A methodology to generate design allowables of composite laminates using machine learning

Furtado, C.; Tavares, R. P.; Gomes Pereira, L.P.; Salgado, M.; Otero, F.; Catalanotti, G.; Arteiro, A.; Bessa, M. A.; Camanho, P. P.

DOI
10.1016/j.ijsolstr.2021.111095

Publication date
2021

Document Version
Final published version

Published in
International Journal of Solids and Structures

Citation (APA)
Furtado, C., Tavares, R. P., Gomes Pereira, L. P., Salgado, M., Otero, F., Catalanotti, G., Arteiro, A., Bessa, M. A., & Camanho, P. P. (2021). A methodology to generate design allowables of composite laminates using machine learning. *International Journal of Solids and Structures*, *233*, Article 111095. https://doi.org/10.1016/j.ijsolstr.2021.111095

International Journal of Solids and Structures 233 (2021) 111095

Contents lists available at ScienceDirect

International Journal of Solids and Structures

journal homepage: www.elsevier.com/locate/ijsolstr

A methodology to generate design allowables of composite laminates using machine learning

C. Furtado [a,b,*], L.F. Pereira [b,c,d], R.P. Tavares [b,e], M. Salgado [b,c], F. Otero [b,f], G. Catalanotti [g], A. Arteiro [b,c], M.A. Bessa [d], P.P. Camanho [b,c]

[a] Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, United States
[b] INEGI, Instituto de Ciência e Inovação em Engenharia Mecânica e Engenharia Industrial, Rua Dr. Roberto Frias, 400, 4200-465 Porto, Portugal
[c] DEMec, Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias, s/n, 4200-465 Porto, Portugal
[d] Department of Materials Science and Engineering, Delft University of Technology, 2628 CD Delft, The Netherlands
[e] Department of Materials, Textiles and Chemical Engineering, Ghent University, Technologiepark-Zwijnaarde 46, B-9502 Zwijnaarde, Belgium
[f] CIMNE, Gran Capita s/n, Barcelona 08034, Spain
[g] Escola de Ciências e Tecnologia, Universidade de Évora, Colégio Luis António Verney, Rua Romão Ramalho, 59, 7000-671 Évora, Portugal

ARTICLE INFO

Keywords:
Polymer-matrix composites (PMCs)
Machine learning
Fracture mechanics
Design allowables

ABSTRACT

This work represents the first step towards the application of machine learning techniques in the prediction of statistical design allowables of composite laminates. Building on data generated analytically, four machine algorithms (XGBoost, Random Forests, Gaussian Processes and Artificial Neural Networks) are used to predict the notched strength of composite laminates and their statistical distribution, associated to the uncertainty related to the material properties and geometrical features. This work focuses not only on the so-called Legacy Quad Laminates ($0°/90°/\pm45°$), typically used in the design of composite aerostructures, but also on the newer concept of double-double (or double-angle ply) laminates. Very good representations of the design space, translating in low generalization relative errors of around $\pm10\%$, and very accurate representations of the distributions of notched strengths around single design points and corresponding B-basis allowables are obtained. All machine learning algorithms, with the exception of the Random Forests, show very good performances, with Gaussian Processes outperforming the others for very small number of data points while Artificial Neural Networks have better performance for larger training sets. This work serves as basis for the prediction of first-ply failure, ultimate strength and failure mode of composite specimens based on non-linear finite element simulations, providing further reduction of the computational time required to virtually obtain the design allowables for composite laminates.

1. Introduction

The generation of design allowables for composite laminates is of utmost importance for the design and certification of the composite structures used in the aerospace industry. The determination of these design allowables, which account for the variability associated with curing/consolidation procedures, geometrical features and defects characteristic of composite structures, usually relies on extensive, expensive and time-consuming experimental test campaigns. With the increase of computational power, and the development of high-fidelity numerical models that accurately represent the response and failure of composite materials, alternatives to generate design allowables based on advanced finite element simulations have also been sought out to reduce the certification costs (Tay et al., 2005; der Meer et al., 2010; Ling et al., 2009; Schuecker and Pettermann, 2006; Camanho et al., 2007a; Vogler et al., 2013; Camanho et al., 2013; Abdi et al., 2016; Zhang et al., 2017; Abumeri et al., 2011; Spendley, 2012). However, these solutions are still computationally expensive, especially if uncertainty is accounted for. The recent advances on machine learning techniques opens a new window of possibilities for the faster prediction of the structural response of composite materials and their optimization (Bessa et al., 2017; Bessa and Pellegrino, 2018; Bessa et al., 2019; Bisagni and Lanzi, 2002; Yvonnet and He, 2007), by allowing the definition of surrogate models that continuously and analytically describe the design space.

* Corresponding author at: Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, United States.
E-mail address: cfurtado@mit.pt (C. Furtado).

https://doi.org/10.1016/j.ijsolstr.2021.111095
Received 5 October 2020; Received in revised form 6 May 2021; Accepted 21 May 2021
Available online 26 May 2021

In recent years, significant effort has been made to use surrogate models to represent the response of composite materials and, consequently, reduce the computation cost of finite element simulations (El Said and Hallett, 2018; Balokas et al., 2018; Yan et al., 2020). In general, the process relies on i) design of experiments, where the descriptors of the design space are defined, ii) data generation, where data to train and test the surrogate models is obtained, and iii) surrogate model definition (or machine learning application) where a model that represents the design space is defined. For instance, El Said and Hallett (2018) and Yan et al. (2020) proposed multiscale modelling approaches based on surrogate models at the mesoscale: the former to predict the elastic response of structures with internal defects (namely wrinkles), and the latter to predict composite structural damage and failure. Both approaches relied on the definition of Representative Volume Elements (RVE) at the mesoscale, which were used to populate the design space and train a surrogate model. The surrogate model was then used to represent the composite response at the macroscale, avoiding the need to run RVE models in parallel with the macroscale simulation, significantly reducing the computational cost of the models. These surrogate models focus on the representation of the material behaviour at the mesoscale level and still rely on numerical simulations at the macroscale. If standard tests for certification of composite materials (plain strength, open-hole strength, bolted strength, among others) can be described parametrically and accurate analytical surrogate models can be built on data from numerical simulations, the virtual certification of composite materials can be greatly simplified. This paper presents the first steps towards that goal.

In this paper, a feasibility study on the application of machine learning techniques for predicting a design allowable, the notched strength of multidirectional composite laminates, is presented, with the main goals of presenting the challenges of applying machine learning techniques for composite laminates, and of evaluating the most appropriate algorithms for the determination of composite design allowables. Even though the data-driven framework is established on data derived from an analytical framework (Furtado et al., 2017; Vallmajó et al., 2019), this work serves as basis and guideline to a more demanding challenge that includes the prediction of first-ply failure strength, ultimate strength and failure mode of composite materials based on non-linear finite element simulations. The data-driven framework is defined following the procedure schematically shown in Fig. 1:

- First, the design of experiments is performed (Section 3), where the input descriptors are selected following a discussion on the representation of stacking sequence of composite laminates compatible with machine learning techniques;
- Secondly, data generation is performed (Section 4) where the analytical framework proposed by Furtado et al. (2017) (Section 2) is used to populate the design space for open-hole strength following the descriptors defined in the design of experiments;
- Then, the prediction of open-hole tensile strength of composite laminates using machine learning techniques is presented (Section 5) to evaluate their ability to capture the overall response of the design space;
- Finally, the generation of design allowables based on machine learning models is explored in Section 6, to verify their ability to capture the variability associated with a given design point, consequence of the uncertainty related to the material properties and geometrical descriptors.

2. Problem definition

2.1. Analytical framework to predict the notched strength of multidirectional composite laminates

Furtado et al. (2017) proposed an analytical framework to predict the notched strength of multidirectional carbon-epoxy laminates based on a combination of three building blocks: a finite fracture mechanics model (Camanho et al., 2012), invariant-based approaches to estimate stiffness and strength (Tsai and Melo, 2014; Tsai and Melo, 2016) and an analytical model based on Fracture Mechanics to estimate the laminate fracture toughness (Camanho and Catalanotti, 2011).

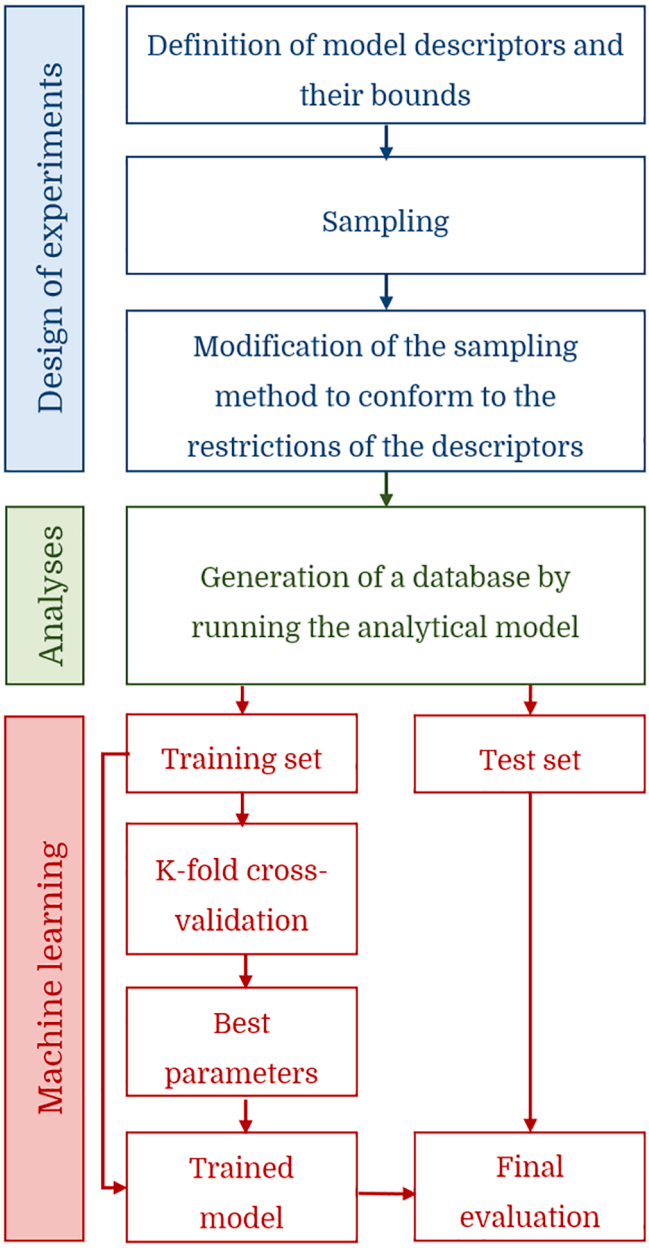

Fig. 1. Schematic representation of the machine learning framework.

The combination of the models provides an efficient framework to predict the laminate elastic, strength and fracture properties required to predict size effects in notched laminates. The model requires the knowledge of:

1. Three ply-level material properties: the longitudinal Young's modulus, E_1, the longitudinal strength, X, and the critical energy release rate, $\mathscr{G}^0$, or longitudinal crack resistance curve, $\mathscr{R}$-curve;
2. The stacking sequence of the laminate.
3. Two geometrical features: the width of the specimen, W, and the diameter of the hole, D;

The proposed methodology (Furtado et al., 2017) can be summarized

as follows (see also Fig. 2):

1. From the longitudinal Young's modulus of the ply, E_1, the Master Ply concept is used to determine Tsai's modulus (Arteiro et al., 2020), Tr, and estimate the ply elastic properties (Tsai and Melo, 2014; Arteiro et al., 2019). From Tsai's modulus, lay-up, and using the universal laminate factors, the elastic properties of the laminate are obtained.
2. From the longitudinal strength of the ply, X, and using the elastic properties of the laminate determined using Tsai's modulus, a maximum allowable strain criterion is applied to predict the laminate unnotched strength, X^L.
3. From the critical energy release rate, $\mathscr{G}^0$, and using the elastic properties of the balanced and 0° sublaminates determined using Tsai's modulus, an analytical model based on fracture mechanics (Camanho and Catalanotti, 2011) is used to predict the laminate fracture toughness, $\mathscr{G}^L$.
4. Finally, the laminate unnotched strength, X^L, and the laminate fracture toughness, $\mathscr{G}^L$, are used in the Finite Fracture Mechanics model (Camanho et al., 2012) to predict the notched strength ($\bar{\sigma}$) of the laminate for any specimen configuration (W and D).

The analytical framework is able to accurately predict the tensile notched strength of composite laminates as shown in references Camanho et al. (2012), Erçin et al. (2013), Arteiro et al. (2013), Arteiro et al. (2014), Vallmajó et al. (2019) and Furtado et al. (2017) where its predictions were compared with experimental results. The original analytical model (Camanho et al., 2012) was applicable for predicting the strength of quasi-isotropic laminates, therefore, a recently proposed generalization for highly orthotropic materials was used here (Catalanotti et al., 2021).[1]

Given its simplicity and efficiency, the analytical framework summarised above is a good candidate to perform a feasibility investigation to appropriately define the input parameters/descriptors, to assess the ability of machine learning algorithms to predict the strength of composite laminates, and to identify the most effective algorithms compatible with structural analysis of composite laminates.

2.2. Strategies for laminate definition

Most of the input parameters required by the analytical framework can be treated as continuous and independent variables, appropriate to build training data to feed machine learning algorithms. However, the laminate stacking sequence requires a more detailed discussion. Considering each ply orientation of a stacking sequence an input parameter results in a high dimensional representation, which is inconvenient in machine learning applications since larger datasets are generally required to accurately capture the design space.

In the following section, two strategies for the definition of laminates are presented: one based on lamination parameters (Tsai and Pagano, 1968), and another based on the double-double (or double angle-ply) laminates recently proposed by Tsai et al. (2017) as a practical alternative to the conventional Legacy Quad laminates[2].

2.2.1. Conventional laminates

The most common way to define a laminate is by its stacking sequence (i.e. assembly of plies with specified fibre orientation angles). This is a simple and convenient representation, but has a major drawback: the number of variables is intrinsically related to the number of plies. This translates in a high-dimensional laminate representation and, more importantly, in a representation with variable dimension.

Lamination parameters, firstly proposed by Tsai et al. in 1968 (Tsai and Pagano, 1968), provide a more compact definition of the laminate: twelve lamination parameters and a thickness variable are sufficient to geometrically define any laminate, independently of the number of plies. However, it is important to note that these twelve lamination parameters are interrelated, i.e. when the values of some parameters are fixed, the other are constrained to a certain feasible region. Significant time and effort have been dedicated to efficiently define the constraints of the 12-dimensional convex (Grenestedt and Gudmundson, 1993) feasible domain of the lamination parameters, both with (Bloomfield et al., 2009) and without (Setoodeh et al., 2006) restrictions on the possible ply orientations. The in-plane, $\zeta^A_{\{1,2,3,4\}}$, coupled, $\zeta^B_{\{1,2,3,4\}}$, and out-of-plane, $\zeta^D_{\{1,2,3,4\}}$, lamination parameters are calculated as:

$$\zeta^A_{\{1,2,3,4\}} = \frac{1}{h}\sum_{i=1}^{N}\begin{Bmatrix}\cos(2\theta_i)\\ \cos(4\theta_i)\\ \sin(2\theta_i)\\ \sin(4\theta_i)\end{Bmatrix}(z_i - z_{i-1}) \tag{1}$$

$$\zeta^B_{\{1,2,3,4\}} = \frac{2}{h^2}\sum_{i=1}^{N}\begin{Bmatrix}\cos(2\theta_i)\\ \cos(4\theta_i)\\ \sin(2\theta_i)\\ \sin(4\theta_i)\end{Bmatrix}(z_i^2 - z_{i-1}^2) \tag{2}$$

$$\zeta^D_{\{1,2,3,4\}} = \frac{4}{h^3}\sum_{i=1}^{N}\begin{Bmatrix}\cos(2\theta_i)\\ \cos(4\theta_i)\\ \sin(2\theta_i)\\ \sin(4\theta_i)\end{Bmatrix}(z_i^3 - z_{i-1}^3) \tag{3}$$

where h is the thickness of the laminate, θ_i is the fibre orientation at height $z \in [z_{i-1}, z_i]$, and N is the number of plies of a laminate. The twelve lamination parameters fully and uniquely represent a laminate, and, therefore, they can be converted to stacking sequences. However, this is a non-trivial process, whose solution generally requires restrictions to the possible permitted ply orientations and the use of optimization algorithms (IJsselmuiden et al., 2009; Irisarri et al., 2011; Meddaikar et al., 2017; Bloomfield et al., 2010) or other techniques (Todoroki and Sekishiro, 2007; Liu et al., 2019; Viquerat, 2020).

For the problem at hand, only two lamination parameters need to be considered, $\zeta^A_{\{1,2\}}$, because:

1. The analytical framework previously described deals with in-plane loading only, and considers that the laminate is homogenized (not necessarily symmetric), i.e. the bending-extension coupling matrix $\mathbf{B} \approx \mathbf{0}$; in other words, it is only capable of distinguishing lay-ups and not stacking sequences. Since laminate organization is not accounted for, the laminates can be fully described by the in-plane lamination parameters, $\zeta^A_{\{1,2,3,4\}}$.

[1] To compute the notched strength using the finite fracture mechanics model proposed in Ref. Camanho et al. (2012), the stress intensity factor for a plate with a central circular hole of radius R and two symmetric cracks emanating from the hole edge needs to be calculated. However, an analytical expression for the stress intensity factor of this configuration only exists for quasi-isotropic laminates (Newman Jr., 1983). The recent generalization mentioned and used in this work stems from a numerical study to determine the stress intensity factor of cracks emanating from circular and elliptical holes in orthotropic plates. The approach is based on the original work from Suo et al. (1991) and a semi-analytical expression for the correction factor, $\phi = \frac{\mathscr{K}_I}{\sqrt{R}\sigma^\infty}$, was used to take into account the effects of orthotropy.

[2] Legacy Quad laminates refers to laminates composed of 0°, 90° and ±45° plies, with a minimum of 10% of each orientation, symmetric at the mid-plane and balanced (same number of 45° and −45° plies). These lamination thumb rules have been used to limit the complexity of composites design.

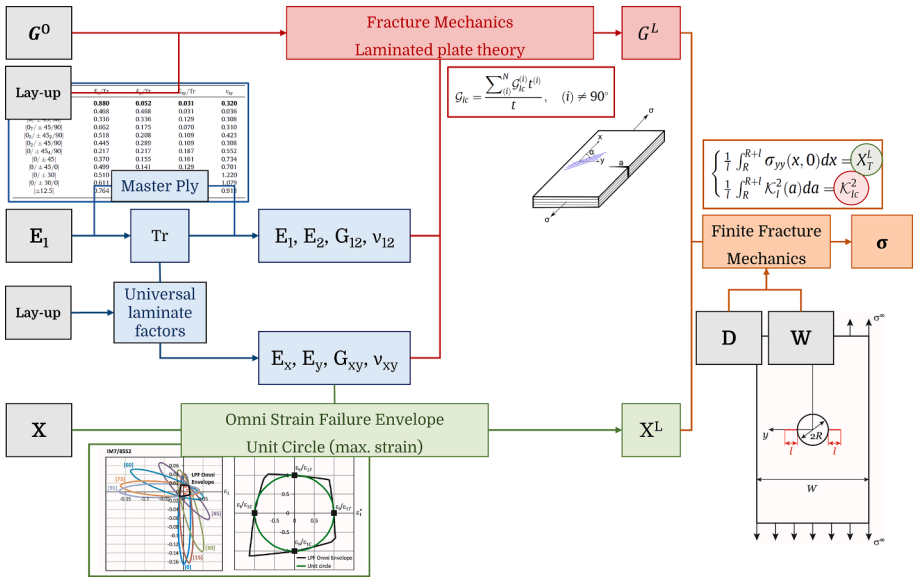

Fig. 2. Schematic representation of the proposed combined framework to predict size effects from the minimum number of properties determined at the ply level.

2. Only balanced laminates can be considered Camanho and Catalanotti, 2011; Furtado et al., 2017, therefore $\zeta^A_{\{3,4\}} = 0$.

As shown in Fig. 3, this definition based on the lamination parameters $\zeta^A_{\{1,2\}}$ results in a non-rectangular (i.e. $\zeta^A_{\{1,2\}}$ are not fully independent) representation of the normalized stiffness. This will be further commented in Section 3.

2.2.2. Double-double laminates

An alternative to the Legacy Quad laminates, where plies with orientation of $0°, 90°$ and $\pm 45°$ are used, was recently proposed by Tsai et al. (2017): the Double-Double (DD) laminates. These laminates are composed of plies of two orientations and are defined as $[\pm\phi/\pm\psi]_n$ or $[+\phi/+\psi/-\phi/-\psi]_n$, where n is the number of repetitions. This bi-angle approach to laminate design results in stronger laminates with higher resistance to micro-cracking and delamination and in other advantages such as faster layup, simpler design and easier tapering through single ply drops (Tsai et al., 2017; Shrivastava et al., 2020).

DD-sublaminate lay-ups can be fully described by two parameters, ϕ and ψ, which can vary continuously from $0°$ to $90°$. This provides a smaller design space, which, in one hand, reduces the design flexibility attributed to composite materials, but, on the other hand, allows for a more efficient and otherwise nonviable stacking sequence optimization.

As shown in Fig. 4, this definition based on the ply orientations, ϕ and ψ, results in a continuous and injective representation of the normalized stiffness (Tsai et al., 2017). Furthermore, the normalized stiffness is diagonal symmetric, a characteristic that will be further commented on and taken advantage of in Section 3.

2.3. Definition of dimensionless parameters

As described in Section 2.1, the notched strength of an open-hole specimen, σ, can be fully described as:

$$\sigma = f_1(E_1,\ X_T,\ \mathcal{G}_0,\ D,\ W/D,\ \text{lay-up}) \tag{4}$$

where E_1, X_T and $\mathcal{G}_0$ are material descriptors, D and W/D are geometric descriptors and the lay-up can be described by:

$$\text{lay-up} = f_2(\alpha,\ \beta) \tag{5}$$

where $\alpha = \zeta^A_1$ and $\beta = \zeta^A_2$ for a Quad laminate and $\alpha = \phi$ and $\beta = \psi$ for a DD laminate (Section 2.2), comprising a total of seven input parameters. The number of input parameters can be reduced using the Buckingham's Π theorem (Buckingham, 1914) that states that, if there is a physical relation involving N physical variables, then the relation can be rewritten as a relation of $(N-K)$ dimensionless products where K is the

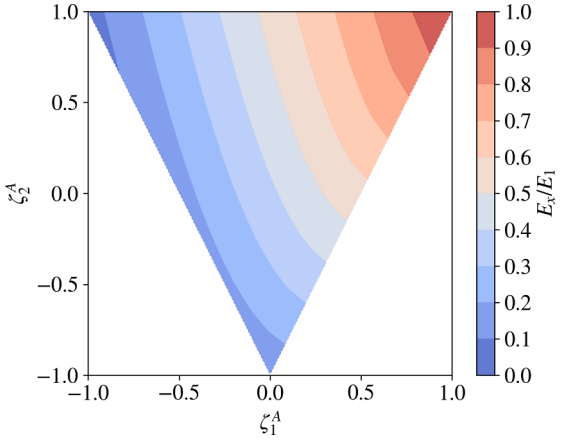

Fig. 3. Normalized stiffness as a function of $\zeta^A_{\{1,2\}}$ (valid for $0°/\pm 45°/90°$ Legacy Quad laminates).

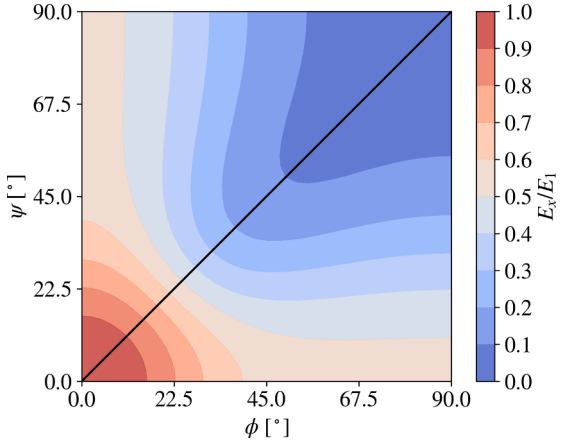

Fig. 4. Normalized stiffness of DD laminates as a function of ϕ and ψ.

number of fundamental dimensions required to describe the physical variables. Following the Buckingham Π theorem, the number of dimensions can be reduced to four dimensions, using the following dimensionless parameters:

$$\frac{\sigma}{X_T} = f\left(\frac{\mathcal{G}_0 E_1}{D X_T^2},\ W \Big/ D,\ \alpha,\ \beta\right) \tag{6}$$

This provides a more compact description of the problem at hand, which is convenient in machine learning applications since high dimensional spaces generally require larger datasets to provide accurate surrogate models.

Note that the design of experiments (presented in Section 3) is performed in the original 7D space (Eq. 4), which is then converted to the dimensionless 4D space (Eq. 6), which is in turn used to train and test the machine learning models. Although other strategies could have been defined, this strategy stems from the fact that the relation between the 4D and 7D space is not necessarily injective. For example, it is possible that two specimens with the same geometry (W/D) but made from two different materials have the same 4D parameters and therefore the same notched strength. For this reason, it was more convenient to define the design space in the 7D space, guaranteeing that each variable is well populated in its defined bounds and then convert them to the 4D space before the training step.

3. Design of experiments

The design of experiments was performed using the implementation of Sobol sequence (Sobol, 2001; Saltelli, 2002; Saltelli et al., 2010) in the SALib Python library (Herman and Usher, 2017): a low discrepancy quasi-random sequence designed to explore the space in a more uniform manner than random sampling[3].

Sampling schemes (Sobol, 2001; Saltelli, 2002; Saltelli et al., 2010; McKay et al., 1979) were designed considering continuous variables with fixed bounds (i.e. no variable dependency is allowed). This attribute is compatible with the material property (E_1 , X_T, $\mathcal{G}_0$) and geometric (D, W/D) input parameters. However, the laminate descriptors require a more careful analysis: on one hand, when considering conventional laminates, it is clear that the lamination parameters (ζ_1^A, ζ_2^A) are not independent (represented by the non-rectangular region in Fig. 3) and depending on the hypothesis initially considered (maximum number of plies, ply angles allowed, etc.), yield discrete combinations of values. On the other hand, when considering DD laminates, and even though ϕ and ψ are independent variables, advantage can be taken of the fact that the stiffness is diagonal symmetric as shown in Fig. 4.

The strategies used to conform the sampling schemes to the laminate descriptors are presented herein.

3.1. Laminate descriptors for conventional laminates

Sampling has to be performed in the lamination parameters space, however, the analytical framework presented in Section 2 requires the lay-up as an input parameter to compute the notched strength. For this reason, there is the need to solve the inverse problem, i.e. obtain the stacking sequence from the lamination parameters. This is a non-trivial problem, that generally involves time-consuming optimization algorithms (IJsselmuiden et al., 2009; Irisarri et al., 2011; Meddaikar et al., 2017; Bloomfield et al., 2010) and can only be performed in an acceptable time frame if restrictions are imposed, such as a fixed number of plies, discrete values of allowed angles, among others. To avoid the in situ computation of the inverse problem, a database that relates all the possible ζ_1^A/ζ_2^A combinations to a corresponding lay-up was created. The following laminate restrictions were imposed:

1. Only ply orientations of 0°, +45°, −45° and 90° are considered (Legacy Quad laminates);
2. The number of plies of each orientation range from 0 to 32;
3. All laminates are balanced.

Since laminate organization (i.e. stacking sequence) is not accounted for in the model, and only balanced laminates can be considered (the number of +45°, −45° plies is equal), this yields $32^3 - 1 = 32767$ laminate permutations (27133 unique orientation-percentage combinations), which configures a database with acceptable size, that can be efficiently accessed to convert lamination parameters to a corresponding lay-up. Note that for less restrictive constraints (more angles considered, larger number of plies, non-balanced laminates, etc.), this will yield a database too large to be accessed efficiently.

Sobol sampling assumes the variables to be continuous and independent from one another. As shown in Fig. 5, where all the 27133 possible pairs of ζ_1^A/ζ_2^A points are plotted in grey, this is not verified in the present case. For this reason, the distance of all the Sobol generated points (in red) to the closest allowed point is calculated. The Sobol generated point is either discarded, in case the computed distance is higher than a given threshold value, or approximated to the closest allowed point (in blue). If the defined threshold distance is too high, all points will be approximated to the closest allowed point, leading to a densely populated region at the allowed ζ_1^A/ζ_2^A boundary (in this case, the triangle boundary shown in Fig. 5). If the threshold distance is too low, most points will be discarded. A sensitivity analysis was made to analyse the effect of the selected threshold distance on the performance of the trained models and no significant dependency was found (the analysis is not shown here for the sake of conciseness). The authors considered that a threshold distance of 0.1, that would guarantee that all points inside the boundary would be included and most points outside the boundary would be discarded, was appropriate as a criterion for discarding generated points. This is a simple and efficient strategy to select the Sobol sequence sampling points that conform with the constraints of the problem. A more complex sampling method that respects the restrictions imposed to the laminate descriptors could be potentially also be envisioned.

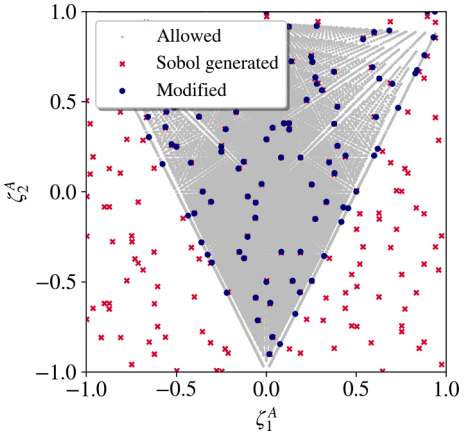

Fig. 5. Sobol sampling modification for Legacy Quad laminates.

[3] Other sampling strategies, namely, Saltelli's extension of Sobol sequence (Saltelli, 2002; Saltelli et al., 2010), Latin hypercube sampling (McKay et al., 1979) and random sampling were tested in this work. However, no significant dependency was found between model performance and the sampling strategy used, i.e., models trained on training sets created using the different sampling methodologies resulted in very similar performances. Therefore, for the sake of conciseness, only the database generated using Sobol sampling is presented here.

3.2. Laminate descriptors for DD laminates

For the DD laminates, sampling is performed in the ply orientation space, which is continuous and composed of independent variables. The two angles are then simply organized to form a laminate: $[\pm\phi/\pm\psi]$. Since laminate organization is not accounted for by the analytical framework, the number of repetitions n does not need to be considered as the percentage of layers in each direction is independent of the number of repetitions. Moreover, the DD laminates are balanced laminates by definition, and therefore, no further restrictions are required.

However, as shown in Fig. 4, the stiffness is diagonal symmetric with respect to ϕ and ψ because the angles are interchangeable. Although this is not required, advantage can be taken of the fact that the stiffness is diagonal symmetric to obtain a more densely packed distribution in the region of interest. As shown in Fig. 6, a ϕ/ψ Sobol generated point for which $\psi > \phi$ is converted into a ψ/ϕ point.

4. Computational analyses

In the previous section, the design of experiments, where the input variables, their bounds and strategies to populate the design space are defined, was described. Here, the analysis, where the output variable (the notched strength) is computed for the defined input data (material, geometric and laminate descriptors), is described. In this work, the output considered is the open-hole strength of a laminate as determined using the analytical framework described in Section 2.

For conventional laminates, for each input data point, $x_i = [E_1,\ X_T,\ \mathscr{G}_0,\ D,\ W/D,\ \zeta_1^A,\ \zeta_2^A]_i$:

1. the laminate parameters, ζ_1^A and ζ_2^A, are converted in a lay-up by accessing the database that configures all the possible $\zeta_1^A/\zeta_2^A \leftrightarrow$ lay-up combinations.
2. The notched strength, σ, is computed.
3. The original 7D parameters are converted to the 4D design space: $x_i^* = [\frac{\mathscr{G}_0 E_1}{D X_T^2},\ W/D,\ \zeta_1^A,\ \zeta_2^A]_i$.
4. The normalized notched strength, σ/X_T, is computed.

For the DD laminates, for each input data point, $x_i = [E_1,\ X_T,\ \mathscr{G}_0,\ D,\ W/D,\ \phi,\ \psi]_i$:

1. The laminate descriptors, ϕ and ψ, are organized in a lay-up: $[\pm\phi/\pm\psi]$
2. The notched strength, σ, is computed.
3. The original 7D parameters are converted to the 4D design space: $x_i^* = [\frac{\mathscr{G}_0 E_1}{D X_T^2},\ W/D,\ \phi,\ \psi]_i$.
4. The normalized notched strength, σ/X_T, is computed.

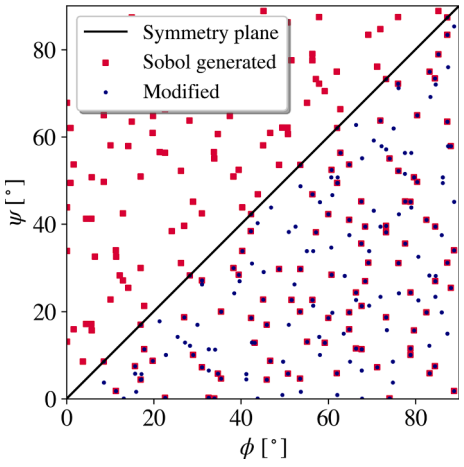

Fig. 6. Sobol sampling modification for DD laminates.

Since the framework is based on an analytical model, the computation of a notched strength is not time-consuming, and therefore, not a limiting factor. For reference, the notched strength of 1000 data points takes 8 s to be computed in a standard laptop. For more computationally expensive problems, e.g. based on finite element simulations, the amount of information available (or that can be generated) is more limited. Therefore, the goal here is to obtain good approximations of the design space with the least amount of data possible and to show how the available number of data points affect performance of the models.

5. Prediction of open-hole tensile strength of composite laminates using machine learning

At this point, a database of well distributed points has been generated and is available to train and test machine learning algorithms. As schematically shown in Fig. 1, in data-rich situations, a given machine learning algorithm, with fixed hyperparameters (parameters that are not optimized during the learning process) is trained using a training set. The trained model is then used to predict the output of the test set points in order to estimate the generalization error of the model. For small number of data points, the randomness associated with the training set used is higher and several repetitions of the evaluation procedure must be performed to obtain a more reliable estimate of the generalization error (or, alternatively, more robust evaluation procedures, such as k-fold cross validation Raschka, 2018a, can be used). It is not trivial to quantify how much data is sufficient, since it depends on several factors, such as the complexity of the problem and the models being fitted to the data (Hastie et al., 2013). Further considerations must be taken into account when several machine learning algorithms and hyperparameter combinations are tested. This will be further commented on in Section 5.2.

Gaussian Processes (Krige, 1951), Artificial Neural Networks (Rosenblatt, 1958) and two tree-ensemble algorithms (Random Forests (Breiman, 2001) and XGBoost (Chen et al., 2016) are considered here. The two tree-ensemble algorithms have a fast learning process, do not require data standardization, have easily tunable hyperparameters and have been reported to be very powerful algorithms when dealing with structured data (Chollet, 2017; Hamidieh, 2018). Furthermore, they have been extensively applied by winning teams in Kaggle competitions (Hamidieh, 2018; Chen et al., 2016), and have been successfully applied in the materials science realm, e.g. for predicting the critical temperature of superconductors (Hamidieh, 2018; Stanev et al., 2018). Artificial Neural Networks are very powerful and extremely scalable, being able to deal with large datasets, but are harder to tune properly, e.g. require adequate hyperparameter tuning to provide good predictions. Gaussian Processes are also very powerful, particularly for regression problems, provided adequate kernel selection (Rasmussen and Williams, 2006; Duvenaud, 2014) is performed, but are poorly scalable. Recently, Gaussian Processes and Artificial Neural Networks have been extensively used in e.g. the design of materials (Bessa et al., 2017; Bessa et al., 2019) and structures (Bessa and Pellegrino, 2018).

The implementations of Random Forests and Gaussian Processes in scikit-learn (Pedregosa et al., 2011), the implementation of XGBoost in xgboost (Chen et al., 2016) and the implementation of Artificial Neural Networks in keras (Chollet et al., 2015) open-source software libraries were used throughout in this work. All the algorithms are summarized hereafter.

5.1. Machine learning algorithms

A Random Forest (Breiman, 2001) is a tree-ensemble technique that combines multiple weak learners (decision trees), each trained with a subset of the training set (this strategy, called *bootstrap aggregation* or *bagging*, improves the stability and accuracy of the trained model), and

Table 1
Material property, geometric and laminate descriptors and their bounds.

	Legacy Quad	Double-Double
Material Property	$E_1 \in [150, 200]$ GPa	$E_1 \in [150, 200]$ GPa
	$X_T \in [2000, 2500]$ MPa	$X_T \in [2000, 2500]$ MPa
	$\mathscr{G}_0 \in [150, 250]$ N/mm	$\mathscr{G}_0 \in [150, 250]$ N/mm
Geometric	$D \in [1, 12]$ mm	$D \in [1, 12]$ mm
	$W/D \in [3, 8]$	$W/D \in [3, 8]$
Layup	$\zeta_1^A \in [-1, 1]$	$\phi \in [0, 90]^\circ$
	$\zeta_2^A \in [-1, 1]$	$\psi \in [0, 90]^\circ$

averages their output in an attempt to produce a strong learner. XGBoost, although also a tree-ensemble technique, is based on *gradient boosting* concept (Friedman, 2001), i.e. the weak learners are trained sequentially in order to predict the error residuals of previous learners. A relevant feature of these tree-based algorithms is that they provide *feature importances*, i.e. they quantify the importance of each parameter for the output prediction.

Gaussian Processes are a Bayesian machine learning method (thus, allow the incorporation of prior knowledge in the learning process) that perform very well for small datasets. In Gaussian Processes each output value is treated as a random variable that follows a Gaussian distribution (Gørtler et al., 2018). The joint distribution of all the output values is also Gaussian (a multivariate Gaussian) and is defined by a mean vector (usually assumed to be zero) and a covariance matrix (Gørtler et al., 2018). The learning process consists in finding the optimal kernel parameters (each component of the covariance matrix is computed based on a kernel function) through maximization of the log-marginal-likelihood (Pedregosa et al., 2011). Being a probabilistic method, this regressor predicts not only a mean output value, but also its variance.

An Artificial Neural Network is a deep learning method consisting on several layers of nodes that perform a (non-linear) operation on their inputs. In the first layer, there are as many neurons as input features and no operation is performed (i.e. the output of these neurons is simply the values of the input features). The last layer contains as many neurons as output variables and its outputs are the model predictions. All the other layers are called *hidden layers* and are intended for learning increasingly meaningful representations of the input data (Chollet, 2017). The learning process consists in learning the *weights* and *biases* of each *neuron* and is usually performed through *backpropagation* (Rumelhart et al., 1986), i.e. the prediction error is propagated backwards in the network.

5.2. Model selection and assessment

In machine learning, a hyperparameter is a parameter whose value controls the learning process. The hyperparameters must ensure that the machine learning model is flexible enough to adapt to the intricacies of the problem (to avoid *underfitting*), but not too flexible to over-adapt to the training set (to avoid *overfitting*). Table 2 shows the possible values established for each hyperparameter (if not mentioned, the defaults of the respective library implementations are used). Due to its simplicity, grid search, a brute-force technique that consists in searching through a manually specified subset of the hyperparameter space of a learning algorithm, is used to optimize the hyperparameters. More automatic approaches, such as Bayesian optimization of the machine learning algorithms (Snoek et al., 2012), could also be used to tune the hyperparameters of the ML algorithms.

The choice of the best hyperparameters for a given algorithm must be performed without access to the test set (Cawley and Talbot, 2010; Varma and Simon, 2006), i.e. the hyperparameter selection must be viewed as an integral part of the learning process (Cawley and Talbot, 2010). This means the obtained generalization error estimate encompasses both the fitted model and the hyperparameter selection strategy. The absence of the test data from the hyperparameter selection procedure ensures no data leakage and, therefore, an unbiased generalization error estimation (Cawley and Talbot, 2010). In this work, *K-fold cross-validation* (Raschka, 2018b), a validation technique that consists in splitting the data into equal-sized k sets, training a model using $k-1$ sets and computing the error estimate using the missing set (and repeat k times), was used. This strategy ensures lower variance estimates, but is computationally intensive. K-fold cross-validation with $k = 5$ was used for model selection. After the selection of the best hyperparameters, the model was retrained using the full training set and a generalization error estimate was computed using the test set.

5.3. Learning curves

The easiest way to improve the performance of an algorithm is to collect more data (although there is normally a plateau, i.e. after a given number of training points the performance does not improve significantly anymore). Nevertheless, gathering data is usually expensive and, since this work intends to serve as a feasibility study, it is relevant to understand how the size of the dataset affects the performance of the model. In order to study such influence, learning curves for Legacy Quad and DD laminates are presented in Figs. 7 and 8, respectively. The root mean squared error (*RMSE*) is chosen as error metric.

For each algorithm and number of points presented in the learning curves, the model selection and assessment procedures described in

Table 2
Grid search hyperparameter values for each algorithm.

Algorithm	Hyperparameter	Search values	DD	Quad
XGBoost	`n_estimators`	100, 500, 1000	1000	1000
	`max_depth`	4, 10	4	10
	`learning_rate`	0.01, 0.1, 0.2	0.1	0.01
	`subsample`	0.5, 1	0.5	0.5
	`colsample_bytree`	0.5, 1	1	1
Random Forest	`n_estimators`	100, 500, 1000, 5000	5000	1000
	`max_depth`	1, 7, 20	20	20
	`min_sample_leaf`	1, 5	1	1
	`max_features`	`auto`, `sqrt`	`auto`	`auto`
Artificial Neural Network	hidden layer 1	8, 16, 64	64	64
	hidden layer 2	8, 16	8	16
	hidden layer 3	8	8	8
Gaussian Processes	Matern kernel (ν)	1/2, 3/2, 5/2	5/2	3/2
	RBF kernel	–	–	–

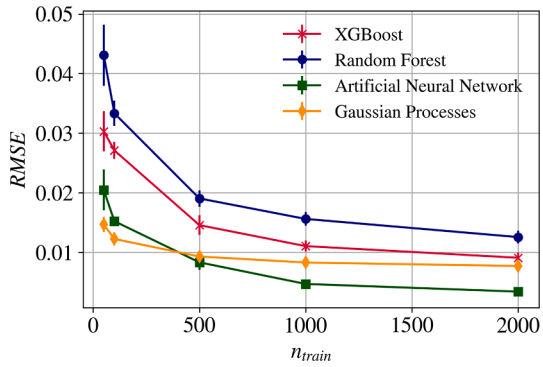

Fig. 7. Legacy Quad laminates: *RMSE* as a function of the size of the training set, n_{train}, for different algorithms.

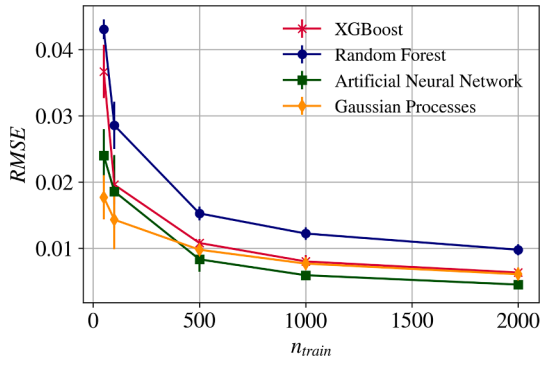

Fig. 8. DD laminates: *RMSE* as a function of the size of the training set, n_{train}, for different algorithms.

Section 5.2 were followed. In order to make the results for different algorithms and number of training points comparable, it was assured all the models were trained using the same points and increasingly larger subsets of the training set and the test set was kept constant.

As shown in Figs. 7 and 8, good overall predictions can be obtained with a reduced number of training points. However, the curves have not yet converged: using a higher number of training points will still translate in a reduction of the generalization error. Moreover, Gaussian processes were able to achieve the best performances for very small number of data points, whereas artificial neural networks outperformed all the algorithms for increasing number of data points (Bessa et al., 2017; Neal, 2012; Williams, 1998). Regarding the tree-ensemble methods, XGBoost demonstrated to be a highly competitive algorithm,

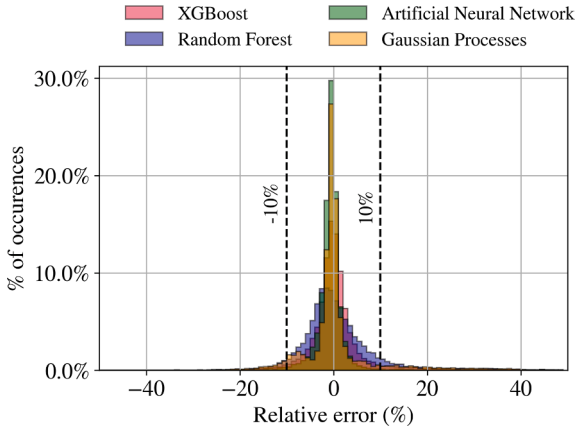

Fig. 9. Legacy Quad laminates: distribution of the relative error for n_{train}=1000 for different algorithms.

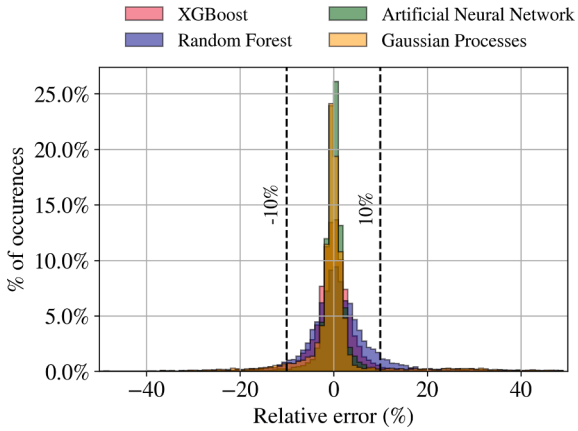

Fig. 10. DD laminates: distribution of the relative error for n_{train}=1000 for different algorithms.

whereas random forests have not achieved satisfactory performances.

5.4. Representations of the design space

Based on the previous results, 1000 training points are considered to be sufficient to accurately capture the overall trend of the design space. This number of data points is also considered to be reasonable for the application of the present framework to more expensive and accurate methodologies for open-hole strength prediction (or related problems). Therefore, several models were trained for this number of training points following the model selection procedure presented in Section 5.2. Afterwards, about 10000 new points were collected and the predictions of the models compared with the ground truth. As shown in Figs. 9 and 10, the obtained relative errors are highly concentrated between the ±10% range.

In the remainder of this section, representations of the design space are shown to assess the ability of the trained models to capture the intricacies of the open-hole strength function as well as its continuity. Here, a sensitivity analysis, where a single parameter is varied ranging from its allowed minimum and maximum, while the remaining six are kept constant, is performed. The results are shown for Legacy Quad and DD laminates in Fig. 11a and b, respectively.

From Fig. 11, it can be concluded that, in fact, all four algorithms are capable of fitting the analytical model with very good accuracy. However, as expected, and given its discontinuous nature, the tree-based models provide a less smooth response. The Gaussian Processes and Artificial Neural Networks provide the smoother prediction curves due to their continuous nature as well as leading to the lower prediction errors and are, therefore, more appropriate to address the present problem.

6. Generation of B-basis allowables using machine learning

In this section, the generation of design allowables, the B-basis allowables (Handbook, 2002), based on machine learning models is explored. The goal is to verify if the machine learning models are able to capture, not only the overall 7D design space, but also the variability associated with a given design point, consequence of the uncertainty related to the material properties and geometrical descriptors.

By taking the variability of the input parameters (material and geometrical) into account, the uncertainty of the input parameters can be propagated to the notched strength, i.e. a statistical distribution of the notched strength can be obtained, which can then be used to compute the statistical design allowables, namely the B-basis allowable. The B-basis allowable is the standard design allowable used in the aeronautical industry for fail safe structures (Handbook, 2002; Spendley, 2012). It is defined as the 95% lower confidence bound on the tenth percentile of a

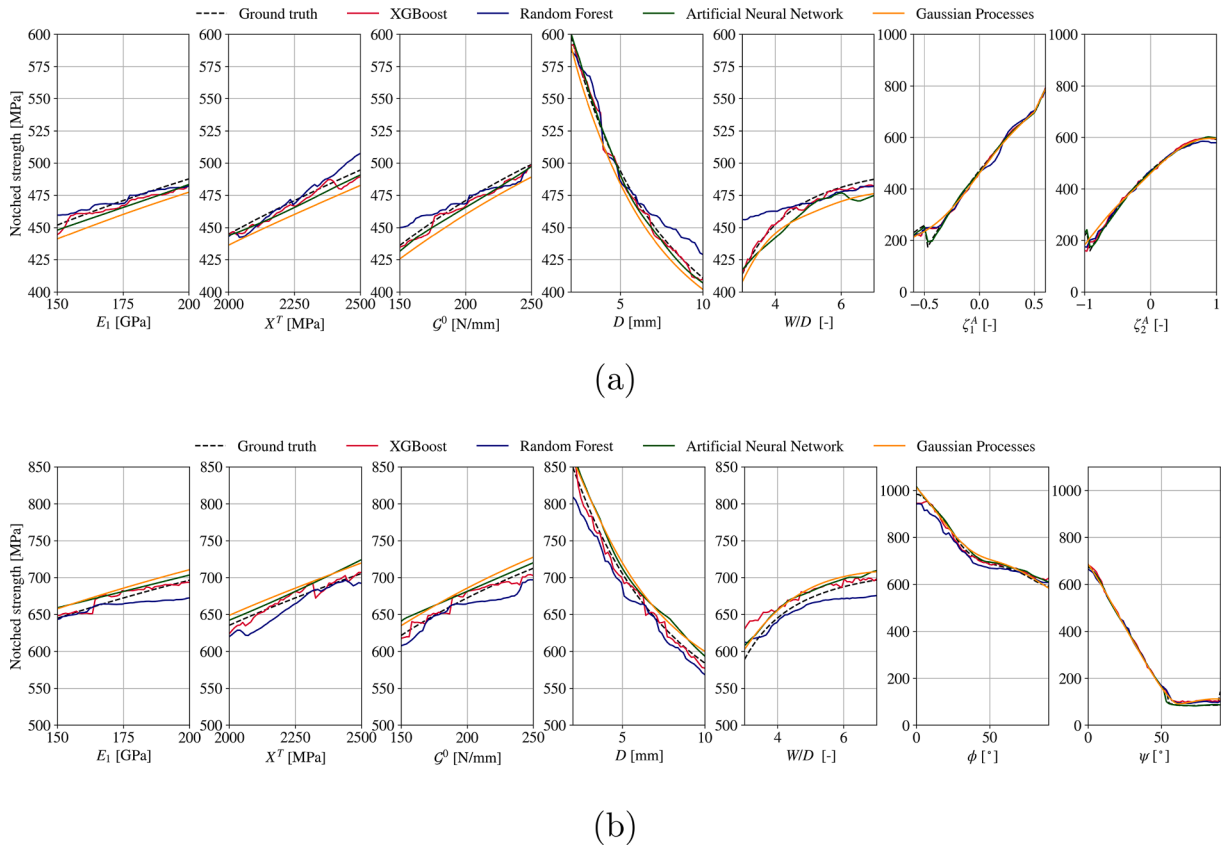

Fig. 11. Sensitivity analysis for a) Legacy Quad and b) DD laminates obtained using the four proposed machine learning models (n_{train}=1000). A single parameter is varied ranging from its allowed minimum and maximum, while the remaining are kept constant at $E_1 = 175$ GPa, $X^T = 2250$ MPa, $\mathcal{G}^0 = 200$ N/mm, $D = 6$ mm, $W/D = 5, \zeta_1^A = 0, \zeta_1^A = 0$ (for QUAD laminates) and $\phi = 60°, \psi = 0°$ (for DD laminates). Note that in the last subfigure, a plateau appears for ψ value above 60°. As the value of ϕ is fixed at 60°, these laminates are very "soft", and present a matrix dominated failure for which the current analytical model predicts very similar behaviours.

specified population of measurements. It is a conservative allowable that ensures with 95% confidence that 90% of the population will have a given property, e.g. strength, higher than the B-basis allowable. Vallmajó et al. (2019) described two methodologies to obtain the B-basis: the CMH-17 approach and a Monte Carlo based approach. The first, is the methodology proposed by the Composite Materials Handbook (Handbook, 2002) and is generally employed for small populations, such as the ones typically obtained experimentally. The second, is a computerized mathematical technique that allows, by repeated, nearly infinite, random sampling of the input parameters, obtaining the distribution of the population of results. Here, given the efficiency of both the baseline analytical model and of the trained machine learning models, the Monte Carlo based approach is used, as described below[4]:

1. 10000 input material ($E_1, X^L, \mathcal{G}_0$) and geometrical (D and W) parameters are generated following a given statistical distribution (normal and uniform distributions, respectively). The lay-up was considered to be fixed, i.e. no laminate rotation/misalignment during the cutting procedure was considered. The baseline analytical model allows accounting for lay-up variability, however the machine learning models do not, since any rotation of the laminate (or any single ply) will yield invalid lay-ups.
2. The distribution of the notched strength is obtained by calculating the notched strength (using the analytical model described in Section 2 and the trained machine learning models described in Section 5).
3. The 10th percentile of the distribution of the notched strengths is obtained and approximated to the B-basis allowable (10th percentile ≈ B-basis Vallmajó et al., 2019).

Table 3
Properties of the IM7/8552 material system (Camanho et al., 2007b) and variability of the geometric parameters.

	E_1 [GPa]	X^T [GPa]	$\mathcal{G}_0$ [N/mm]
Mean	171.42	2323.47	206.75
Standard deviation	2.38	127.45	23.64
	D [mm]	W [mm]	
Mean	D	W	
Tolerance	±2%	±2%	

In a representative example, which allows the comparison of the distribution of the notched strengths, B-basis values and mean values obtained using the baseline analytical model and the trained machine learning models is shown below. An IM7/8552 $[90/0/-45/45]_{3s}$ quasi-isotropic lay-up and a [45/−45/20/−20] DD lay-up, and specimens with hole diameter-to-width ratios of $3 < W/D < 8$ and hole diameters of 2, 4, 6, 8 and 10 mm were considered (Camanho et al., 2007b; Vallmajó et al., 2019). The material parameters were considered to

[4] This process should be repeated N times to obtain the distribution of the 10th percentiles, allowing the calculation of the 5% percentile of the 10th distribution, i.e. the B-basis allowable. However, as shown in Ref. Vallmajó et al. (2019), sample size larger than 10000 are representative of the whole population, and therefore, the 10th percentile can be directly approximated to the B-basis allowable, as the variability is minimal.

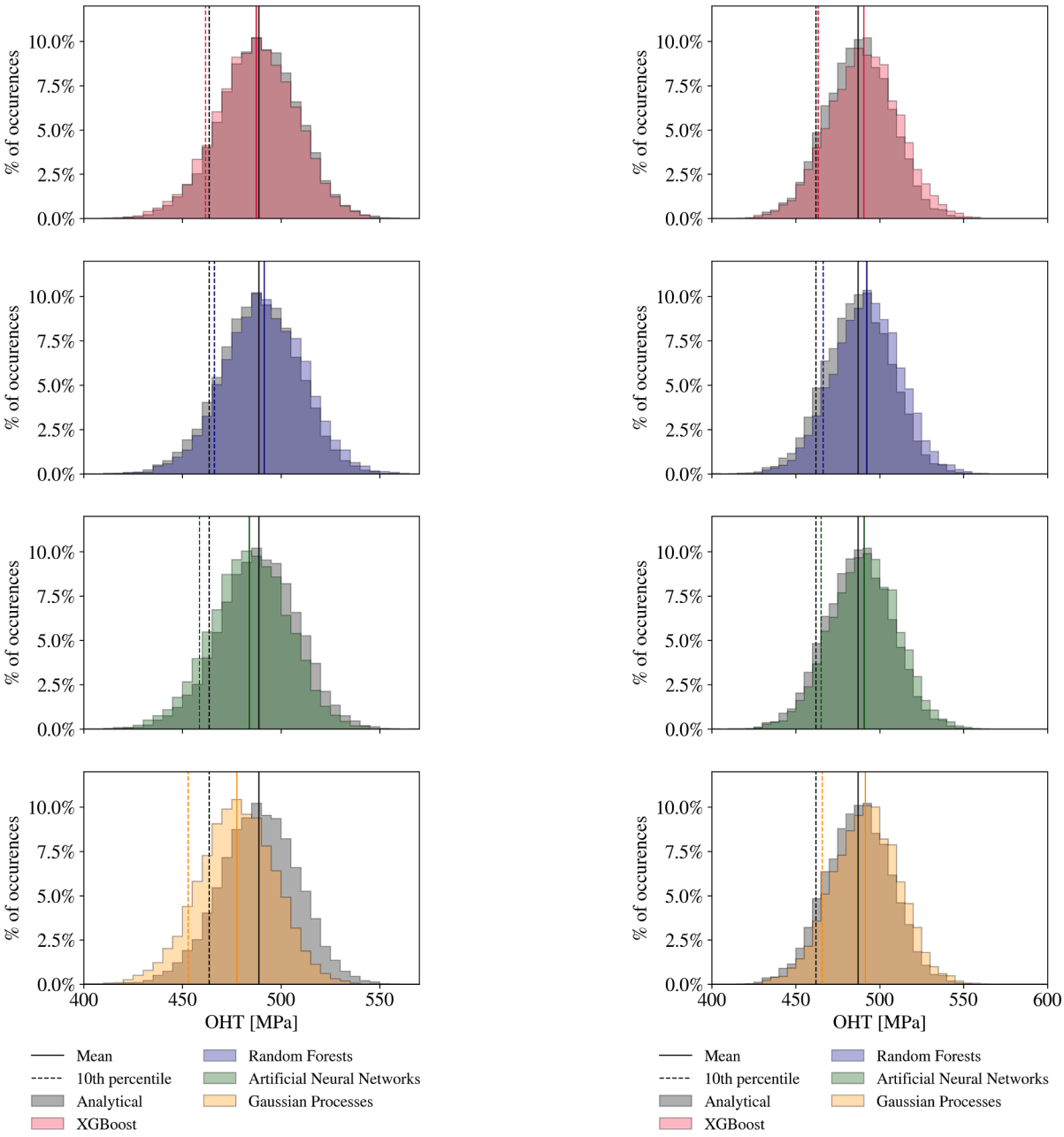

Fig. 12. Legacy Quad: distribution of the notched strengths, B-basis (dashed lines) and mean value (full lines) provided by the analytical model and the trained machine learning models (IM7/8552 $[90/0/-45/45]_{3s}$, $D = 6$ mm and $W = 36$ mm).

Fig. 13. Double-Double: distribution of the notched strengths, B-basis (dashed lines) and mean value (full lines) provided by the analytical model and the trained machine learning models (IM7/8552 DD [45/−45/20/−20] laminate, $D = 6$ mm and $W = 36$ mm).

follow a normal distribution with means and standard deviation shown in Table 3 and the geometrical descriptors were considered to follow an uniform distribution with a tolerance of ±2%, related to tolerances allowed during specimen cutting.

The distributions of the notched strength, B-basis and mean value provided by the analytical model and the trained machine learning models are presented in Figs. 12 and 13 for a specimen with $D = 6$ mm and $W = 36$ mm and the Empirical Cumulative Distribution Function (ECDF) for all the geometries considered ($D = 2$–6 mm and $W = 12$–60 mm) is shown in Figs. 14 and 15, for the Legacy Quad and DD laminate, respectively.

As shown in Figs. 12–15, for both the Legacy Quad and DD lay-up, the XGBoost, Artificial Neural Networks and Gaussian Processes models provide very accurate strength distributions compared to those obtained using the analytical model, allowing an accurate determination of design allowables related to material and geometrical variability. It should be noted that for a different number of training samples, some ML algorithms may lead to more accurate distributions than others. Even though no clear distinction between the accuracy of the models on the determination of B-basis allowables was found, the Gaussian Processes model has the advantage of having a continuous nature and a fast training process, in contrast to the XGBoost and Artificial Neural Network models, respectively. The Gaussian Processes models having i) a fast learning process, ii) less hyperparameters to optimize, iii) having provided a continuous accurate representation of the design space, iv) the lowest relative error and v) good B-basis allowable predictions, and vi) being more powerful model for small training sets, were considered the most interesting and convenient Machine Learning models for this

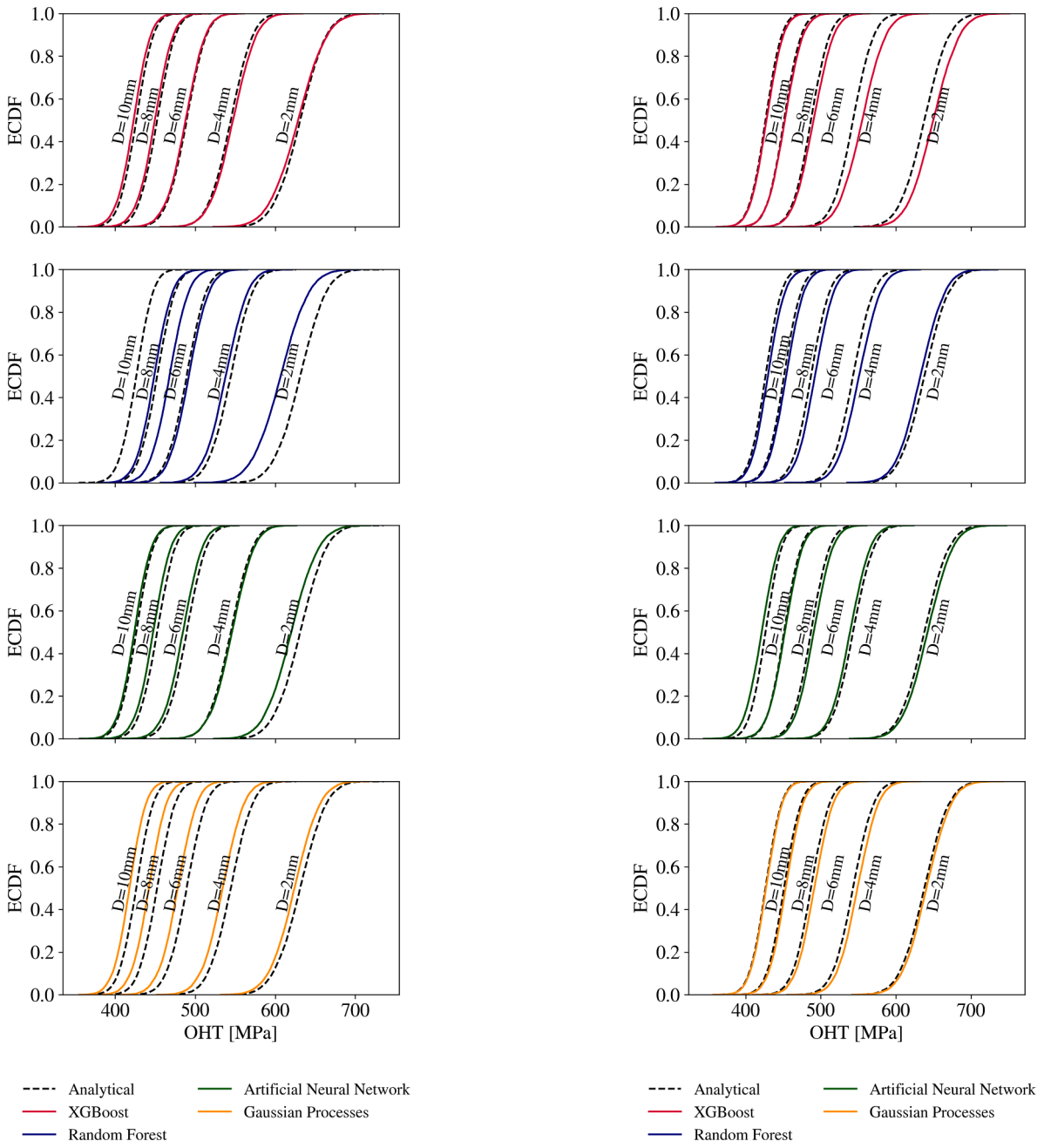

Fig. 14. Legacy Quad: ECDF obtained using the analytical model and the trained machine learning models (IM7/8552 $[90/0/-45/45]_{3s}$, D = 2–10 mm and W = 12–60 mm).

Fig. 15. Double-Double: ECDF obtained using the analytical model and the trained machine learning models (IM7/8552 DD [45/−45/20/−20] laminate, D = 2–10 mm and W = 12–60 mm).

type of analysis. Particularly the fact that Gaussian Process models provide better performances for smaller training sets can be an extremely important characteristic when training models based on numerical data (e.g. from finite element simulations), whose generation is much more time-consuming.

To further test the Gaussian processes models, a design chart, where the diameter-to-width ratio, D/W, is varied between the bounds (1/8 and 1/3) for three hole diameters (D = 2, 4, 10 mm) was generated (Figs. 16 and 17). The mean and B-basis values for each geometry are presented. The Gaussian processes models provide very accurate predictions of both the mean values and B-basis and correctly capture the effect of varying the hole diameter as well as the diameter-to-width ratio (D/W).

Note that the curves diverge slightly near the boundaries (for D/W = 1/8 and D/W = 1/3). This can be explained by the fact that when variability is considered, the generated input parameters can be outside of the training space (see bounds used to train the machine learning models in Table 1). This causes some loss in accuracy in the machine learning algorithms, as they are known to be very powerful as interpolation tools, i.e. working inside the bounds in which they were established, and very ineffective when extrapolating outside the bounds of the design space used for training. This drawback can be circumvented by increasing the design space and retraining the models, which would be trivial using the analytical model, but may otherwise be impossible when dealing with more computationally expensive models or experimental data. This further highlights the need to accurately define the design space used to train the machine learning models and the limitations of such techniques.

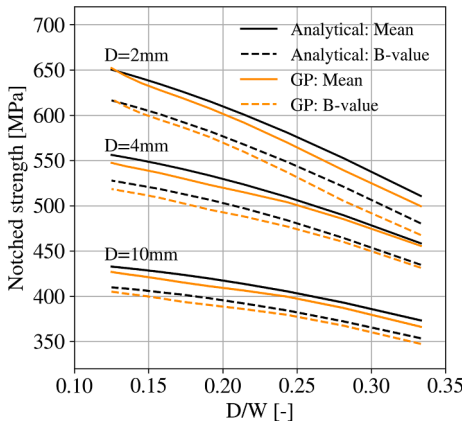

Fig. 16. Design chart for diameter-to-width ratio, $1/8< D/W <1/3$ and hole diameters D = 2,4,10 mm for an IM7/8552 $[90/0/-45/45]_{3s}$ Legacy Quad laminate obtained using the analytical framework and the Gaussian Processes model.

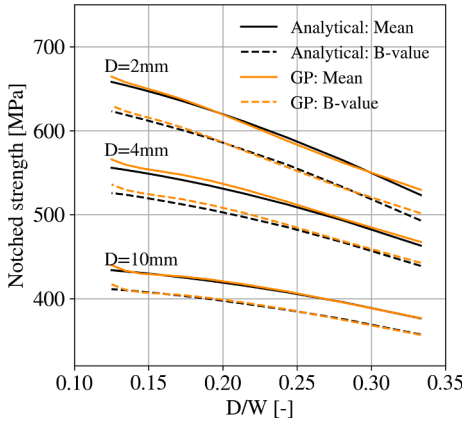

Fig. 17. Design chart for diameter-to-width ratio, $1/8< D/W <1/3$ and hole diameters D = 2,4,10 mm for an IM7/8552 [45/−45/20/−20] DD laminate obtained using the analytical framework and the Gaussian Processes model.

The analysis presented above, where the ability of the trained machine learning algorithms to capture the variability associated with a given design point, consequence of the uncertainty related to the material properties and geometrical descriptors, provides a more clear rationale for model selection. The Gaussian processes algorithm used here provides the more interesting formulation for this particular problem, given the low generalization error obtained for a low number of training points (which provides a general overview of the models performance in the whole design space), their continuous nature (that provide a more accurate representation of the design space) and their ability to capture the distribution of notched strengths and B-basis allowable.

7. Conclusions

This work represents the first step towards the application of machine learning techniques in the prediction of design allowables of composite laminates. First, a discussion on the representation of stacking sequence of composite laminates, based on lamination parameters (for conventional laminates) and on ply angles (for double angle-ply laminates), compatible with machine learning techniques, is presented. The design space of conventional laminates was limited to the Legacy Quad laminates, where only 0/90/±45° plies are considered. This definition is generally compatible with industry requirements, however, the present framework should potentially be expanded for more complex laminates by considering more ply orientations than the baseline 0/90/±45°. The use of two angles that vary continuously (ϕ and $\psi \in [0,90]$) as laminate descriptors of the double angle-ply laminates provides a more comprehensive description of the design space.

Then, the ability of machine learning algorithms to estimate the notched strength of composite laminates based on material, geometric and laminate descriptors is evaluated. To select the most appropriate ML algorithm for the problem at hand, different algorithms (Random Forests, XGBoost, Artificial Neural Networks and Gaussian Processes) were trained on the same datasets. From those analyses, Gaussian Processes were able to achieve the best performances for very small number of data points, whereas Artificial Neural Networks outperformed all the algorithms for increasing number of data points. The continuous nature of these methods provided a more accurate representation of the design space. Even though the tree-ensemble methods have the disadvantage of predicting highly discontinuous responses, XGBoost still proved to be a reliable method.

To further explore the potential and understand the limitations of the trained ML algorithms, the generation of design allowables based on machine learning models is explored to verify their ability to capture not only the overall response of the design space, but also the variability associated with a given design point (consequence of the uncertainty related to the material properties and geometrical descriptors). The models were shown to accurately represent the statistical distribution of open-hole strength, thus giving good estimation for the B-value design allowable. The Gaussian Processes models proved to be the most reliable and convenient Machine Learning models for this type of analysis, considering their i) continuous and accurate representation of the design space, ii) low relative errors of the predictions iii) good B-basis allowable predictions, iv) fast learning process, v) low number of hyperparameters to optimize and vi) better performance for small sized training sets.

While the data-driven framework proposed here is established on data derived from an analytical model, this work serves as the basis to tackle a more demanding future challenge: the prediction of first-ply failure strength, ultimate strength and failure mode of composite materials based on non-linear finite element simulations. This process will comprise a similar sampling, computational analysis and machine learning methodologies to those presented in this work. The envisioned framework is intended to be compatible with more complex test configurations, including in-plane loadings (open-hole tension/compression, plain strength, bolt bearing) and out-of-plane loadings (pull-through and low-velocity impact), providing further reduction of the computational time required to virtually obtain the design allowables for composite laminates.

8. Data availability

The raw data required to reproduce these findings cannot be shared at this time as the data also forms part of an ongoing study.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The first author acknowledges the support of CIT – Financiamento de Base dos centros Interface – Aviso 01/FITEC/2018. The second author acknowledges the support of the Portuguese Government's Fundação para a Ciência e Tecnologia, under the Grant SFRH/BD/146146/2019. The authors are also grateful to Prof. S.W. Tsai (Department of Aeronautics and Astronautics, Stanford University, CA USA) for the inspiring discussions and fruitful collaboration.

References

Tay, T., Tan, S.H.N., Tan, V.B.C., Gosse, J.H., 2005. Damage progression by the element-failure method (EFM) and strain invariant failure theory (SIFT). Composites Science and Technology 65 (6), 935–944.

der Meer, F.P., Oliver, C., Sluys, L.J., 2010. Computational analysis of progressive failure in a notched laminate including shear nonlinearity and fiber failure. Composites Science and Technology 70 (4), 692–700.

Ling, D., Yang, Q., Cox, B., 2009. An augmented finite element method for modeling arbitrary discontinuities in composite materials. International Journal of Fracture 156 (1), 53–73.

Schuecker, C., Pettermann, H.E., 2006. A continuum damage model for fiber reinforced laminates based on ply failure mechanisms. Composite Structures 76 (1), 162–173.

Camanho, P.P., Maim, P., Mayugo, J.A., Dávila, C.G., 2007a. A continuum damage model for composite laminates: Part I - Constitutive model. Mechanics of Materials 39 (10), 897–908.

Vogler, M., Rolfes, R., Camanho, P.P., 2013. Modeling the inelastic deformation and fracture of polymer composites - Part I: Plasticity model. Mechanics of Materials 59, 50–64.

Camanho, P.P., Bessa, M.A., Catalanotti, G., Vogler, M., Rolfes, R., 2013. Modeling the inelastic deformation and fracture of polymer composites–Part II: smeared crack model. Mechanics of Materials 59, 36–49.

Abdi, F., Clarkson, E., Godines, C., DorMohammadi, S., 2016. AB basis allowable test reduction approach and composite generic basis strength values. In: 18th AIAA Non-Deterministic Approaches Conference, p. 951.

Zhang, Y., Schutte, J., Meeker, J., Palliyaguru, U., Kim, N.H., Haftka, R.T., 2017. Predicting B-basis allowable at untested points from experiments and simulations of plates with holes. In: 12th World Congress on Structural and Multidisciplinary Optimization, Braunschweig, Germany.

Abumeri, G., Abdi, F., Raju, K.S., Housner, J., Bohner, R., McCloskey, A., 2011. Cost effective computational approach for generation of polymeric composite material allowables for reduced testing. In: Advances in Composite Materials-Ecodesign and Analysis. InTech.

Spendley, P.R., 2012. Design Allowables for Composite Aerospace Structures. Ph.D. thesis;. University of Surrey.

Bessa, M.A., Bostanabad, R., Liu, Z., Hu, A., Apley, D.W., Brinson, C., et al., 2017. A framework for data-driven analysis of materials under uncertainty: Countering the curse of dimensionality. Computer Methods in Applied Mechanics and Engineering 320, 633–667.

Bessa, M.A., Pellegrino, S., 2018. Design of ultra-thin shell structures in the stochastic post-buckling range using bayesian machine learning and optimization. International Journal of Solids and Structures 139–140, 174–188.

Bessa, M.A., Glowacki, P., Houlder, M., 2019. Bayesian machine learning in metamaterial design: Fragile becomes supercompressible. Advanced Materials 1904845.

Bisagni, C., Lanzi, L., 2002. Post-buckling optimisation of composite stiffened panels using neural networks. Composite Structures 58 (2), 237–247.

Yvonnet, J., He, Q.C., 2007. The reduced model multiscale method (r3m) for the non-linear homogenization of hyperelastic media at finite strains. Journal of Computational Physics 223 (1), 341–368.

El Said, B., Hallett, S.R., 2018. Multiscale surrogate modelling of the elastic response of thick composite structures with embedded defects and features. Composite Structures 200, 781–798.

Balokas, G., Czichon, S., Rolfes, R., 2018. Neural network assisted multiscale analysis for the elastic properties prediction of 3d braided composites under uncertainty. Composite Structures 183, 550–562 (In honor of Prof. Y. Narita).

Yan, S., Zou, X., Ilkhani, M., Jones, A., 2020. An efficient multiscale surrogate modelling framework for composite materials considering progressive damage based on artificial neural networks. Composites Part B: Engineering 194, 108014.

Furtado, C., Arteiro, A., Bessa, M., Wardle, B., Camanho, P.P., 2020. Prediction of size effects in open-hole laminates using only the Young's modulus, the strength, and the R-curve of the 0° ply. Composites Part A: Applied Science and Manufacturing 101.

Vallmajó, O., Cózar, I.R., Furtado, C., Tavares, R., Arteiro, A., Turon, A., et al., 2019. Virtual calculation of the b-value allowables of notched composite laminates. Composite Structures 212, 11–21.

Camanho, P.P., Erçin, G.H., Catalanotti, G., Mahdi, S., Linde, P., 2012. A finite fracture mechanics model for the prediction of the open-hole strength of composite laminates. Composites Part A: Applied Science and Manufacturing 43 (8), 1219–1225.

Tsai, S.W., Melo, J.D.D., 2014. An invariant-based theory of composites. Composites Science and Technology 100, 237–243.

Tsai, S.W., Melo, J.D.D., 2016. A unit circle failure criterion for carbon fiber reinforced polymer composites. Composites Science and Technology 123, 71–78.

Camanho, P.P., Catalanotti, G., 2011. On the relation between the mode I fracture toughness of a composite laminate and that of a 0 ply: Analytical model and experimental validation. Engineering Fracture Mechanics 78 (13), 2535–2546.

Arteiro, A., Sharma, N.D., Melo, J.D., Ha, S.K., Miravete, A., Miyano, Y., et al., 2020. A case for Tsai's modulus, an invariant-based approach to stiffness. Composite Structures, 112683.

Arteiro, A., Pereira, L., Bessa, M., Furtado, C., Camanho, P., 2019. A micro-mechanics perspective to the invariant-based approach to stiffness. Composites Science and Technology 176, 72–80.

Erçin, G.H., Camanho, P.P., Xavier, J., Catalanotti, G., Mahdi, S., Linde, P., 2013. Size effects on the tensile and compressive failure of notched composite laminates. Composite Structures 96, 736–744.

Arteiro, A., Catalanotti, G., Xavier, J., Camanho, P.P., 2013. Notched response of non-crimp fabric thin-ply laminates: Analysis methods. Composites Science and Technology 88, 165–171.

Arteiro, A., Catalanotti, G., Xavier, J., Camanho, P.P., 2014. Large damage capability of non-crimp fabric thin-ply laminates. Composites Part A: Applied Science and Manufacturing 63, 110–122.

Catalanotti, G., Salgado, R.M., Camanho, P.P., 2021. On the Stress Intensity Factor of cracks emanating from circular and elliptical holes in orthotropic plates. Engineering Fracture Mechanics 252, 107805.

Newman Jr., J.C., 1983. A nonlinear fracture mechanics approach to the growth of small cracks. In: Proceedings of the AGARD Conference, vol. 328, no. 6, pp. 1–26.

Suo, Z., Bao, G., Fan, B., Wang, T.C., 1991. Orthotropy rescaling and implications for fracture in composites. International Journal of Solids and Structures 28 (2), 235–248.

Tsai, S.W., Pagano, N.J., 1968. Invariant properties of composite materials. Tech. Rep.; Air force materials lab Wright-Patterson AFB Ohio.

Tsai, S.W., Melo, J.D.D., Sihn, S., Arteiro, A., Rainsberger, R., 2017. Composite Laminates: Theory and practice of analysis, design and automated layup. Stanford Aeronautics & Astronautics.

Grenestedt, J., Gudmundson, P., 1993. Layup optimization of composite material structures. Optimal Design with Advanced Materials 311–336.

Bloomfield, M., Diaconu, C., Weaver, P., 2009. On feasible regions of lamination parameters for lay-up optimization of laminated composites. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 465 (2104), 1123–1143.

Setoodeh, S., Abdalla, M., Gurdal, Z., 2006. Approximate feasible regions for lamination parameters. In: 11th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, p. 6973.

IJsselmuiden, S.T., Abdalla, M.M., Seresta, O., Gürdal, Z., 2009. Multi-step blended stacking sequence design of panel assemblies with buckling constraints. Composites Part B: Engineering 40 (4), 329–336.

Irisarri, F.X., Abdalla, M.M., Gürdal, Z., 2011. Improved shepard's method for the optimization of composite structures. AIAA Journal 49 (12), 2726–2736.

Meddaikar, Y.M., Irisarri, F.X., Abdalla, M.M., 2017. Laminate optimization of blended composite structures using a modified shepard's method and stacking sequence tables. Structural and Multidisciplinary Optimization 55 (2), 535–546.

Bloomfield, M.W., Herencia, J.E., Weaver, P.M., 2010. Analysis and benchmarking of meta-heuristic techniques for lay-up optimization. Computers & Structures 88 (5–6), 272–282.

Todoroki, A., Sekishiro, M., 2007. New iteration fractal branch and bound method for stacking sequence optimizations of multiple laminates. Composite Structures 81 (3), 419–426.

Liu, X., Featherston, C.A., Kennedy, D., 2019. Two-level layup optimization of composite laminate using lamination parameters. Composite Structures 211, 337–350.

Viquerat, A., 2020. A continuation-based method for finding laminated composite stacking sequences. Composite Structures 238, 111872.

Shrivastava, S., Sharma, N., Tsai, S.W., Mohite, P., 2020. D and dd-drop layup optimization of aircraft wing panels under multi-load case design environment. Composite Structures 112518.

Buckingham, E., 1914. On physically similar systems; illustrations of the use of dimensional equations. Physical Review 4 (4), 345.

Sobol, I., 2001. Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. Mathematics and Computers in Simulation 55(1), 271–280 (The Second IMACS Seminar on Monte Carlo Methods).

Saltelli, A., 2002. Making best use of model evaluations to compute sensitivity indices. Computer Physics Communications 145 (2), 280–297.

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., Tarantola, S., 2010. Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index. Computer Physics Communications 181 (2), 259–270.

Herman, J., Usher, W., 2017. Salib: An open-source python library for sensitivity analysis. The Journal of Open Source Software 2.

McKay, M., Beckman, R., Conover, W., 1979. A comparison of three methods for selecting vales of input variables in the analysis of output from a computer code. Technometrics 21, 239–245.

Raschka, S., 2018. Model evaluation, model selection, and algorithm selection in machine learning. arXiv preprint arXiv:181112808.

Hastie, T., Tibshirani, R., Friedman, J., 2013. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York, ISBN 9780387216065.

Krige, D., 1951. A statistical approach to some basic mine valuation problems on the witwatersrand. Journal of the Southern African Institute of Mining and Metallurgy 52 (6), 119–139.

Rosenblatt, F., 1958. The perceptron: A probabilistic model for information storage and organization in the brain 65, 386–408.

Breiman, L., 2001. Random forests. Machine Learning 45 (1), 5–32.

Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794.

Chollet, F., 2017. Deep Learning with Python. Manning Publications Company. ISBN 9781617294433.

Hamidieh, K., 2018. A data-driven statistical model for predicting the critical temperature of a superconductor. Computational Materials Science 154, 346–354.

Stanev, V., Oses, C., Kusne, A.G., Rodriguez, E., Paglione, J., Curtarolo, S., et al., 2018. Machine learning modeling of superconducting critical temperature. npj Computational Materials 4(1), 29.

Rasmussen, C., Williams, C., 2006. Gaussian Processes for Machine Learning. University Press Group Limited. ISBN 9780262182539.

Duvenaud, D., 2014. Automatic model construction with gaussian processes.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al., 2011. Scikit-learn: Machine learning in python. Journal of Machine Learning Research 12.
Chollet, F., et al., 2015. Keras. https://keras.io.
Friedman, J.H., 2001. Greedy function approximation: A gradient boosting machine. The Annals of Statistics 29 (5), 1189–1232.
Gørtler, J., Kehlbeck, R., Deussen, O., 2018. A visual exploration of gaussian processes. In: Proceedings of the Workshop on Visualization for AI Explainability 2018 (VISxAI).
Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 323 (6088), 533–536.
Snoek, J., Larochelle, H., Adams, R.P., 2012. Practical bayesian optimization of machine learning algorithms.
Cawley, G., Talbot, N., 2010. On over-fitting in model selection and subsequent selection bias in performance evaluation. Journal of Machine Learning Research 11, 2079–2107.
Varma, S., Simon, R., 2006. Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics 7, 91.
Raschka, S., 2018. Model evaluation, model selection, and algorithm selection in machine learning. arXiv preprint arXiv:181112808 2018b;.
Neal, R.M., 2012. Bayesian Learning for Neural Networks, vol. 118. Springer Science & Business Media.
Williams, C.K., 1998. Computation with infinite neural networks. Neural Computation 10 (5), 1203–1216.
Handbook, M., 2002. Mil-hdbk-17-1f: Composite materials handbook, volume 1-polymer matrix composites guidelines for characterization of structural materials.
Camanho, P.P., Maimí, P., Dávila, C.G., 2007b. Prediction of size effects in notched laminates using continuum damage mechanics. Composites Science and Technology 67 (13), 2715–2727.