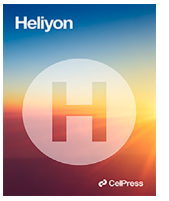
Heliyon 10 (2024) e29764

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Iterative experimental design and identifiability analysis of composite material failure models

Ádám Ipkovich [a], Alex Kummer [a,*], László Kovács [b], Balázs Fodor [c], János Abonyi [a]

[a] ELKH-PE Complex Systems Monitoring Research Group, University of Pannonia, Egyetem u. 10, H-8200 Veszprém, Hungary
[b] eCon Engineering Kft., Kondorosi Str., Budapest, H-1116, Hungary
[c] BMW Group, Research and Innovation Center, Knorrstraße 147, 80788 Munich, Germany

ARTICLE INFO

Keywords:
Fisher information matrix
Model identification
Composite failure models
Design of experiments
Iterative parameter optimization

ABSTRACT

The parameter identification of failure models for composite plies can be cumbersome, due to multiple effects as the consequence of brittle fracture. Our work proposes an iterative, nonlinear design of experiments (DoE) approach that finds the most informative experimental data to identify the parameters of the Tsai-Wu, Tsai-Hill, Hoffman, Hashin, max stress and Puck failure models. Depending on the data, the models perform differently, therefore, the parameter identification is validated by the Euclidean distance of the measured points to the closest ones on the nominal surface. The resulting errors provide a base for the ranking of the models, which helps to select the best fitting. Following the validation, the sensitivity of the best model is calculated by partial differentiation, and a theoretical surface is generated. Lastly, an iterative design of the experiments is implemented to select the optimal set of experiments from which the parameters can be identified from the least data by minimizing the fitting error. In this way, the number of experiments required for the identification of a model of a composite material can be significantly reduced. We demonstrate how the proposed method selected the most optimal experiments out of generated data. The results indicate that if the dataset contains enough information, the method is robust and accurate. If the data set lacks the necessary information, novel material tests can be proposed based on the optimal points of the parameters' sensitivity of the generated failure model surface.

1. Introduction

More and more industries are being penetrated by composite materials. The increasing popularity can be credited to their high strength-to-weight ratio. Gradually, the subject gained importance and immense efforts were made to understand the failure behavior of composite products. All composite materials exert anisotropy and complex fracturing mechanisms with various modes, which increase the difficulty of composite material failure analysis [1]. Failure and strength analysis is a fundamental step in the design process to determine the risk of failure and provide guidance on mitigating its occurrence. Some examples include wind turbine

* Corresponding author.
E-mail addresses: ipkovich.adam@mk.uni-pannon.hu (Á. Ipkovich), kummer.alex@mk.uni-pannon.hu (A. Kummer), laszlo.kovacs@econengineering.com (L. Kovács), Balazs.Fodor@bmw.de (B. Fodor), janos@abonyilab.com (J. Abonyi).
URL: https://www.abonyilab.com/ (J. Abonyi).

https://doi.org/10.1016/j.heliyon.2024.e29764
Received 28 November 2023; Received in revised form 11 April 2024; Accepted 15 April 2024

Available online 23 April 2024

strength analysis [2], composite joints modeling [3], 3D simulation of metal bearings failure [4], plywood bending [5], glass fiber drilling [6], *etc*.

The behavior of the composite materials is modeled with several orthotropic failure models, most of which are often available in commercial Finite Element Method (FEM) software packages. The most popular are the Max Stress [7], Tsai-Wu [8], Hoffman [9], Tsai-Hill [10], Hashin [11], and Puck [12] failure criteria. In our research, we investigated classical failure models for a shell-like composite structure, represented by orthotropic constitutive behavior with translaminar (*e.g.* fiber cleavage, buckling, and shearing) and intralaminar failure modes (through-thickness failures).

Selecting the best model for a specific application depends on the circumstances, however, before making the decision, one should always consider the available validation data set and the required quality of prediction. It is not the most complicated model the engineer should look for, but rather the reliable and best validated one. Even using a rudimentary and conservative failure model calibrated on a proper data set may provide a more reliable estimation rather than an underrepresented complex model. Our study focuses only on failure models defined in the stress state space, however, from a model identification point of view, confining our investigation does not influence the practicability. Nevertheless, in some cases, the usage of strain related models can be even more purposeful. The choice of failure model best suiting for purpose in our concept means that it satisfies two major conditions: the available test data contains information about the model parameters (identifiability) and the model can be fitted in a robust way.

Machine learning has already been used in material model identification. The Tsai-Wu model was employed to simulate unidirectional composite lamina, which was optimized by a neural network [13]. Soft computing has also been used in material fatigue estimation [14]. [15] proposed a Micromechanics-based approach for Static Failure (M-SaF) and Bayesian statistical framework to predict the failure in heterogeneous materials. The identification of the model parameters is performed using Bayesian Inference method. We see that there is some work in the literature where global optimization algorithms (e.g. genetic algorithms) are used to determine the model parameters, but no work can be found where the identification task is supported by experimental design. Despite the fact, that Fisher Information Matrix (FIM) has been successfully applied in parameter optimization in many cases [16], the parameter estimation of material failure models with the design of experiments is unknown in the literature, hence we propose to use FIM-based optimal design to support the model identification problem.

The identification of the model parameters is not straightforward. The most common practice is to perform experiments, provoking characteristic stress-strain states, exploring the failure characteristic along a few radial trajectories of the failure envelope and use these test data to calibrate numerical prediction models. Accurate identification requires informative experiments contributing to the holistic description of material failure parameters. Different models are expressed by different parameter sets that can be characterized with dissimilar experiments. Inaccurate test data inevitably adds error and uncertainty, mainly from limited knowledge of the material behavior, indirect phase space (stress space), scatter, and noise in the experimental data [17]. Therefore, our goal is to a) find the right failure model with the best possible fitting quality under the given data set and b) find the most informative material tests for the model fitting. This can lead to a minimum test matrix to be defined for any pre-selected failure model that provides the most information about all parameters and thus, results in robust and reliable model fit including uncertainty quantification and survival probability dependent design.

For a robust and fast algorithm development, we used virtual tests to generate the data set to develop the methodology. It gave us the opportunity to directly validate the methodology by knowing the real sought parameters upfront. Moreover, extending the nominal virtual tests with systematic and unsystematic scatter could give us valuable feedback about the behavior of the algorithm and quantify its robustness. However, in a future work, the algorithm should be tested using real test data.

Identifiability analysis is closely related to sensitivity analysis. If the objective function does not vary with a model parameter, it follows that the parameter in question cannot be readily estimated from the available tests [18]. The set of experiments must contain information about all the model parameters; otherwise, some of the model parameters may not be identifiable, leading to a false model prediction. To reduce the number of experiments and retain accuracy, we propose the application of experimental design [19]. It allows us not only to consider identifiability but also to perform the most informative experiments possible. The advantage is that only a set of predefined (minimum) experiments are required, improving the time and cost-efficiency of data-driven modeling. The use of the optimal design of experiments (DoE) is recommended as not only the selection of the most informative experiments but ignoring bad/inconsistent or redundant data is also possible with its help [20].

In this work, the identifiability analysis is based on the trace of the FIM, performed in local points of the ply failure stress state, along with optimal designs criteria (*i.e.* A-, E- and D-optimality) [21]. The most informative experiments are chosen based on their A-optimality, as the contribution of all experiments can be linearly decomposed into the contribution of the individual experiments. The key idea is to accurately identify the model that best describes the material with the least amount of data. The best model is selected based on how the measured data fits the failure surface generated by the identified parameters, and the best one is carried over to the identification process.

FIM-based sensitivity analysis is proposed to investigate the identifiability of the parameters. We propose an iterative experimental design and parameter fitting method to identify model parameters. In each iteration, the experiments are selected based on their contribution to identifying the model parameters, and the iteration stops if the model performance improvement becomes insignificant. Moreover, if the characteristic behavior of the composite ply failure models is required for a certain material, the experimental data points must contain information on all parameters. However, this is seldom the case in the industry environment, and the available data set should define which model can be calibrated best for the specific application.

Various experimental data sets have been tested with the models. However, a particular model may dominate the other in the same material with a different condition. For example, the stack-up design, amount of layers, and their orientation may alter the governing failure model [22]. Therefore, there is a need to select the model that best fits the material behavior and its nominal

values. In most cases, the model parameters are determined by specific experiments, *i.e.* the aim is to determine a model parameter as precisely as possible. However, some parameters cannot be measured from simple experiments, and an optimization algorithm is needed.

Hence, our work provides the following contribution:

- We present a compact methodology to decide which composite failure model should be used for simulation based on the available experimental data. As such, the failure models can be systematically analyzed and evaluated in terms of model accuracy and model parameter identifiability (see Section 2);
- We evaluate the performance of the failure models to select the best one based on the Euclidean distance of the experimental points from the failure surface estimated by data reconciliation (see Section 2.1);
- We performed a Fisher Information Matrix-based sensitivity analysis to investigate the identifiability of model parameters using experimental data (see Section 2.2);
- We select the most informative subsets of the experimental data using a mixed integer nonlinear programming experimental design, and propose experiments that can be measured by a simple experiment resulting in a state of uniaxial failure stress state (if possible) (see Section 2.3);
- We prove the performance of the proposed method using an experimental data set generated algorithmically. The visualization of the sensitive regions of the failure surface is presented with the recommended and most informative experiments for parameter identification. The feasibility of the parameter identification algorithm is demonstrated in Section 3.

The structure of the paper is as follows:

The theoretical background for the iterative identifiability analysis of composite material failure models is introduced in Section 2. Section 2.1 describes the generalized form of the failure models, the experiments, and parameters, whereas Section 2.2 includes the tools, such as the Fisher Information Matrix and optimal design. The selection problem is described in Section 2.3, and the algorithm is detailed in Section 2.4. The results are presented in Section 3, where both the analysis of the models and the iterative DoE are presented on algorithmically generated data. In Section 4 concluding remarks are given. An Appendix is also provided for the failure models (Appendix A), shapes of the failure surface (Appendix B and C).

2. Experimental design-based analysis of failure models

During the introduction of the methodology of the novel approach, first, the identification problem is formalized in Section 2.1, which is followed by the theoretical background of optimal selection (Section 2.2), including the sensitivity, Fisher Information Matrix, and optimal design. The selection problem is defined in Section 2.3, and the algorithm is described in Section 2.4.

2.1. Formulation of the identification problem

The general formula of 1st-ply-failure model describes material breakage with failure stresses and specific parameters. The goal of the models is to approximate material failure through parallel, perpendicular (to fiber direction), and shear stress. Damage occurs when the failure model value (Inverse Reserve Factor, Failure Index) reaches one, calculated from the actual stress state if failure tests and specific parameters present accurate values. As various failure models are available, the particular models are denoted by f_m, where m represents the index of a failure model. In addition, each model bolsters a set of parameters, some of which are unique to a specific physical phenomenon. The general notation for a failure model is as follows:

$$f_m\left(\mathbf{x}, \theta^{(m)}\right) = 1 \tag{1}$$

$$\mathbf{x} = \left[\sigma_1, \sigma_2, \tau_{12}\right] \tag{2}$$

where $\theta^{(m)}$ is considered the parameter vector of the failure model f_m, which equals to one when the material would break; therefore, $f_m(\mathbf{x}, \theta^m) = 1$ denotes the marginal state of the failure, the fracture of fiber or matrix in the composite material. Eq. (2) introduces $\mathbf{x}$ that is the vector of state variables that consists of three stress components: normal in-plane stress parallel (σ_1) and perpendicular to (σ_2) the fiber direction, and shear stress (τ_{12}).

Parameters are imperative to correctly predict breakage and are often unknown. A model is identified through the optimization of its parameters, which is formalized in Eq. (3):

$$\theta^{(m)} = \left[\theta_1^{(m)}, \theta_2^{(m)}, \ldots, \theta_n^{(m)}\right] \tag{3}$$

where $\theta^{(m)}$ defines the parameter vector with length n.

The Tsai-Wu model is considered one of the most popular failure criteria [8]. The model consists of a single continuous function and six parameters. In the case of the Tsai-Wu failure model, the parameter vector is defined as:

$$\theta^{(\text{Tsai-Wu})} = \left[X_t, X_c, Y_t, Y_c, S_{12}, F_{12}\right] \tag{4}$$

where $\theta^{(\text{Tsai-Wu})}$ denotes the parameter vector of the Tsai-Wu model, X_t and X_c stand for the tensile and compressive strength in the direction of the fibers, respectively, Y_t and Y_c describe the tensile and compressive strength perpendicular to the fiber direction, whereas S_{12} is the pure shear strength. F_{12} is a parameter unique to the Tsai-Wu model.

As such, the model is formulated as in Eq. (1) with its parameters defined in Eq (4), resulting in Eq (5):

$$f_{\text{Tsai-Wu}}\left(\mathbf{x}_k, \theta^{(\text{Tsai-Wu})}\right) = 1 = \frac{{\tau_{12}}^2}{{S_{12}}^2} - \sigma_2\left(\frac{1}{Y_c} - \frac{1}{Y_t}\right) - \sigma_1\left(\frac{1}{X_c} - \frac{1}{X_t}\right) + \frac{{\sigma_1}^2}{X_c X_t} + \frac{{\sigma_2}^2}{Y_c Y_t} + \frac{2 F_{12} \sigma_1 \sigma_2}{\sqrt{X_c X_t Y_c Y_t}} \tag{5}$$

All other models incorporate $[X_t, X_c, Y_t, Y_c, S_{12}]$ parameters, the intersection points of the axes with the failure envelope representation in the in-plane stress space. Some models, such as the Puck, Hashin, and Tsai-Wu failure criterion, also contain unique parameters. The Max Stress (Eq. (22) and (23)), Hoffman, Hashin, Tsai-Hill, and Puck failure models are described in Appendix A.

Model parameters require fine-tuning to identify the highest accuracy model. The goal is to estimate the parameters as accurately as possible. N_p number of experiments is available that contains as much information as possible while keeping the number at a minimum so that the parameters are efficiently identified. One realization of failure (experiment) takes the form of $\mathbf{x}_k$, which consists of the principal stresses $\left[\sigma_{1,k}, \sigma_{2,k}, \tau_{12,k}\right]$. The experiments are organized into a matrix $\mathbf{X}^{(N_p)}$, and the optimized parameters are identified based on the information content of available tests. As such, the model errors stemming from the parameters are minimized based on their difference from the breakage status ($f_m = 1$):

$$\mathbf{X}^{(N_p)} = \left[\mathbf{x}_1^T, \ldots, \mathbf{x}_{N_p}^T\right] \tag{6}$$

$$\epsilon = \frac{1}{N_p}\sum_{k=1}^{N_p}\left(f_m(\mathbf{x}_k, \theta^{(m)}) - 1\right)^2;\ k = 1, \ldots, N_p \tag{7}$$

$$\theta^{(m,0)} = \arg\min_{\theta^{(m)}} \epsilon \tag{8}$$

where $\theta^{(m,0)}$ in Eq. (8) denotes the optimized parameters of the model, $\mathbf{x}_k$ in Eq. (7) stands for the kth experiment $\mathbf{x}_k = \left[\sigma_{1,k}, \sigma_{2,k}, \tau_{12,k}\right]$ from matrix $\mathbf{X}^{(N_p)}$, with $k = 1, \ldots, N_p$ (in Eq. (6)) denoting the running index of the experiments. ϵ defines the mean squared error of the predicted model failure to breakage.

Although the mean squared error of a model provides information on how well it fits, these error values cannot be compared between models as these denote different functionalities in behavior. Therefore, a robust measure is required to evaluate the performance of the models. The quality of the fitting can be evaluated by calculating the Euclidean distances of the experiments and their corresponding value during breakage returned by the nominal failure model. The values of the experiments and projected ones do differ: most test results contain measurement errors that heavily depend on the actual state of the material, and also the material strength has a natural variance. The projection is also known as data reconciliation [23], the goal of which is to calculate the Euclidean distance of the experiments to the closest point on the fitted theoretical model while simultaneously projecting the data. The method, described in Eq. (9), resembles the least squares method [24]:

$$\begin{aligned} &\arg\min_{\hat{\mathbf{x}}_k}\ e_k = \left(\mathbf{x}_k - \hat{\mathbf{x}}_k\right)^T \mathbf{V}\left(\mathbf{x}_k - \hat{\mathbf{x}}_k\right) \\ &\quad \text{s.t.} \qquad f_m(\hat{\mathbf{x}}_k, \theta^{(m,0)}) = 1 \end{aligned} \tag{9}$$

where e_k is considered as the distance between the experiments and the closest failure model value, $\mathbf{x}_k$ denotes the stress state on failure, $\hat{\mathbf{x}}_k$ is considered as a projected point on the mth model with the nominal parameters $\theta^{(m,0)}$ and $\mathbf{V}$ describes the weight matrix. The function minimizes the distance value of the experiments to the nominal failure surface, whose median indicates the performance of the identified model.

Additionally, the argmin calculates the closest projected point to the experiment. Fig. 1 depicts the individual Euclidean distance error (e_k) of the kth experiment. The Euclidean distance between the experiments and the identified failure model is minimized, while simultaneously, the experiment is reconciled to the optimized failure model and added to the set of errors $E^{(m)}$, whose median defines the fitness of the nominal failure.

Performing failure tests are time-consuming and costly, and the number of available experiments is often insufficient. Yet, if enough points are selected to identify the parameters, the variable stress space can be reproduced and examined to propose the necessary experiments that may help in the identification process. Thus, a failure surface is generated to present which part of the failure surface contributes the most to parameter identification. The potential experiments ($\mathbf{X}^{(N_s)}$) with N_s amount of experiments are sampled from the failure surface, which also contains the optimal experiments ($\mathbf{X}^{(N_p)}$). The number of potential experiments must be one order of magnitude more than the optimal amount ($N_s >> N_p$). The failure surface may be generated with a grid-like allocation of two failure stress states $\mathbf{x'}_k$, in this case, the $[\sigma_1, \sigma_2]$ variable pairs. The last one (τ_{12}) is derived analytically from the model equations of the failure criteria (see Appendix A) so that the substitution into the model results in a one ($f_m(\mathbf{x}_k, \theta^{(m)} = 1)$).

There are several advantages of generating a failure surface, it is possible to:

- Determine the most informative experiments without having to conduct more failure tests.
- Visualize the failure of the composite material in any eligible state.
- Calculate the information content of the potential experiments for the entire surface.

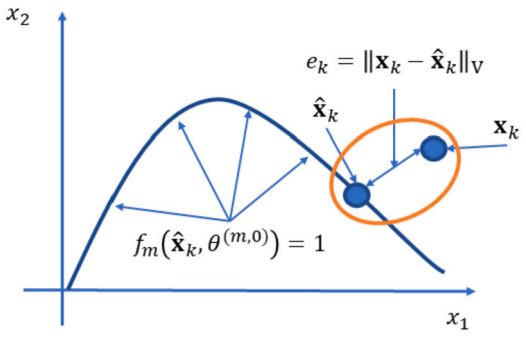


Fig. 1. Data reconciliation-based evaluation of the failure surfaces. The fitness of failure surface (f_m) is evaluated by measuring distances between the original experiment $\mathbf{x}_k$ and the closest experiments $\hat{\mathbf{x}}_k$ on the nominal failure model ($f_m(\hat{\mathbf{x}}_k, \theta^{(m,0)})$).

Numerous theoretical composite ply failure models have been developed based on different continuum mechanical assumptions that are not always consistent. The failure surfaces heavily depend on the parameter identification process. Moreover, the experiments include errors that change with the dataset, so two datasets of the same material may provide different nominal parameters. As such, there is a need to measure how an experiment may contribute to the identification of a model. Performing sensitivity analysis and experiment design help describe the information content of the failure surfaces.

2.2. Experimental design-based identifiability analysis

Any experiment may provide a certain contribution to the identifiability of the parameters. As the models are based on physical behavior, each parameter has its particular meaning *e.g.* the Tsai-Hill model has five elementary parameters that can be best identifiable using the corresponding unique stress components. Each experiment has a contribution to the parameter identification process, but that value depends on where the point is located in the stress space. Some parameters may require a multiaxial stress state (*e.g.* the Tsai-Wu coefficient), and a group of parameters may be simultaneously identifiable with only one specific experiment. One should calculate the contribution of each parameter so that the set of optimal solutions can be determined.

The local sensitivity $\mathbf{s}_k$ of the *kth* experiment in the stress state space is an excellent measure for the contribution of the point, which can be calculated by partially differentiating the model by the parameter vector:

$$\mathbf{s}_k = \left.\frac{\partial f_m(\mathbf{x}_k, \theta)}{\partial \theta}\right|_{\theta=\theta^{(m,0)}} = \left[\frac{\partial f_m(\mathbf{x}_k, \theta)}{\partial \theta_1}, \ldots, \frac{\partial f_m(\mathbf{x}_k, \theta)}{\partial \theta_n}\right]^T ; k = 1, \ldots, N_p \tag{10}$$

where $\mathbf{s}_k$ denotes the sensitivity of the parameters regarding the *kth* experiment using the *mth* model with the nominal parameters $\theta^{(m,0)}$.

Also, the sensitivities of the experiments can be organized into a matrix, as shown in Eq. (11):

$$\mathbf{S}^{(N_p)} = [\mathbf{s}_1, \ldots, \mathbf{s}_{N_p}] \tag{11}$$

where $\mathbf{S}^{(N_p)}$ stands for the sensitivity matrix of $\mathbf{X}^{N_p}$ optimal experiments.

Although the sensitivities provide information separately on the contributions to the parameters, standard tools are available to calculate their aggregated information content. Aggregation is performed with optimal design. The Fisher Information Matrix (FIM) functions as the basic structure for the aggregation: it presents the possible information content of a set of experiments for both individual and joint sensitivities [21]. The FIM of the *kth* experiment ($\hat{\mathbf{F}}_k\left(\theta^{(m)}\right)$) is calculated by the matrix product of the parameter sensitivity vector found in Eq (10), a user-defined orthogonal matrix $\mathbf{W}_{\mathbf{n}\times\mathbf{n}}$ that defines which parameters are the priority, and the transpose of the sensitivity vector. The FIM ($\hat{\mathbf{F}}\left(\theta^{(m)}\right)$) of the set of experiments is considered the sum of the individual FIMs, which can be formalized as in Eq. (12):

$$\begin{aligned}\hat{\mathbf{F}}\left(\theta^{(m)}\right) &= \sum_{k=1}^{N_p} \hat{\mathbf{F}}_k\left(\theta^{(m)}\right) = \sum_{k=1}^{N_p} \mathbf{s}_k^T \mathbf{W} \mathbf{s}_k = \\ &= \sum_{k=1}^{N_p} \left.\frac{\partial f_m\left(\mathbf{x}_k, \theta^{(m)}\right)}{\partial \theta^{(m)}}\right|_{\theta^{(m)}=\theta^{(m,0)}}^T \mathbf{W} \left.\frac{\partial f_m\left(\mathbf{x}_k, \theta^{(m)}\right)}{\partial \theta^{(m)}}\right|_{\theta^{(m)}=\theta^{(m,0)}}\end{aligned} \tag{12}$$

where $\hat{\mathbf{F}}(\theta^{(m)})$ denotes the Fisher Information Matrix of a set, $\hat{\mathbf{F}}_k(\theta^{(m)})$ stands for the Fisher Information of the *kth* experiment, $\mathbf{s}_k$ describes the sensitivity of an experiment, while $\mathbf{W}$ represents an orthogonal weight matrix. The priority of the model parameters can be varied so that more information is extracted by modifying the weight values of the $\mathbf{W}$ matrix. In this work, equal priorities are used. However, weights can be chosen to put a stronger emphasis on model parameters when the uncertainty is greater.

Optimal design finds the most informative points in the set of available experiments. Each design is connected to minimizing the variance of the parameters by optimizing the FIM. Several approaches exist, such as the E-, D-, and A-optimality [25]; each requires

the FIM matrix to qualify the experiment. E-optimality maximizes the minimum eigenvalues of the FIM, while D-optimality maximizes the determinant. A-optimality maximizes the sum of the diagonal. The advantage of A-optimality is that it can be decomposed into the individual A-optimality values of the experiments. As such, Eq. (13) introduces the individual contribution of the experiments as a simple addition:

$$\hat{A} = tr(\hat{\mathbf{F}}(\theta^{(m)})) = \sum_{k=1}^{N_p} \sum_{i=1}^{n} s_{k,i}^2 \tag{13}$$

where $\hat{A}$ denotes the A-optimality of the set of experiments, k denotes the index of the experiment, i stands for the index of the parameter. The total A-optimality value is calculated as the trace of the FIM, in other words, the sum of the diagonal.

If the parameters, the models, and two of the stress variable vectors are known (or generated), it is possible to calculate the information content of the failure surface and determine its sensitivity concerning the parameters. Also, the theoretical maximum of the parameter sensitivities can be defined, which can be used to normalize the sensitivity values. By understanding how the experiments contribute to the identification process, we can select the optimal ones and recommend them to determine and provide pointers on the behavior of the material.

2.3. Clustering and linear programming-based optimal experiment selection

The goal of this section is twofold, the first one is to define the optimal set of measurement set (N_p) for parameter identification from N experiments (see Section 2.3.1). In the case when a model parameter cannot be identified we present a methodology to propose experiment points to generate a measurement set from N_s nominal surface stress points (see Section 2.3.2).

2.3.1. Linear programming-based optimal experiment selection – iterative design of experiments

In order to determine the parameters properly, an N_p amount of optimal experiments should be chosen from the N_s potential experiments. As newly added experiments may not contain additional information, the goodness of an individual point must be established so that only the necessary optimal ones are selected. The A-optimality provides the required information on the contribution of the experiments; thus, the remaining problem is to select the ones that contribute the most to each parameter. One of the advantages of the Fisher Information Matrix is that the contributions to the identification of each model parameter can be defined and investigated individually. However, A-optimality has an advantage over the FIM: the parameter identification can be solved as a linear programming problem due to the additive nature of the A-optimality. Note that the identifiability of the parameters must be ensured, for which several methods have been developed [26]. Therefore, we assume that the set of potential experiments N_s contains N_p optimal experiments that provide enough information to identify the parameters.

The selection of the most informative points requires an optimization algorithm, which is a nonlinear mixed-integer solver [27]. A binary integer y_k is assigned to each experiment determining whether the experiment is chosen. N_p represents the maximum number of experiments to be selected, so the sum of y_k must not exceed the total number of experiments. This way, it is ensured that only the N_p most informative points are selected. The optimization problem itself is to maximize the total A-optimality ($\sum_{k=1}^{p} y_k \sum_{i=1}^{n} u_{k,i} s_{k,i}^2$) of the optimal set of experiments, while the squared sensitivity associated with the ith parameter of the chosen points is more than or equal to a b_i constraint value. Moreover, the sensitivity values can be weighted based on the importance of the parameter (or its accuracy), $u_{k,i}$ can be used to represent the weight of the kth experiments by the ith parameter.

$$\begin{aligned} &\max \sum_{k=1}^{N} y_k \sum_{i=1}^{n} u_{k,i} s_{k,i}^2 \\ &\text{s.t.} \sum_{k=1}^{N} y_k u_{k,i} s_{k,i}^2 \geq b_i \quad \forall i \in 1, \ldots, n \text{ and } \; y_k \in \{0,1\} \text{ and } \sum_{k=1}^{N} y_k \leq N_p \end{aligned} \tag{14}$$

where y_k represents whether the kth experiment is selected, $s_{k,i}^2$ stands for the squared sensitivity vector of the parameters that is compared to a predefined b_i constraint. N_p represents the number of feasible experiments to select from the set of experiments. $u_{k,i}$ represents the weight of the ith parameter sensitivity of the kth experiment.

The optimization determines which experiments have a significant impact on parameter identification. The design of experiments is expected to be performed in the appropriate variable space, where the kth experiment contains information to identify at least one parameter. It is also possible to tweak the algorithm to propose mechanical experiments. By selecting the theoretical point with the maximum sensitivity constrained to one of the principal stresses, it can be associated with uniaxial failure tests.

2.3.2. Experiment selection for measurement set

An experiment set may not contain enough information to reliably define the model parameters of a failure model. In this case, the goal is to propose experiments that have the highest information content possible for the specific parameters and are experimentally feasible. This problem is solved in a two-step method, first, we choose N_c amount of feasible experiments from the N_s set, and the second step is to define the optimal N_p experiments from N_c using Equation (14).

The feasibility analysis of experiments is performed in a normalized state-space, where a normalized experiment $\hat{\mathbf{x}}_\mathbf{k}$ consists of the normalized principal stresses ($\left[\hat{\sigma}_{1,k}, \hat{\sigma}_{2,k}, \hat{\tau}_{12,k}\right]$), the normalized stress components are defined in Eq. (15):

$$\hat{\sigma}_{1,k} = \begin{cases} \left|\frac{\sigma_{1,k}}{X_c}\right| & \texttt{if } \sigma_{1,k} < 0 \\ \frac{\sigma_{1,k}}{X_t} & \texttt{otherwise} \end{cases} \quad \hat{\sigma}_{2,k} = \begin{cases} \left|\frac{\sigma_{2,k}}{Y_c}\right| & \texttt{if } \sigma_{2,k} < 0 \\ \frac{\sigma_{2,k}}{Y_t} & \texttt{otherwise} \end{cases} \quad \hat{\tau}_{12,k} = \frac{\tau_{12,k}}{S_{12}} \tag{15}$$

where $[X_t, X_c, Y_t, Y_c, S_{12}] \in \theta^{(m,0)}$, and k is the kth experiment.

The normalized local sensitivity $\hat{\mathbf{s}}_\mathbf{k}$ of the kth experiment is represented by Eq. (16):

$$\hat{\mathbf{s}}_k = \left[\frac{s_{k,1}}{s_{max,1}}, \ldots, \frac{s_{k,n}}{s_{max,n}}\right]; s_{max,i} = max([s_{1,i}, \ldots, s_{N_s,i}]) \tag{16}$$

The failure surface is clustered based on the normalized local sensitivities of the N_s experiments forming a set of clusters, $C = \{C_1, \ldots, C_{N_c}\}$, where N_c also denotes the number of clusters, which is exactly the same as the number of selected feasible experiments. Any clustering algorithm can be used, so we decided to use the hierarchical clustering algorithm.

The number of clusters is defined iteratively, where the maximal of within-cluster variances ($H(C)$) must be lower than a threshold value (th). The problem can be written as a typical min-max problem, as shown in Eq. (17):

$$\begin{aligned} &\arg\min_{N_c} \arg\max_{c} H(C_c); \quad c = 1, \ldots, N_c \\ &\quad s.t. H(C) = \arg\max_{c} H(C_c) < th \end{aligned} \tag{17}$$

where $H(C_c)$ denotes the within variance of the cth cluster, and th is a user-defined threshold.

The within-variance of the cth cluster $H(C_c)$ is defined based on Eq. (18):

$$H(C_c) = \frac{1}{|C_c|} \sum_{k \in C_c} (s_k - \bar{s}_c)^2; \quad c = 1, \ldots, N_c \tag{18}$$

The resulting clusters represent the failure surface where the stress points in a cluster are contributing to the identifiability of the model parameters similarly. One possible experimental point is chosen from each cluster which is physically measurable and feasible. A measurement can be easily carried out if only one stress component is active, so an ideal measurement matrix ($\mathbf{I}$) can be formulated in a normalized stress space ($[\hat{\sigma}_1, \hat{\sigma}_2, \hat{\tau}_{12}]$), where each row presents an ideal experimental point (Eq. (19)).

$$\mathbf{I} = \begin{matrix} \hat{\sigma}_1 & \hat{\sigma}_2 & \hat{\tau}_{12} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{matrix} \tag{19}$$

An experiment is $e_c \in C_c$ selected from each cluster based on the minimum of the Euclidean distance between the normalized stress-values found in the cth cluster ($\hat{\mathbf{x}}_j^{(c)} = [\hat{\sigma}_{1,j}^{(c)}, \hat{\sigma}_{2,j}^{(c)}, \hat{\tau}_{12,j}^{(c)}]$) and the rows of the ideal matrix ($\mathbf{I}$). Eq. (20) defines the singular cluster of experiments with similar information content:

$$e_c = \arg\min_j \left(\min \left(\|\hat{\mathbf{x}}_j^{(c)} - \mathbf{I}_{(1)}\|_2, \|\hat{\mathbf{x}}_j^{(c)} - \mathbf{I}_{(}2)\|_2, \|\hat{\mathbf{x}}_j^{(c)} - \mathbf{I}_{(3)}\|_2 \right) \right) \Big| j \in C_c \tag{20}$$

$$E = \left\{ e_1, \ldots e_{N_c} \right\} \tag{21}$$

where e_c denotes the selected experiment from cluster c, j defines the running index of the cluster, and $\mathbf{I}_{(1)}$, $\mathbf{I}_{(2)}$, $\mathbf{I}_{(3)}$ defines the row vectors of the identity matrix. Set E, defined in Eq. (21), comprises of the indices of the selected experiments from each cluster. The final goal is to select p amount of optimal experiments from E experiment set in such a way that all the model parameters can be identified. It is solved using the presented nonlinear mixed-integer solver (see Equation (14)).

2.4. The proposed iterative parameter identification algorithm

This section describes the iterative identification process. Fig. 2 and Algorithm 1 and Algorithm 2 provide visual and textual drafts of the algorithm; the content of both is discussed in the following paragraphs.

The inputs of the algorithm include a starting set of failure tests $\mathbf{X}^{(0)}$ and the model - parameter pairs ($F = \{(f_1, \theta^{(1)}) \ldots, (f_M, \theta^{(M)})\}$) to extract the optimized (nominal) parameter $\theta^{(m,0)}$. The algorithm currently contains six failure models, namely the Max Stress [7], Tsai-Wu [8], Hoffman [9], Tsai-Hill [10], Hashin [11], and Puck [12] failure criteria. The input data are reconciled $\hat{\mathbf{X}}^{(m)}$ to evaluate the degree to which the identified failure model fits the data. A failure model f_m is considered the best if the median of the set $E^{(m)}$ of distances between the experimental data and failure models is the least, and all its parameters are identifiable. The best model is referred to as f_b, and its nominal parameters are denoted by $\theta^{(b,0)}$.

In order to select the optimal experiments, the failure surface is generated to provide the potential experiments $\mathbf{X}^{N_s}$, with a specific resolution. The sensitivity matrix $\mathbf{S}^{(N_s)}$ is then determined to examine which experiment contributes to a particular parameter. A $\mathbf{X}^{N_c}$ set of experiments contributes to the parameter identification similarly, therefore we cluster the experiments and select the best. From the N_c amount of clusters, we select the optimal points $\mathbf{X}^{(N_p)}$. We apply an iterative design of experiments approach to select only the most informative points. It is inevitable to obtain slightly different sensitivity values during the identification process

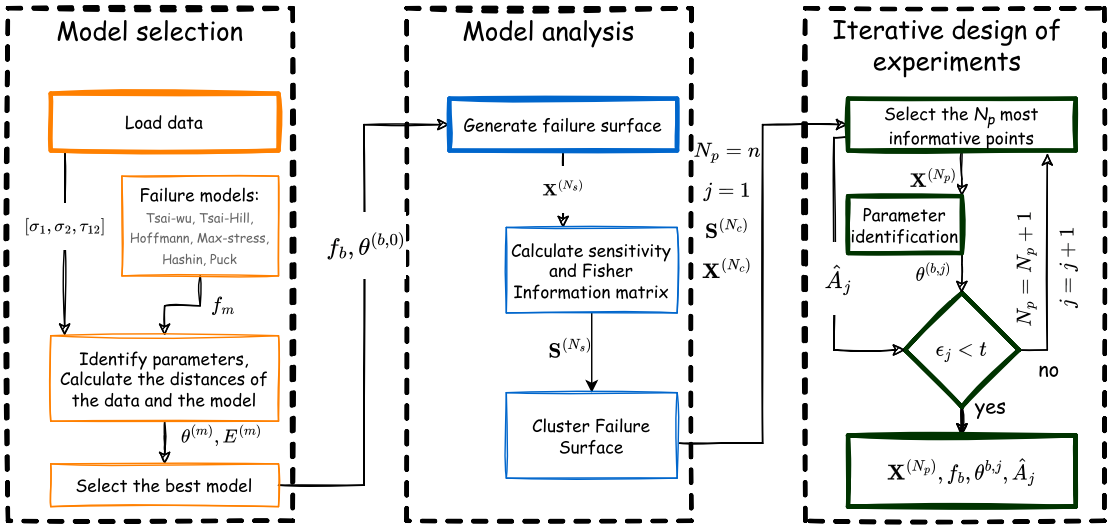


Fig. 2. Flowchart of the iterative identifiability algorithm. The algorithm requires the data and the failure models to be loaded and forwarded to the model selection function. The chosen model has the best fit for the data. The parameters are identified, and a failure surface is generated before the sensitivity of the experiments is calculated. The Fisher Information matrix of the individual experiments is determined so that the N_p most informative points can be selected through the fitting error and A-optimality ($\hat{A}_j$). If the results are inadequate, N_p is incremented by one, and the parameters, MSE (ϵ_j), and A-optimality are recalculated regarding the added optimal experiment. If adequate, the algorithm stops.

if the resolution of the mesh is altered. If the grid is underpopulated, then it is possible not to find sensitivity regions on the failure surface. On the contrary, overpopulated grids produce lengthier runtime and do not provide additional information to establish the sensitivity regions.

The idea of replication would be immensely powerful if this problem was a classical DoE problem. However, the selection process, in this case, is deterministic, and therefore the result of the identification from a set of experiments remains the same as the information content does not change. Also, please note that the sensitivity regions are dependent on the models, and do not change if the parameters stay constant. If the data is replicated, the randomization alters the information content, and may influence the model identification process heavily, may even distort the nominal failure model.

It is important to note that, assuming that each parameter is independent of the other, N_p needs to be at least the length of the parameter vector (n) as each parameter requires at least one experiment with high sensitivity (regarding that parameter) to be chosen. In some cases, where the parameters are interconnected, the N_p number of experiments may be decreased.

The DoE part of the algorithm begins by selecting the optimal points. The iteration starts with $j = 1$, where j is the number of iterations. During the selection of optimal points, the constraint vector $\mathbf{b}$ requires comparable values, so the sensitivities are normalized by the theoretical maximum of the failure surface. Then, the optimization problem described in Eq. (14) is solved with $\mathbf{U}$ weight matrix to prioritize parameters (if required).

After the optimal points are selected, the parameters are re-estimated, from which the A-optimality $\hat{A}_j$ is calculated by summing the diagonal of the FIM of the selected experiments. Lastly, the mean squared error ϵ is calculated from the reidentified models and the selected experiments. Stopping the iteration requires selecting all points or the error to be less than a threshold value t. The A-optimality and error values of the selected experiments are always recalculated. The parameter estimates constantly improve over each iteration, and the selection order defines the importance of the new set of experiments. It is also possible to modify the algorithm to only select the necessary experiments, and stop. The output returns the best model (f_b), the selected experiments ($\mathbf{X}^{(N_p)}$), the corresponding parameters ($\theta^{(b,j)}$) and their A-optimality ($\hat{A}_j$).

3. Results and discussion

The performance of the described algorithm is illustrated through two case studies using a procedurally generated dataset with 5% standard deviation errors using the Tsai-Hill failure model, and with 20% standard deviation errors using Puck failure model. The strength parameters are typical values for unidirectional carbon fibre reinforced epoxy plies with a fibre volume content $>$ 50%, the nominal parameters were [$X_t = 2134$ MPa, $X_c = 541$ MPa, $Y_t = 87.5$ MPa, $Y_c = 214$ MPa, $S_{12} = 62.4$ MPa] for Tsai-Hill and [$X_t = 2134$ MPa, $X_c = 541$ MPa, $Y_t = 87.5$ MPa, $Y_c = 214$ MPa, $S_{12} = 62.4$ MPa, $p^+ = 0.35, p^- = 0.3, s = 0.5, m = 0.5$] for Puck model. A total of fifty marginal states of failure in the stress space σ_1-σ_2-τ_{12} were sampled from the entire failure surface for both cases, so it is ensured that all parameters are identifiable. The fifty experiments act as the starting matrix for the identification process. In the first case, due to the low noise, we expect that the identifiability analysis will determine the model correctly and identify the parameters with a small error. However, in the second case, due to the high noise, other types of failure models can compete with the nominal Puck model. The models include the Max-Stress, Tsai-Wu, Hoffmann, Tsai-Hill UD, Hashin, and Puck failure criteria.

Algorithm 1 Description of iterative DoE-based experiment selection.

Input: $\mathbf{X}^{(0)}, F$ ▷ $\mathbf{X}^{(0)}$ denotes the set of experiments, $\mathbf{x}_k = \left[\sigma_{1,k}, \sigma_{2,k}, \tau_{12,k}\right]; k = 1...N_0$.
▷ The failure model set F contains the M amount of failure models f_m and their starting parameters $\theta^{(m,0)}$

for $m = 1, \ldots, M$ **do**
$\theta^{(m)}$ = ParameterIdentification(f_m, $\mathbf{X}$, $\theta^{(m,0)}$) ▷ Identify the parameters from a number of starting experiments.

$\hat{\mathbf{X}}^{(m)}$ = ReconcileDataToFailureSurface$\left(f_m, \mathbf{X}, \theta^{(m)}\right)$ ▷ Reconcile $\left(\hat{\mathbf{X}}^{(m)}\right)$ the starting experiments $\mathbf{X}^{(0)}$ so that they fit on the failure surface $f_m\left(\cdot, \theta^{(m)}\right)$.

$E^{(m)} = \text{GetDistance}\left(\mathbf{X}, \mathbf{V}, \hat{\mathbf{X}}^{(m)}\right)$
▷ Calculates the distances of all experiments $\mathbf{X}^{(0)}$ from the nominal surface $\hat{\mathbf{X}}^{(m)}$, and returns the set of the distances $E^{(m)}$.
▷ $\hat{\mathbf{X}}^{(m)}$ stands for the reconciled data for failure model f_m
end for

f_b, $\theta^{(b,0)}$ = SelectBestModel$\left(\left[\theta^{(1)}, \ldots, \theta^{(M)}\right], \left[E^{(1)} \ldots, E^{(M)}\right]\right)$
▷ Returns the best performing model and its parameters based on the median of the Euclidean distances

$\mathbf{X}^{(N_s)}$ = GenerateFailureSurface$\left(\theta^{(b,0)}, \text{resolution}\right)$ ▷ Generates the failure surface of the model with $\theta^{(b,0)}$ parameters.
▷ The resolution of the surface is also required (*e.g.* 10×10 grid)

$\mathbf{S}^{(N_s)}$ = CalculateSensitivity$\left(\mathbf{X}^{(N_s)}, \theta^{(b)}\right)$ ▷ Calculates the sensitivity of the experiments (individually) of the sampled surface $\mathbf{X}^{(N_s)}$.

$\mathbf{X}^{(\mathbf{N_c})}, \mathbf{S}^{(\mathbf{N_c})}$ = ClusterSurface($\mathbf{X}^{(N_s)}, \mathbf{S}^{(N_s)}$)
$N_p = n$
▷ The starting experiment number equals to the number of parameters so that each has at least one experiment to be identified from.

$j = 1$ ▷ Defines the running index for the while cycle.

while $\epsilon_j < t$ and $N_p \leq N_c$ **do** ▷ t stands for a user-defined threshold value.

$\mathbf{X}^{(N_p)}$ = SelectOptimalPoints$\left(\mathbf{S}^{(N_c)}, \mathbf{U}, \mathbf{b}, N_p\right)$ ▷ Calculates the FIM, and selects optimal points, where N_p number of experiments are selected.
▷ The nonlinear mixed integer optimizer is balanced by a constraint vector $\mathbf{b}$. The total contribution of the parameters must exceed the corresponding constraint.

$\theta^{(b,j)}$ = ParameterIdentification$\left(f_b(), \mathbf{X}^{(N_p)}, \theta^{(b,j-1)}\right)$
▷ The model based on the optimal points is reevaluated by identifying the parameters and recalculating errors.
▷ $j = 0$ is defined as the starting parameters $\theta^{(b,start)}$.

$\hat{A}_j$ = $tr\left(\hat{\mathbf{F}}\left(\theta^{(b,j)}\right)\right)$ ▷ Calculating the A-optimality

ϵ_j = MSE$\left(f_b, \theta^{(b,j)}, \mathbf{X}^{(N_p)}\right)$ ▷ Calculate the mean squared error of the optimal points.
$j = j + 1$
$N_p = N_p + 1$
end while ▷ Returns the best model, optimal points, parameters and the A-optimality values
Output: $f_b(), \mathbf{X}^{(N_p)}, \theta^{(b,j)}, \hat{A}_j$

Algorithm 2 Description of Iterative Identifiability

Input: $\mathbf{X}, f_m, \text{folds}, \theta_{start}$ ▷ $\mathbf{X}$ as set of experiments, $\mathbf{x}_k = [\sigma_1, \sigma_2, \tau_{12}]$

for $m = 1, \ldots, M$ **do**
θ_{f_m} = EvaluateParameters ($f_m(), \mathbf{X}, \theta_{start}$)

$\hat{\mathbf{X}}^m$ = ReconcileData($\mathbf{X}$, θ_{f_m}, f_m)

e_m = Median(EuclideanDistance($\mathbf{X}$, $\hat{\mathbf{X}}^m$)) ▷ Project the data on the identified nominal failure surface and calculate the Euclidean distance. The median is taken to obtain the fitness of the failure surface to the data.
end for
f_b, θ = SelectBestModel ($[\theta_{f_1}, \ldots, \theta_{f_M}], [\mathbf{e}_1, \ldots, \mathbf{e}_v]$) ▷ upper quartiles or median
$\mathbf{s}_{max}$ = GenerateFailureSurface (θ, resolution) ▷ over *e.g.* 30×30 grid
$p = n$ ▷ Starts with as many experiments as the number of parameters
$\mathbf{S}$ = CalculateSensitivity ($\mathbf{X}, \theta, \mathbf{s}_{max}$) ▷ sensitivity of the experiments (individually)
E = ProposeExperiments($\mathbf{S}$, $\mathbf{X}$, θ)
$\mathbf{X}_p$, ϵ_0 = SelectOptimalPoints ($\mathbf{S}$, $\mathbf{W}$, $\mathbf{b}$, p) ▷ calc. FIM, contribution should exceed $\mathbf{b}$
$j = 1, p = p + 1$
while $d_j <$ std($\mathbf{d}$) and $p \leq N$ **do**
$\mathbf{X}_p$ = SelectOptimalPoints ($\mathbf{S}$, $\mathbf{W}$, $\mathbf{b}$, p) ▷ Selects $p + 1$ optimal points
θ, ϵ_j = EvaluateModel ($f_b(), \mathbf{X}_p, \theta$) ▷ parameter identification and error calculus
a_j = $tr(\hat{\mathbf{F}}p(\theta))$ ▷ A-optimality
d_j = $\epsilon_{j-1} - \epsilon_j$ ▷ Change in error
end while
Output: $f_b(), \mathbf{X}_p, \theta, \mathbf{a}$ ▷ *e.g.* six parameters for the Tsai-Wu model

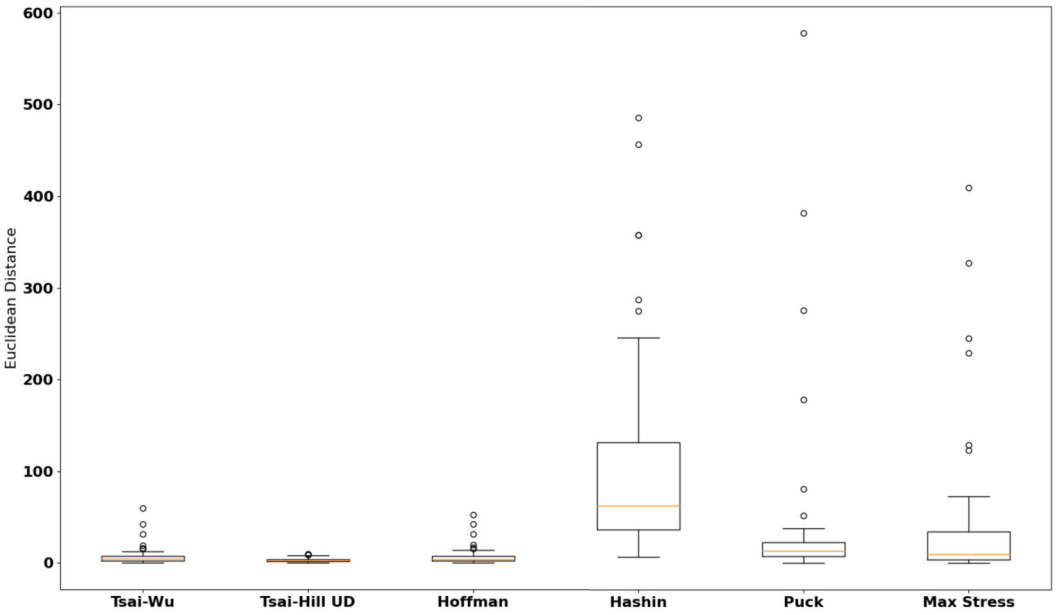


Fig. 3. Performance boxplots of composite failure models. The Tsai-Hill model performed the best, while the Hashin and the Max Stress were the worst.

In addition, the algorithm is also tasked to recommend experiments by highlighting the most sensitive experiments on the failure surface clusters.

We have programmed the algorithm in Python. Therefore, several different Python packages have been used (*e.g.* pandas and NumPy) [28]. The Euclidean distance error (data reconciliation) and optimization algorithms were implemented with scipy/lmfit optimizer [29,30]. As the model selection is a simple ranking, we calculate the medians with pandas and NumPy. The generation of the failure surface uses the NumPy linspace function for a 2D grid of $[\sigma_1, \sigma_2]$, and τ_{12} is optimized through scipy/lmfit with the help of the calculated parameters, implemented failure models and generates stress values. The partial differentiation to establish the sensitivity matrix is provided by the numdifftools package [31]. The selection of optimal experiments happens through the SCIP mixed-integer nonlinear solver of the ortools package [32].

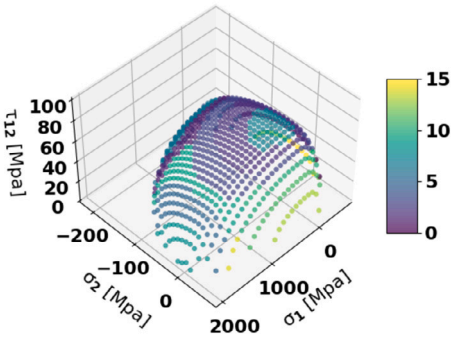
3.1. Case #1

If the actual parameters and governing principles behind the behavior of the material are unknown, best-fitting models should be selected first. The evaluation process requires the optimal parameters and uses the minimum Euclidean distance of the measured point from the optimal failure surface. The median of the distances provides information on which model performs the best with a given set of experiments. Fig. 3 presents the boxplots of the Euclidean distances shown on the y-axis. Most models (located on the x-axis) perform comparably well due to their similar failure surfaces. The Tsai-Wu, Tsai-Hill, and Hoffman models are half ellipsoids but somewhat different in their orientation. Appendix B (Figure 15) depicts the shapes of the failure surfaces. It is important to note that while the previously mentioned failure models consistently provide ellipsoids, the Puck is flexible in its shape due to the four additional parameters. One can also configure it to be an ellipsoid. Meanwhile, the Max-stress is considered a cuboid and can very well cover the stress space where the experiments can be located. As such, Max-stress may not be the best choice; if the failure model is an ellipsoid, then there is a possibility that the max-stress does cover non-populated parts of the stress space. The Hashin failure surface takes up a curved pyramid-like shape (simply put) and performs the worst, as the experimental data was sampled from an ellipsoid with fairly low noise.

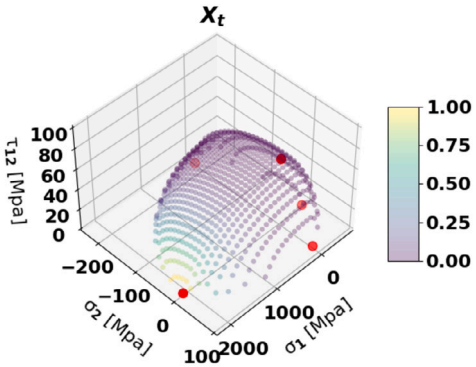
The best model is chosen based on how well the data fits on the model: the median of the Euclidean distances to the failure surface is considered. In this case, the Tsai-Hill criterion performs with the lowest mean distance, and Hashin and Max Stress are the worst. Most non-ellipsoid models have significant outliers that highlight that the models cannot describe various experiments and have blank spaces that cannot be fitted onto the data. The outcome is not surprising as the stress pairs were generated by the Tsai-Hill model, and the deviation of the data from the model is relatively low (5%). It is important to note that the Puck, Tsai-Wu, and Hoffman models perform quite well in terms of the median distances.

Following the model performance evaluation, we explore the parameter-sensitive stress regions of the Tsai-Hill failure surface. Therefore, we performed the model parameter sensitivity analysis and presented the sensitivity of the model parameters (X_t, X_c, Y_t, Y_c and S_{12}) on the failure surface in Fig. 4.

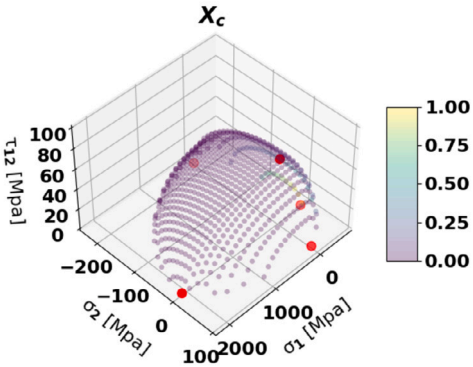
The color bar presents the sensitivity values on the failure surface. Each parameter sensitivity was normalized with the corresponding maximum value. The stress regions are shown for each investigated model parameter. Note that the sensitivities are perpendicular to the related axis. The nominal parameter X_t presents where the material would break upon tensile tests performed



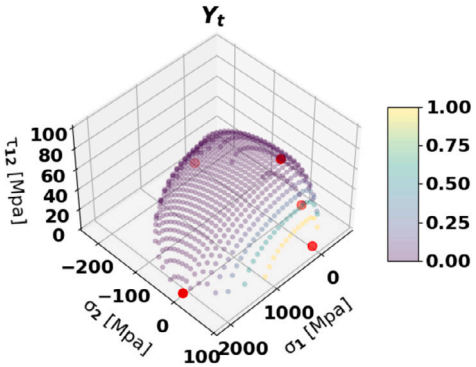
(a) Clustered regions



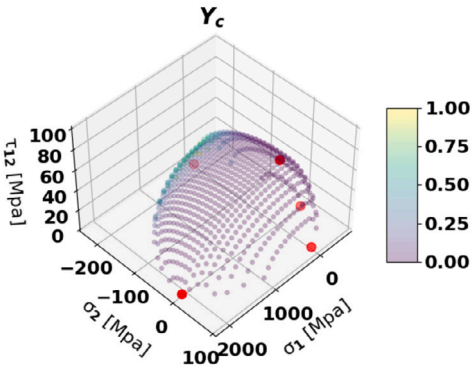
(b) Sensitivity of the X_t parameter



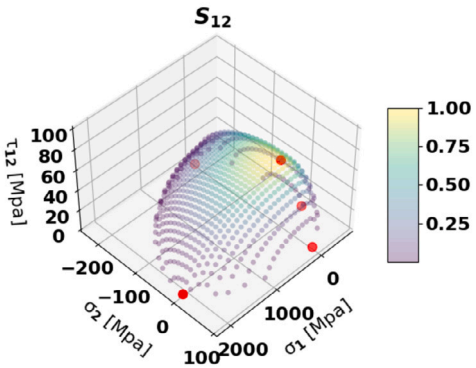
(c) Sensitivity of the X_c parameter



(d) Sensitivity of the Y_t parameter



(e) Sensitivity of the Y_c parameter



(f) Sensitivity of the S_{12} parameter

Fig. 4. Sensitivity of Tsai-Hill model parameters (X_t, X_c, Y_t, Y_c, S_{12}). The $[X_t, X_c, Y_t, Y_c]$ parameters are identifiable jointly, with only two experiments. S_{12} depicts the shear force (τ_{12}), so it requires an experiment where the shear force is significant. Colorbars represent the normalized sensitivity with regards to maximum possible sensitivity, except for subfigure 4a, in which it denotes the color of numbered clusters.

in fibre direction. The σ_1 has to be at least the nominal parameter to break the material. As the stresses change, the relevancy (information provided to the identification) of one parameter begins to falter. So to identify all experiments that may cause failure, a segment including the information content for a parameter should be selected. The sensitivity plots of Fig. 4 show how strongly the given stress influences the parameters. The relevant area is the yellowish-greenish segment of the failure surface. This way, all the stress regions can be explored where the specific model parameters can be identified. Some points may contain information about more than one parameter, *e.g.* the subfigures contain points on the mesh that contribute to both X_c and Y_t. It is preferable to execute the experiments on the vertices of the failure mesh. Experimental design can be used to define theoretical stress pairs with the highest sensitivity to a model parameter.

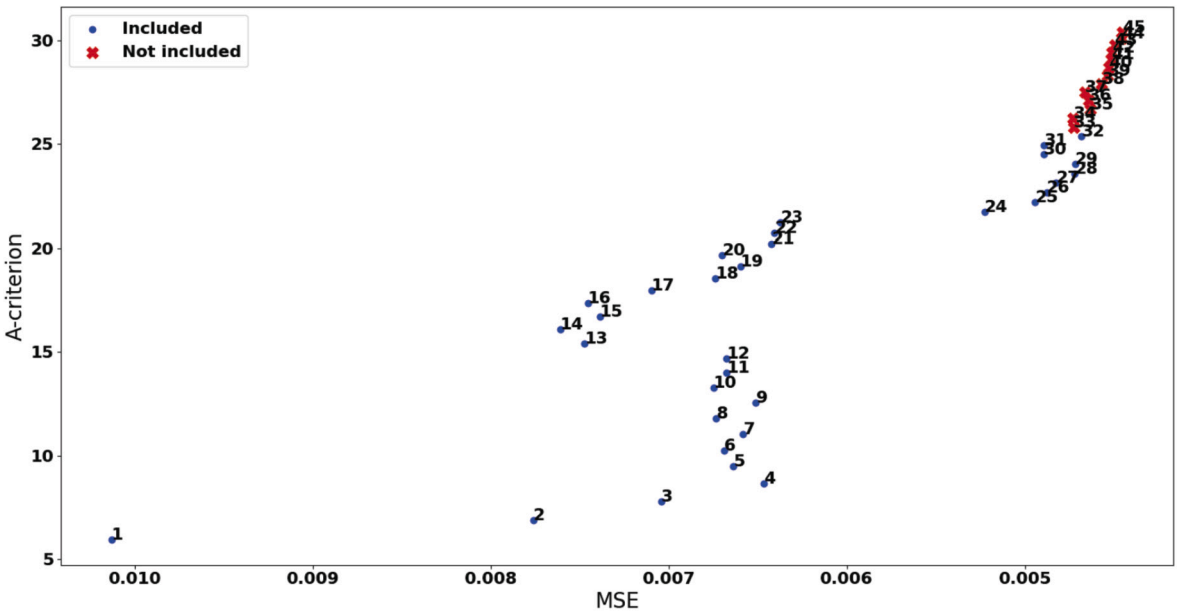


Fig. 5. MSE and A-optimality values during the iterative experiment design and parameter fitting. While the A-optimality increases monotonically, the error may deviate. If the change in error is less than the standard deviation, the experiments are not selected shown with red crosses. The blue points are the ones that are chosen.

The result of the iterative experimental design and parameter fitting is presented in Fig. 5. The depicted values present how the A-optimality increases monotonically from iteration to iteration while the mean squared error (MSE) drops significantly. The MSE is insignificant (≈ 0.035) from the first iteration due to the low noise of the experimental data, yet, the MSE, in general, decreases further until the last iteration. However, based on the model accuracy of the values of MSE, there is no need to include additional experiments after the 32*nd* iteration step as the model accuracy does not improve significantly. Although the A-optimality is monotonous, it may not necessarily be the case for the error value. If the uncertainty is high or too many uninformative experiments are selected, the MSE may begin to rise again. After some additional inclusions, it may also drop, so stopping the algorithm at the right time may be difficult for the analyst. Fig. 5 shows how the new experimental points were added during the iterations. The blue dots present the experimental points up to the iteration stop condition (so until 32*nd* step), and the red crosses can be defined as not included.

The iterative experiment design algorithm aims to cover the entire stress field and collect information from most locations, as it is presented in Fig. 6. The iterative DoE algorithm selects experiments in a manner that promotes obtaining the maximum information. The selection seems to be in a zigzag pattern; in other words, the algorithm often chooses a point far from the last included experiment. The selector considers the whole failure surface, not just one region, and in doing so identifies the parameters with accuracy. However, with the inclusion of non-informative experiments, disruptive, noisy information can also be selected, and therefore the parameter values may shift significantly towards the assigned limits. Higher error values may appear during identification, or the improvement of the error becomes negligible. Hence, monitoring MSE during iterative DOE is necessary.

Fig. 7 presents the fitted model parameters in every iteration step, and the black vertical line presents the 32*nd* step, where the model parameters are accepted as nominal parameters. During the identification process, we aim to find experiments that provide the most information in identifying the model parameters. As the experimental points were generated with the same model (Tsai-Hill model), the known nominal parameter can also be depicted in the figure, which is presented by a black horizontal line. As such, comparing the nominal and identified parameters shows that the algorithm determines the parameters in the vicinity of the nominal parameters, with a minor difference. Parameters that fluctuate may be interconnected with other parameters, so that the other parameters may react and change their values significantly. Moreover, the parameters may find a local extreme, so they fail to change significantly. If the algorithm stops at the right time, overfitting can be avoided. With the inclusion of a point, the multivariate optimization may be significantly impacted by the additional information content and therefore may shift the parameter value unexpectedly. Selecting experimental points that contain uncertainty may distort (regarding the nominal surface) the result of the iterative parameter identification process. For example, in Fig. 7, X_c fluctuates by adding specific points (especially between iteration numbers 15, and 25). The parameters can suddenly change value, back and forth, however, after enough variance, the parameter value may not differ significantly from previous iterations. Therefore, it is important to include as many points as possible so that the model MSE remains negligible and the deviation of parameters is not significant.

Fig. 8 presents the identified failure surface in the σ_1-σ_2-τ_{12} plane, where the blue dots denote the experimental data stress points that were considered in the parameter identification, and the red crosses represent that the experimental points were uninvolved. The identified Tsai-Hill surface describes the experimental data well. The red crosses are mostly those that significantly deviate from the failure surface. The blue dots seem to fit perfectly as the model describes their behavior rather than those with heavy noise.

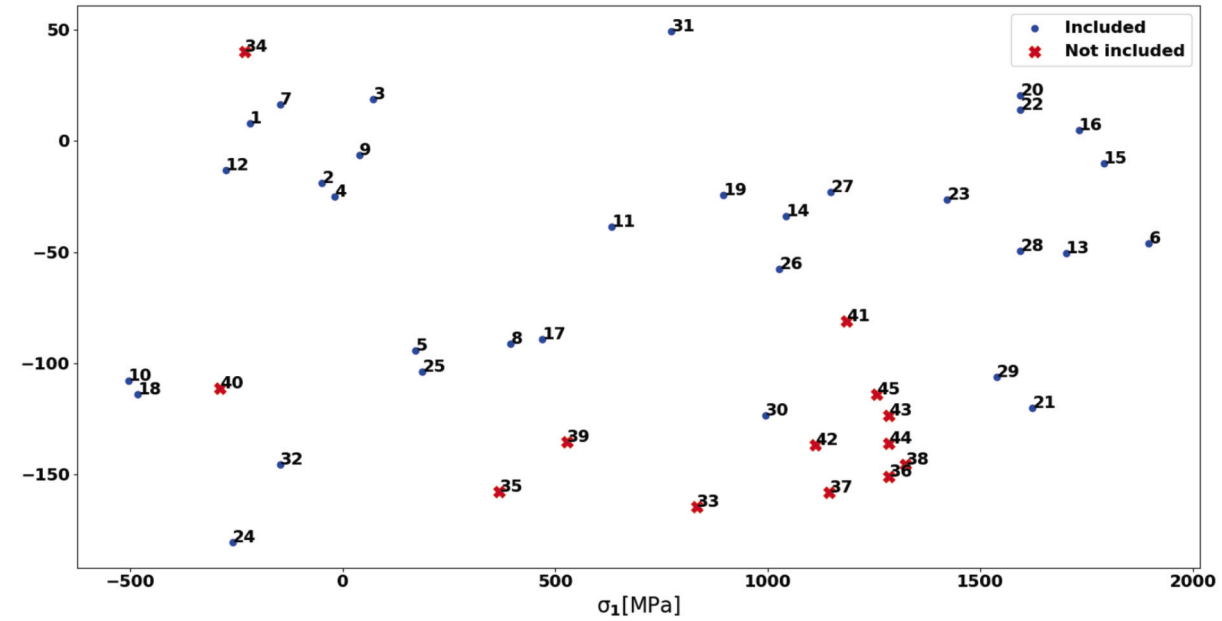


Fig. 6. Result of iterative experiment design in the σ_1-σ_2 plane The optimal selection algorithm walks a zigzag path as it chooses points from all over the failure surface. The next point is often at the other side of the surface and contributes to other parameters than the previous one. Blue points were selected, while the red crosses were selected after the algorithm stopped.

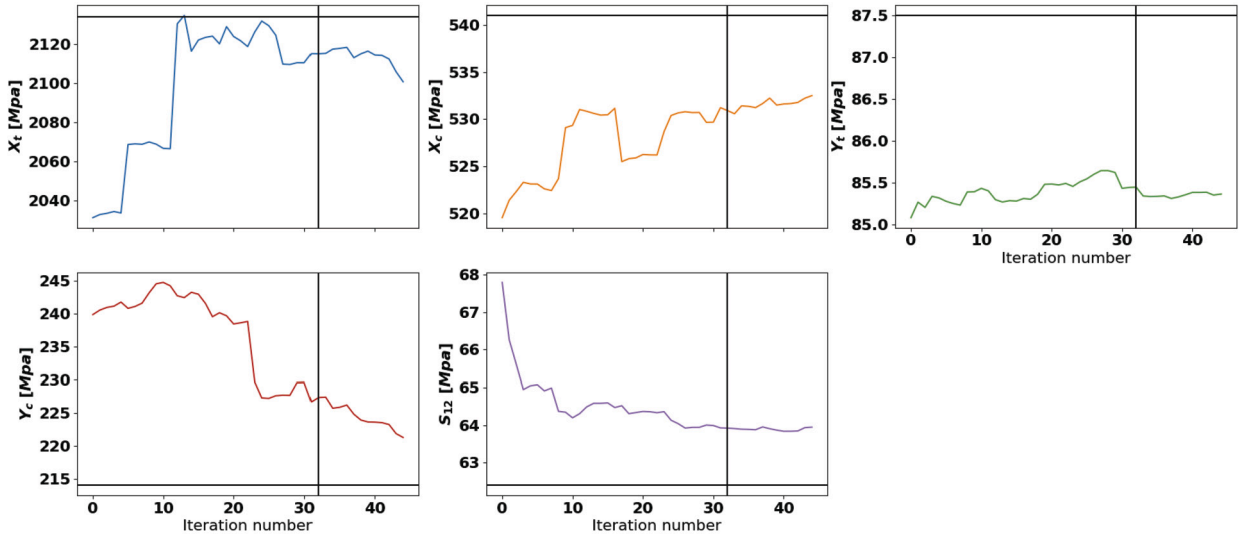


Fig. 7. Fitted model parameters in the iteration process. This figure presents the fitted model parameters in every iteration step, and the black vertical line presents the 32*nd* step, where the model parameters are accepted as nominal parameters. The horizontal black line presents the nominal model parameters used in experimental data generation. All horizontal axes represent the iteration number. The southernmost axis labels describe all other vertically aligned subfigures.

Therefore, it is important to stop before the surface is fitted on the red crosses, which may significantly disturb the identification process and cause overfitting, *i.e.* the noise modeling.

3.2. Case #2

The first step in the algorithm is to define the most suitable failure model based on existing experimental data points. For this purpose, the Euclidean distance of the measurement points from the identified failure surface is calculated with a data reconciliation technique using each composite failure model (Tsai-Wu, Tsai-Hill, Hoffmann, Hashin, Puck, and Max-Stress models). The generation of the points includes a set of experiments randomly generated on a stress surface, which is then altered by adding a 20% standard deviation error on each stress axis. Hence, the result of the best-performing model is not unambiguous. Fig. 9 presents the boxplots of the results of the distance calculation, where the Euclidean distance is shown on the *y*-axis. The best model is chosen based on

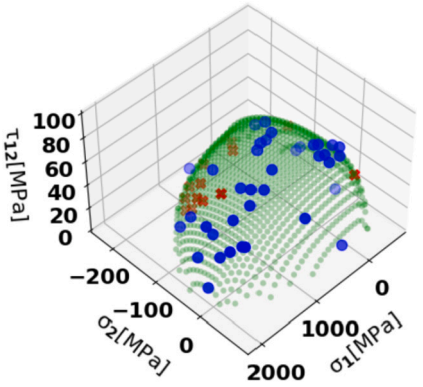


Fig. 8. Tsai-Hill failure surface with the generated experiments. The surface is identified in the iteration process with a nominal set of parameters. The blue dots indicate the experimental points used for identification and the red crosses were unused.

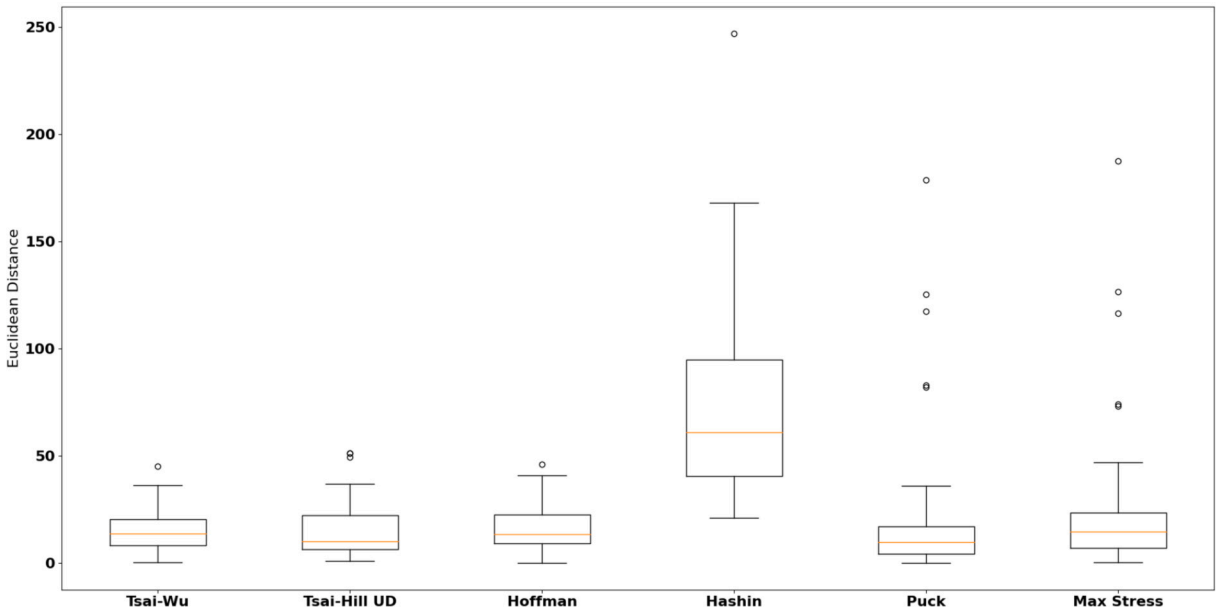
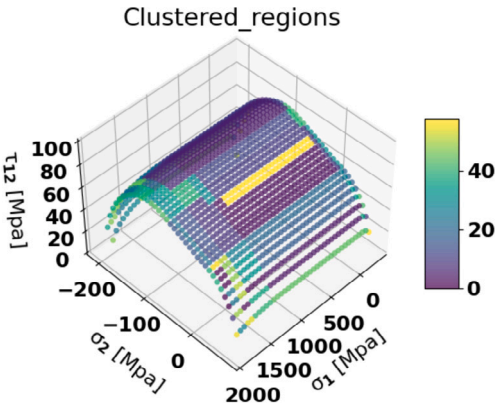


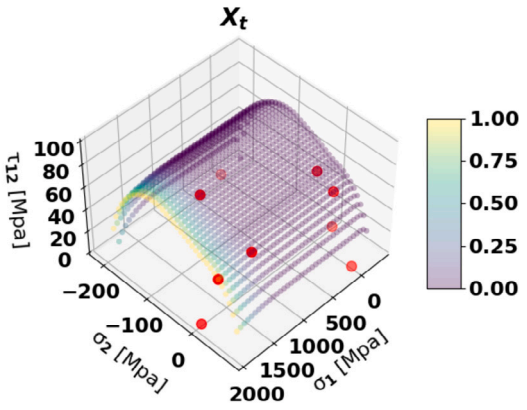
Fig. 9. Performance boxplots of composite failure models. The Tsai-Hill model performed the best, while the Hashin and Max Stress were the worst.

their accuracy, where the mean distance is considered. Despite the high deviation, the developed algorithm found the Puck criterion to be the best performing, while the Hashin is the worst due to its pyramid-like shape. Yet, the Tsai-Wu, Tsai-Hill UD, and Hoffman models perform quite well. Appendix C (Figure 16) presents the failure surfaces of all identified failure models. The ground truth is not necessarily reflected by the experiments due to errors. Case #2 has a 20% standard deviation of errors added to the stresses, and therefore, it is not evident from the data that the nominal parameters were used for sampling. However, if the model identification process does not improve significantly over iterations, we assume there is no more information we can use in the process, therefore, it is advised to stop performing experiments at this point. Selecting the stopping criterion is a gruesome task, but the aim is to find a proper surface that is justified by the data. The quality of the experiments directly influences the process. Moreover, if any parameters cannot be identified, the algorithm can propose experiments, depending on the maximum sensitivity. A good example is the potential experiments denoted by red dots on the sensitivity meshes of Fig. 4.

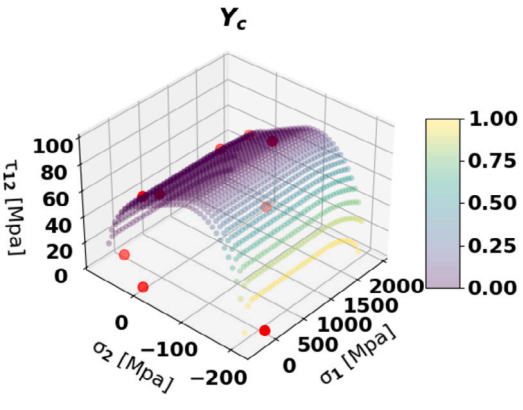
Following the evaluation of model performance, we explore the parameter-sensitive stress regions on the failure surface. For this purpose, we performed the model parameter sensitivity analysis. We present the sensitivity of the model parameters (X_t, Y_c, S_{12}, p^+ and s) on the failure surface in Fig. 10 and the sensitivity of (X_c, Y_t, p^-, m) in Appendix D. The color bar presents the sensitivity values on the failure surface, signaling that the sensitivity values require normalization before optimal selection. A-optimality requires normalization to equally select experiments from the different stress values during the integer optimization. D-optimality does not require normalization, but it is not additive monotonically, so there is a possibility of ambiguous change in the optimization score. The stress regions are shown for each investigated model parameter. For instance, X_t can be identified if the σ_1 stress is high ($\sigma_1 > 1800$), or Y_c can be identified if the absolute value of σ_2 stress is high ($\sigma_2 < -150$). In this way, all the stress regions can be



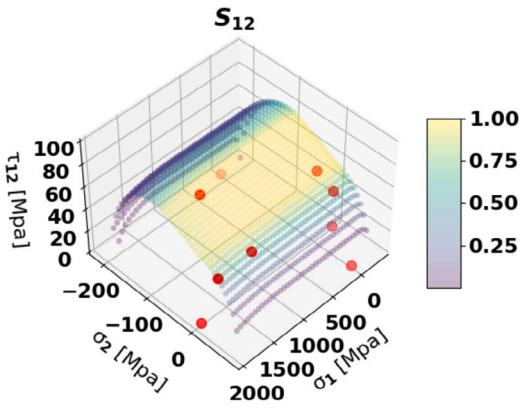
(a) Clustered regions



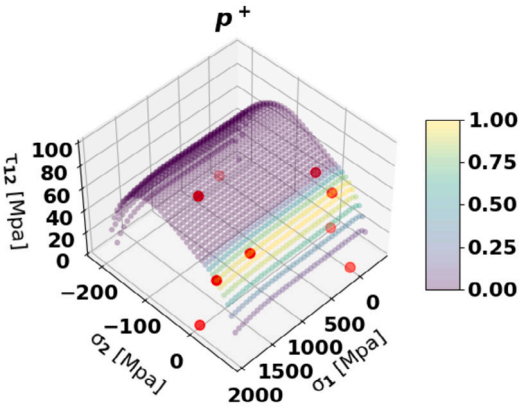
(b) Sensitivity of the X_t parameter



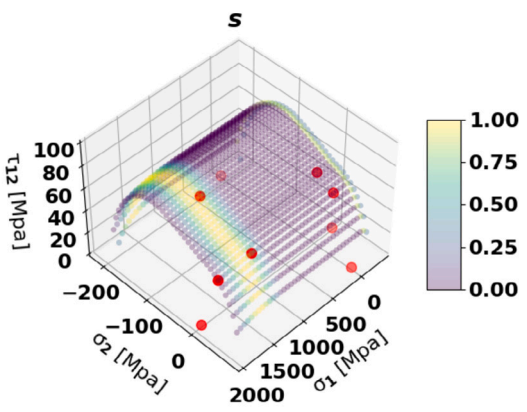
(c) Sensitivity of the Y_c parameter.



(d) Sensitivity of the $S12$ parameter



(e) Sensitivity of the s parameter



(f) Sensitivity of the S_{12} parameter

Fig. 10. Sensitivity of some of the parameters of the Puck model (X_t, Y_c, S_{12}, p^+, s). The axes of Fig. 17c are rotated to present the relevant section of the mesh. Colorbars represent the normalized sensitivity with regards to maximum possible sensitivity, except for subfigure 10a, in which it denotes the color of numbered clusters.

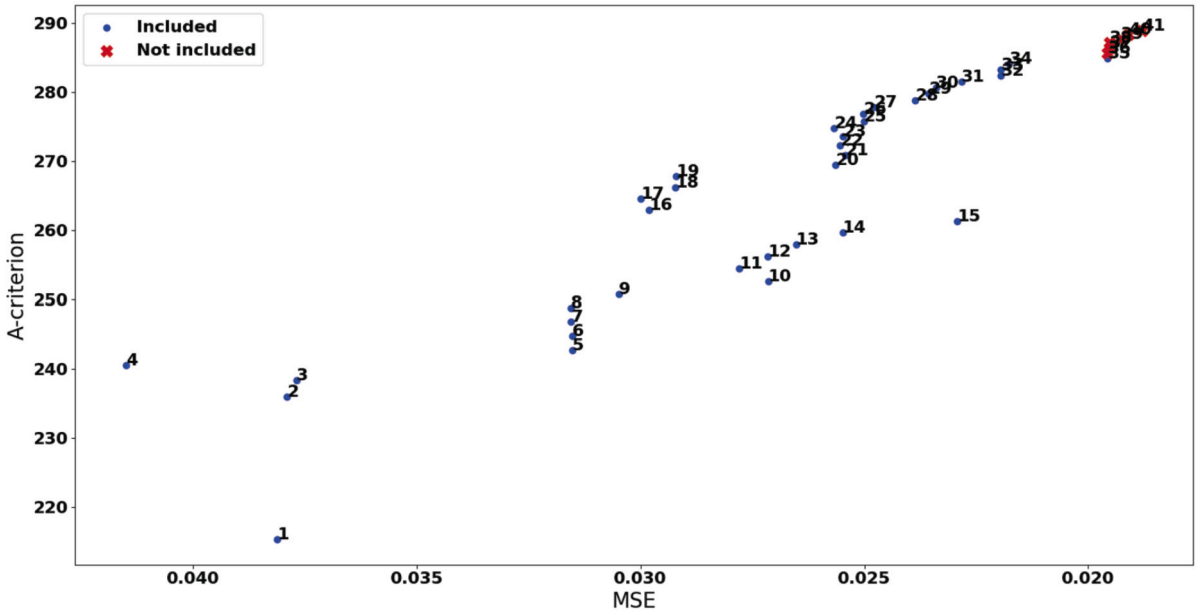


Fig. 11. MSE and A-optimality values during the iterative experiment design and parameter fitting. While the A-optimality increases monotonically, the error may deviate. If the change in error is less than the standard deviation, the experiments are not selected and are represented with red crosses. The blue points are the ones that are chosen.

explored where the specific model parameters can be identified. Some points may contain information about multiple parameters, *e.g.* the subfigures contain points on the mesh that contribute to both S_{12} and s, p^{+}. It is preferable to execute the experiments on the vertices of the failure mesh. Experimental design can be used to define theoretical stress pairs with the highest sensitivity to a model parameter. For this purpose, we clustered the failure surface into such groups, as presented in Fig. 10a, to find out which areas would contribute the most to specific/multiple parameters. 14 clusters are defined based on our method, and from these clusters, 9 experimental points (red dots) were proposed to identify the model parameters. These red dots in Fig. 10 present the stress pairs which include the highest information concerning the identifiability of the model parameters. (X_t, X_c, Y_t, Y_c and S_{12}) parameters can be identified using uniaxial tests. For the identification of p^{-} parameters the experiment should contain σ_2 (≈ -65 MPa) and τ_{12} (≈ 78 MPa) stress components. (p^{+} and m) model parameters can be identified using a similar experiment, it needs the three stress components ($\sigma_1 \approx 1890$ MPa, $\sigma_2 \approx 30-50$ MPa and $\tau_{12} \approx 45-50$ MPa), and the identification of (s) model parameter also needs three stress component (($\sigma_1 \approx 1100$ MPa, $\sigma_2 \approx -66$ MPa and $\tau_{12} \approx 77$ MPa)).

The result of the iterative experiment design and parameter fit is presented in Figs. 11 and 12. Fig. 11 presents how the value of the A-criterion increases monotonically from iteration to iteration as a function of mean squared error. The MSE is low (≈ 0.04) from the first iteration despite the high variance in the experimental data, and still decreases significantly further until the last iteration. However, based on the model accuracy values measured by MSE, there is no need to include additional experiments after the 35*th* iteration step. From this point on, the model accuracy does not improve significantly. Fig. 12 shows the new experimental points selected in the iterations, where the blue dots show the experimental points up to the iteration stop condition (so until 35*th* step), and the red crosses are after this. The iterative experiment design algorithm tries to cover the entire stress field and collect information from most locations. The order in the selection of the points has no importance, as the process is deterministic. The information content will always decide which points will be selected. This is also credited to the behavior of the additive A-optimality, as no reevaluation is required. As the identification process focuses on including as much information as possible to identify all parameters, it is preferable to select points from all over the stress surface of the experiments. Therefore, it should not concentrate on a few points, but on those that contain the necessary information to execute multivariate parameter optimization, which will heavily depend on the number of parameters, their relationship, and the variance of the data.

Fig. 13 presents the parameters of the fitted model in each iteration step, and the black vertical line presents the 35*th* step, where the model parameters are accepted as nominal parameters. The horizontal line presents the nominal Puck parameters. Due to the complex model structure and high standard deviation, it is not granted to identify the nominal parameters in all cases. There may be similar surfaces, with different parameter sets that can accurately represent the experimental data. The Puck model itself is "elastic", and can change its shape due to the additional parameters that govern the roundness of the edges. Some parameters may not be properly identified during the iteration process, however, we can propose experiments to correct for that absence of information.

Fig. 14 presents the identified (green surface) and the nominal failure surface (red surface) in the σ_1-σ_2-τ_{12} plane, where the blue dots mean the stress points of the experimental data that are considered in the parameter identification. Due to the high variance of the experimental points, the identified failure surface is slightly different from the nominal Puck failure surface but fits the experimental points well.

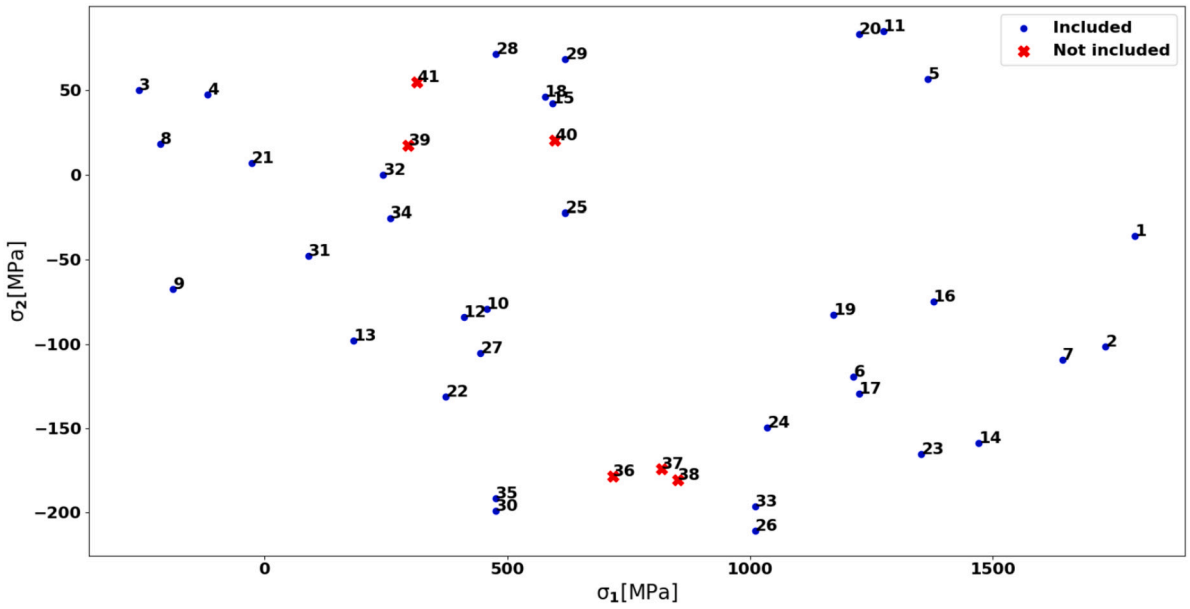


Fig. 12. Result of iterative experiment design in the σ_1-σ_2 plane The optimal selection algorithm walks a zigzag path as it chooses points from all over the failure surface. This is due to the DoE aiming to select experiments with the most information about the parameter identification process. The next point is often at the other side of the surface and contributes to other parameters than the previous one. Blue points were selected, while the red crosses were selected after the algorithm stopped.

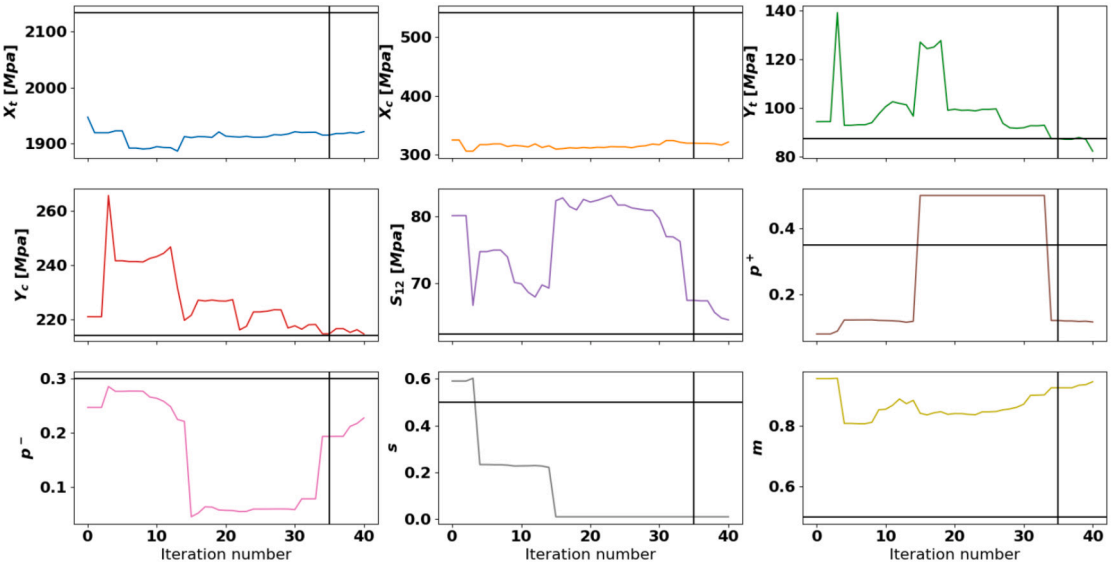


Fig. 13. Fitted model parameters in the iteration process. This figure presents the fitted model parameters in every iteration step, and the black vertical line presents the 35*th* step, where the model parameters are accepted as nominal parameters. The horizontal black line presents the nominal model parameters used in experimental data generation. All horizontal axes represent the iteration number. The southernmost axis labels describe all other vertically aligned subfigures.

4. Conclusions

In this work, we proposed a methodology to identify the composite material failure models and select experiments. Material behavior may be described with different failure models, such as the Tsai-Wu, Max Stress, Hoffman, Hashin, Tsai-Hill, and Puck models. We proposed a robust method to evaluate which model best fits the experimental data through the Euclidean distances between the experiments and failure surfaces.

When the model is selected, the failure surface is generated to provide information on where the parameters are best identifiable. Some of the parameters of more complex models, such as the Puck or Hashin, may need to be identified from more than one stress. Selecting from the clustered groups yields experiments that contain different information, which helps identify more complex models.

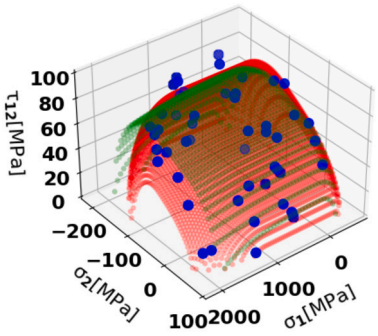


Fig. 14. Identified Puck failure surface (green) and the nominal Puck failure model (red) with the generated experiments. The surface is identified in the iteration process with a nominal set of parameters. The blue dots indicate the experimental points.

Therefore, we generate a set of potential experiments that describe the failure surface, from which several groups are clustered, so that the optimal points are selected from the clusters, and include points from all over the surface.

To properly apply the different failure models, the model parameters must be determined with as little uncertainty as possible, and all parameters must be identifiable. However, the available experimental data set may not contain information to identify all the model parameters, and all the experimental data points may not include additional information for parameter estimation. Therefore, we propose an experimental design-based identifiability analysis and optimal experiment selection. As from the sensitivity analysis, the identifiability of the model parameters can be inferred, we propose the application of Fisher Information Matrix-based sensitivity analysis to search for the most informative test results. The diagonal of the matrix describes the identifiability of each model parameter. On the other hand, A-optimality refers to the maximization of the trace of the Fisher Information Matrix, hence only by calculating the diagonal of the matrix we can conclude the information content to select the most representative experimental data points. The model parameter fitting is performed in an iterative mode, so in this way, the contribution of each of the experimental points to the model accuracy can be evaluated one by one. If the accuracy of the failure model does not improve significantly, there is no need to take into account any new experimental points.

Two cases were investigated to prove the applicability of the proposed methodology, where the experimental data sets were generated using the Tsai-Hill UD and Puck model with a standard deviation for the stress values of 5% and 20% respectively. In the first case, due to the relatively low deviation, the algorithm found that the Tsai-Hill model best describes the experimental set, and we successfully identified its model parameters using the proposed iterative experiment design and identification algorithm. In the second case, despite the relatively high deviation of the experimental set, most of the failure models performed well, but our methodology successfully found that the Puck model best describes the experimental data. We presented that the Fisher-matrix-based identifiability analysis can be applied to define those regions in the stress state-space where the unique model parameters can be identified, and with the proposed clustering and linear programming-based methodology, we can propose guided and feasible experiments to perform.

CRediT authorship contribution statement

Ádám Ipkovich: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Alex Kummer:** Writing – review & editing, Writing – original draft, Software, Methodology, Conceptualization. **László Kovács:** Writing – review & editing, Supervision, Formal analysis, Conceptualization. **Balázs Fodor:** Writing – review & editing, Supervision, Conceptualization. **János Abonyi:** Writing – review & editing, Supervision, Methodology, Formal analysis, Conceptualization.

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.

Data availability

Data will be made available upon request.

Appendix A. Description of the studied models

Max-Stress criterion is one of the simplest composite failure model [7]. The interactions between the three stress components are not taken into account, but a distinction between the compressive and tensile loading of the material is considered. The criterion

consists of five independent states, and five model parameters $\left(\theta = [X_t, X_c, Y_t, Y_c, S_{12}]\right)$. X_t and X_c denote the tensile and compressive strength parameters in the direction of the fibers, respectively. Y_t and Y_c stand for tensile and compressive strength parameters perpendicular to the direction of fibers, S_{12} represents the in-plane shear strength parameter of the material.

$$f_{\sigma_1} = \begin{cases} \frac{\sigma_1}{X_t} & \texttt{if } \sigma_1 \geq 0 \\ \frac{|\sigma_1|}{X_c} & \texttt{if } \sigma_1 < 0 \end{cases} \quad f_{\sigma_2} = \begin{cases} \frac{\sigma_2}{Y_t} & \texttt{if } \sigma_2 \geq 0 \\ \frac{|\sigma_2|}{Y_c} & \texttt{if } \sigma_2 < 0 \end{cases} \quad f_{\tau_{12}} = \frac{|\tau_{12}|}{S_{12}} \tag{22}$$

A material failure occurs if any of the three function (f_{σ_1}, f_{σ_2}, $f_{\tau_{12}}$) reaches value one.

$$f = max(f_{\sigma_1}, f_{\sigma_2}, f_{\tau_{12}}) \tag{23}$$

Tsai-Wu model is considered one of the most popular failure criteria [8]. The model consists of a single continuous function and six parameters $\left(\theta = [X_t, X_c, Y_t, Y_c, S_{12}, F_{12}]\right)$. F_{12} is considered as a coefficient of the Tsai-Wu model. Each model requires unidirectional parallel (σ_1) and perpendicular (σ_2) compression/tension and shear stress (τ_{12}) measurements.

The Tsai-Wu model is described in Eq. (24):

$$f = \frac{{\tau_{12}}^2}{{S_{12}}^2} - \sigma_2 \left(\frac{1}{Y_c} - \frac{1}{Y_t}\right) - \sigma_1 \left(\frac{1}{X_c} - \frac{1}{X_t}\right) + \frac{{\sigma_1}^2}{X_c X_t} + \frac{{\sigma_2}^2}{Y_c Y_t} + \frac{2 F_{12} \sigma_1 \sigma_2}{\sqrt{X_c X_t Y_c Y_t}} \tag{24}$$

where f denotes failure of the material that theoretically takes up the value of one.

The Tsai-Wu model can be further simplified [33]; the **Hoffman criterion** is similarly quadratic polynomial in nature. The Hoffman failure criterion is almost the equivalent of the Tsai-Wu failure criterion except for the disappearing Tsai-Wu coefficient as is shown in Equation (25), being equivalent to Tsai-Wu model with $F_{12} = 0$. Hoffman-criterion consists of a continuous function and five parameters $\left(\theta = [X_t, X_c, Y_t, Y_c, S_{12}]\right)$.

$$f = \frac{{\tau_{12}}^2}{{S_{12}}^2} + \sigma_2 \left(\frac{1}{Y_t} - \frac{1}{Y_c}\right) + \sigma_1 \left(\frac{1}{X_t} - \frac{1}{X_c}\right) + \frac{{\sigma_1}^2}{X_c X_t} + \frac{{\sigma_2}^2}{Y_c Y_t} - \frac{\sigma_1 \sigma_2}{X_c X_t} \tag{25}$$

where f denotes failure of the material that theoretically takes up the value of one.

In the case of the **Tsai-Hill criterion** two independent functions describe the failure, depending on the type of the reinforcement structure (unidirectional or fabric type). Uni-directional composites are materials whose layers consist of parallel fibres. The Tsai-Hill failure criterion is defined in Eq. (26) for unidirectional composites:

$$f = \left(\frac{\sigma_1}{X}\right)^2 + \left(\frac{\sigma_2}{Y}\right)^2 + \left(\frac{\tau_{12}}{S_{12}}\right)^2 - \frac{\sigma_1 \sigma_2}{X^2} \tag{26}$$

where f is the failure index, X denotes the strength of the material along the fiber direction, Y stands for the strength of the composite perpendicular to the fiber direction, S is considered the shear strength.

In order to incorporate differing tensile and compressive strengths X and Y are considered based on the values of σ_1 and σ_2 as in Eq (27).

$$X = \begin{cases} X_t & \texttt{if } \sigma_1 \geq 0 \\ X_c & \texttt{if } \sigma_1 < 0 \end{cases} \quad Y = \begin{cases} Y_t & \texttt{if } \sigma_2 \geq 0 \\ Y_c & \texttt{if } \sigma_2 < 0 \end{cases} \tag{27}$$

Hashin-model considers fiber and matrix as different entities (f_f and f_m), and each has two discontinuous functions that describe the onset of failure. The maximum of f_f and f_m values decide when the material breaks. Hashin-model considers six parameters $\left(\theta = [X_t, X_c, Y_t, Y_c, S_{12}, Q]\right)$. In fibers, the failure behaves as presented in Equation (28).

$$f_f = \begin{cases} \left(\frac{\sigma_1}{X_t}\right)^2 + \left(\frac{\tau_{12}}{S_{12}}\right)^2 & \texttt{if } \sigma_1 \geq 0 \\ \left(\frac{\sigma_1}{X_c}\right)^2 & \texttt{if } \sigma_1 < 0 \end{cases} \tag{28}$$

where f_f denotes the failure of the fiber of the composite. Whereas in the matrix, the material failure can be described by Equation (29).

$$f_m = \begin{cases} \left(\frac{\sigma_2}{Y_t}\right)^2 + \left(\frac{\tau_{12}}{S_{12}}\right)^2 & \texttt{if } \sigma_2 > 0 \\ \frac{1}{Y_c}\left(\left(\frac{Y_c}{2Q}\right)^2 - 1\right)\sigma_2 + \left(\frac{\sigma_2}{2Q}\right)^2 + \left(\frac{\tau_{12}}{S_{12}}\right)^2 & \texttt{if } \sigma_2 < 0 \end{cases} \tag{29}$$

where f_m denotes failure in the matrix of the composite, and Q stands for interlaminar shear stress. When either the fibres or the matrix fails, the material is considered broken. Therefore the use of maximum is recommended, as shown in Eq. (30):

$$f = max(f_f, f_m) \tag{30}$$

The **Puck criterion** is one of the most complex models, consisting of five different models describing the axial fiber failure (f_{ff}) and inter-fiber failure f_m. The simplified representation of fiber failure is the following (Eq (31)):

$$f_{ff} := \begin{cases} \frac{\sigma_1}{X_t} \ \texttt{if} \ \ \sigma_1 \geq 0 \\ \left|\frac{\sigma_1}{X_c}\right| \ \ \texttt{if} \ \ \sigma_1 < 0 \end{cases} \tag{31}$$

A fracture occurs if f equals one in case of f_{ff} or if it equals weakening factor (f_w) in case of f_m. The weakening factor is calculated based on the following:

$$\begin{aligned} f_w &= \begin{cases} \sqrt{1 - c\left(|\sigma_1| - sX\right)^2} \ \ \texttt{if} \ \ |\sigma_1| > sX \\ 1 \ \ \texttt{if} \ \ |\sigma_1| \leq sX \end{cases} \\ c &= \frac{1 - m^2}{[(1-s)X]^2}; \ \texttt{where} \ \ 0 \leq m \ 1, 0 \leq s < 1 \\ X &= X_t \ \ \texttt{if} \ \ \sigma_1 \geq 0 \\ X &= X_c \ \ \texttt{if} \ \ \sigma_1 < 0 \end{aligned} \tag{32}$$

Inter-fiber failure (f_m) can be calculated based on different conditions differing in three modes (Mode A, Mode B, and Mode C). Matrix tension dominated (A) mode happens when the unidirectional stress perpendicular to the fiber direction is tension or non-existent ($\sigma_2 \geq 0$):

$$\begin{aligned} f_A &= \sqrt{\left(\frac{\tau_{12}}{S_{12}}\right)^2 + \left(1 - p^+ \frac{Y_t}{S_{12}}\right)^2 \left(\frac{\sigma_2}{Y_t}\right)^2} + p^+ \frac{\sigma_2}{S_{12}} + (1 - f_w) \ \ \texttt{if} \ \ \sigma_2 \geq 0 \\ p^+ &= -\left.\frac{d\tau_{12}}{d\sigma_2}\right|_{\sigma_2 = 0}; \sigma_2 \geq 0 \end{aligned} \tag{33}$$

where p^+ is the slope of the failure envelope in $\tau_{12} - S_2$ plane, considered from $S_2 > 0$ stress quadrant.

Matrix shear dominant (B) mode is the following:

$$\begin{aligned} f_B &= \frac{1}{S_{12}} \left(\sqrt{\tau_{12}^2 + (p^- \sigma_2)^2} + p^- \sigma_2\right) + (1 - f_w) \ \ \texttt{if} \ \ \sigma_2 < 0 \ \ \texttt{and} \ \ 0 \leq \left|\frac{\sigma_2}{\tau_{12}}\right| \leq \frac{R^A}{|\tau_{12,c}|} \\ p^- &= -\left.\frac{d\tau_{12}}{d\sigma_2}\right|_{\sigma_2 = 0}; \sigma_2 \leq 0 \\ R^A &= \frac{S_{12}}{2p^-} \left(\sqrt{1 + 2p^- \frac{Y_c}{S_{12}}} - 1\right) \\ \hat{p}^- &= p^- \frac{R_A}{S_{12}} \\ \tau_{12,c} &= S_{12} \sqrt{1 + 2\hat{p}^-} \end{aligned} \tag{34}$$

where p^- is a slope similar to p^+ but considered from the $S_1 < 0$ quadrant, $\hat{p}^-$ stands for transverse-transverse slope, R^A is the transverse-transverse shear fracture resistance $\tau_{12,c}$ turning point shear stress.

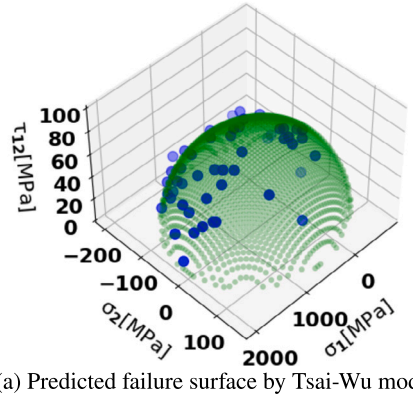
Matrix compression dominant failure (Mode C) when the shear stress is not significant:

$$f_C = \left[\left(\frac{\tau_{12}}{2(1 + \hat{p}^-) S_{12}}\right)^2 + \left(\frac{\sigma_2}{Y_c}\right)^2\right] \frac{Y_c}{-\sigma_2} + (1 - f_w) \ \ \texttt{if} \ \ \sigma_2 < 0 \ \ \texttt{and} \ \ 0 \leq \left|\frac{\tau_{12}}{\sigma_2}\right| \leq \frac{|\tau_{12,c}|}{R^A} \tag{35}$$

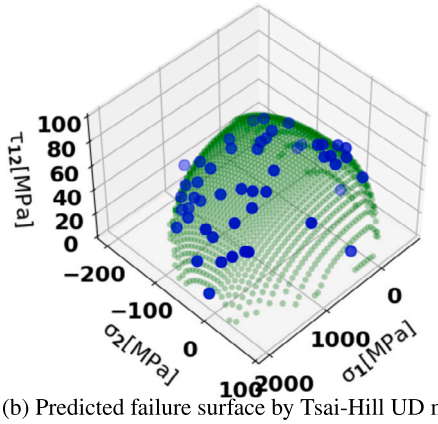
A failure occurs in Eq. (36) when the maximum of f_{ff} (consisting of Eq. (31)) and f_m (consisting of weakening factor, Mode A, Mode B, and Mode C in Eq. (32), (33), (34) and (35) respectively) reaches value one.

$$f = max(f_{ff}, f_m) \tag{36}$$

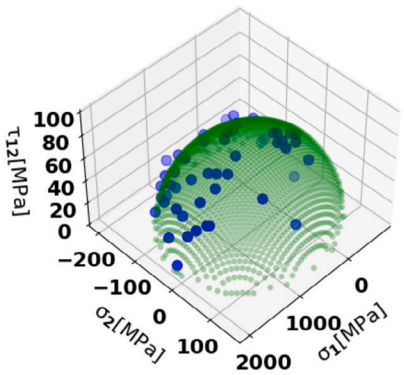
Appendix B. Failure surface of the models for Case #1



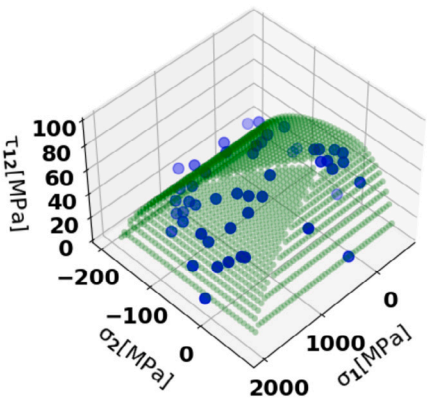
(a) Predicted failure surface by Tsai-Wu model



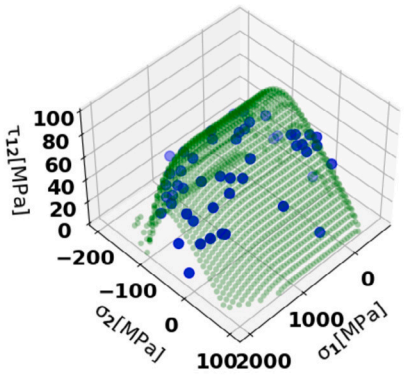
(b) Predicted failure surface by Tsai-Hill UD model



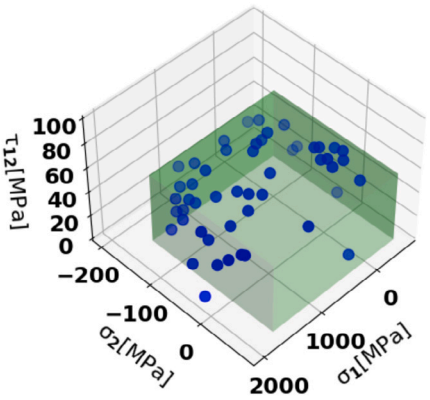
(c) Predicted failure surface by Hoffman model



(d) Predicted failure surface by Hashin model



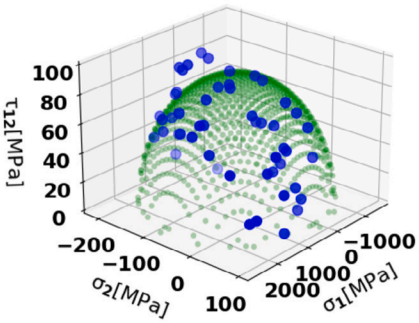
(e) Predicted failure surface by Puck model



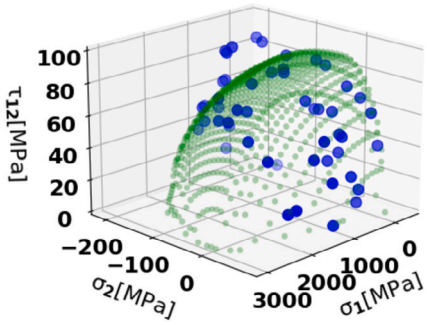
(f) Predicted failure surface by Max-Stress-Strain model

Fig. 15. Predicted failure surfaces by Tsai-Wu, Tsai-Hill UD, Hoffman, Hashin, Puck and Max-Stress-Strain models in the model selection phase.

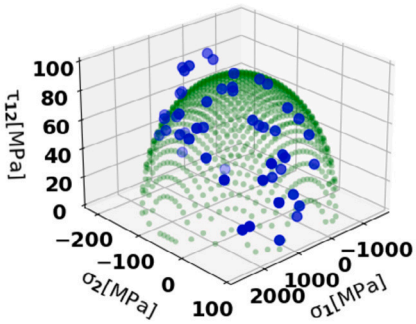
Appendix C. Failure surface of the models for Case #2



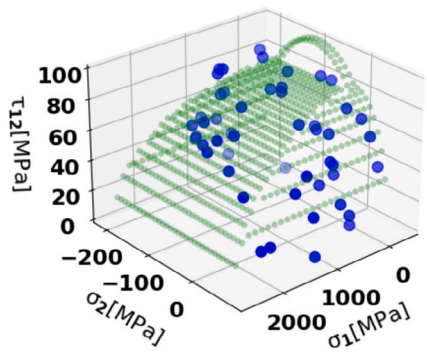
(a) Predicted failure surface by Tsai-Wu model



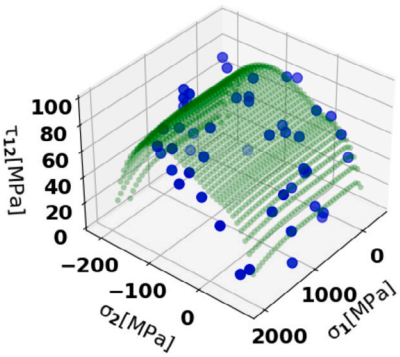
(b) Predicted failure surface by Tsai-Hill UD model



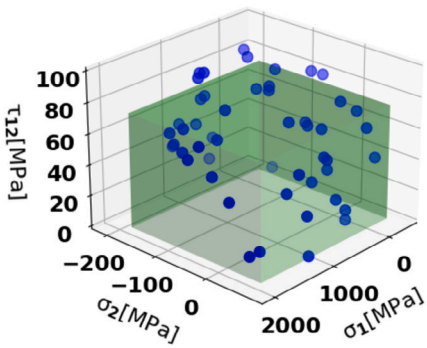
(c) Predicted failure surface by Hoffman model



(d) Predicted failure surface by Hashin model



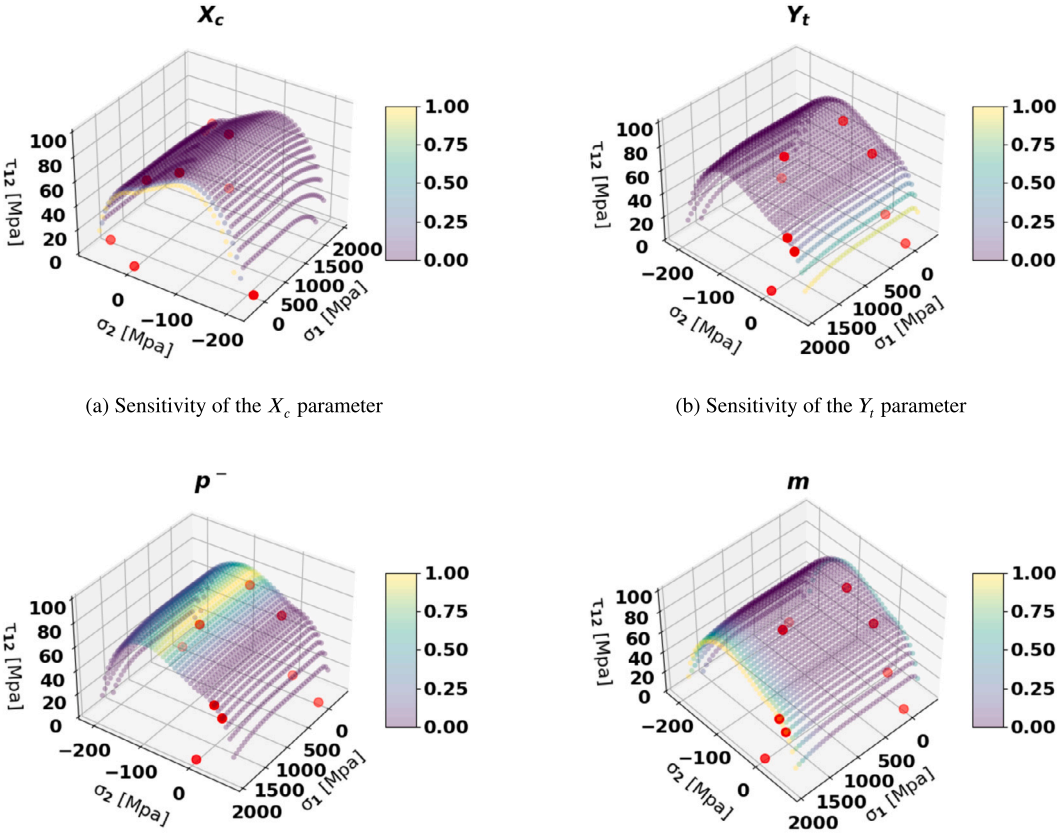
(e) Predicted failure surface by Puck model



(f) Predicted failure surface by Max-Stress-Strain model

Fig. 16. Predicted failure surfaces by Tsai-Wu, Tsai-Hill UD, Hoffman, Hashin, Puck and Max-Stress-Strain models in the model selection phase.

Appendix D. Sensitivity plots of model parameters for Case #2



(a) Sensitivity of the X_c parameter

(b) Sensitivity of the Y_t parameter

(c) Sensitivity of the p^- parameter

(d) Sensitivity of the $S12$ parameter

Fig. 17. Sensitivity of the (X_c, Y_t, p^-, m) Puck model parameters.

References

[1] P.L. Stumpff, Introduction to failure analysis, in: D.B. Miracle, S.L. Donaldson (Eds.), Composites, ASM Handbook, ASM International, Materials Park, 2001, pp. 949–950.

[2] M. Ozyildiz, C. Muyan, D. Coker, Strength analysis of a composite turbine blade using puck failure criteria, J. Phys. Conf. Ser. 1037 (2018) 042027, https://doi.org/10.1088/1742-6596/1037/4/042027.

[3] L. Zhao, W. Yang, T. Cao, H. Li, B. Liu, C. Zhang, J. Zhang, A progressive failure analysis of all-c/sic composite multi-bolt joints, Compos. Struct. 202 (2018) 1059–1068, https://doi.org/10.1016/j.compstruct.2018.05.029.

[4] M.S. Rouhi, D. Ramantani, T.E. Tay, 3d explicit simulation of bearing failure in metal–composite bolted joints, Compos. Struct. 284 (2022) 115108, https://doi.org/10.1016/j.compstruct.2021.115108.

[5] M. Merhar, Application of failure criteria on plywood under bending, Polymers 13 (2021), https://doi.org/10.3390/polym13244449.

[6] C. Prakash, K.S. Vijay Sekar, Influence of friction coefficient and failure model in 3d fea simulation of drilling of glass fiber reinforced polymer composites, in: K.S. Vijay Sekar, M. Gupta, A. Arockiarajan (Eds.), Advances in Manufacturing Processes, Springer, Singapore, Singapore, 2019, pp. 81–90.

[7] S. Li, The maximum stress failure criterion and the maximum strain failure criterion: their unification and rationalization, J. Compos. Sci. 4 (2020), https://doi.org/10.3390/jcs4040157.

[8] S.W. Tsai, E.M. Wu, A general theory of strength for anisotropic materials, J. Compos. Mater. 5 (1971) 58–80, https://doi.org/10.1177/002199837100500106.

[9] O. Hoffman, The brittle strength of orthotropic materials, J. Compos. Mater. 1 (1967) 200–206, https://doi.org/10.1177/002199836700100210.

[10] S.W. Tsai, Strength Theories of Filamentary Structures Fundamental Aspects of Fiber Reinforced Plastic Composites, Wiley-Interscience, 1968.

[11] Z. Hashin, Failure criteria for unidirectional fiber composites, J. Appl. Mech. 47 (1980) 329–334, https://doi.org/10.1115/1.3153664.

[12] A. Puck, H. Schürmann, Failure analysis of frp laminates by means of physically based phenomenological models, Compos. Sci. Technol. 62 (2002) 1633–1662, https://doi.org/10.1016/S0266-3538(01)00208-1.

[13] F.H. Bhuiyan, L. Kotthoff, R.S. Fertig, A machine learning technique to predict biaxial failure envelope of unidirectional composite lamina 3 (2018) 1451–1463.

[14] C.B. Kalayci, S. Karagoz, O. Karakas, Soft computing methods for fatigue life estimation: a review of the current state and future trends, Fatigue Fract. Eng. Mater. Struct. 43 (2020) 2763–2785, https://doi.org/10.1111/ffe.13343.

[15] G. Mustafa, A. Suleman, C. Crawford, Probabilistic first ply failure prediction of composite laminates using a multi-scale m-saf and Bayesian inference approach, J. Compos. Mater. 52 (2018) 169–195.

[16] D. Gürgünoğlu, B. Dulek, S. Gezici, Power adaptation for vector parameter estimation according to Fisher information based optimality criteria, Signal Process. 192 (2022) 108390, https://doi.org/10.1016/j.sigpro.2021.108390.
[17] O. Allix, P. Ladeveze, E. Vittecoq, Modelling and identification of the mechanical behavior of composite laminates in compression, Compos. Sci. Technol. 51 (1994) 35–42.
[18] J.H. Guillaume, J.D. Jakeman, S. Marsili-Libelli, M. Asher, P. Brunner, B. Croke, M.C. Hill, A.J. Jakeman, K.J. Keesman, S. Razavi, et al., Introductory overview of identifiability analysis: a guide to evaluating whether you have the right type of data for your modeling purpose, Environ. Model. Softw. 119 (2019) 418–432.
[19] P. Yu, M.Y. Low, W. Zhou, Design of experiments and regression modelling in food flavour and sensory analysis: a review, Trends Food Sci. Technol. 71 (2018) 202–215, https://doi.org/10.1016/j.tifs.2017.11.013.
[20] G. Franceschini, S. Macchietto, Model-based design of experiments for parameter precision: state of the art, Chem. Eng. Sci. 63 (2008) 4846–4872, https://doi.org/10.1016/j.ces.2007.11.034.
[21] J.C. Spall, Monte Carlo computation of the Fisher information matrix in nonstandard settings, J. Comput. Graph. Stat. 14 (2005) 889–909, https://doi.org/10.1198/106186005X78800.
[22] D. Chatterjee, A. Ghosh, D. Chakravorty, Nonlinear first ply failure study of laminated composite skew plates, Mater. Today Proc. 45 (2021) 4925–4930, https://doi.org/10.1016/j.matpr.2021.01.370.
[23] D. Dubois, H. Fargier, M. Ababou, D. Guyonnet, A fuzzy constraint-based approach to data reconciliation in material flow analysis, Int. J. Gen. Syst. 43 (2014) 787–809, https://doi.org/10.1080/03081079.2014.920840.
[24] I. Markovsky, S. Van Huffel, Overview of total least-squares methods, Signal Process. 87 (2007) 2283–2302, https://doi.org/10.1016/j.sigpro.2007.04.004.
[25] A.C. Atkinson, A.N. Donev, R.D. Tobias, Optimum Experimental Designs, with SAS, Oxford University Press, Oxford, 2007, pp. 134–137.
[26] L. Pronzato, A. Pazman, Optimal Design for Nonlinear Response Models, Chapman and Hall/ CRC Press, Boca Raton, 2013, pp. 187–235.
[27] M.R. Bussieck, A. Pruessner, et al., Mixed-integer nonlinear programming, SIAG/OPT Newsletter: Views & News 14 (2003) 19–22.
[28] W. McKinney, et al., Data structures for statistical computing in python, in: Proceedings of the 9th Python in Science Conference, Austin, TX, vol. 445, 2010, pp. 51–56.
[29] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S.J. van der Walt, M. Brett, J. Wilson, K.J. Millman, N. Mayorov, A.R.J. Nelson, E. Jones, R. Kern, E. Larson, C.J. Carey, İ. Polat, Y. Feng, E.W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E.A. Quintero, C.R. Harris, A.M. Archibald, A.H. Ribeiro, F. Pedregosa, P. van Mulbregt, SciPy 1.0 contributors, SciPy 1.0: fundamental algorithms for scientific computing in Python, Nat. Methods 17 (2020) 261–272, https://doi.org/10.1038/s41592-019-0686-2.
[30] M. Newville, T. Stensitzki, D.B. Allen, A. Ingargiola, LMFIT: non-linear least-square minimization and curve-fitting for Python, https://doi.org/10.5281/zenodo.11813, 2014.
[31] J. Perktold, numdiff package, 2019-7-19, http://statsmodels.sourceforge.net/0.6.0/_modules/statsmodels/tools/numdiff.html.
[32] L. Perron, V. Furnon, Or-tools, 2019-7-19, https://developers.google.com/optimization/.
[33] S.R. Soni, A comparative study of failure envelopes in composite laminates, J. Reinf. Plast. Compos. 2 (1983) 34–42, https://doi.org/10.1177/073168448300200104.