2023.3267280>.

Digital Object Identifier 10.1109/TEMC.2023.3267280

TABLE I
FUEL TANK ELEMENTS CONDUCTIVITIES

	Ribs and wedges	Spars (σ_1)	Panels (σ_2)	Panels + ECF195 (σ_3)	Panels + ECF815 (σ_4)
Conductivity ($S \cdot m^{-1}$)	$2.400 \cdot 10^7$	$5.500 \cdot 10^3$	$1.040 \cdot 10^5$	$3.647 \cdot 10^5$	$1.194 \cdot 10^6$

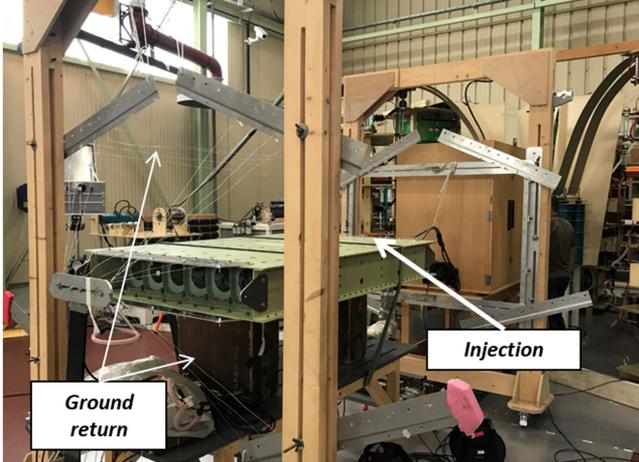


Fig. 1. Photography of the fuel tank under test in an attachment mode [9].

described in further details in Section II-C. We then compare the performances of a variety of standard machine learning models for regression, and discuss their respective advantages in Section III-A. Finally, we propose a detailed analysis of the failure modes and behavior of the most promising method, and we discuss its practical potential in predicting lightning currents on fastening assemblies in Section III-B.

II. PROBLEM PRESENTATION

A. Fuel Tank Overview

The studied object, presented in Fig. 1, is a fuel tank designed for lightning certification by Dassault Aviation. This fuel tank has been yet studied and has been depicted in [9]. Hence, only the useful details for this article are presented. This fuel tank of $174 \times 117 \times 21 \text{ cm}^3$ dimensions is built with several elements as ribs, panels, wedges, spars, fasteners, etc. This fuel tank is composed with five different materials: a composite material for the panels, another composite material for the spars, 6062 aluminium alloy for the wedges and ribs, a $815 \text{ g} \cdot \text{m}^{-2}$ copper mesh (ECF815), and a $195 \text{ g} \cdot \text{m}^{-2}$ copper mesh (ECF195) for the extended copper foils (ECF). The mapping of these elements is depicted in Fig. 2. Table I sums up the standard conductivities associated to each fuel tank element. Moreover, the fuel tank is constituted with 260 fasteners. All the fasteners used to screw together the elements are identical. At each interface (panel/rib, spar/panel, etc.), the current can flow only through the fasteners. Furthermore, the lightning injection is applied with an attachment mode, which means that the current is directly injected on a specific fastener.

As mentioned in Section I, the aim is to predict the simulated lightning currents. Therefore, a modeling of the fuel tank is

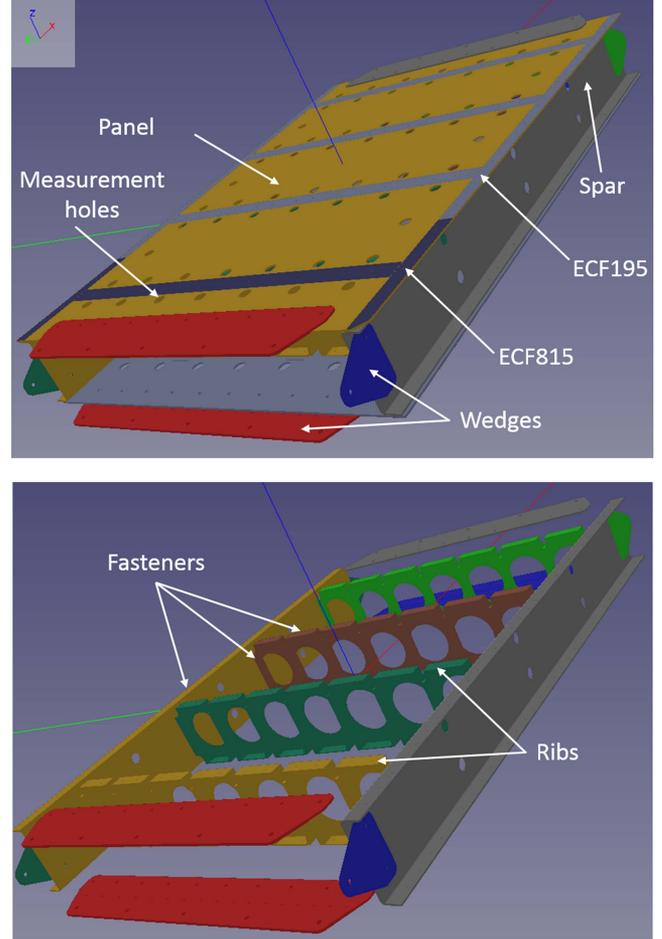


Fig. 2. Fuel tank CAD [9]. At top, full CAD. At bottom, CAD without panels.

proposed. The FDTD method has been chosen as simulation method because of its reliability in this context [8], [9], [11], [12]. The FDTD mesh of the fuel tank is shown in Fig. 3. As we can see in this figure, the fuel tank and the test bench equipment are represented. Furthermore, we choose the same simulation parameters as [9] because they have proven their relevance for this fuel tank modelling. Thus, the computational volume without the perfectly matched layers (PML) is $185 \times 152 \times 122$ cells with a 1.5 cm cubic step mesh, the PML [13], [14], [15] are used as boundary conditions, the plates are represented with the low frequency thin-plate-model [16], [17], [18]. The ground wires radius is 4.5 mm and the wires resistance per unit length is fixed to $0.26 \text{ m}\Omega/\text{m}$ while the lumped resistance between the ground wires and the plates is fixed to $5 \text{ m}\Omega$. In addition, the same fastener model used in [9], which is a wire with a lumped resistance inside a low-conductive plate, is applied. Moreover, for the fastener that receives the lightning injection, we use

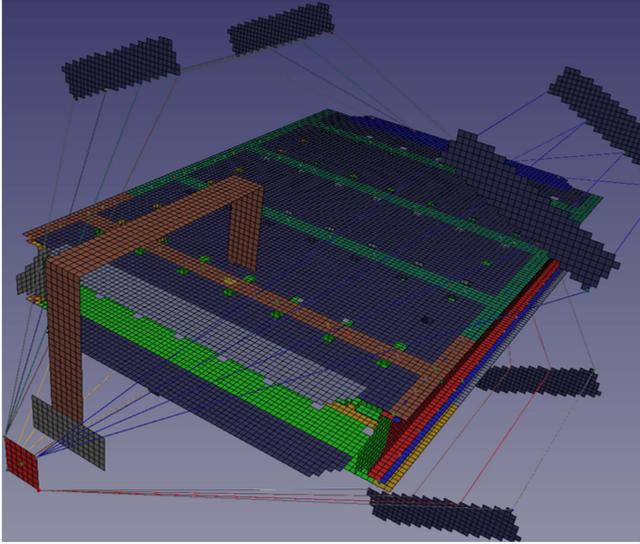


Fig. 3. Fuel tank FDTD mesh in an attachment mode [9].

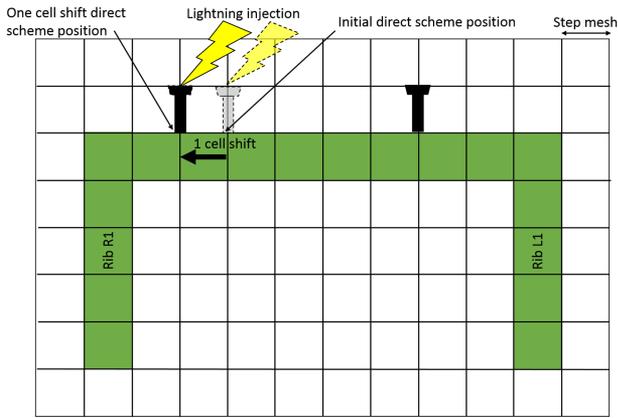


Fig. 4. Illustration of the direct scheme with one cell shift.

the most efficient fastener model called direct scheme with one cell shift in [9], which represent better the real position of this fastener. Indeed, this model has been demonstrated as the most relevant for the lightning current prediction in comparison with measurement. With this model, all the injected current flows through the fastener that receives this injection and the location of this fastener is one FDTD cell shift, as shown in Fig. 4. Finally, the FDTD simulation is performed with the homemade code TEMSI-FD [19] using a Gaussian waveform as source for 1000 frequencies between 100 Hz and 1 MHz, which is an appropriate frequency range for lightning issue. The Gaussian waveform is used rather than an A-type waveform in order to reduce long time simulation. Furthermore, we note that the simulated currents in the frequency domain are rescaled in the range [0,1]. Indeed, TEMSI-FD provides the normalized responses $S_{\text{normalized}}^{\text{simulation}}(f)$ in the spectral domain computed as follows:

$$S_{\text{normalized}}^{\text{simulation}}(f) = \frac{S_{\text{solution}}^{\text{simulation}}(f)}{S_{\text{source}}^{\text{simulation}}(f)} \quad (1)$$

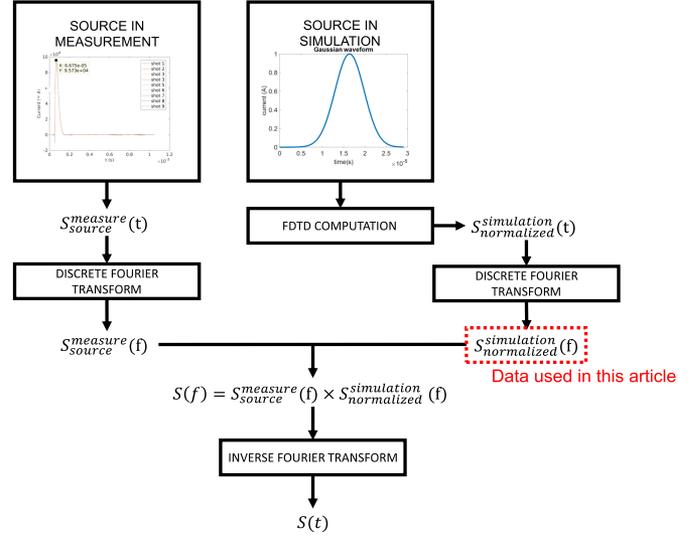


Fig. 5. Flowchart of the post-treatment used in [9].

where $S_{\text{solution}}^{\text{simulation}}(f)$ is the initial solution in the frequency domain and $S_{\text{source}}^{\text{simulation}}(f)$ is the Fourier transform of the source time waveform. Thus, the normalized solution $S_{\text{normalized}}^{\text{simulation}}(f)$ corresponds to the calculation provided by a frequency solver that would have applied a source of unity amplitude over the entire frequency band. In the following, the normalized data $S_{\text{normalized}}^{\text{simulation}}(f)$ are considered as theoretical results are not compared with measurements. However, if we want to get back to realistic lightning data (with specific lightning waveforms), the same post-treatment process used in [9] can be performed. Indeed, we can apply a Fourier transform on a realistic injected current, which is usually measured in the time domain. Hence, the resulting signal $S_{\text{source}}^{\text{measure}}(f)$ in the frequency domain is multiplied by $S_{\text{normalized}}^{\text{simulation}}(f)$ to derive

$$S(f) = S_{\text{source}}^{\text{measure}}(f) \times S_{\text{normalized}}^{\text{simulation}}(f) \quad (2)$$

with $S(f)$ the simulated current signal with the measured lightning source in the frequency domain. Finally, the time domain signal $S(t)$ can be obtained by applying an inverse Fourier transform of $S(f)$. This allows a direct comparison between numerical results and measurements in the time domain. This procedure applied in [9] is illustrated in Fig. 5.

B. Machine Learning Interest

As mentioned in Section I, previous works [7], [8], [9] proposed a stochastic lightning current estimation surrounding measurement with simulations. This approach is interesting when the number of different parameters (fastener resistance, fastener location, material conductivity, etc.) is not important, otherwise the number of simulations must be huge in order to obtain an accurate prediction. Therefore, machine learning methods are more appropriate to solve a parametric issue. In particular, the aim of this article is to propose a straightforward machine learning model, which can predict accurately the lightning fastener currents, which are the outputs, from several inputs

depicted in the following section. Obviously, such a model main benefit over a simulation is to give a fast response regardless of the inputs. Hence, the model response should be much faster than a simulation and less expensive in terms of computing resources. Thus, as long as we are confident in the simulation modeling, such a model could be really useful for the lightning fastener currents analysis, as for instance, to determine the most important parameter for the lightning currents distribution, to evaluate current paths, to detect local effects, etc. Nevertheless, in order to obtain this kind of model, a simulation database should be performed. The following section depicts the proposed database.

C. Database Description

The database is constituted with 2000 simulations with 264 variables, which are the 260 fastener resistances and 4 material conductivities. Thus, for each simulation, each lumped resistance takes a stochastic value drawn from the log-normal law used in [9] because this law has been proven as appropriate for this modeling. Moreover, except for the ribs and wedges, the elements conductivities take a random value drawn from a uniform law in the range $\pm 10^2 \text{ S}\cdot\text{m}^{-1}$ from the nominal values of Table I. Indeed, we consider that the 6062 aluminium of the ribs and wedges is much more stable than the composite elements conductivities. For each simulation, we would like to predict each fastener current module at each frequencies, i.e., 260×1000 outputs (number of fastener, number of frequency). We split the 2000 simulations into 1600 training and validation simulations, and 400 test simulations that are reserved only for the final performance evaluation.

D. Machine Learning Approach

From a machine learning perspective, this problem has the following two particularly significant properties.

- 1) The extremely low-data regime, with a database comprised of only 1600 training simulations despite relatively high-dimensional inputs (264 dimensions).
- 2) The unusually large number of output dimensions: 260 000.

While such high-dimensional output may be challenging from a computational perspective, we show that it remains reasonably feasible. Moreover, for practical purposes it may be sufficient to predict only a subset of the output, for instance, by subsampling the frequencies.

We explore the performances of the following four well-known machine learning models.

1) *K-Nearest Neighbors*: k -nearest neighbors [20] is not particularly suited to this problem due to its high sensitivity to the input dimension and dataset size. However, we use it as a simple baseline and indicator of the difficulty of the problem. Through cross-validation, we find $k = 3$ to be the best tuning.

2) *Support-Vector Regression (SVR)*: Support-vector machines [21] are some of the most robust supervised classification algorithms. Although they are linear models, the kernel allows us to use them for nonlinear classification. We propose to use their adaptation to regression problems, i.e., SVR [22]. Due to the sensibility of algorithms to scale, we normalize both the

training inputs and outputs. At test time, we apply the same transformation to test inputs, and apply the inverse transformation to the model output, to recover the original scale. Through cross-validation, we find that the radial basis function kernel is the most suited among usual kernels, and find $\epsilon = 0.1$ and the regularization parameter $C = 1.0$ to be optimal. Due to SVR being naturally a single-dimension predictor, we independently train a separate model for each desired output dimension.

3) *Xgboost*: XGBoost [23] is a decision tree ensemble learning library that can be leveraged for either classification or regression. It uses gradient boosting to learn an ensemble of decision trees. XGBoost became well-known for winning a large number of machine learning competitions, and has been particularly dominant on tabular and structured data. In our setting, the input is tabular and corresponds to independent parameters with interpretable physical meaning, which makes XGBoost particularly suited and likely to perform competitively. Additionally, while tree forests are more difficult to interpret than single trees, recent tools have been developed to analyze the impact of input parameters with boosted tree models, such as SHapley Additive exPlanations (SHAP) [24]. Similarly to SVR, XGBoost is not naturally suited for multidimensional prediction, so we train an independent model for each output dimension.

4) *Neural Networks*: In recent years, deep learning has achieved spectacular breakthroughs in a wide range of applications [25], [26], [27], [28], in particular with high-dimensional unstructured data, such as images and text. Despite these successes, neural networks have generally remained behind boosted tree approaches for structured tabular data [29], and thus, may not be well-suited for our problem. However, they present the significant advantage of naturally performing multidimensional regression with negligible additional compute, and are capable of sharing learnt features across outputs. Similarly to SVR, we normalize training inputs and outputs. At test time, we apply the same transformation to input and the inverse transformation to the model output in order to recover the original scale. Using 400 of the training samples as validation, we tune the architecture of a fully connected multi-layer perceptron (MLP) with batch normalization [30], ReLU activations [31], and optimized with stochastic gradient descent [32] with momentum (SGD-m). The architecture search spans from 2 to 8 hidden layers, 1.2×10^5 to 5×10^6 hidden layer parameters and 128 to 1024 last layer parameters per output dimension. Unsurprisingly given the small number of training samples, we find that a two-layer architecture with 2×10^5 hidden layer parameters works best. We also tune the learning rate $\alpha = 5 \times 10^{-3}$ and momentum $\beta = 0.9$. We optimize over 100 epochs with early stopping on the validation set.

III. RESULTS

A. Methods Performances Comparison

In order to get interpretable results, we report the following two metrics of performances for output j

- 1) the relative error:

$$\text{RE}_j = \frac{\sum_{i \leq n} |\hat{y}_i^{(j)} - y_i^{(j)}|}{\sum_{i \leq n} |y_i^{(j)}|} \quad (3)$$

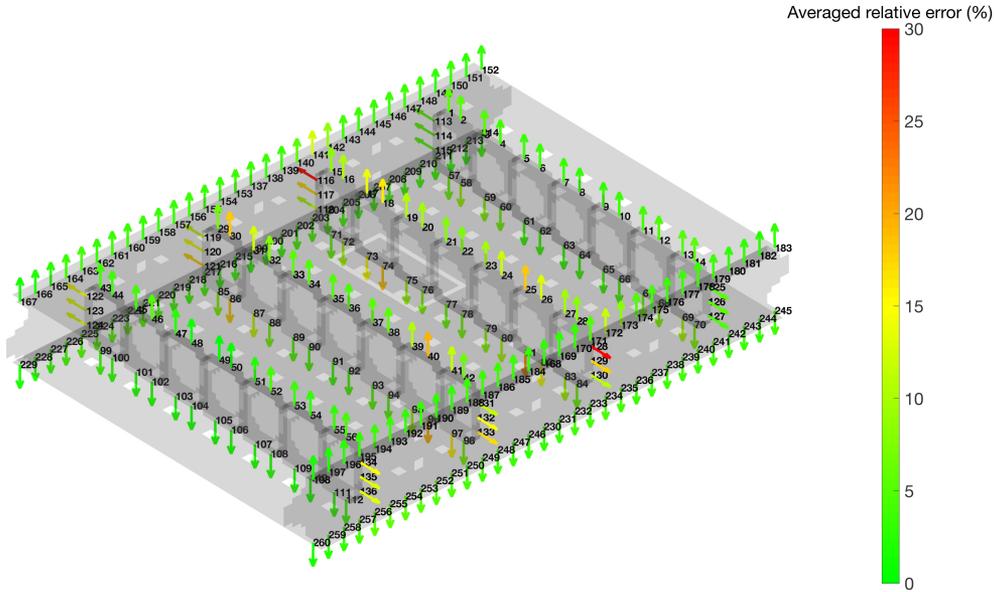


Fig. 6. Relative error averaged over frequencies for each fastener.

TABLE II
AVERAGE COEFFICIENT OF DETERMINATION AND RELATIVE ERROR

	R^2	RE
k-nn	-0.382	34.1%
SVR	0.524	19.0%
XGBoost	0.948	5.9%
MLP	0.224	25.2%

where $\hat{y}_i^{(j)}$ and $y_i^{(j)}$ are, respectively, the predicted value and the ground truth of output j of the i th sample where n is the total number of experiments. In the following, the currents computed by the FDTD simulations are taken as the ground truth.

2) The coefficient of determination

$$R_j^2 = 1 - \frac{\sum_i (y_i^{(j)} - \hat{y}_i^{(j)})^2}{\sum_i (y_i^{(j)} - \bar{y}^{(j)})^2} \quad (4)$$

with $\bar{y}^{(j)}$ the mean value of output j across samples and the same notations.

The relative error gives an indication of how close predictions are to target values and, thus, of the practical potential of a model. Perfect prediction corresponds to a relative error of 0. In contrast, the coefficient of determination is a measure of the proportion of the target variance that is captured by the model. In particular, perfect prediction corresponds to $R^2 = 1$, while consistently predicting the mean value \bar{y} corresponds to $R^2 = 0$. It is possible for a model to have negative R^2 , which means we are not even predicting \bar{y} . These two metrics are complementary as the relative error gives an indication of prediction accuracy across different scales, while the coefficient of determination gives an indication of how much information the model is able to extract from the input. For each considered model, the R^2

and the RE averaged over all output dimensions (i.e., over all frequencies and fasteners) are reported in Table II.

Unsurprisingly, XGBoost achieves significantly better performances on our task of interest. This is consistent with the behavior observed on other tasks with structured tabular data. We now propose to thoroughly analyze the performances of XGBoost across fasteners, frequencies, and experiments.

B. XGBoost Performance Analysis

To preserve brevity and clarity, we will now focus solely on relative error as a performance metric.

In order to better understand XGBoost performances, we illustrate in Fig. 6, a 3-D representation of the fuel tank with the relative error for each fastener, averaged over frequencies. In order to compare Fig. 6 to Fig. 3, both figures have been drawn in the same axis, which means that the lightning injection is performed in the fastener 49. Furthermore, for the following figures in this article, we do not present the results for the fastener 49, which is not sensitive to parameter variation due to the modeling. First, we note that the quasisymmetry of the physical issue obviously occurs in the results. Furthermore, Fig. 6 shows that the prediction of the currents close to the injection (fasteners 43 to 56 and 99 to 112) or close to the ground return (fasteners 1 to 14 and 57 to 70) are very well predicted with a relative error less than 5%. This is also usually the case for the currents of the fasteners between the panels and the spars while the fasteners between the ribs and the spars are generally the hardest to predict. It could be due to the current waveforms (usually with several resonances), which are more difficult to learn on these fasteners because of their particular positions. In addition, Fig. 7 shows the relative error as a function of fastener number and frequency, and as a function of either one when averaged over the other. This figure highlights the ability of the model to predict well the currents for most fasteners except for the very low frequencies.

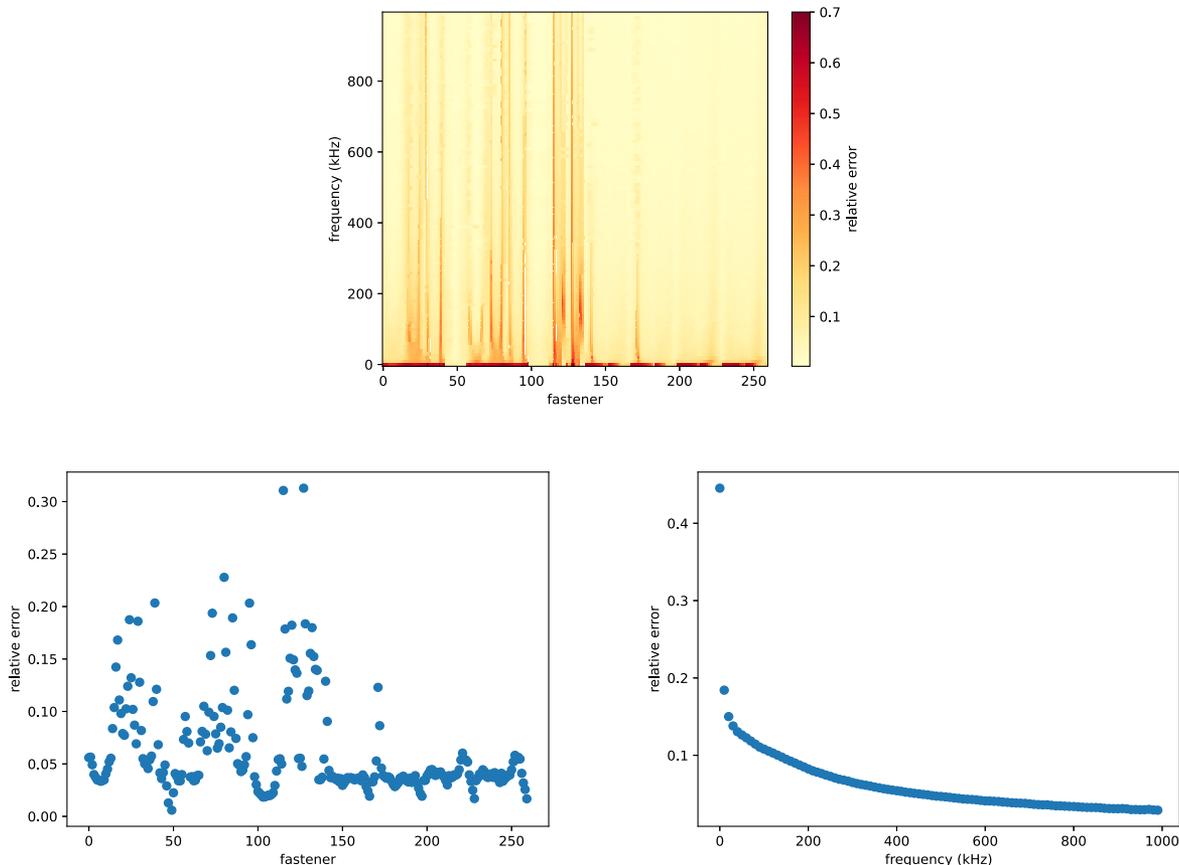


Fig. 7. Relative error as a function of fastener number and frequency (top), as a function of fastener averaged over frequencies (bottom left), as a function of frequency averaged over fasteners (bottom right).

We suppose that is due to the complex relationship between the input parameters and the output at low frequency. Indeed, at very low frequency, the current paths are mainly led by the resistive effect. As a result, many input parameters have significant impact on the output. On the other hand, at higher frequency the current is mainly driven by the inductive effects, which concentrates the impact on fewer input parameters. This assumption based on the physics is explored in the following of this article (see Section III-C). XGBoost is capable of focusing most of its model capacity toward highly impactful input dimensions and to ignore irrelevant ones, resulting in a problem of lower effective dimensionality at higher frequency, and thus, in better predictions, especially in such a low data setting. Conversely, it is very hard for the model to generalize well the behavior at very low frequencies with our few number of training data. Increasing the amount of training data could greatly improve accuracy especially for the very low frequencies.

To go into further details, we also focus on the following three specific frequencies: 100 Hz, 500 kHz, and 990 kHz, respectively, representing the lowest frequency, the middle of the frequency range, and one of the highest frequencies. Similarly, we focus on three specific fasteners, corresponding to positions 75, 130, and 160. These fasteners were chosen for the variety of their positions within the fuel tank. Indeed, as shown in Fig. 8,

the fastener 75 represents the link between panel and rib while fastener 130 links rib and spar, and, fastener 160 screws together panel and spar. Fig. 9 shows the relative error over positions for each of these three frequencies, and the relative error over frequency for each of these three fasteners. For the 3 fasteners taken into account, we note the same behavior with an accuracy in terms of prediction increasing rapidly with frequency. This behavior is independent of the chosen fastener as illustrated in the bottom plots of Fig. 9. Indeed, this figure highlights the difficulty to predict the current at 100 Hz with a relative error of less than 10% except for the fasteners close to the injection. However, this accuracy is achieved increasing the frequency for most fasteners. Overall, we obtain relative errors of less than, 15% at 500 kHz, and 10% at 990 kHz, for 250 fasteners of 260 fasteners.

Finally, we directly illustrate the quality of our model's predictions. Fig. 10 shows the predicted current versus ground truth for each test experiment. Points closer to the line correspond to better predictions. Obviously as shown previously, it is particularly hard to predict at 100 Hz. Fig. 10 shows this is mostly due to the model failing to predict unusually large target values at this frequency. At higher frequencies, XGBoost performs very well and successfully approximates target current. In particular, we note that the prediction is accurate regardless of the current

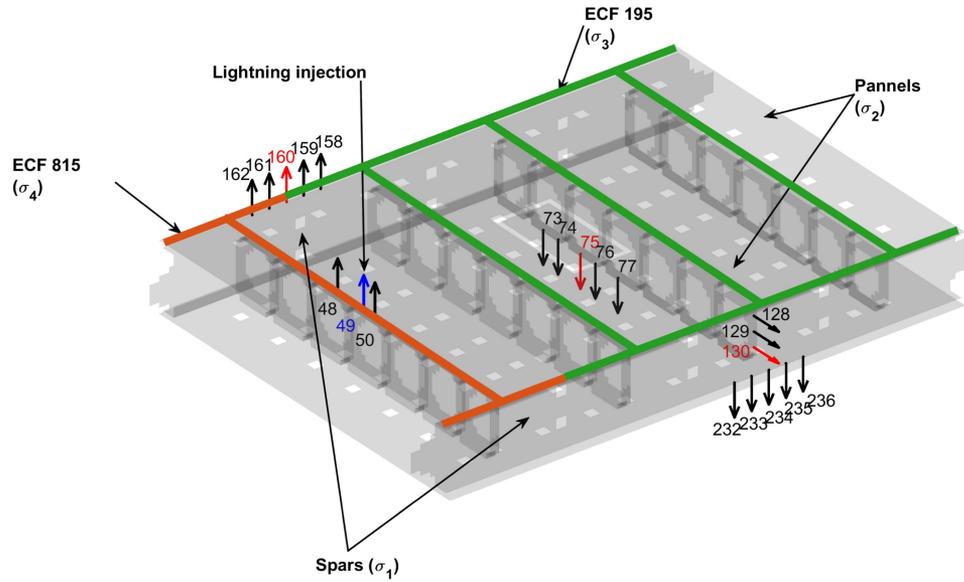


Fig. 8. Illustration of the main parameters around the observation points (in red).

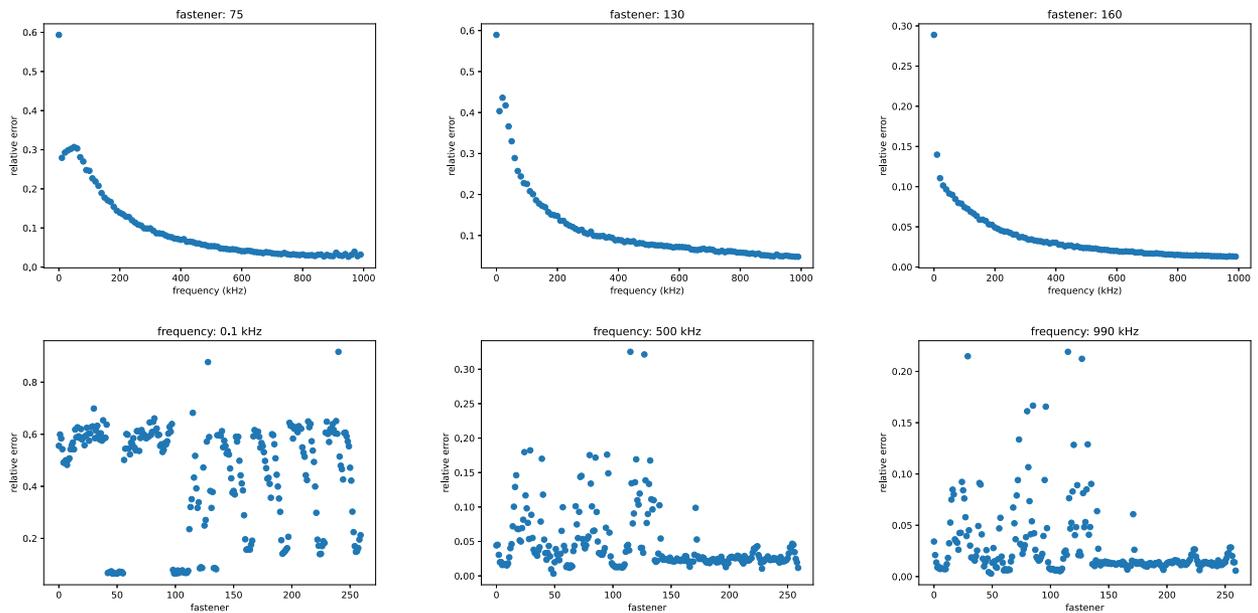


Fig. 9. Relative error as a function of frequency (top), for fasteners 75 (left), 130 (middle), and 160 (right) and as a function of position (bottom), for frequencies 100 Hz (left), 500 kHz (middle), 990 kHz (right).

amplitude at high frequencies, unlike very low frequencies. In addition, Fig. 11 shows the predicted current versus ground truth for a specific test sample chosen at random (here experience 179), either as a function of frequency for a specific fastener, or as a function of position for a specific frequency. For the three chosen fasteners (top plots), the predicted and simulated currents are in a good agreement except for the first frequency. Moreover, over all fasteners for the three selected frequencies (bottom plots), we obtain a very good agreement between the predicted and simulated currents. Furthermore, we also note that obviously the high currents are mainly located close to the injection as it has been still mentioned in [10].

C. Parameters Impact

The interpretable nature of regression trees allows us to analyze the impact of the input parameters on the model's predictions. Since forests are more difficult to directly interpret than single trees, we follow SHAP [24]. Fig. 12 shows the SHAP figures for fasteners 75, 130, and 160, at frequencies 100 Hz, 500 kHz, and 990 kHz. Each figure describes the impact of the 8 most impacting input parameters on the model predictions. For each feature F , each point corresponds to a single experiment. The color of the point describes whether the value of F is higher or lower than average in that specific experiment. Its position

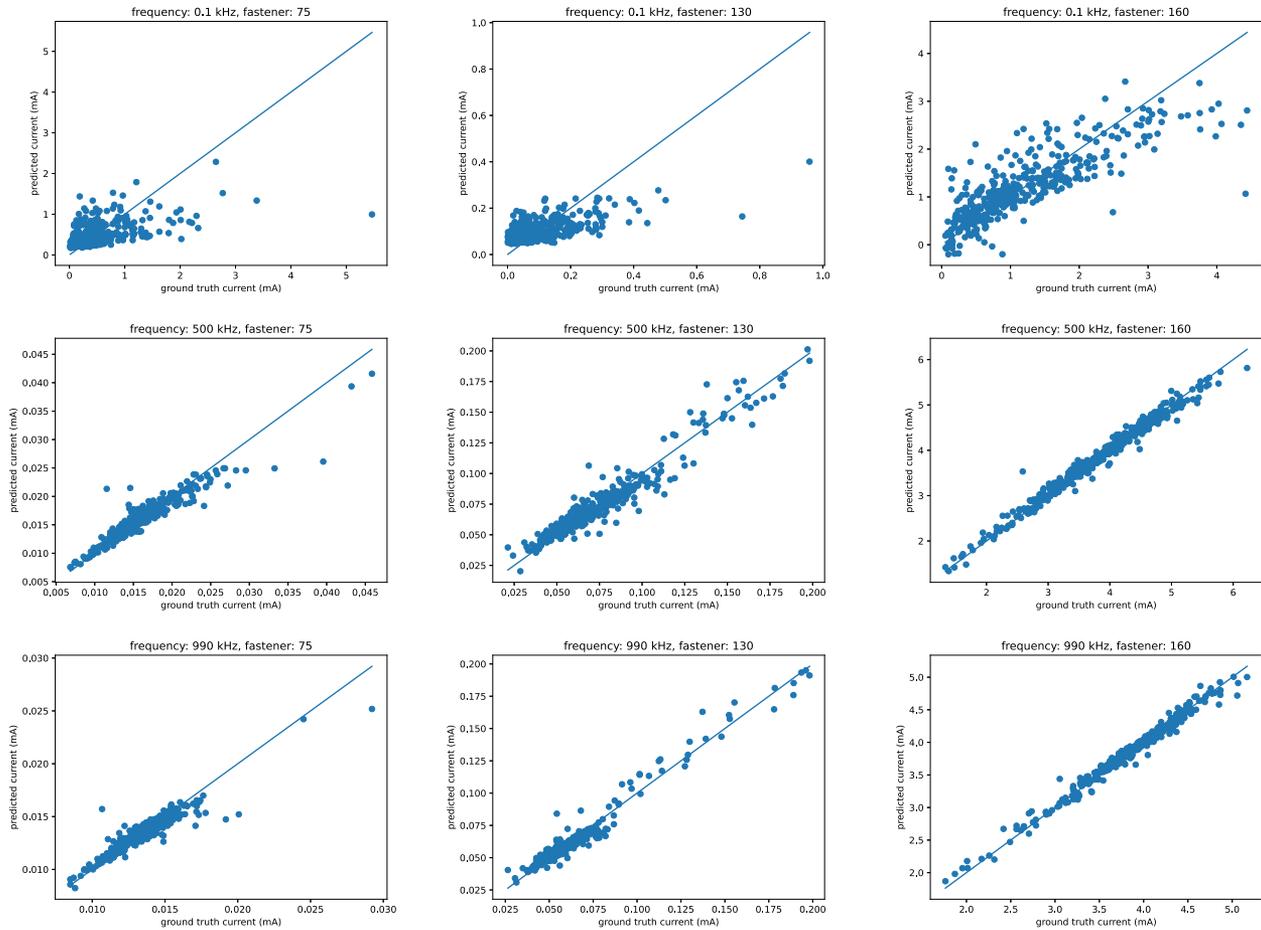


Fig. 10. Predicted values versus current ground truth for fasteners 75 (left), 130 (middle), 160 (right), and for frequencies 100 Hz (top), 500 kHz (middle), 990 kHz (bottom). The line corresponds to perfect predictions.

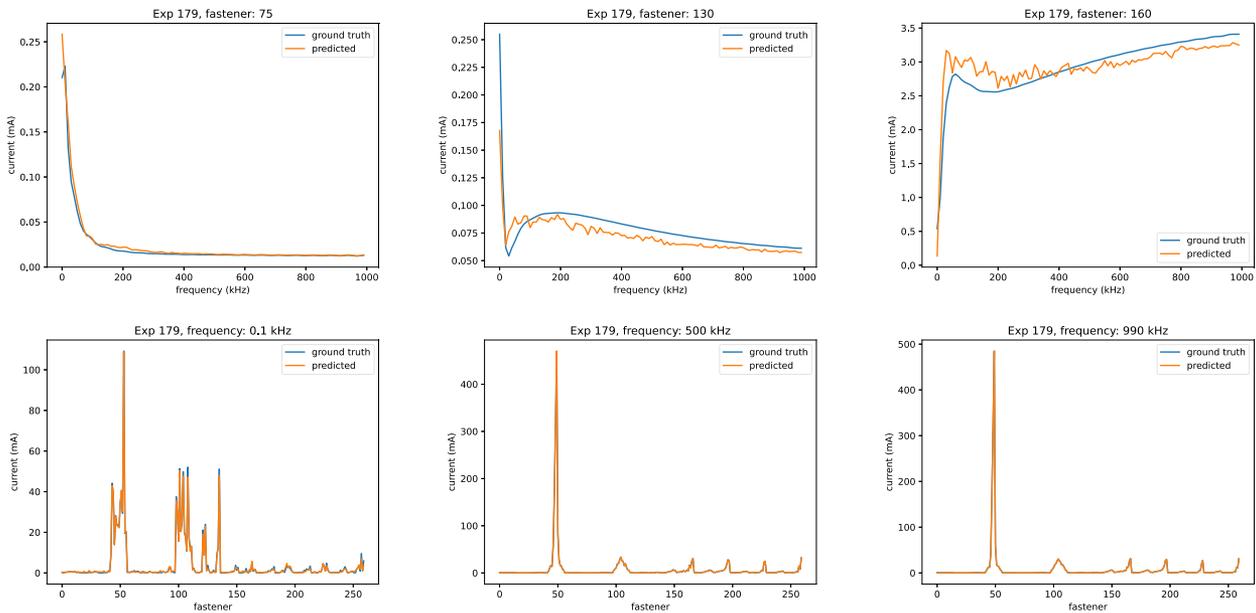


Fig. 11. Predicted value versus ground truth for test sample 179 as a function of frequency (top) for fastener 75 (left), 130 (middle), 160 (right), and as a function of position (bottom) for frequency 100 Hz (left), 500 kHz (middle), 990 kHz (right).

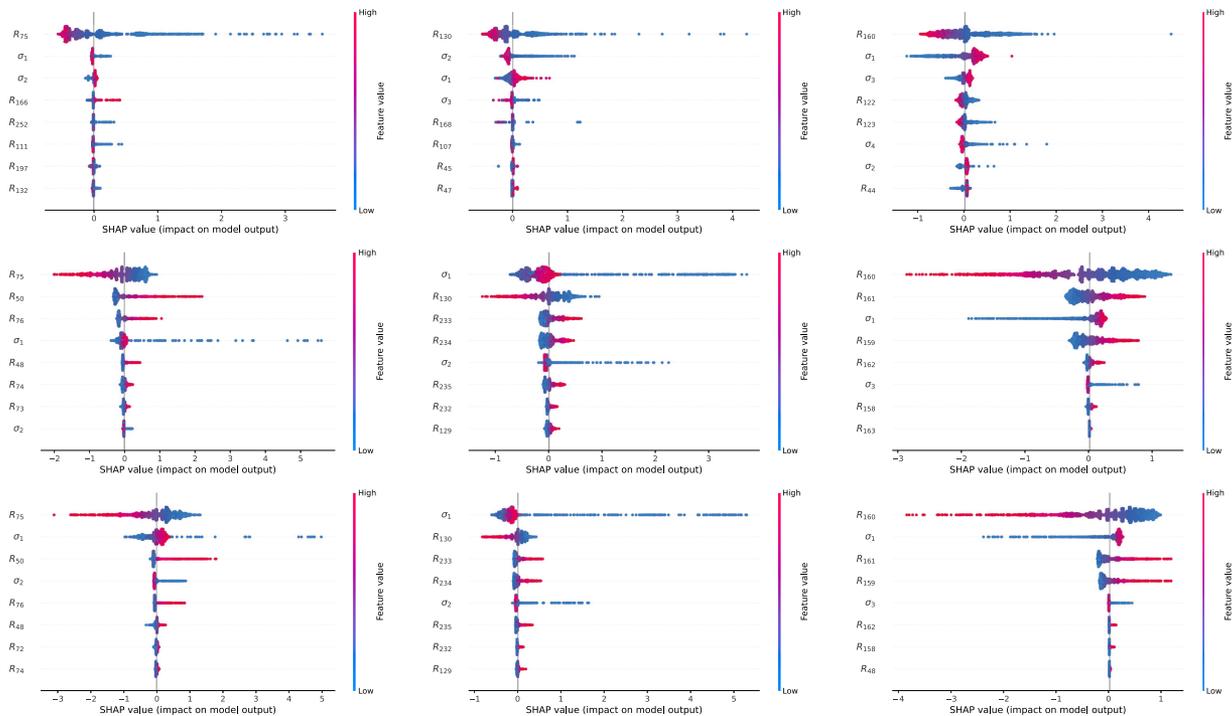


Fig. 12. SHAP plots for fasteners 75 (left), 130 (middle), 160 (right), and for frequencies 100 Hz (top), 500 kHz (middle), 990 kHz (bottom).

on the x -axis (called shapley value) describes how much the knowledge of that feature has impacted the model's output (for a rigorous description of the computation, which averages the marginal contributions of \mathbf{F} over all coalitions of features containing it, we refer to [24]). Intuitively, the more a point is on the right (resp. left) of the x -axis, the more the knowledge of that feature has increased (resp. decreased) the model's prediction.

For instance, we observe that unsurprisingly, the resistance of a fastener is strongly negatively correlated with the current measured in this fastener: low resistance (blue points) increases the model prediction (high x -axis value) while high resistance (red points) decreases the model prediction (low x -axis value). From a physical standpoint, it means that the lower the fastener resistance, the higher the current, and vice versa.

Fig. 12 highlights that the primary fastener parameter (its own resistance) always plays a major role in the current distribution in the fastener. Indeed, it is the most impacting parameter for the fasteners 75 and 160 whatever the frequency. While for the fastener 130, we note that the resistance is the most impacting parameter at low frequency whereas at higher frequency, it is the second impacting parameter after the conductivity σ_1 , which corresponds to the conductivity of the spars as shown in Tables I. In addition, it can be seen that the impact of σ_1 is more important for the lower values of σ_1 . We explain these results by the location of the fastener 130 and several physical phenomena (resistive effect, inductive effect, and skin effect). Indeed, the fastener 130 ensures the current flow between a rib and a spar, so this fastener is directly dependent on the conductivity of the spar. Moreover, at high frequencies, the inductive effect increases, the current lines repel each other and the current spreads to the ends of the conductor. Inductive effects of the conducting surfaces

bring back dominant reactance on the distributed impedance of these ones at high frequencies. This inductive reactance become widely superior to surface resistivity and also skin effect. Then, it has the strongest influence on the distributed current and above all attenuate the effect of local fastener contact resistances [9]. Since the fastener 130 is located on the edge of the fuel tank and the inductive effect is correlated to σ_1 , σ_1 has a major impact at high frequencies.

Furthermore, Fig. 12 shows that, overall, the conductivity of the materials has a significant impact on the current paths even at low frequency. The associated conductivities taken into account for each fastener are of course related to the considered fastener location. For instance, for the fastener 160 located on the edge of the fuel tank in the copper mesh ECF815 with σ_4 as conductivity, we can interpret the results at 100 Hz as follow: if the panel conductivity (σ_2) and the copper mesh ECF195 conductivity (σ_3) are high while the copper mesh ECF815 conductivity (σ_4) is low, then the current in fastener 160 will be higher. This interpretation is intuitive from a physical point of view. Moreover, Fig. 12 highlights also some nonmonotonous effects due to the complexity of the issue. For instance, regarding the impact of σ_1 on the fastener 130 at 500 MHz, we note that, depending on the set of parameters, a low σ_1 (blue points) can lead to decrease the current (negative values) as well as increase the current (positive values) in the fastener 130. This result is very promising about the accuracy of our model.

In addition, Fig. 12 illustrates that another influential parameter is the resistance of the fasteners adjacent to the one concerned. Obviously, the resistance of an adjacent fastener is strongly positively correlated with the current measured in the considered fastener. It means that the higher the adjacent fastener

resistance, the higher the current in the considered fastener, and vice versa. For example, looking at the results of fastener 160 at 500 MHz, we observe the positive correlation on the adjacent fasteners 161, 159, 162, 158, and 163. Unsurprisingly, the closer the neighbor, the greater the impact. However, this local effect is milder at low frequency. Indeed, as mentioned in the previous section, more parameters should impact the output at very low frequency than at higher frequencies (which Fig. 12 does not highlight because our model is not able to perform well at very low frequencies, and Fig. 12 is only based on model predictions). Since the decision boundaries of XGBoost are aligned with the coordinate system, the effective dimension at higher frequencies is much smaller than the original feature dimension (as the model is able to ignore irrelevant features). At low frequency, due to a higher effective dimension and the curse of dimensionality, a higher number of samples is required to learn. Thus, we believe this effect can be confirmed and the quality of model predictions can be drastically improved at low frequencies by increasing the amount of training data.

Finally, Fig. 12 also shows that the current in specific fasteners can be strongly influenced by the resistance of the fasteners close to the injection. This is illustrated in the case of the fastener 75 where fasteners 48 and 50 present an important correlation at 500 and 990 MHz. Indeed, the higher their resistances, the higher the current in fastener 75. It can be explained by the fact that if the resistances of fasteners 48 and 50 are high, then the current will mainly flow in the panel instead of privileging a rib path. As a result, more current will flow close to fastener 75, and, consequently, more current will flow into fastener 75.

D. Computational Cost

On a personal device, XGBoost requires in average 3.72 s to train for a given output dimension, which corresponds to 11.2 days for all 260 000 target dimensions. This computation only needs to be performed once. After training, a model requires in average 10.8 ms to perform a single prediction. This corresponds to 2.81 s on a personal device for all 260 000 target dimensions. Both training and predicting are easily parallelizable, and the costs can be further brought down by subsampling output dimensions. Such computational cost is negligible compared to performing simulations and justifies the use of XGBoost as a cost-effective approximation of fastener currents.

IV. CONCLUSION

This article proposes a machine learning approach based on the decision tree to predict the lightning currents on the fasteners of an aircraft fuel tank. From an FDTD modeling of the fuel tank, we completely describe the system with 264 input parameters, which corresponds to the 260 fastener resistances and the 4 material conductivities, as well as the 260 lightning currents sampled in 1000 frequencies between 100 Hz and 1 MHz. We build the database from 2000 simulations, where each fastener resistance takes a stochastic value drawn from the log-normal law (as shown in [9]), and where each material conductivity is drawn from a uniform law. As a result, we aim to predict 260 000 outputs, which correspond to each fastener current module at each frequencies. Then, several machine learning approaches

(k -nearest neighbors, SVR, XGBoost, and a neural network) are explored that lead to significantly better performance for XGBoost decision tree ensembles based on relative error and coefficient of determination. Indeed, XGBoost outperforms the other proposed models with an averaged coefficient on all output dimensions of 0.948 and an averaged relative error on all output dimensions of 5.9% while the second best model has only an averaged coefficient of 0.524 and an averaged relative error of 19%. Thus, we focus our work on XGBoost. Further analyzes show the following:

- 1) for most fasteners, we predict well the currents (with a relative error average over frequencies less than 10%);
- 2) the prediction accuracy increases with frequency;
- 3) the prediction accuracy is independent on the current amplitude to predict (except for the lowest frequencies).

The difficulty of the model to predict the lowest frequencies is assumed to be due to the large number of parameters impacting the current path at these frequencies rather than at higher frequencies where the current is driven by the inductive effects. Moreover, using SHAP, we analyze the impact of the input parameters. The results illustrate the following.

- 1) the primary fastener parameter (its own resistance) always plays a major role in the current distribution of the considered fastener;
- 2) other parameters may also play an important role, especially at higher frequencies depending on the position of the fastener under consideration, such as the conductivities of the associated materials due to the current path induced by inductive effects at these frequencies or the resistance of the fastener close to the injection;
- 3) our model is able to take into account nonmonotonous effects where a specific parameter at the same value can have a positive or negative impact on the considered fastener depending on the set of parameters. All these results are very promising.

Furthermore, a discussion about the computational cost is performed and highlights one of the main advantages of the proposed approach with a prediction in 2.81 s on a personal device for all 260 000 target dimensions (after training) that obviously outperforms the computational cost of the FDTD simulation. In addition, this kind of machine learning model allows us to perform further parametric studies with a low computational cost. Finally, we obtain very good results with few training data (relative to the question at hand). We believe there is much potential from improving the quality of predictions (especially at low frequencies) simply by increasing the number of training simulations.

REFERENCES

- [1] J. A. Plumer and J. D. Robb, "The direct effects of lightning on aircraft," *IEEE Trans. Electromagn. Compat.*, vol. EMC-24, no. 2, pp. 158–172, May 1982.
- [2] B. Burrows, "Scientific aspects of prevention of lightning induced sparks, fire, and explosion in above ground steel petroleum storage tank," in *Proc. Int. Conf. Lightning Static Electricity*, Paris, France, 2007.
- [3] L. Chemartin, P. Lalande, and F. Tristant, "Modeling and simulation of sparking in fastening assemblies," in *Proc. Int. Conf. Lightning Static Electricity*, Seattle, WA, USA, 2013. [Online]. Available: <https://hal-onera.archives-ouvertes.fr/hal-01058554>

- [4] F. Fustin, F. Tristant, J.-P. Moreau, and F. Terrade, "Fuel tank safety—3D distributions on fasteners/assemblies outside and inside fuel tanks," in *Proc. Int. Conf. Lightning Static Electricity*, 2013, Seattle, WA, USA, 2013.
- [5] I. Revel, G. Peres, B. Lepetit, and F. Flourens, "Understanding of sparking phenomenon in CFRP assemblies," in *Proc. Int. Conf. Lightning Static Electricity*, Paris, France, 2007.
- [6] P. Monferran, C. Guiffaut, A. Reineix, F. Fustin, and F. Tristant, "Fastening assemblies modelling in finite difference time domain," in *Proc. Int. Symp. Electromagn. Compat.*, 2018, pp. 521–526.
- [7] P. Monferran, C. Guiffaut, A. Reineix, F. Fustin, and F. Tristant, "Lightning currents on fastening assemblies of an aircraft fuel tank, Part I: Uncertainties assessment with statistical approach," *IEEE Trans. Electromagn. Compat.*, vol. 62, no. 3, pp. 807–817, Jun. 2020.
- [8] P. Monferran, C. Guiffaut, A. Reineix, F. Fustin, and F. Tristant, "Lightning currents on fastening assemblies of an aircraft fuel tank, Part II: FDTD modeling merged with a circuit model supplemented by a statistical model," *IEEE Trans. Electromagn. Compat.*, vol. 62, no. 3, pp. 818–828, Jun. 2020.
- [9] P. Monferran, C. Guiffaut, A. Reineix, F. Fustin, and F. Tristant, "Lightning currents on fastening assemblies of an aircraft fuel tank—Part III: Validation of the current assessment method with measurement," *IEEE Trans. Electromagn. Compat.*, vol. 62, no. 5, pp. 2174–2185, Oct. 2020.
- [10] H. Mulazimoglu and L. Haylock, "Recent developments in techniques to minimize lightning current arcing between fasteners and composite structure," in *Proc. Int. Conf. Lightning Static Electricity*, Oxford, U.K., 2011.
- [11] D. Li, A. Luque, F. Rachidi, and M. Rubinstein, "The application of the finite-difference time-domain (FDTD) technique to lightning studies," *Adv. Time Domain Model. for Elect. Eng.*, London, U.K.: IET, 2022, pp. 335–370. [Online]. Available: <http://infoscience.epfl.ch/record/296476>
- [12] F. Fustin, F. Tristant, F. Terrade, M. Latorré, and E. Pierré, "Numerical modeling - lightning response of a falcon 7X wing fuel tank: Test results and validation of simulation model," in *Proc. Int. Conf. Lightning Static Electricity 2017*, Nagoya, Japan, Sep. 2017.
- [13] J. P. Béranger, "A perfectly matched layer for the absorption of electromagnetic waves," *J. Comput. Phys.*, vol. 114, no. 2, pp. 185–200, 1994.
- [14] J. P. Béranger, "Perfectly matched layer for the FDTD solution of wave-structure interaction problems," *IEEE Trans. Antennas Propag.*, vol. 44, no. 1, pp. 110–117, Jan. 1996.
- [15] J. P. Béranger, "Perfectly matched layer (PML) for computational electromagnetics," *Synth. Lectures Comput. Electromagnetics*, vol. 2, no. 1, pp. 1–117, 2007.
- [16] J. P. Béranger, "Plaques minces AUX différences finies," in *Proc. Géme Colloq Int. Expos. Compat. Electromagn.*, 1992, pp. 298–303.
- [17] R. J. Luebbers and K. Kunz, "FDTD modeling of thin impedance sheets (radar cross section calculation)," *IEEE Trans. Antennas Propag.*, vol. 40, no. 3, pp. 349–351, Mar. 1992.
- [18] C. Railton and J. P. McGeehan, "An analysis of microstrip with rectangular and trapezoidal conductor cross sections," *IEEE Trans. Microw. Theory Techn.*, vol. 38, no. 8, pp. 1017–1022, Aug. 1990.
- [19] Time ElectroMagnetic Simulator—Finite Difference software, TEMSI-FD. CNRS, University of Limoges, Limoges, France, 2006.
- [20] E. Fix and J. L. Hodges, "Discriminatory analysis. nonparametric discrimination: Consistency properties," *Int. Statist. Rev. / Revue Int. de Statistique*, vol. 57, no. 3, pp. 238–247, 1989. [Online]. Available: <http://www.jstor.org/stable/1403797>
- [21] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
- [22] V. N. Vapnik, *The Nature of Statistical Learning Theory*. Berlin, Germany: Springer, 1995.
- [23] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. and Data Mining*, 2016, pp. 785–794. [Online]. Available: <https://doi.org/10.1145%2F2939672.2939785>
- [24] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 4768–4778. [Online]. Available: <http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf>
- [25] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learn.*. Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: <http://www.deeplearningbook.org>
- [26] M. A. Nielsen, *Neural Networks and Deep Learning*, vol. 25. San Francisco, CA, USA: Determination, 2015.
- [27] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, 2015. [Online]. Available: <https://www.science.org/doi/abs/10.1126/science.aaa8415>
- [28] C. M. Bishop and N. M. Nasrabadi, *Pattern Recognition and Machine Learning*, vol. 4, no. 4. Berlin, Germany: Springer, 2006.
- [29] V. Borisov, T. Leemann, K. Sebler, J. Haug, M. Pawelczyk, and G. Kasneci, "Deep neural networks and tabular data: A survey," 2021, *arXiv:2110.01889*.
- [30] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *CoRR*, vol. abs/1502.03167, 2015. [Online]. Available: <http://arxiv.org/abs/1502.03167>
- [31] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Int. Conf. Mach. Learn.*, 2010, pp. 807–814.
- [32] D. Saad, "Online algorithms and stochastic approximations," *Online Learn.*, vol. 5, no. 3, 1998, Art. no. 6.



Paul Monferran was born in Paris, France, in 1992. He received the master's degree in sensors, measurement and instrumentation from the University Pierre-et-Marie Curie, Paris, France, in 2015 and the Ph.D. degree in electronics from the XLIM Laboratory, University of Limoges, Limoges, France, in 2018.

He is currently a Postdoctoral Fellow with the LEOST Laboratory, University Gustave Eiffel, Villeneuve d'Ascq, France. His current research interests include the electromagnetic characterization of

several materials, lightning effects in aircraft for electromagnetic compatibility issues, and machine learning applied to EMC issues.



Charles Guille-Escuret was born in Paris, France, in 1994. He received a master's degree in machine learning from École Normale Supérieure de Paris-Saclay, Cachan, France, in 2016. He is currently working toward the Ph.D. degree in computer science and statistics with Mila, Quebec AI Institute, Montréal, QC, Canada.

His research interests focus mainly on neural networks and their optimization, with additional interests in machine learning applications.



Christophe Guiffaut (Member, IEEE) was born on March 14, 1973 in Chateaubriant, France. He received the master's and Ph.D. degrees in electronics and telecommunications from the University of Rennes, Rennes, France, in 1997 and 2000, respectively.

He joined the CNRS Research Center in 2002 and he integrated in the same year the XLIM Laboratory, University of Limoges, Limoges, France. His research interests include the numerical methods domain for application areas in electromagnetic compatibility, antennas, and ground penetrating radar.



Alain Reineix (Member, IEEE) was born in Limoges, France, in 1961. He received the master's degree in electronics and telecommunications and the Ph.D. degree in electronics from the University of Limoges, Limoges, France, in 1984 and 1986, respectively.

In 1986, he joined Centre National de la Recherche Scientifique (CNRS) as a Researcher with the IRCOM Laboratory, University of Limoges. Since 1991, he has been a Research Director with CNRS and also the Head of a Research Group on EMC with IRCOM Laboratory and now (since 2006) XLIM laboratory,

University of Limoges. His current research interests include improvements on numerical modelings particularly in FDTD and ground penetration studies in the radar domain.

Dr. Reineix is a URSI Correspondent and a member of the French Electrical Engineering Society (SEE).