Data-driven intelligent optimisation of discontinuous composites

James M. Finley[a], Milo S.P. Shaffer[b], Soraia Pimenta[a,*]

[a] *Department of Mechanical Engineering, South Kensington Campus, Imperial College London, SW7 2AZ, United Kingdom*

[b] *Department of Chemistry, South Kensington Campus, Imperial College London, SW7 2AZ, United Kingdom*

Abstract

Fibre composites, and especially aligned discontinuous composites (ADCs), offer enormous versatility in composition, microstructure, and performance, but are difficult to optimise, due to their inherent variability and myriad permutations of microstructural design variables. This work combines an accurate yet efficient virtual testing framework (VTF) with a data-driven intelligent Bayesian optimisation routine, to maximise the mechanical performance of ADCs for a number of single- and multi-objective design cases. The use of a *surrogate model* helps to minimise the number of optimisation iterations, and provides a more accurate insight into the expected performance of materials which feature significant variability. Results from the single-objective optimisation study show that a wide range of structural properties can be achieved using ADCs, with a maximum stiffness of 505 GPa, maximum ultimate strain of 3.94%, or a maximum ultimate strength of 1.92 GPa all possible. A moderate trade-off in performance can be achieved when considering multi-objective optimisation design cases, such as an optimal ultimate strength & ultimate strain combination of 982 MPa and 3.27%, or an optimal combination of 720 MPa yield strength & 1.91% pseudo-ductile strain.

Keywords: Discontinuous reinforcement, Optimisation, Mechanical properties, Microstructures

1. Introduction

Composite materials are often used in the aerospace and automotive industries due to their high specific strength and specific stiffness [1]. However, composites often feature a complex microstructure, with many permutations of design variables (such as lay-up configurations, or constituent(s)), and many sources of variability (such as variability in local fibre spacing or fibre strength/modulus). These levels of variability and complexity make composites difficult to simulate, and even more difficult to optimise. This problem becomes even more difficult when trying to optimise multiple competing performance characteristics (e.g. strength and ductility), or when sensitivity analyses must be made to ensure robustness of mechanical performance.

Bayesian Optimisation provides a principled technique based on Bayes' Theorem to direct a

*Corresponding author.
Email address: soraia.pimenta@imperial.ac.uk (Soraia Pimenta)

Acronyms and initialisms

ADC	Aligned Discontinuous Composite
ARD	Automatic Relevance Determination
CF	Carbon Fibre
CoV	Coefficient of Variation
NGF	Nippon Graphite Fibre corporation
PAN	Poly-AcroNitrile
RVE	Representative Volume Element
SFPO	Single Fibre Pull-Out
VTF	Virtual Testing Framework

search of a global optimisation problem that is efficient and effective[2]. This method is suitable for problems where an optimal solution needs to be determined with the minimum number of *objective function* evaluations, such as selection of hyperparameters for machine learning algorithms [3], robot gait optimisation [4], or traffic congestion planning [5]. This routine works by building a probabilistic model of the *objective function*, called the *surrogate model*, which is computationally inexpensive to evaluate, and can therefore be used to efficiently estimate the *objective function* in unexplored regions of the design space; these estimations are then fed into an acquisition function, which is used to decide which are the best points to evaluate next, in order to maximise optimisation efficiency [6]. As the optimisation progresses, not only is the design space explored extensively, but also more and more training data becomes available to the *surrogate model*, improving its accuracy with each design iteration.

The use of surrogate models in data-driven intelligent optimisation routines (such as Bayesian optimisation) ensures that the next sampling point(s) can be carefully selected to minimise the number of objective function iterations; by reducing the number of objective function evaluations, data-driven intelligent optimisation routines run much faster than non-data-driven routines (such as monte-carlo, particle swarm, or genetic algorithm methods), particularly when the objective function is computationally expensive to evaluate [7]. Bayesian optimisation uses Gaussian process regression [8, 9] to create the *surrogate model*; this method fits a collection of statistical distributions to the input data to predict the objective function (which is also output as a distribution function), thus making this method well suited to stochastic modelling (such as predicting the mechanical performance of composites) [8]. Other methods (such as artificial neural networks) may also be used to create a surrogate model for data-driven optimisation [10], although these methods are traditionally deterministic, making these methods poorly suited to applications with significant variability. Modern Bayesian deep learning methods can be used to capture the effects of variability using neural networks [11], although they are computationally expensive, and thus not well-suited for most optimisation applications.

Bayesian optimisation has been used to construct reduced order models for micromechanical analysis of composites [12], and has been used to optimise the manufacturing process for short

polymer fibre composite materials [13]; however, Bayesian optimisation has never been used to perform microstructural design of composite materials. More specifically, data-driven optimisation has never been combined with efficient virtual testing methods to enable the optimal design of aligned discontinuous composite microstructures for multiple objectives, while simultaneously considering a large number of design parameters and the stochastic nature of such materials.

Aligned discontinuous composites (ADCs) are a sub-class of composite material that feature versatility in their mechanical behaviour: for example, Henry et al. [14] showed an ultimate strength of 1.4 GPa could be achieved, due to the fibres being well aligned to the loading direction; on the other hand, Finley et al. [15] demonstrated pseudo-ductile behaviour, with a maximum pseudo-ductile strain (defined by Fuller and Wisnom [16]) of 1.6 %, which was achieved through controlled fibre fragmentation and fibre-matrix debonding within the ADC, without premature fracture of the material. However, with added mechanical versatility comes a particularly complex microstructure: ADCs feature many different constituents, each with many stochastic properties [17]. The myriad permutations of constituent materials (such a hybrid fibre-types and variety of compatible matrices), and the variability in constituent properties (such as fibre overlap lengths, hybrid fibre-type arrangements, and local matrix strength variability), make it difficult to find the optimum microstructure to maximise structural performance. These extreme levels of versatility and complexity make ADCs the perfect vehicle to demonstrate the potential for data-driven intelligent optimisation of composite materials.

An accurate, yet efficient virtual testing framework (VTF) [14, 15, 17–19] has been developed to predict the structural response of hybrid [15, 17–19] or non-hybrid [14] ADC material systems, which may also feature many sources of variability and defects [17]. The VTF is significantly faster than alternative analysis methods [20], with VTF runtimes for full ADC specimens taking just minutes. However, while the VTF is an efficient analysis tool, the complexity of the material system means that an unreasonably large number of *virtual tests* would be required to fully explore the design space and find optimal microstructural designs, unless a more intelligent optimisation routine is used.

This paper combines a data-driven intelligent Bayesian optimisation routine with the efficient VTF to explore various ADC microstructural designs and maximise a range of structural performance requirements. Section 2 provides an overview of the VTF used to evaluate different material systems, and describes the Bayesian optimisation routine. Two separate optimisation campaigns will be studied: firstly, a single-objective optimisation campaign will focus on individually optimising the initial stiffness, ultimate strain, pseudo-ductile strain, ultimate strength, or yield strength; the results and discussion of the single-objective optimisation campaign are shown in Sections 3.1 and 3.2 respectively. Next, a multi-objective optimisation campaign will focus on real-world material design cases, by optimising selected combinations of the above material properties (see Section 2.2.3 for details); the results and discussion for the multi-objective optimisation

campaign are shown in Sections 4.1 and 4.2. Finally, conclusions will be drawn in Section 5.

2. Optimisation strategy

In this section, an overview of the VTF is given in Section 2.1, while the optimisation routine (and associated optimisable parameters, constraints, and data management structure) is described in Section 2.2.

2.1. The virtual testing framework

A VTF is used to simulate the stress-strain curves of various ADC material systems. The VTF is a culmination of several years of research [14, 15, 17–19], and is based on the following features required for the analysis of versatile ADC material systems:

- The VTF is a multi-scale model that simulates material behaviour at the fibre-fibre interaction, fibre, RVE, and complete specimen length scales, without the use of inaccurate homogenisation techniques [14, 17].
- Hybridisation of fibre-types is supported, including hybridisation of fibre Young's moduli, fibre diameters, and the Weibull parameters used to define stochastic fibre strengths [15, 17–19]. While the VTF can consider hybrid composites with any combination of two fibre-types (e.g. high-strength carbon and high-modulus carbon [19] or high-modulus carbon and E-glass fibres [15, 17]), it should be noted that the two different fibre-types will be referred to as the 'glass' and 'carbon' fibres throughout this paper, with the 'glass' fibre-type denoting the fibre-type with the lowest Young's modulus of the two.
- The most relevant sources of variability are modelled in the ADC specimens, including variability in matrix strength, inter-fibre distance, fibre strength, fibre stiffness, and hybrid fibre-type arrangement [14, 17, 19].
- Various sub-critical damage events are predicted to occur during the simulation, such as matrix softening, matrix debonding, fibre fragmentation, and fibre failure [17]. Tracking of these damage events helps to determine which sources of damage are contributing to final failure of the material system.
- Final failure is governed by a non-linear fracture mechanics failure criterion, which has been proven to accurately determine final failure of a variety of ADC material system designs, both for ductile and brittle failure modes [17].

Each ADC microstructure is defined by the material properties shown in Table 1 and 2. Carbon and glass fibre properties are defined by selecting one of 15 commercially available fibre-types, details of which are shown in Table 3. Each fibre-type is characterised by its fibre diameter ϕ_{f}, expected Young's modulus $\mathbf{E}\,(E_{\mathrm{f}})$, Weibull reference strength X_{f}, Weibull modulus m, and Weibull

reference length l_w; each fibre-type also belongs to a fibre group (i.e. Pitch-CF, PAN-CF, E-glass, or S-glass), which denotes the constituent materials and method of manufacture.

The VTF outputs a stress-strain curve of the mechanical response of each material system that is analysed. In this work, each stress-strain curve is characterised by five *key performance characteristics* that are used to establish the optimality of any material system:

1. Initial stiffness, defined as the gradient of the stress-strain curve at the start of the simulation.
2. Ultimate strength, defined as the maximum stress achieved by the material.
3. Yield strength, defined as the 0.1% strain offset yield strength, as per Fuller et al. [16].
4. Ultimate strain, defined as the maximum strain achieved by the material.
5. Pseudo-ductile strain, defined as the difference between the ultimate strain and the elastic strain (the elastic strain is calculated as the ultimate strength divided by the initial stiffness) [21].

In addition to these five *key performance characteristics*, the maximum stress drop in the stress-strain curve simulated by the VTF is also recorded, as will be justified in Section 2.2.2. These five *key performance characteristics* (and the maximum stress drop) will be the output of the VTF in the scope of the optimisation process, as detailed in the next section.

2.2. Bayesian optimisation algorithm

The VTF is used as the *objective function* for the Bayesian optimisation routine, which attempts to find the best combination of optimisable input variables to maximise one (or more) of the *key performance characteristic(s)*. Seven optimisable variables are used to modify the ADC microstructure, as per Table 1; each of the continuous optimisable variables has a range defined by typical values seen in commercially available material systems.

2.2.1. Single-objective Bayesian optimisation routine

The Bayesian optimisation routine uses a *surrogate model* to approximate the *objective function* and hence find optimal material system designs with far fewer *objective function* evaluations than alternative methods [37–39]. The Bayesian optimisation routine follows the flowchart shown in Figure 1 with the steps described in detail below:

1. Before the start of the optimisation process there is no *surrogate model* for the *objective function* (as there is no observed data yet). The first step of the Bayesian optimisation process is to define the *prior*, which in this case is a collection of Gaussian distributions that may fit the *objective function* within the specified ranges of the optimisable variables [8, 9].
2. If at the start of the optimisation process the observed data contains between zero and four evaluations of the VTF, the VTF should sample random combinations of inputs to generate observed data, until the size of the observed data is sufficiently large. Otherwise, continue to Step 3.

Table 1: Optimisable input variables that were used in this study.

Optimisable variable	Variable type	Range
Carbon fibre type, F_c	Categorical	15 fibre types (see Table 3)
Glass fibre type, F_g	Categorical	15 fibre types (see Table 3)
Carbon volume ratio, V_c	Double	0.0 - 1.0
Fibre length, l_f (mm)	Double	0.5 - 12.0
Matrix shear strength, $\mathbf{E}(S_m)$ (MPa)	Double	40 - 100
Matrix shear modulus, G (GPa)	Double	1.0 - 1.8
Mode-II fracture toughness, $\mathcal{G}^c_{II_m}$ $\left(kJ/m^2\right)$	Double	0.6 - 1.0

Table 2: Constant settings used for simulation with the VTF.

Setting	Value	Reference
Fibre volume fraction, V_f	0.55	-
Specimen length, l_s (mm)	50	-
No. fibres in cross-section, n_f	6400	-
Fibre modulus CoV, CoV (E_f)	0.05	[17]
Matrix strength CoV, CoV (S_m)	0.15	[17]
Cross-section shape	Square	-
Pull-out friction stress (SFPO), τ^0_μ (MPa)	10	[22, 23]

Table 3: Fibre material data used for this optimisation study. All reference strength values are scaled [24] to a Weibull reference length of 6 mm (which is approximately the midpoint in the range of fibre length values).

Manufacturer	Fibre-type	Fibre group	Expected Young's modulus (GPa)	Weibull reference strength (MPa)	Weibull modulus	Diameter (μm)	Reference
Tenax	C124	PAN-CF	225	4939	5.72	7.00	[14, 25, 26]
Sigri Great Lakes	C320	PAN-CF	230	3273	5.89	3.90	[27, 28]
Tenax	HTA5131	PAN-CF	238	2829	5.65	7.00	[29, 30]
Toray CA	M40B	PAN-CF	359	1714	6.80	7.05	[31]
Toray CA	M60JB	PAN-CF	521	2202	5.80	5.13	[31]
Toray CA	T1000GB	PAN-CF	291	3734	5.90	5.03	[31]
Toray CA	T300	PAN-CF	221	2244	7.00	7.39	[31]
Toray CA	T800H	PAN-CF	294	3437	5.50	5.00	[32, 33]
Mitsubishi	K13D	Pitch-CF	940	1777	4.20	11.72	[31]
Amoco	P120J	Pitch-CF	820	1972	3.85	9.60	[27, 34]
Amoco	P75S	Pitch-CF	517	1837	6.45	5.50	[27, 35]
NGF	XN-05	Pitch-CF	41	803	7.90	9.64	[31]
NGF	XN-90	Pitch-CF	876	2105	5.00	10.03	[31]
Vetrotex	C100	E-glass	73	2726	5.76	7.00	[15, 26]
Owens Corning	FliteStrand S ZT	S-glass	88	1969	4.00	9.00	[36]

3. The observed data consists of a matrix of sampled *key performance characteristics* (and the maximum stress-drop, see Section 2.2.2) $\boldsymbol{Y}(\boldsymbol{X})$, where $\boldsymbol{X}$ is the associated inputs that are used to achieve each observed data point $\boldsymbol{Y}$. The observed *objective function* output(s) can then be determined by extracting the relevant *key performance characteristic(s)* (i.e. the sub-set of key performance characteristics to be optimised) from the matrix of sampled outputs.
4. Gaussian process regression [8, 9] is then used to assign probabilities that each of the Gaussian distributions in the *prior* fits with the *objective function*; these probabilities are called the *marginal likelihood*. The product of the *prior* (i.e. the collection of Gaussian distributions that may fit the *objective function* in the required range) and the *marginal likelihood* (i.e. the probabilities that each of the Gaussian distributions within the *prior* fit the *objective function*, given the observed data) forms the *surrogate model* (otherwise known as the posterior distribution), which can be used to approximate the *objective function* (and hence predict the relevant *key performance characteristic(s)*). Several authors [8, 9] describe Gaussian process regression in great detail, and specific settings used for this paper can be found in Section 2.2.4.
5. Once the *surrogate model* is created, it is sampled a large number of times (in this paper, $10,000$ times per iteration) to estimate both the values of the *objective function* and the levels of uncertainty in the *surrogate model*, over the whole design space (which is defined by the range of the optimisable input parameters in Table 1). The *surrogate model* runs many times faster than the *objective function*, hence the design space can be explored rapidly.
6. An *acquisition function* is then used to determine the *best next point* to sample, so as to maximise the efficiency of the optimisation routine. In order to find global optima, a balance between exploitation and exploration of the sampled space must be made; for this reason the 'expected improvement plus' *acquisition function* [40] is used, because this *acquisition function* looks for points that maximise the predicted performance, yet adds a penalty to any prospective sampling point if it lies too close to points that have been previously sampled.
7. The *best next point* is then sampled via a *virtual test* using the more computationally expensive, yet more accurate VTF; the inputs and outputs for the best estimated point are added to the matrices of observed inputs $\boldsymbol{X}$ and observed *key performance characteristics* $\boldsymbol{Y}(\boldsymbol{X})$.
8. Steps 3 to 7 are repeated using the additional observed data from the most recent *virtual test*. As the optimisation routine progresses, not only are more areas of the design space explored, but also the accuracy of the *surrogate model* improves as more observed data becomes available. The *best observed point* $\boldsymbol{Y}^{\star}(\boldsymbol{X}^{\star})$ is determined by finding the maximum *key performance characteristic* being optimised from the observed *key performance characteristcs* $\boldsymbol{Y}$, where $\boldsymbol{X}^{\star}$ are the associated inputs used to achieve the *best observed point*. The *best estimated point* $\dot{\boldsymbol{Y}}^{\star}\left(\dot{\boldsymbol{X}}^{\star}\right)$ is found by sampling the *surrogate model* using the inputs in $\boldsymbol{X}$ to find the best point (which in this paper is defined as the highest expected *key performance*

characteristic) from the estimated *key performance characteristics* $\dot{\boldsymbol{Y}}$, and where $\dot{\boldsymbol{X}}^{\star}$ is the associated inputs used to achieve the *best estimated point*.

9. The optimisation routine finishes when convergence is achieved, i.e. when either (i) a set number of iterations are achieved, or (ii) no improvement in the best estimated point(s) can be found after a set number of iterations.

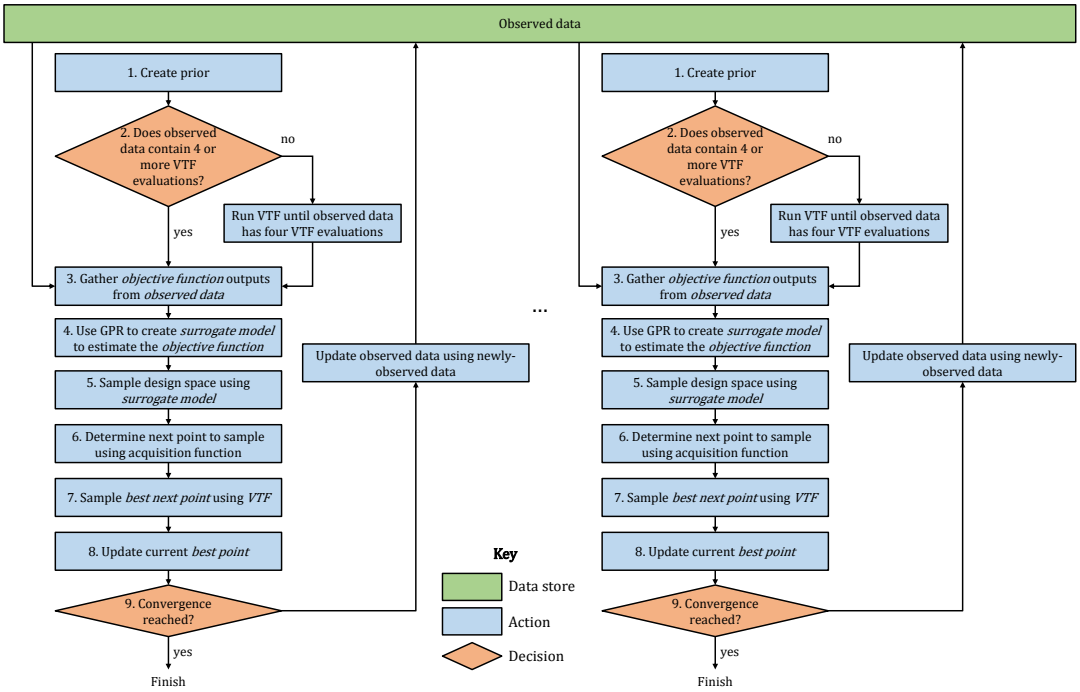


Figure 1: A flowchart showing the Bayesian optimisation routine described in Section 2.2.1, and how the Bayesian optimisation routine can be run in parallel using a common observed data store.

2.2.2. Constraints

The stress-strain curve of an ADC is not necessarily monotonically increasing; in fact, hybrid ADCs may feature a significant stress-drop after the pseudo-yield point [15], which makes the structural response of the material unstable under real-world loading conditions. In some cases, the optimisation routine might yield a design which has one or more optimal *key performance characteristic(s)*, but features a sudden large stress-drop in the stress-strain curve, as per the design shown in Figure 2f. The maximum permissible stress-drop in this study is therefore constrained to 50 MPa to ensure a stable stress-strain response.

In order to constrain the maximum stress-drop in the stress-strain curve, the maximum stress-drop for each *virtual test* is recorded (using the output from the most recent *virtual test*), and applied as a non-deterministic coupled constraint as part of the bayesopt() [41] function in Matlab. This type of constraint uses Gaussian process regression [8, 9] (with specific settings used for this paper detailed in Section 2.2.4) to model the *probability of feasibility* (i.e. the probability that a proposed combination of inputs will have a sufficiently small stress-drop), so that if the

probability of feasibility at the proposed combination of inputs is low, then a penalty is applied to the *acquisition function*. This process ensures that points with a low *probability of feasibility* are not evaluated by the VTF, thus avoiding expensive evaluation of points that are likely to feature large stress-drops.

2.2.3. Multi-objective optimisation

In many real-world structural applications there are several competing *key performance characteristics* which need to be optimised concurrently. Therefore, three multi-objective optimisation design cases are considered in this work; each multi-objective design case aims to maximise two *objective functions* (each relating to a different *key performance characteristic*), inspired by real-world requirements for applications in lightweight structures:

1. Maximum ultimate strength and initial stiffness (a conventional requirement for general aerospace applications [42]).
2. Maximum ultimate strength and ultimate strain (for design cases that require a high work of fracture, e.g. under crash conditions [43]).
3. Maximum yield strength and pseudo-ductile strain (for damage tolerant design [44]).

In this work, a modified version of the ε-constraint method is developed to perform the multi-objective optimisation campaign, in order to support the non-convex and stochastic nature of the Pareto fronts associated with the above multi-objective design cases. This method works by constraining the first *objective function* to be larger than an imposed minimum value, while maximising the second *objective function*; the minimum limit of the constrained *objective function* is then incrementally increased to enable a Pareto optimal set of solutions to be found, even for convex Pareto fronts. More details on Pareto optimality can be found in Section 4.1.1, while more details on the modified ε-constraint method can be found in Appendix A.

2.2.4. Optimisation details and data management structure

In this paper, categorical variables are used to define the fibre-types that are available. In order to perform Gaussian process regression using categorical variables, a process similar to one-hot encoding [45] is used to select the properties of the posterior distribution according to the fibre-types.

Another feature of this paper is that the *objective function* is non-deterministic, i.e. for the same inputs, the VTF will always produce a distribution of stress-strain responses. The stochastic nature of the *objective function* is preserved by adding an independent noise component to the ARD Matérn 5/2 kernel function [8], and is implemented in Matlab by setting the *IsObjectiveDeterministic* and *AreCoupledConstraintsDeterministic* flags to 'false' as part of the bayesopt package [41].

All settings for the Bayesian optimisation routine described in this paper are summarised in Table 4, including a reference to the Section in which they are used.

Table 4: Settings used for the Bayesian optimisation routine in this paper. The exploration was set to the default value of 0.5, although the results were insensitive to changes in the exploration ratio between values of 0.2 - 0.9. Settings in italics are for the bayesopt package in Matlab [41].

Single- and multi-objective optimisation settings		
Setting name	Setting value	Section
Kernel function	ARD Matérn 5/2 [8]	2.2.4
Acquisition function	Expected improvement plus [40]	2.2.1
Exploration ratio	0.5*	2.2.4
IsObjectiveDeterministic	False	2.2.4
AreCoupledConstraintsDeterministic	False	2.2.4
Multi-objective optimisation settings		
Setting name	Setting value	Section
Increment of guiding constraint	$1/40^{\text{th}}$ of secondary objective range	Appendix A

Every time the VTF is used to evaluate the *objective function*, the input material parameters, damage events, fracture information, and stress-strain curve (including the resulting *key performance characteristics* and stress-drop data) for each virtual test are stored in a user-defined *virtual specimen* object. The *virtual specimen* data is collected for all design optimisation cases, post-processed, and stored in the observed data store. The observed data can then be used as observed data to improve the accuracy of *surrogate models* in future Bayesian optimisation iterations.

The advantage of this data management structure is that, if multiple Bayesian optimisation tasks are run concurrently, the data can be shared between the optimisations at regular intervals (as per Figure 1), enabling faster training of the *surrogate models*, and hence a more efficient Bayesian optimisation process. Sharing of the observed data is particularly powerful when combined with the modified ε-constraint multi-objective optimisation method (see Section 2.2.3 and Appendix A), as optimisation can be performed with different levels of constraint on the first *objective function* simultaneously, all whilst using the same observed data.

3. Single-objective optimisation

This section relates to the single-objective optimisation campaign, which focuses on optimising the *key performance characteristics* (the initial stiffness, ultimate strain, pseudo-ductile strain, ultimate strength, or yield strength) individually. All five optimisations were performed in parallel (using the data management structure described in Section 2.2.4), with observed data being shared between parallel optimisation tasks every 5 *virtual tests*. The results and discussion of the single-objective optimisation campaign are shown in Sections 3.1 and 3.2 respectively.

3.1. Single-objective optimisation results

3.1.1. Single-objective optimal designs

The single-objective optimisation process is used to find the *best estimated points* $\dot{\boldsymbol{Y}}^{\star}\left(\dot{\boldsymbol{X}}^{\star}\right)$ for all five *key performance characteristics*; the *best estimated points* are favoured for this optimisation

task, as opposed to the *best observed points* $\boldsymbol{Y}^{\star}(\boldsymbol{X}^{\star})$, as the *best estimated points* are found to be a more accurate method for identifying optimal microstructural designs when the material systems feature inherent variability (this will be discussed in more detail in Section 4.2.2). Figure 2 shows the optimal stress-strain curves for the five single-objective optimisation design cases, while Table 5 shows the associated input parameters $\dot{\boldsymbol{X}}^{\star}$ that are used to create the optimal designs. Figure 2f shows how the optimal stress-strain curve of the maximum pseudo-ductile strain design case changes when the maximum stress-drop constraint (see Section 2.2.2 for details) is relaxed.

Table 5: Optimal inputs $\dot{\boldsymbol{X}}^{\star}$ for the five single-objective optimisation design cases. Values in brackets show where each continuous optimisable variable sits within the design range defined in Table 1 (i.e. values of 0%, 50%, and 100% mean that an optimisable variable lies at the lower bound, middle, or upper bound of the range).

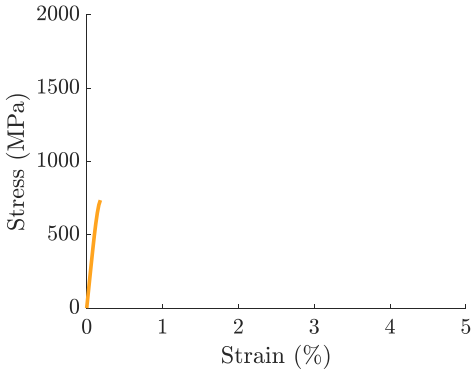
	Single-objective optimal design case				
Optimisable variable	Initial stiffness	Ultimate strain	Pseudo-ductile strain	Ultimate strength	Yield strength
Carbon Fibre, F_{c} (see Table 3 for details)	K13D	K13D	K13D	C124	C124
Glass fibre, F_{g} (see Table 3 for details)	-	C100	C100	-	-
Carbon volume ratio, V_{c}	1.00 (100%)	0.46 (46%)	0.60 (60%)	1.00 (100%)	1.00 (100%)
Fibre length, l_{f} (mm)	11.7 (97%)	4.42 (34%)	4.52 (35%)	12.0 (100%)	10.7 (89%)
Matrix shear strength, $\mathbf{E}(S_{\mathrm{m}})$ (MPa)	52.3 (21%)	40.7 (0%)	77.4 (62%)	47.5 (13%)	98.9 (98%)
Matrix shear modulus, G_{m} (GPa)	1.75 (94%)	1.11 (14%)	1.73 (91%)	1.73 (91%)	1.08 (10%)
Matrix mode-II fracture toughness, $\mathcal{G}^{\mathrm{c}}_{\mathrm{IIm}}$ (kJ/m^2)	0.77 (43%)	0.75 (38%)	0.84 (60%)	0.93 (83%)	0.78 (45%)

3.2. Single-objective optimisation discussion

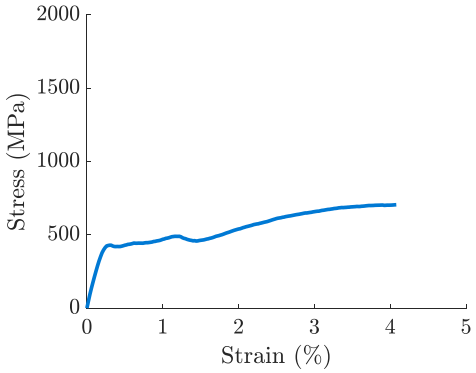
A large range of mechanical properties can be achieved through the use of ADCs (see Figures 2a to 2e), with a maximum stiffness of 505 GPa, maximum ultimate strain of 3.94%, maximum pseudo-ductile strain of 3.64%, maximum ultimate strength of 1.92 GPa, or a maximum yield strength of 1.83 GPa; all of these optimal cases have been found using input parameters that are representative of commercially-available fibres and matrix systems. The different inputs required to achieve these different mechanical properties are shown in Table 5, with the general trends outlined below:

- Long-fibre, non-hybrid ADCs are ideal for maximising stiffness (see Table 5). This is because longer fibres tend to feature longer fibre-fibre overlap lengths, which maximises load transfer between the fibres. The stiffness of a fibre-fibre interaction is also maximised when the diameter and Young's modulus of both fibres are the same [18], hence why maximum initial stiffness is achieved with non-hybrid high modulus fibre reinforcement (the K13D fibre, see Table 3). Maximising the matrix shear modulus further increases the stiffness of the fibre-fibre interactions, which explains why this input is so close to the upper bound of the allowable shear stiffness range (see Table 1).

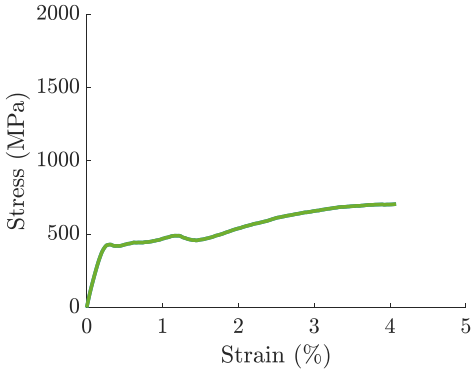
- Long-fibre non-hybrid ADCs are also useful when maximising either ultimate strength or yield strength (see Table 5). This is because a combination of long, yet high strength fibres (such as the C124 fibre, see Table 3) help to prevent both debonding damage and fibre fragmentations, and hence maximise the strength of the ADC; moreover, hybrid ADCs often fail prematurely in areas where the lower failure strain fibres are grouped together [15], hence



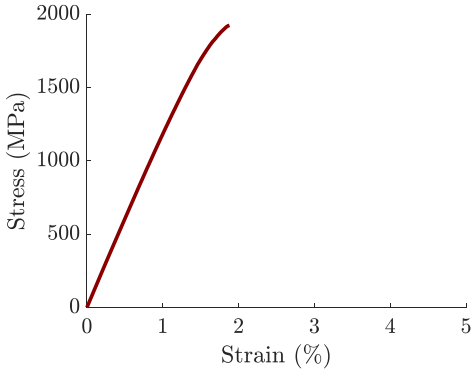
(a) Stress-strain curve for the specimen design with the max. initial stiffness.



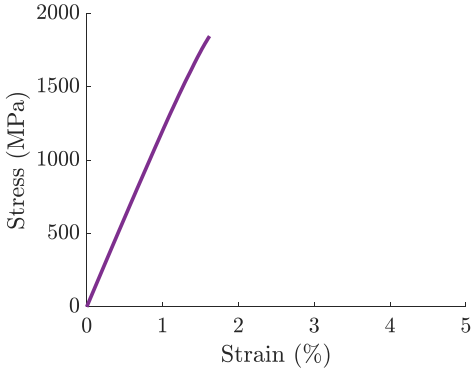
(b) Stress-strain curve for the specimen design with the max. ultimate strain.



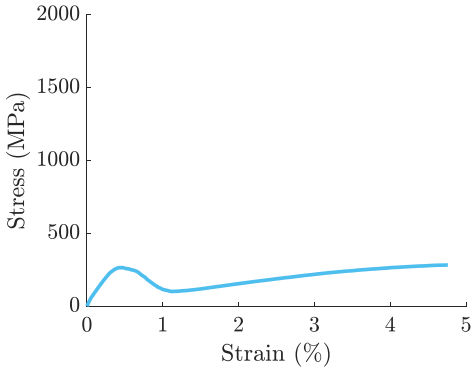
(c) Stress-strain curve for the specimen design with the max. pseudo-ductile strain.



(d) Stress-strain curve for the specimen design with the max. ultimate strength.



(e) Stress-strain curve for the specimen design with the max. yield strength.



(f) Stress-strain curve for the specimen design with the max. pseudo-ductile strain, without the maximum stress-drop coupled constraint applied.

Figure 2: Stress-strain curves for the specimen designs that maximise the expected initial stiffness, ultimate strain, pseudo-ductile strain, ultimate strength, and yield strength. The stress-strain curve in Figure 2f shows how the maximum pseudo-ductility solution would look if the maximum stress-drop coupled constraint was not applied.

hybridisation is avoided when maximising strength alone. Trends in matrix properties tend to diverge between high ultimate strength and yield strength designs: ultimate strength requires low matrix strength and high mode-II fracture toughness (to maximise translaminar fracture toughness and prevent premature final failure of the virtual specimen, see Appendix B), while the maximum yield strength design requires a low matrix stiffness and high matrix strength to delay debonding and prevent yielding of the ADC for as long as possible.

- Hybrid ADCs are best suited for maximising ultimate strain and pseudo-ductile strain (see Table 5). This is because hybridisation of fibre-types promotes debonding and fragmentation of the high modulus, low-elongation fibres, hence promoting maximum levels of sub-critical damage, and increasing the non-linearity of the material stress-strain response. The maximum ultimate strain design again features a low matrix shear strength: this maximises translaminar fracture toughness (see Appendix B) and hence delays final failure; a low matrix shear strength also increases debonding, which further increases the non-linearity of the stress-strain response. On the other hand, the matrix strength is much higher for the maximum pseudo-ductility design, and this design also features a higher proportion of high modulus fibres and a higher matrix shear stiffness; these differences in the microstructural design all work to increase the initial stiffness of the material, and hence maximise the pseudo-ductile strain by minimising the proportion of elastic strain in the material response (as per the definition of pseudo-ductile strain in Section 2.1).

It should be noted that a further increase in the maximum ultimate strain and pseudo-ductile strain could be achieved if the maximum stress-drop coupled constraint (see Section 2.2.2) was relaxed (see Figure 2f). However, relaxation of the maximum stress-drop constraint results in a significant stress-drop, which could result in premature failure if the specimen were loaded under real-world conditions (e.g. non-uniform or load-controlled conditions, instead of under displacement control).

4. Multi-objective optimisation

This section relates to the multi-objective optimisation campaign, which focuses on optimising combinations of the *key performance characteristics* (maximum ultimate strength & initial stiffness, maximum ultimate strength & ultimate strain, and maximum yield strength & pseudo-ductile strain), as per Section 2.2.3. The results and discussion of the multi-objective optimisation campaign are shown in Sections 4.1 and 4.2 respectively.

4.1. Multi-objective optimisation results

4.1.1. Observed Pareto fronts

Multi-objective optimisation often yields multiple solutions that could be considered optimal. The Pareto-optimal front of solutions is defined as the set of solutions that cannot improve any

objective without degrading one or more of the other objectives [46]. The *observed Pareto fronts* (which contain all of the *best observed points* $\boldsymbol{Y}^\star(\boldsymbol{X}^\star)$) were found for all three multi-objective design cases, by following the domination algorithm [47] described below:

1. Load all of *key performance characteristics* from the observed data (see Section 2.2.4) and extract the two *objective function* outputs that are relevant for the multi-objective design case in question (e.g. ultimate strength & initial stiffness). This loaded data becomes the *working data set*.
2. Find the optimal data point that maximises the first *objective function* in the *working data set* (e.g. maximum initial stiffness) and add it to the *observed Pareto-optimal front*.
3. Delete any data points from the *working data set* that have a lower value for the second *objective function* (e.g. ultimate strength) than the optimal data point in Step 2, and also delete the optimal data point from the *working data set*.
4. Repeat Steps 2 and 3 until there are no data points left in the working data set. The *observed Pareto front* $\boldsymbol{Y}^\star(\boldsymbol{X}^\star)$ is now complete for this multi-objective design case.
5. Repeat Steps 1 to 4 for all three multi-objective design cases to create the three *observed Pareto fronts* .

The *observed Pareto fronts* are plotted in Figures 3a, 3c, and 3e.

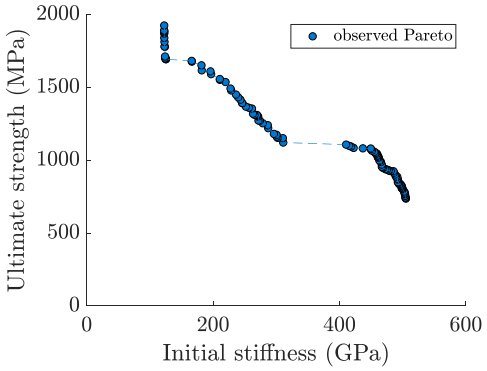
4.1.2. Variability analysis and estimated Pareto fronts

The output of the VTF is non-deterministic, and therefore will produce different results every time it is run with the same inputs. In order to establish the influence of variability on the *observed Pareto fronts*, all points in the *observed Pareto front* $\boldsymbol{Y}^\star(\boldsymbol{X}^\star)$ are re-run 25 times using the VTF and the associated inputs to the *observed Pareto front* $\boldsymbol{X}^\star$; the results of this study are plotted next to the *observed Pareto front*, as shown by the 'variability study' in Figures 3b, 3d, and 3f.

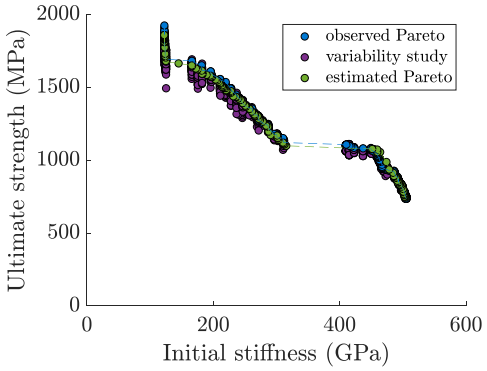
Next, the Pareto front process from Section 4.1.1 is used again, except that for Step 1 the *surrogate models* (created during the Bayesian optimisation routine, see Section 2.2) of the relevant *objective functions* (e.g. maximum ultimate strength & initial stiffness) are sampled at the same inputs as the observed data $\boldsymbol{X}$, to give the *estimated working set*. The domination algorithm from Section 4.1.1 is used again, except with the *estimated working set* instead of the *observed working set* to give the *estimated Pareto front* $\dot{\boldsymbol{Y}}^\star\left(\dot{\boldsymbol{X}}^\star\right)$ at the observed input points (from here-on described as the *estimated Pareto front*). The *estimated Pareto fronts* for the three multi-objective design cases are shown in Figures 3b, 3d, and 3f.

4.1.3. Optimal inputs for Pareto fronts

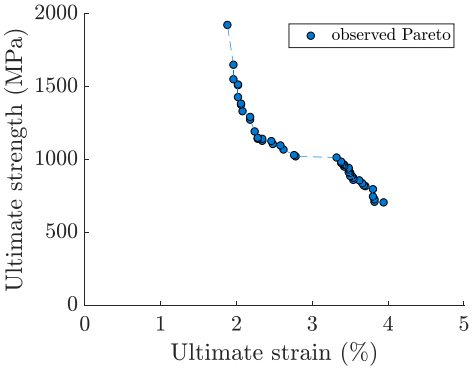
The sub-plots in Figures 4 to 6 collectively show the combination of input parameters $\dot{\boldsymbol{X}}^\star$ that are required to generate each Pareto-optimal solution on the corresponding *estimated Pareto front* $\dot{\boldsymbol{Y}}^\star\left(\dot{\boldsymbol{X}}^\star\right)$ for all three multi-objective design cases. The categorical inputs for the fibre-types are arranged by their fibre group (defined in Table 3) in an effort to simplify the plots in



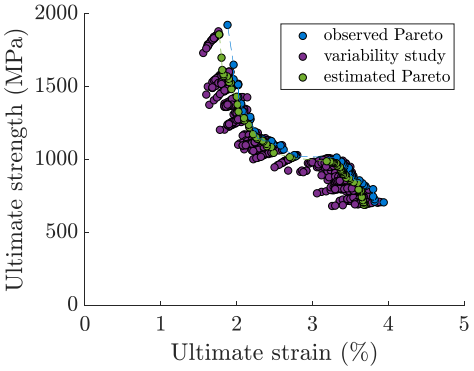
(a) Observed Pareto front for the ultimate strength & initial stiffness design case.



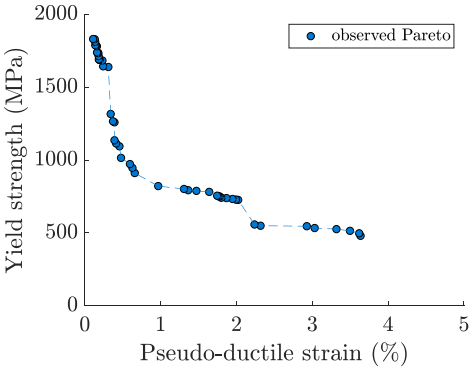
(b) Variability study of the *observed Pareto front* , for the ultimate strength & initial stiffness design case. The *estimated Pareto front* is also shown.



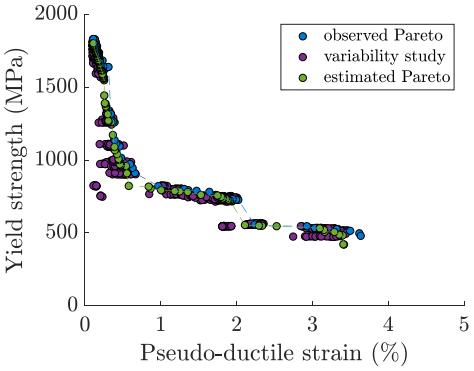
(c) Observed Pareto front for the ultimate strength & ultimate strain design case.



(d) Variability study of the *observed Pareto front* , for the ultimate strength & ultimate strain design case. The *estimated Pareto front* is also shown.



(e) Observed Pareto front for the yield strength & pseudo-ductile strain design case.



(f) Variability study of the *observed Pareto front* , for the yield strength & pseudo-ductile strain design case. The *estimated Pareto front* is also shown.

Figure 3: Observed Pareto fronts (see Section 4.1.1), variability studies, and *estimated Pareto fronts* (see Section 4.1.2) for the three multi-objective design cases considered.

Figures 4a, 5a. and 6a, as the properties of the different fibre-types within the same fibre group are generally similar.

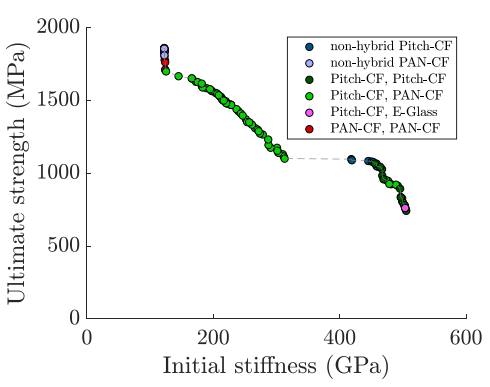
4.1.4. Sensitivity analysis

A sensitivity analysis is performed on the optimal designs from all three *estimated Pareto fronts* $\dot{\boldsymbol{Y}}^{\star}\left(\dot{\boldsymbol{X}}^{\star}\right)$, in order to determine the robustness of each of the design solutions to small changes (or uncertainty) in the associated input parameters $\dot{\boldsymbol{X}}^{\star}$. Direct calculation of many evaluations of the VTF with small differences in input parameters would be intractable, therefore the authors propose an alternative approach that exploits an intrinsic advantage of the Bayesian optimisation approach: as well as being able to estimate the mean Pareto front for the multi-objective design cases, the *surrogate model* can also be used to perform a local sensitivity analysis without the need to further evaluate the *objective function(s)*. A proposed method for the sensitivity analysis is shown out for all three multi-objective design cases, taking the maximum initial stiffness & maximum ultimate strength multi-objective design case as an example:

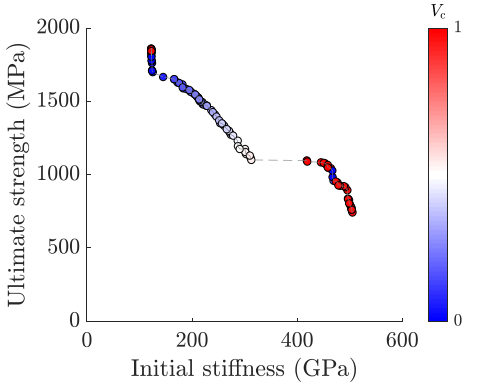
1. Load the associated inputs $\dot{\boldsymbol{X}}^{\star}$ that are used to create the *estimated Pareto front* $\dot{\boldsymbol{Y}}^{\star}\left(\dot{\boldsymbol{X}}^{\star}\right)$ for the maximum initial stiffness & maximum ultimate strength multi-objective design case.
2. Modify the input data $\dot{\boldsymbol{X}}^{\star}$ to individually change each of the continuous optimisable variables (i.e carbon fibre volume ratio, fibre length, expected matrix shear strength, shear modulus, or mode-II interlaminar fracture toughness) by $\pm 10\%$, thus producing the *sensitivity input data set* (which, since there are 5 continuous optimisable variables which experience a positive and negative modification, contains 10 times the original number of input data points).
3. Use the *surrogate models* to predict the expected ultimate strength and initial stiffness of the *sensitivity input data set*; this creates the *sensitivity working data set*.
4. Multiply the *sensitivity working data set* by -1 and apply the same domination algorithm as in Section 4.1.1, to calculate the worst-case set of solutions from the sensitivity analysis.
5. Multiply the resulting Pareto-optimal set by -1. This will determine the *sensitivity analysis Pareto front*, i.e. the front of points that are most detrimentally affected by the sensitivity analysis.
6. Repeat Steps 1 to 5 for all three multi-objective design cases.

The sensitivity analysis is shown in Figures 7a, 7c, and 7e, whereby the *sensitivity working data set* is plotted for each *estimated Pareto front* to show how individual modifications to the input parameters affects each point on the *estimated Pareto front*. Figures 7b, 7d, and 7f show the *sensitivity analysis Pareto fronts*, i.e. the least optimal solutions that might arise after the sensitivity analysis is performed.

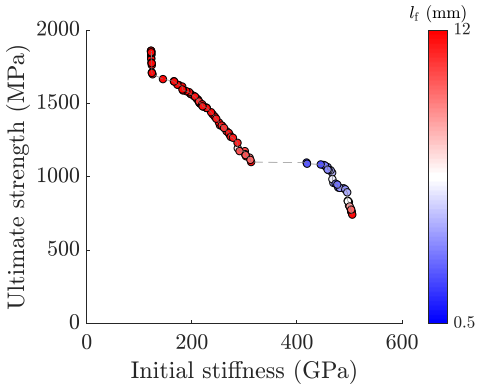
The sensitivity analysis for the maximum ultimate strength & initial stiffness design case indicates a significant reduction in both ultimate strength and initial stiffness in the lower-right portion of the Pareto front (see Figure 7a); this region of the Pareto front features a significant number



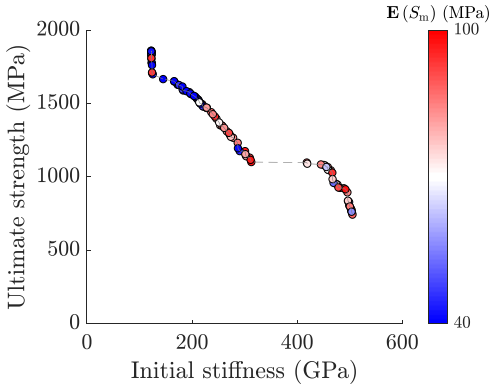
(a) Material system.



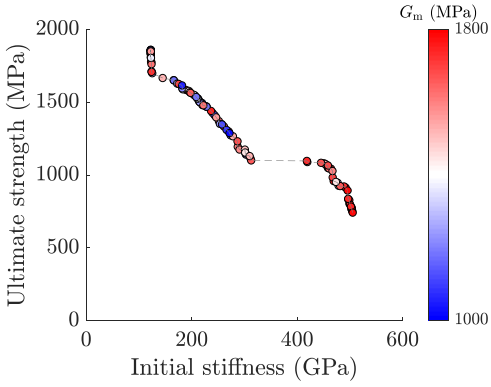
(b) Carbon volume ratio.



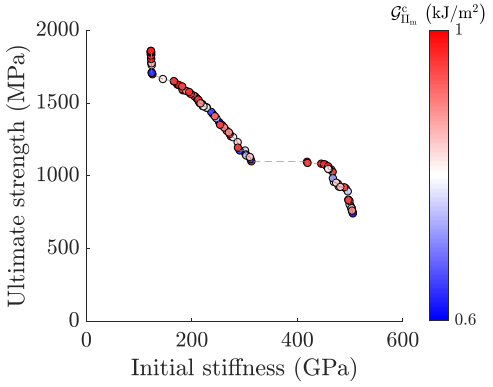
(c) Fibre length.



(d) Expected matrix strength.

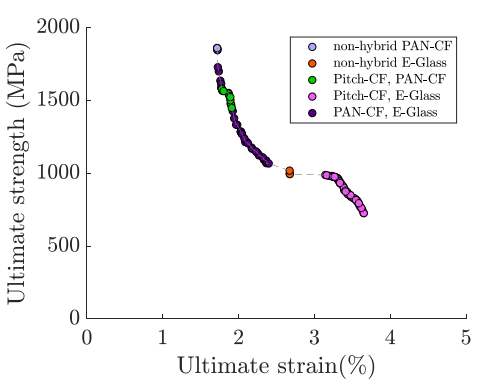


(e) Matrix shear modulus.

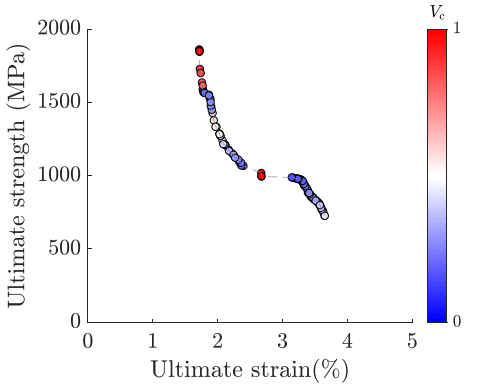


(f) Mode-II interlaminar fracture toughness.

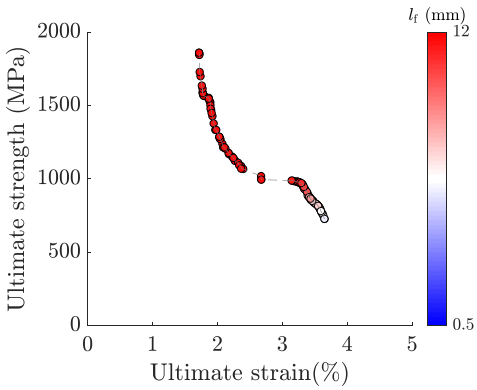
Figure 4: The *estimated Pareto front* for the maximum ultimate strength & initial stiffness design case, with each point coloured according to the values of specific inputs.



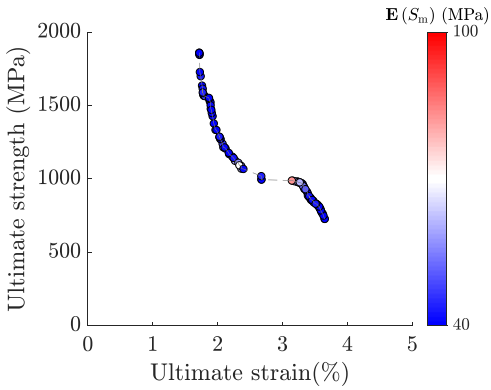
(a) Material system.



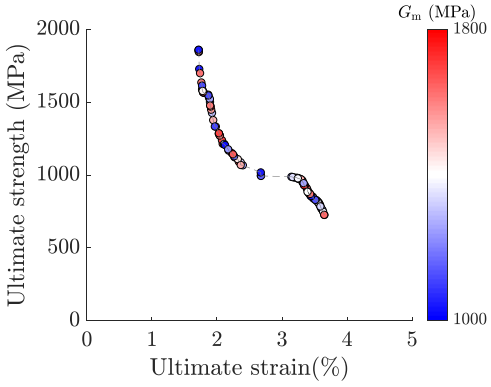
(b) Carbon volume ratio.



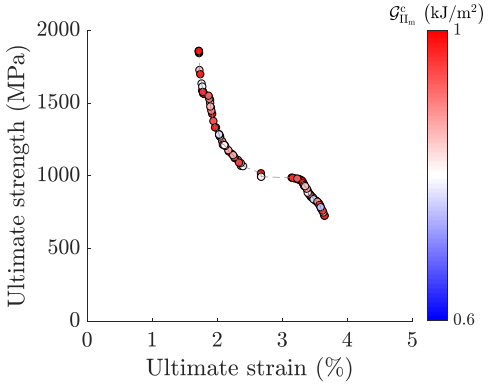
(c) Fibre length.



(d) Expected matrix strength.

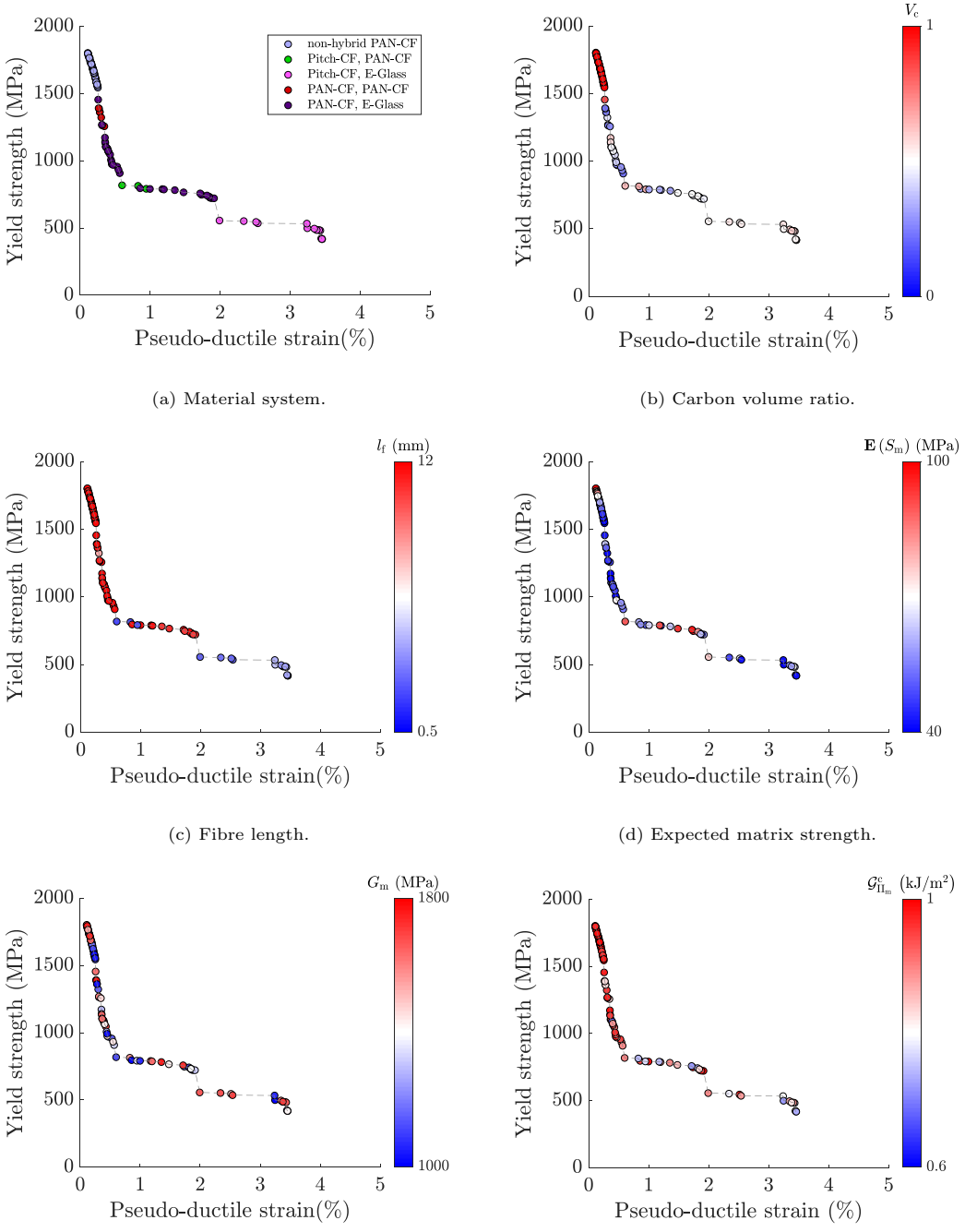


(e) Matrix shear modulus.



(f) Mode-II interlaminar fracture toughness.

Figure 5: The *estimated Pareto front* for the maximum ultimate strength & ultimate strain design case, with each point coloured according to the values of specific inputs.



(a) Material system.

(b) Carbon volume ratio.

(c) Fibre length.

(d) Expected matrix strength.

(e) Matrix shear modulus.

(f) Mode-II interlaminar fracture toughness.

Figure 6: The *estimated Pareto front* for the maximum yield strength & pseudo-ductile strain design case, with each point coloured according to the values of specific inputs.

of designs which feature hybridisation of fibre-types with the Pitch-CF fibre-group, with a carbon volume ratio of greater than 95% (this is discussed further in Section 4.2.1). It can be argued that such tight control over the carbon ratio cannot be consistently achieved in real-life processes, and it may be prudent to avoid these design cases. To tackle this, all specimen designs on the *estimated Pareto fronts* that feature a carbon ratio of greater than 95% had their carbon ratios changed to 100%, making these specimen designs non-hybrid. The sensitivity analysis was then repeated for these modified optimal sets, and the resulting Pareto fronts (labelled the *robust Pareto fronts*) plotted in Figure 7b, 7d and 7f.

It should be noted that the carbon and glass fibre-types (which are categorical variables in the scope of this work) were not subject to the sensitivity analysis, as it would be difficult to mix-up different fibre types during a typical ADC manufacturing process (such as the HiPerDiF method [15, 25, 26, 48–50]), and variability in the fibre strength and Young's modulus are already captured directly within the VTF [17].

4.2. Multi-objective optimisation discussion

4.2.1. Optimal designs and trade-offs

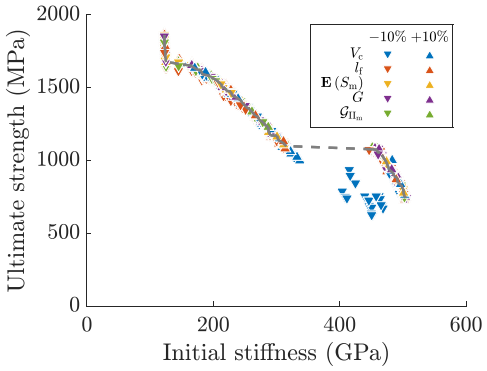
All three *observed Pareto fronts* are convex (see Figure 3), indicating that there is a limited trade-off between each of the multi-objective design cases; this Pareto front convexity also confirms the need for the modified ε-constraint multi-objective optimisation method proposed in this paper (as per Appendix A). Despite the convexity of the Pareto fronts, there are some regions in which useful combinations of material properties can be found; for example, an optimal ultimate strength & ultimate strain combination of 1127 MPa and 2.48% was achieved (see Figure 3c), or an optimal combination of 760 MPa yield strength and 1.84% pseudo-ductile strain (see Figure 3e).

The Pareto front for the maximum ultimate strength & initial stiffness multi-objective design case is far less convex than the other two multi-objective design cases (see Figure 3a); this reduced convexity indicates that less of a trade-off between these two competing *key performance characteristics* has to be made when selecting an optimal design. However, the Pareto fronts for the remaining two multi-objective design cases do feature a stepped shape; Engineers may have a preference towards the peak(s) of these stepped regions of the Pareto front, because while all points on the Pareto front can be considered optimal, the stepped regions represent areas where a severe sacrifice in one *key performance characteristic* must be made to enable a marginal increase in the other *key performance characteristic*.

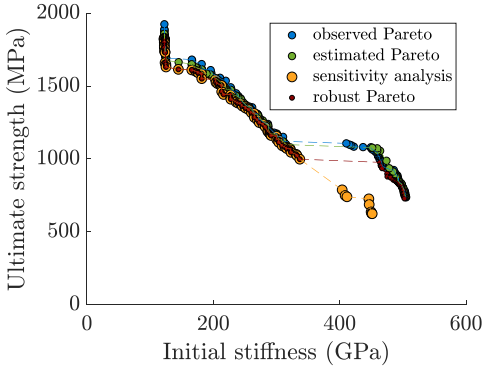
Maximum ultimate strength & initial stiffness.

From the analysis of Figure 4, two trade-off strategies can be identified to create a balance between optimal ultimate strength & initial stiffness:

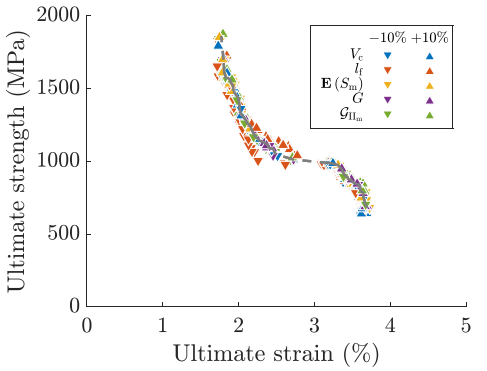
1. To maintain a near-optimal initial stiffness while increasing ultimate strength, one should (a) reduce the fibre length (which reduces the number of co-planar fibre fragmentations, hence



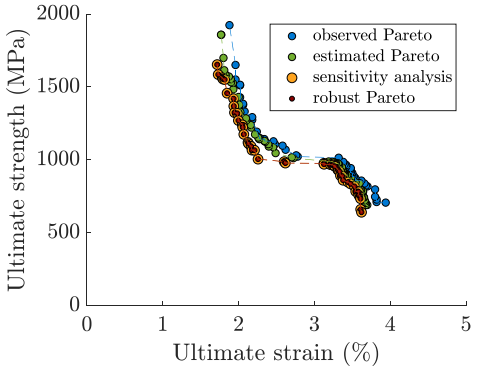
(a) Estimated Pareto front sensitivity analysis (maximum ultimate strength & initial stiffness design case). The grey dashed line represents the *estimated Pareto front*.



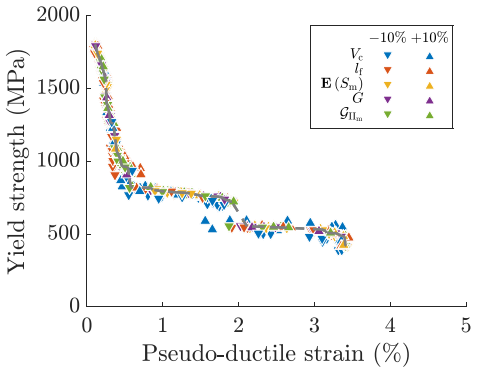
(b) The *sensitivity analysis Pareto front* and the *robust Pareto front* for the ultimate strength & initial stiffness design case.



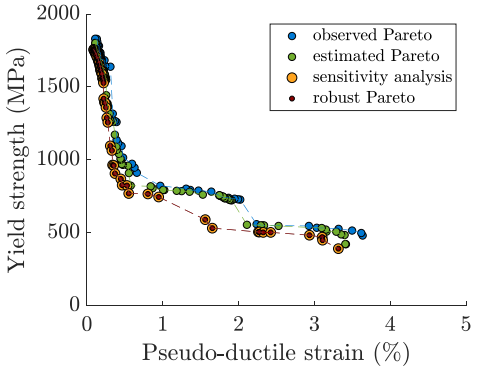
(c) Estimated Pareto front sensitivity analysis (maximum ultimate strength & ultimate strain design case). The grey dashed line represents the *estimated Pareto front*.



(d) The *sensitivity analysis Pareto front* and the *robust Pareto front* for the ultimate strength & ultimate strain design case.



(e) Estimated Pareto front sensitivity analysis (maximum yield strength & pseudo-ductile design case). The grey dashed line represents the *estimated Pareto front*.



(f) The *sensitivity analysis Pareto front* and the *robust Pareto front* for the yield strength & pseudo-ductile strain design case.

Figure 7: Sensitivity analyses for all three multi-objective design cases are shown on the left; plotted points that are far away from the *estimated Pareto front* and towards the bottom-left corner of the plot indicate optimal solutions that are most detrimentally sensitive to the inputs, with the colour and shape of the point indicating which sensitivity case produces each result. The *sensitivity analysis Pareto fronts* and *robust Pareto fronts* (see Section 4.2.3 for details on the *robust Pareto fronts*) are shown in the right.

reducing the strain energy release rate associated with broken clusters of fibres and delaying final fracture), while (b) hybridising the Pitch-CF carbon fibre with a small volume (i.e. less than 5% volume) of E-glass or PAN-CF fibres (which prevents the formation of large clusters of fibre breaks, hence delaying final fracture of the *virtual specimen*).

2. To achieve a further increase in ultimate strength while maintaining a reasonable initial stiffness, one should (a) hybridise the Pitch-CF fibres with PAN-CF fibres (which reduces the number of fibre fragmentations and delays final fracture); further increases in ultimate strength can be achieved by reducing the Pitch-CF volume ratio further from strategy 2(a) above (which increases the number of high-strength PAN-CF fibres, hence preventing the formation of large clusters of fragmented Pitch-CF fibres and further delaying fracture of the *virtual specimen*).

Maximum ultimate strength & ultimate strain.

From the analysis of Figure 5, three trade-off strategies can be identified to create a balance between optimal ultimate strength & ultimate strain:

1. To improve the ultimate strength and maintain a nearly-optimal ultimate strain, one should (a) maintain hybridisation between Pitch-CF and E-glass fibres (which promotes a non-linear, pseudo-ductile response, as per Section 3.2), while (b) increasing the fibre length (which improves the stress-transfer between neighbouring fibres, and hence increases ultimate strength), and (c) decrease the carbon volume ratio (which reduces the number of carbon-fibre fragmentations, and hence delays final failure).
2. To create a material that features a near-optimal ultimate strength and a relatively low ultimate strain, a Pitch-CF / PAN-CF hybrid ADC should be used (which combines two relatively high-strength fibre-groups to maximise strength).
3. To achieve a balanced level of ultimate strength and ultimate strain, moving from a nearly-optimal ultimate strain (as per Point 1) to a nearly-optimal ultimate strength (as per Point 2), one should (a) switch the Pitch-CF fibres with PAN-CF fibres to make a PAN-CF / E-glass hybrid ADC (which reduces the number of carbon-fibre fragmentations, thus delaying final fracture), while (b) increasing the carbon volume ratio (which progressively reduces the number of carbon-fibre fragmentations and further increases strength at the expense of ultimate strain).

Maximum yield strength & pseudo-ductile strain.

From the analysis of Figure 6, three trade-off strategies can be identified to create a balance between optimal yield strength & pseudo-ductile strain:

1. To improve the yield strength and maintain a large amount of pseudo-ductility, one should reduce the matrix shear strength (which increases the amount of debonding and reduces the

amount of fragmentation, simultaneously leading to both an increase in translaminar fracture toughness, as per Appendix B, and a reduction in pseudo-ductility).

2. To create a material that features a near-optimal yield strength and a relatively low pseudo-ductile strain, one should (a) use a PAN-CF non-hybrid fibre-group (which minimises the number of fibre fragmentations and delays final fracture), while (b) selecting a high matrix strength (which prevents matrix debonding and softening damage events from occurring, thus preventing non-linear stress-strain behaviour for as long as possible).
3. To achieve a balanced level of yield strength and pseudo-ductile strain, moving from a nearly-optimal pseudo-ductile strain (as per Point 1) to a nearly-optimal yield strength (as per Point 2), one should (a) hybridise fibre-types using PAN-CF and E-glass fibres (which reduces the number of carbon-fibre fragmentations, thus delaying final fracture), while (b) increasing the fibre length (to maximise the number of fibre fragmentations, hence promoting pseudo-ductility).

4.2.2. The influence of variability on the Pareto front

The *observed Pareto fronts* $\boldsymbol{Y}^\star$ ($\boldsymbol{X}^\star$) in Figures 3b, 3d, and 3f often lie at the upper tail-end of the distribution of results from the variability study; this is because the *observed Pareto front* takes the best observed points during the optimisation, without considering whether the observed point is a true representation of the mean performance (for that specific set of inputs) or not. In some cases the *observed Pareto front* may consist of designs that have a sub-optimal mean performance, but have sufficient variability such that the upper tail-end of the distribution of the design performance may indicate it is one of the best observed points. These results therefore indicate that the *observed Pareto front* is a poor representation of the expected performance of the optimal designs (and perhaps even a poor representation of which designs may be optimal themselves).

The *estimated Pareto front* $\dot{\boldsymbol{Y}}^\star \left(\dot{\boldsymbol{X}}^\star \right)$ (generated as per Section 4.1.2, and shown in Figures 3b, 3d, and 3f) gives a much better representation of the true expected Pareto front, as the estimated points lie close to the mean value of the variability analysis points. Designs with sub-optimal mean performance, but high levels of variability are not mistakenly considered optimal in this case. It should be noted that the *surrogate model* is capable of outputting the standard deviation of the predicted material performance, thus enabling confidence intervals for the mechanical performance of various material systems to be predicted (if so desired).

4.2.3. Testing and improving the robustness of optimal designs

For all three sensitivity analyses in Figures 7a, 7c, and 7e, most points lie very close to the *estimated Pareto front*, indicating that the *estimated Pareto fronts* generally offer a robust set of optimal solutions. However, there are three cases where the sensitivity analyses indicate that the robustness of the solutions could be improved:

1. On the bottom-right of the maximum ultimate strength & initial stiffness sensitivity analysis Pareto chart in Figure 7a there is a significant decrease in performance when the carbon ratio is decreased by 10%, with both the ultimate strength and initial stiffness reducing by up to 16%. Most designs in this region feature non-hybrid material systems (which are insensitive to changes in carbon ratio), although several designs feature hybridisation with a small number of other fibre-types and a carbon ratio of over 95%, and are thus sensitive to changes in carbon volume ratio, despite being 'almost non-hybrid' (see Section 4.2.1 for details on the associated inputs $\dot{\boldsymbol{X}}^{\star}$ used to achieve these points on the *estimated Pareto front*). While the 'almost non-hybrid' material designs generate a strong and stiff response, there is a significant reduction in their ultimate strength and initial stiffness as a result of a 10% change in carbon ratio. As a result, the robustness of these particular solutions is poor (suggestions for improving the robustness of these solutions will be discussed in Section 4.2.3).
2. A small decrease in performance can be found if the fibre length is decreased in the central region of the maximum ultimate strength & ultimate strain sensitivity analysis Pareto front in Figure 7c. In this region, a combination of high strength ductility is achieved via careful balance of all of the input parameters, so as to produce a large amount of sub-critical damage events (so as to maximise ductility), whilst also preventing premature fracture of the *virtual specimen*. Consequently, small changes to the input parameters can result in a large loss of strength or ductility.
3. A significant departure from the estimated Pareto front can be seen in the middle of the maximum yield strength & pseudo-ductile strain sensitivity Pareto chart (see Figure 7e). For the optimal designs in this region, PAN-CF fibres are combined with E-glass, with an intermediate value of the carbon ratio selected to maximise the pseudo-ductility of the response. However, a small increase to the carbon ratio causes the once ductile stress-strain curve to become brittle, resulting in a result with poor robustness; this lack of robustness can also be seen in Figure 3f, whereby there is a large variability in the pseudo-ductility of the optimal solutions in the middle of the Pareto front.

By changing the 'almost non-hybrid' ADC specimen designs that feature over 95% volume of one fibre-type into non-hybrid ADCs (as per Section 4.1.4), a clear improvement in robustness can be achieved, as shown by the *robust Pareto front*, which is never further way from the *estimated Pareto front* than the previous *sensitivity analysis Pareto front*, and is in some cases closer (see Figures 7b, 7d, and 7f). This robustness improvement is most apparent for the maximum ultimate strength & initial stiffness multi-objective optimisation case (see Figure 7b), where the region which previously was shown to be very sensitive to changes in inputs now only features a reduction of ultimate strength and initial stiffness of up to 1% and 2% respectively.

These improvements in robustness were determined only through interrogation of the *surrogate model*, and as the *surrogate model* can analyse thousands of designs in a few minutes (compared

to around 30 minutes runtime to analyse one design using the VTF), the proposed method can be used to accurately improve the robustness of a Pareto optimal set of solutions many times faster than alternative techniques. Changing the continuous optimisable variables by ±10% is a common way to investigate sensitivity in many Engineering applications, and this ±10% change is useful to demonstrate the applicability of the new proposed sensitivity analysis method. This sensitivity analysis method could be made even more powerful if uncertainty in the individual inputs is characterised more accurately with individual measures of confidence, before coupling this data with the proposed sensitivity analysis technique.

5. Conclusion

This paper combines an efficient virtual testing framework (VTF) with a data-driven intelligent Bayesian optimisation routine to maximise the mechanical performance of aligned discontinuous composites (ADCs) for a variety of real-world design cases. Several novel features of this optimisation routine have been developed as part of this work:

- For the first time in the literature, an optimisation routine is used to maximise the micromechanical performance of composite materials which feature significant levels of both complexity and variability.
- A modified version of the ε-constraint method is created, which enables easier multi-objective optimisation of systems with high levels of variability at a lower computational cost.
- Variability analysis shows that composite materials which feature significant variability are difficult to optimise when using the observed data directly from analysis tools (such as the VTF used in this work). Instead, the use of a *surrogate model* (created using Gaussian process regression) more accurately captures the expected performance of variable composite materials.
- The advantages of the *surrogate model* are exploited further as it was used to improve the robustness of optimal ADC designs with minimal added computational cost.

These novel features enable several conclusions to be drawn about optimisation of ADCs:

- Single-objective optimisation shows that a wide range of structural properties can be achieved using ADCs, with a maximum stiffness of 505 GPa, maximum ultimate strain of 3.94%, maximum pseudo-ductile strain of 3.64%, maximum ultimate strength of 1.92 GPa, or a maximum yield strength of 1.83 GPa all possible when using this type of versatile composite material system.
- Multi-objective optimisation studies are performed for three design cases: maximum ultimate strength & initial stiffness, maximum ultimate strength & ultimate strain, or maximum yield

strength & pseudo-ductile strain. A moderate trade-off in performance can be achieved when considering multi-objective optimisation design cases, with an optimal ultimate strength & ultimate strain combination of 982 MPa and 3.27%, or an optimal combination of 720 MPa yield strength & 1.91% pseudo-ductile strain all possible.

- Scrutiny of the inputs required to create the *estimated Pareto fronts* show how careful control of the fibre-type combinations, fibre length, carbon volume ratio, or matrix shear strength (or combinations thereof, depending on the design case) enable a controlled trade-off between required mechanical performance.

- A sensitivity analysis shows that the use of hybrid ADCs with carbon volume ratios over 95% were not sufficiently robust to changes in the carbon volume ratio, with both their ultimate strength and initial stiffness being reduced by up to a 16%. By limiting designs to ensure that small proportions of one fibre-type are avoided, this sensitivity deficit is reduced to 1% initial stiffness loss and 2% ultimate strength loss.

The combination of this VTF and a Bayesian optimisation routine is proven to be a robust and complete tool for the optimisation and design of composite materials with significant levels of complexity and variability; the authors welcome other researchers to use this VTF to design new materials or explore other avenues of research. A live GitHub repository of the working code can be found here: https://github.com/meComposites/Aligned-Discontinuous-Composites/releases/tag/Data-driven_intelligent_optimisation_of_discontinuous_composites.

6. Acknowledgements

This work was funded under the UK Engineering and Physical Sciences Research Council (EPSRC) programme grant EP/I02946X/1 on *High Performance Ductile Composite Technology*. Soraia Pimenta acknowledges the support from the Royal Academy of Engineering for her Research Fellowship on *Multiscale discontinuous composites for large scale and sustainable structural applications* (2015-2019). James M. Finley thanks Mike Flynn of Deep Learning Machine Ltd. for his guidance on Bayesian optimisation and machine learning.

All underlying data to support the conclusions are provided within this paper.

7. References

References

[1] A.N. Netravali and S. Chabba. Composites get greener. *Mater. Today*, 6(4):22–29, 2003.

[2] J. Brownlee. How to implement Bayesian optimization from scratch in Python, 2019. [Online; accessed 06-Jan-2020].

[3] J. Snoek et al. Practical Bayesian optimization of machine learning algorithms. In F. Pereira, C.J.C. Burges, and L. Bottou, editors, *Neural Information Processing Systems*, pages 2951–2959, Lake Tahoe, 2012.

[4] D. Lizotte et al. Automatic gait optimization with Gaussian process regression. In *IJCAI International Joint Conference on Artificial Intelligence*, pages 944–949, Hyderabad, 2007.

[5] N. Srinivas et al. Gaussian process optimization in the bandit setting: no regret and experimental design. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, pages 1015–1022, Haifa, 2010.

[6] J. Močkus. On bayesian methods for seeking the extremum. In *Optimization Techniques IFIP Technical Conference*, volume 27 LNCS, pages 400–404, Novosibirsk, 1975. Springer, Berlin, Heidelberg.

[7] L. Sun et al. A Bayesian optimization-based evolutionary algorithm for flexible job shop scheduling. In *Procedia Comput. Sci.*, volume 61, pages 521–526, 2015.

[8] C.E. Rasmussen and C.K.I. Williams. *Gaussian processes for machine learning*. MIT Press, Cambridge, MA, 2006.

[9] K.P. Murphy. Kernels. In *Machine learning : a probabilistic perspective*, chapter 14, pages 479–514. MIT Press, Cambridge, MA, 2012.

[10] G. Villarrubia et al. Artificial neural networks used in optimization problems. *Neurocomputing*, 272:10–16, jan 2018.

[11] Y. Gal and Z. Ghahramani. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In *33rd International Conference on Machine Learning, ICML 2016*, volume 3, pages 1661–1680, 2016.

[12] O. Goury et al. Automatised selection of load paths to construct reduced-order models in computational damage micromechanics: from dissipation-driven random selection to Bayesian optimization. *Comput. Mech.*, 58:213–234, 2016.

[13] C. Li et al. Rapid Bayesian optimisation for synthesis of short polymer fiber materials. *Sci. Rep.*, 7(1):5683, 2017.

[14] J. Henry and S. Pimenta. Semi-analytical simulation of aligned discontinuous composites. *Compos. Sci. Technol.*, 144:230–244, 2017.

[15] J.M. Finley et al. Exploring the pseudo-ductility of aligned hybrid discontinuous composites using controlled fibre-type arrangements. *Composites Part A*, 107:592–606, 2017.

[16] J.D. Fuller and M.R. Wisnom. Pseudo-ductility and damage suppression in thin ply CFRP angle-ply laminates. *Composites Part A*, 69:64–71, 2015.

[17] J.M. Finley et al. The influence of variability and defects on the mechanical performance of tailorable composites. *Journal of Composite Materials*, pages 1–25, 2019.

[18] J. Henry and S. Pimenta. Modelling hybrid effects on the stiffness of aligned discontinuous composites with hybrid fibre-types. *Compos. Sci. Technol.*, 152(1):275–289, 2017.

[19] J. Henry and S. Pimenta. Virtual testing framework for hybrid aligned discontinuous composites. *Compos. Sci. Technol.*, 159:259–272, 2018.

[20] C. Qian et al. Establishing size effects in discontinuous fibre composites using 2D finite element analysis. *Computational Materials Science*, 64:106–111, nov 2012.

[21] M.R. Wisnom. Mechanisms to create high performance pseudo-ductile composites. *IOP Conference Series: Materials Science and Engineering*, 139(1):1–9, 2016.

[22] L.-M. Zhou et al. On the single fibre pull-out problem: effect of loading method. *Compos. Sci. Technol.*, 45(2):153–160, 1992.

[23] X. Zhang et al. On steady-state fibre pull-out. I: The stress field. *Compos. Sci. Technol.*, 59:2179–2189, 1999.

[24] H.F. Wu and A.N. Netravali. Weibull analysis of strength-length relationships in single Nicalon SiC fibres. *Journal of Materials Science*, 27(12):3318–3324, 1992.

[25] H. Yu et al. A novel manufacturing method for aligned discontinuous fibre composites (High Performance-Discontinuous Fibre method). *Composites Part A*, 65:175–185, 2014.

[26] H. Yu et al. Pseudo-ductility in intermingled carbon/glass hybrid composites with highly aligned discontinuous fibres. *Composites Part A*, 73:35–44, 2015.

[27] M.C. Paiva et al. A comparative analysis of alternative models to predict the tensile strength of untreated and surface oxidised carbon fibers. *Carbon*, 39(7):1091–1101, 2001.

[28] A. Bismarck et al. Adhesion: comparison between physico-chemical expected and measured adhesion of oxygen-plasma-treated carbon fibers and polycarbonate. *J. Adhes.*, 73(1):19–42, 2000.

[29] K.L. Pickering and T.L. Murray. Weak link scaling analysis of high-strength carbon fibre. *Composites Part A*, 30(8):1017–1021, 1999.

[30] Toho. Toho Tenax HTA40 technical data sheet, 2011.

[31] K. Naito et al. Tensile properties of ultrahigh strength PAN-based, ultrahigh modulus pitch-based and high ductility pitch-based carbon fibers. *Carbon*, 46:189–195, 2008.

[32] P. Bertrand et al. On the effects of surface treatments on the mechanical strength of carbon fibres. *J. Mater.*, 3(8):5029–5036, 1998.

[33] Toray. TorayCA T800H technical data sheet, 2018.

[34] M.A. Montes-Morán et al. Effects of plasma oxidation on the surface and interfacial properties of ultra-high modulus carbon fibres. *Composites Part A*, 32(3-4):361–371, 2001.

[35] J.K. Wessel. *Handbook of Advanced Materials: Enabling New Designs*. J. Wiley, 2004.

[36] M.R. Wisnom et al. Hybrid effects in thin ply carbon/glass unidirectional laminates: Accurate experimental determination and prediction. *Composites Part A*, 88:131–139, 2016.

[37] J. Mockus. Application of Bayesian approach to numerical methods of global and stochastic optimization. *J. Global Optim.*, 4:347–365, 1994.

[38] D.R. Jones. A taxonomy of global optimization methods based on response surfaces. *J. Global Optim.*, 21:345–383, 2001.

[39] D.R. Jones et al. Efficient global optimization of expensive black-box functions. *J. Global Optim.*, 13:455–492, 1998.

[40] A.D. Bull. Convergence rates of efficient global optimization algorithms. *J. Mach. Learn. Res.*, 12:2879–2904, 2011.

[41] MathWorks. bayesopt - matlab r2019a documentation, 2019. [Online; accessed 03-Sept-2019].

[42] P.D. Mangalgiri. Composite materials for aerospace applications. *B. Mater. Sci.*, 22(3):657–664, 1999.

[43] G. Gerard. Structural significance of ductility in aerospace pressure vessels. *ARS Journal*, 32(8):1216–1221, 2012.

[44] A.F. Grandt. *Fundamentals of Structural Integrity: Damage Tolerant Design and Nondestructive Evaluation*. J. Wiley, 2003.

[45] MathWorks. How does bayesopt() optimise using categorical optimizable variables? - MATLAB Answers, 2019. [Online; accessed 03-Sept-2019].

[46] G. Chiandussi et al. Comparison of multi-objective optimization methodologies for engineering applications. *Comput. Math. Appl.*, 63(5):912–942, 2012.

[47] K. Deb. Multi-objective optimization. In Sheldon Ross and Richard Weber, editors, *Multi-objective optimization using evolutionary algorithms*, chapter 2, pages 13–48. John Wiley & Sons, Ltd, Chichester, 1 edition, 2002.

[48] M.L. Longana et al. Aligned discontinuous intermingled reclaimed/virgin carbon fibre composites for high performance and pseudo-ductile behaviour in interlaminated carbon-glass hybrids. *Compos. Sci. Technol.*, 143:13–21, may 2017.

[49] M.L. Longana et al. Quasi-isotropic and pseudo-ductile highly aligned discontinuous fibre composites manufactured with the HiPerDiF (High Performance Discontinuous Fibre) technology. *Materials*, 12(11):1794, jun 2019.

[50] H. Yu et al. Hierarchical pseudo-ductile hybrid composites combining continuous and highly aligned discontinuous fibres. *Compos. Part A Appl. Sci. Manuf.*, 105:40–56, feb 2018.

[51] A. Osyczka and B.J. (Ed.) Davies. *Multicriterion optimization in engineering with fortran programs*, volume 1. E. Horwood, Chichester, 1984.

[52] Z. Lounis and M.Z. Cohn. Multiobjective optimization of prestressed concrete structures. *J. Struct. Eng.*, 119(3):794–808, 1993.

Appendices

Appendix A Modified ε-constraint method for multi-objective optimisation

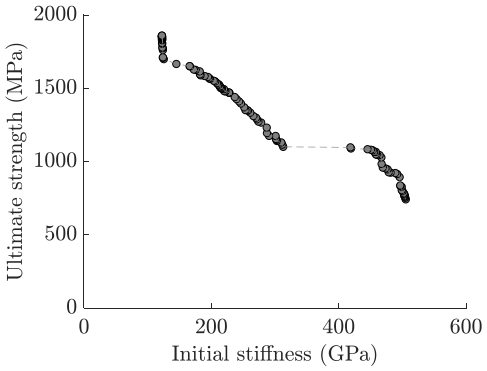
The optimisation campaigns in this paper are carried out using bayesopt function in Matlab [41]; this function optimises a single *objective function*, meaning the it must be modified to support the multi-objective optimisation campaign. Several methods for optimising a multi-objective function are available, each with their own disadvantages:

- The most well-known multi-objective optimisation routine for this situation is the weighted-sum approach [46, 51], which combines linear combinations of two or more *objective functions* to find a global optimum. However, a trial optimisation study showed that the Pareto fronts for all three design cases included in this paper were non-convex, indicating that the weighted-sum approach was unsuitable for this optimisation campaign, as the weighted-sum approach is unable to generate large portions of a non-convex Pareto front due to its mathematical formulation [46, 51].

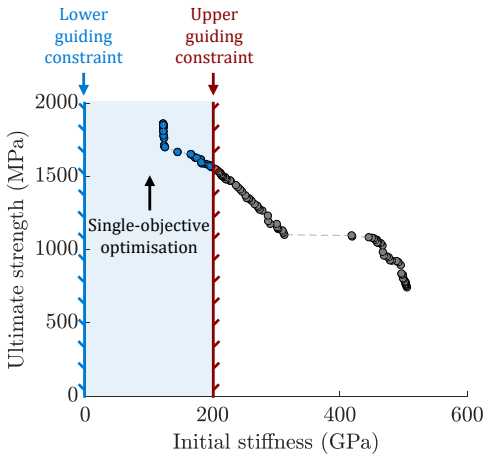
- The ε-constraint method [46, 51, 52] is an alternative to the weighted-sum approach which is able to generate all Pareto optima on a generalised Pareto front [46]. The first set of Pareto optimal solutions are found by optimising one *primary objective function* $f_1(\boldsymbol{x})$, while constraining the output of a *secondary objective function* $f_2(\boldsymbol{x})$ between upper- and lower-bounded constraints (~~ε_1 and ε_2, respectively,~~ from now on referred to as guiding constraints, thus ensuring that secondary objective value is greater than the lower guiding constraint, and

lower than the upper guiding constraint), as shown in Figure 8b. All Pareto optimal solutions can be found by moving both of the guiding constraint bounds to cover the achievable range of values of $f_2(\boldsymbol{x})$ (which in this case were determined via the single-objective optimisation campaign), as per Figure 8c. The ε-constraint method is well-suited to this optimisation campaign, as both *objective functions* are calculated using the same VTF model, and the Bayesian optimisation routine can treat the upper and lower guiding constraints as coupled constraints, and hence create Gaussian process regression models to ensure only feasible solutions are evaluated (as per Section 2.2.2). However, optimisation of a non-deterministic system using Bayesian optimisation and the ε-constraint method is extremely difficult if the range between the upper and lower guiding constraints is close to the level of variability in the system, as the optimiser will often predict a low probability of feasibility (due to the relatively large degree of uncertainty) and may in some cases be unable to find any feasible points.

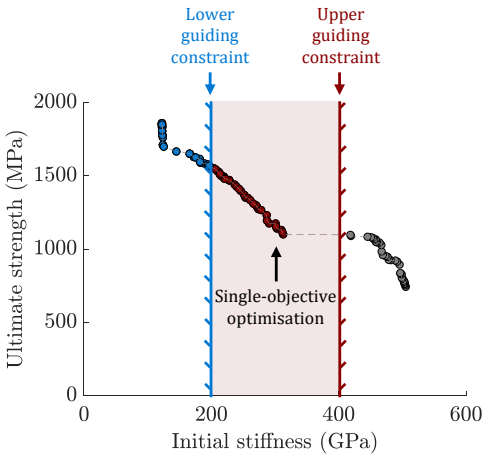
A modified version of the ε-constraint method is proposed in this paper, which tackles the issue of combining a non-deterministic *objective function* with upper and lower guiding constraints as part of the ε-constraint method. The process follows the conventional ε-constraint method as described above, but in this case the upper guiding constraint is removed (as shown in Figure 9b). The removal of the upper constraint increases the predicted probability of feasibility (found by the *acquisition function*), as the non-deterministic system no longer has to lie within two guiding constraints. Secondly, the *acquisition function* guides the optimiser to the feasible region that has the highest expected value for the *primary objective function* $f_1(\boldsymbol{x})$, without the need for the upper guiding constraint (as shown in Figure 9b). The value of the lower guiding constraint can be incrementally increased (as per the conventional ε-constraint method) to find the complete Pareto front, as shown in Figure 9c (in this paper the guiding constraint is incremented by $^1/_{40}{}^{\text{th}}$ of the range of the secondary objective function, to ensure the Pareto front has a high resolution). This process above is for the situation where both *objective functions* are to be maximised, however it can be adapted for when both objectives are to be minimised simply by removing the lower constraint instead of the upper constraint.



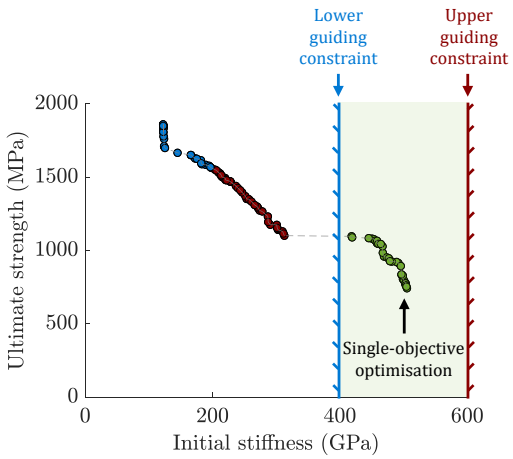
(a) At the start of the multi-objective optimisation, no optimal points are discovered yet (and are hence displayed in grey).



(b) Single-objective optimisation is performed within the first guiding constraint range.

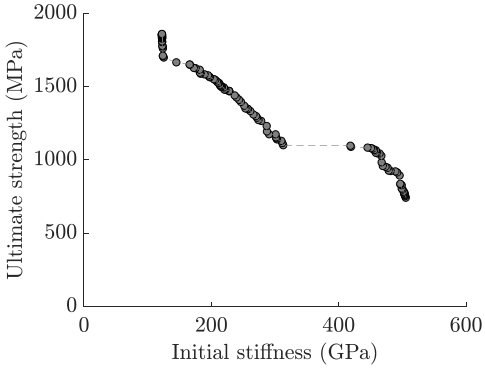


(c) Single-objective optimisation is performed within the second guiding constraint range.

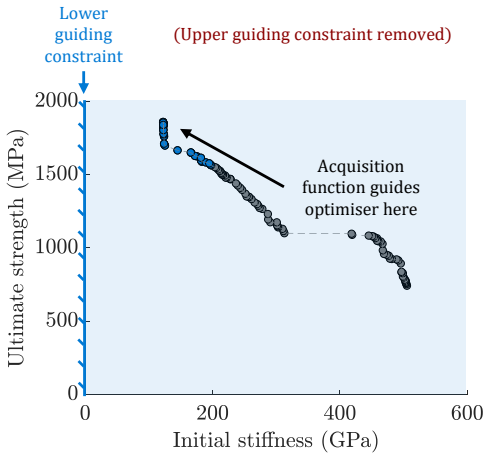


(d) Single-objective optimisation is performed within the third guiding constraint range.

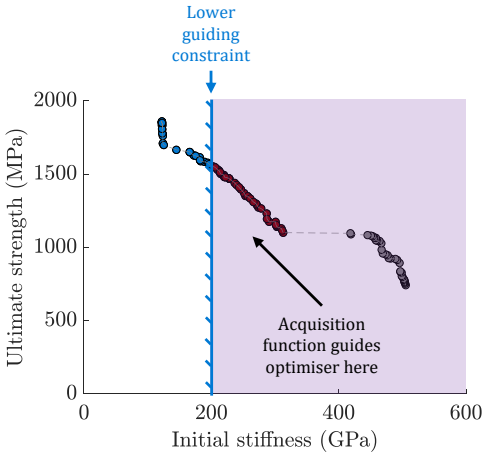
Figure 8: The ε-constraint method optimises the primary objective function (in this case the ultimate strength), while constraining the secondary objective function (in this case the initial stiffness) between a upper and lower guiding constraints. The whole Pareto front is found by moving the range of the guiding constraints across the range of achievable values for the secondary objective function (as shown in Figures 8b to 8d). In this case, the range of the guiding constraints is incremented by $1/3^{\text{rd}}$ of the total secondary objective range, although in practice this range is reduced to $1/40^{\text{th}}$ of the secondary objective function range to increase the Pareto front resolution.



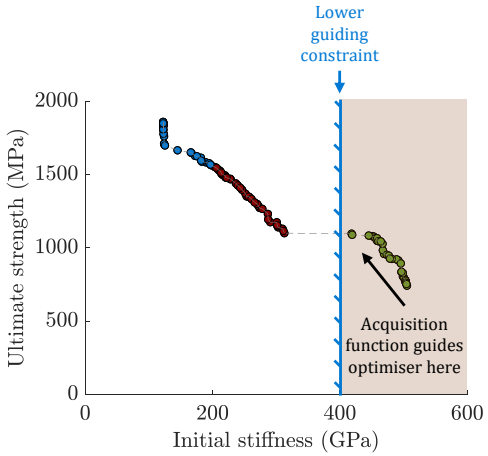
(a) At the start of the multi-objective optimisation, no optimal points are discovered yet (and are hence displayed in grey).



(b) The upper guiding constraint is removed, and single-objective Bayesian optimisation is performed within the first range specified by the lower guiding constraint.



(c) Single-objective Bayesian optimisation is performed within the second range specified by the lower guiding constraint.



(d) Single-objective Bayesian optimisation is performed within the third range specified by the lower guiding constraint.

Figure 9: The modified ε-constraint method is implemented the same way as the ε-constraint method, but the upper guiding constraint is removed. This technique works when implementing Bayesian optimisation, as the *acquisition function* guides the optimiser to the top-left corner of the feasible region (highlighted in colour), without the need of the upper guiding constraint. In this case, the range of the guiding constraints is incremented by $1/3^{\text{rd}}$ of the total secondary objective range, although in practice this range is reduced to $1/40^{\text{th}}$ of the secondary objective function range to increase the Pareto front resolution.

Appendix B The influence of matrix shear strength on translaminar fracture toughness

The matrix shear strength may have a significant influence on the translaminar fracture toughness of some ADC designs. To investigate this, the translaminar fracture toughness is calculated using the method proposed by Henry et al. [14] for the designs which maximised each of the single-objective design cases; next, the matrix shear strength is adjusted within the range of possible values to show how the matrix strength influences the toughness for each of the design cases. Figure 10 shows that the translaminar fracture toughness increases as the matrix strength is reduced.

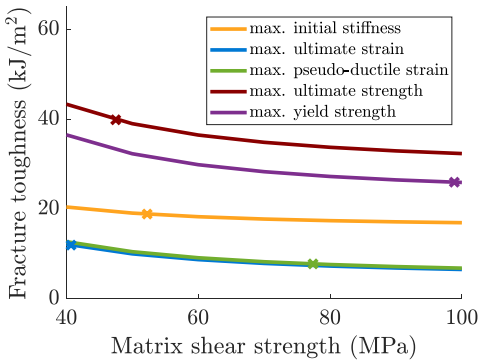


Figure 10: The influence of matrix shear strength on the translaminar fracture toughness of the five designs that maximise each of the single-objective design cases. Crosses indicate the actual translaminar toughness values for each of the five optimal single-objective designs.