Multi-Objective Optimisation of Variable Stiffness Cylindrical Shells: An Artificial Intelligence Approach

S. Prasing

MSc. Thesis

Multi-Objective Optimisation of Variable Stiffness Cylindrical Shells: An Artificial Intelligence Approach

by

S. Prasing

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Wednesday November 25, 2020 at 09:30 o'clock

Student number: 4911687

Project duration:	November 29, 2019 – November 25,2020	
Thesis committee:	Dr. C. Bisagni,	TU Delft, supervisor
	Dr. D. Peeters,	TU Delft, examiner
	Dr. D. Zarouchas,	TU Delft, external examiner

TU Delft

Preface

The current work was initiated with the observation that the prominence of artificial intelligence techniques, as constantly displayed in numerous fields, is likely to infiltrate all the engineering disciplines for good. This observation has now turned into a belief that the time has come where these techniques have advanced so far as to become formidable tools in an engineer's arsenal. When the current topic of the optimisation of the novel concept of fibre steering of composite structures using artificial intelligence was proposed to me, I was immediately intrigued. Part of the allure was the fact that my familiarity on the subjects of machine learning, optimisation, and fibre steering was quite limited, and thus an opportunity was granted to delve into unknown territory and obtain new-found knowledge.

To this end, I must say that the completion of this thesis would be impossible without sitting on the shoulders of giants in both academia but also in the form of online communities who through their selfless efforts, they have created free-to-access computational packages that have proved to be indispensable for this project.

I would like to warmly thank my faculty supervisor, professor Chiara Bisagni, for giving me the opportunity to undertake this project, for her positive attitude, and for the support she provided. I also extent my gratitude to all the members of her group, since our group meetings have been a great source of academic information and a much-needed social touch during the COVID-19 quarantine.

I would also like to thank all of my academic colleagues for their interest and contributions to my work, and I mainly thank them for our countless discussions about life beyond academic matters. Their friendship was an essential component towards the completion of this thesis.

I want to thank my family for their unwavering support in all of my endeavours. Without them I would not be able to accomplish whatever I have, and I am forever grateful to them.

Lastly, I am thanking my partner in life, Angeliki, who granted me the mental fortitude to overcome the hardships I encountered and push onwards in all aspects of life.

S. Prasing
Delft, November 2020

Abstract

The structural properties of a composite laminate can be tailored by allowing the fibres to vary their orientation within the individual plies. This class of composites is termed variable stiffness and they are able to enhance the structural performance over traditional laminates by redistributing the internal load of the structure. Therefore, it is of interest to pinpoint the structural properties that may benefit most from this concept and the extent thereof.

To that end, a multi-objective optimisation of a variable stiffness cylinder was performed which identifies the Pareto front for objectives such as buckling, natural frequency, stiffness and strength. Manufacturing considerations were also incorporated by accounting for the allowable tow curvature and the resulting overlaps during the process. The stiffness variability was formed by means of a Lagrange polynomial, the non-linearity of which is determined by a specified amount of control points. An investigation on the suitable degree of non-linearity of the functional representation of the cylinder was thus carried out.

In order to complete the optimisation in a practical time frame, machine learning algorithms were utilised to act as surrogate models and perform predictions on the optimisation objectives. Different algorithms were implemented for that task, namely Radial Basis Function, Kriging and Artificial Neural Network models. The Artificial Neural Network surrogate model outperformed the others when the dimensionality of the problem was high, although the opposite was observed for low dimensional problems.

Cylinders with and without cut-outs were optimised separately and compared with their optimised constant stiffness counterparts. It is found that the performance gains for the cylinder without cut-outs are not significant when the additional mass due to the overlaps is accounted for. On the other hand, the cut-out cylinder greatly benefits from the stiffness variation for multiple objectives. Additional structural improvements were witnessed for higher degrees of non-linearity of the utilised Lagrange polynomial.

Contents

Nomenclature

AFP	Automated Fibre Placement
ANN	Artificial Neural Network
AV	Axial Variation
CNN	Convolutional Neural Network
CP	Control Point
CS	Constant Stiffness
CTS	Continuous Tow Shearing
CV	Circumferential Variation
DE	Differential Evolution
DoF	Degrees of Freedom
EA	Evolutionary Algorithm
ELU	Exponential Linear Unit
FEM	Finite Element Method
FPF	First Ply Failure
GA	Genetic Algorithm
HL	Hidden Layer
LHS	Latin-Hypercube Sampling
MAE	Mean Absolute Error
MLP	Multi-Layer Perceptron
MSE	Mean Squared Error
$Nadam$	Nesterov adaptive moment estimation
$NSGA$	Non-Dominated Sorting Genetic Algorithm
$NURBS$	Non-Uniform Rational B-spline
RBF	Radial Basis Function
$ReLU$	Rectified Linear Unit
RNN	Recurrent Neural Network
SBX	Simulated Binary Crossover
$SELU$	Scaled Exponential Linear Unit
VS	Variable Stiffness

List of Figures

List of Tables

1 Introduction

1.1. Research motivation

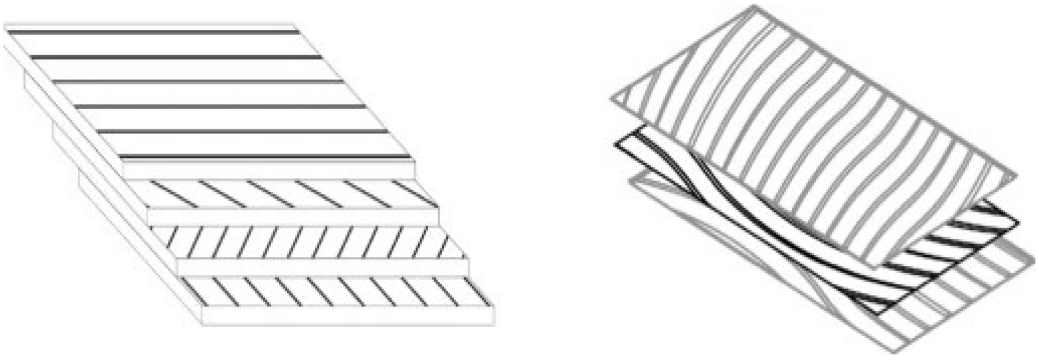
An objective that is ever-present in aerospace engineering is the weight reduction of the designed structure. Various efforts have been made for the realisation of that goal, with a relatively novel approach being the incorporation of variable stiffness to enhance the structural properties of the design. This field has already yielded promising results of improved structural performance, albeit at the expense of increased complexity.

A composite material is made from the combination of two or more constituent materials, which often have significantly different properties. The end result is a tailored material with superior properties in relation to those of its individual components. The majority of composite materials contain fibres enclosed within a matrix. The fibres act as reinforcements to the matrix, providing strength and rigidity, whereas the matrix serves as a binder for the reinforcement, transferring the mechanical load and protecting it from the environment. Many such plies of composite material can be stacked together to form a composite laminate, allowing the designer to further adapt the structural properties by aligning the fibres of each ply in certain directions. In that way, the structural loads that act in specific orientations can be carried in an effective manner without increasing the structural weight.

A composite laminate structure can be optimised even further by allowing the fibres to vary their orientation within the ply. Therefore, this class of composites is termed variable stiffness (VS) composites in contrast to the aforementioned constant stiffness (CS) composites. A pictorial difference between the two concepts is shown in Figure 1.1. This angle variability can be effectively utilised to redistribute the internal load of the structure and tailor the structural properties in a way that they best suit the application requirements. For instance, it has been reported that the buckling load can be increased by modifying the stiffness in key locations [1]. This is especially beneficial for cylindrical structures who are typically very sensitive to geometric imperfections, as the stiffness variation can lead the internal load away from these locations [2]. Additionally, further structural improvements are possible such as an enhancement in structural strength, particularly in the presence of cut-outs. In that case the load can be redistributed away from the cut-out, resulting in a reduced stress concentration factor in the surrounding area [3].

However, this additional angle variability only increases the complexity of the structure and therefore more refined analytical expressions are required, or more computational time in the case of numerical analyses. It is fortunate then that the emergence of the machine learning field offers powerful engineering tools to aid in tackling such complicated problems. The rapid advance of the field has potentially led to the underutilisation of what engineering solutions are capable with the help of artificial intelligence. The advent of artificial effective neural networks marks a unique opportunity to take advantage of their capabilities, which can serve as the efficient means to achieve the much coveted improvement in structural performance.

Arguably, the most important numerical approach in the structural engineer's arsenal is the *Finite Element Method* (FEM). The method has revolutionised the field and has found widespread use for decades now, leading in a major decrease in the time required to perform analysis of complicated structures. However, with the persistent drive for more optimised structures, FEM is now regarded as a rather computationally expensive approach (recognised as a high-fidelity simulation) with respect to other models that approximate the structural behaviour [5–7]. Such approximate models are termed surrogate models and their aim is to approximate a function to a sufficient degree for the sake of computational efficiency [8].



(**a**) Constant stiffness laminate (**b**) Variable stiffness laminate

Figure 1.1: An illustration of a (a) constant stiffness composite laminate and a (b) variable stiffness composite laminate side by side. Recreated from: [4].

Various surrogate models that can perform the function approximation exist, but they all require training. The training of a surrogate model is achieved by essentially "feeding" it enough data that have been acquired via the true function to be approximated, or through other means of data acquisition if the true function is unknown. Therefore, the surrogate model can produce a vector of predictions $\boldsymbol{y}$ according to

$$\mathbf{y} = f\left(\mathbf{x}, \mathbf{x}_t, \mathbf{y}_t\right) \tag{1.1}$$

where $\boldsymbol{x}$ is the vector of inputs, $\boldsymbol{x_t}$ is the vector of input training data, and $\boldsymbol{y_t}$ are the corresponding outputs of the training data [9]. Once enough data have been gathered for the model it will be able to make accurate predictions. The amount of data that is required for training depends on the model chosen however, and in certain cases it may take an unrealistic amount of time to collect enough data to train a surrogate model that makes predictions that are deemed acceptable in terms of accuracy.

The advances made in the field of artificial neural networks have rendered them great candidates to act as surrogate models in place of expensive higher order simulations, such as FEM. Their preference over other surrogate models is further complemented by their increasing accessibility, ease of use, and great versatility, allowing for imaginative solutions to problems, which the engineer would not consider otherwise.

Such a surrogate model can aid in the mathematical optimisation of the variable stiffness structures. In practical applications, multiple objectives have to be optimised simultaneously, in order to tailor the structural response of these objectives. This is termed as a multi-objective optimisation problem, where the improvement of the objectives might be conflicting with one another and often-times certain optimisation constraints need to be satisfied at the same time. As a result, many computational iterations may be required until satisfying designs are identified. That process will ultimately yield information on what structural gains are possible for the selected optimised objectives. It is up to a decision-maker to then determine the best design based on the structural requirements. Thus, this operation is crucial for realising the full potential of variable stiffness structures.

The contribution of the present thesis is towards the efficient optimisation of a fibre steered composite material cylinder, for which at the moment, the research regarding its optimisation is ongoing. It is the hope of the author that the work undertaken will be a useful addition to the body of literature surrounding such composite materials by identifying the potential benefits and drawbacks that are inherent in these structures and showcase the methodology that was employed to model them and efficiently perform their optimisation.

1.2. Research objectives

The main objective of this thesis is to develop a numerical tool set to optimise a variable stiffness composite cylinder. To ground this work in a realistic framework, certain limitations imposed by the current manufacturing methods are taken into account. Additionally, the aspect of defects or deviations from the mathematically ideal design of the cylinder are considered. The aim of this work can be formalised with the following research objective:

Within the time-span of this thesis, the objective is to explore the extent of the applicability of artificial intelligence techniques in the multi-objective optimisation of variable stiffness cylinders, carried out in an efficient manner to yield enhanced structural performance, over a basis constant stiffness cylinder.

Thus, to achieve the objective, a number of research questions have been formulated to help guide this work into successful completion:

1. What is an effective modelling technique to create a variable stiffness cylinder?
   (a) What kind of fibre path formulation is best suited for the optimisation problem in terms of performance and efficiency?
   (b) To what extent can a surrogate model be exploited to include the manufacturing defects in the model in a more efficient manner?
2. What is the optimal approach to the training of the artificial neural network architecture?
   (a) What is the impact of a deep neural network architecture on the multi-objective optimisation problem?
   (b) What is the most effective network architecture on average?
   (c) How much data is needed from the high-fidelity simulations in order to acquire an accurate surrogate model?
3. What is the difference in the objective space between a variable stiffness cylinder and a reference constant stiffness cylinder?
4. Which artificial intelligence techniques should be included in the comparison and which one yields the best results?

Answering some of these research questions serves as a requirement for the fulfilment of the main research objective, while some are complementary to fill the gaps that are present in the current literature concerning the optimisation of variable stiffness structures.

1.3. Thesis outline

The thesis is comprised of five chapters. In this chapter, the motivation behind the current research has been provided, along with the research objectives and questions that define it. A literature review is conducted in Chapter 2, which strives to identify the approaches that lead to an efficient optimisation of a fibre steered composite material cylinder by reviewing the most current research and techniques that are employed in the optimisation of variable stiffness structures. The methodology that was followed in the undertaking of the current work can be found in Chapter 3. Chapter 4 contains all the optimisation results following the implemented methodology and finally, conclusions are drawn based on the obtained results and recommendations for future work are presented in Chapter 5.

2

Literature review

2.1. Surrogate modelling

2.1.1. Design of experiments

When it comes to the gathering of the surrogate model data, one may ask, what is a time-efficient way of obtaining data? Ideally, this gathering process should explore enough data points in the space of feasible solutions as dictated by the true function and any imposed constraints. This procedure of filling the design space with enough data points is called *sampling* or *Design of Experiments*. Perhaps the first approach that one would consider is to utilise a full factorial sampling plan, where the data points are equally spaced within the space domain [10]. Thus, all conceivable combinations are explored during each reproduction of the experiment. However, as one can imagine this lacks computational efficiency since a significant portion of the design space needs to be explored, and thus it is not the preferred sampling approach.

Many alternatives exist to the full factorial sampling plan, such as fractional factorial, orthogonal arrays, optimal designs, etc. [11], but for the sake of brevity only two common ones will be introduced next. The first is the random sampling plan, where pseudo-random samples are drawn from the design space, in hopes of filling the design space in an efficient manner [12]. Given the bounds that limit the variables, the samples can be generated from a distribution, such as the uniform distribution. However, a significant downside of this method is attributed to the randomness of the point generation, excessive clustering of data points can occur, which results in poor coverage of the design space. The second common method is able to circumvent the problem of data clustering. This sampling plan is called *Latin-hypercube sampling* (LHS). This approach divides the design space in multiple strata of equal marginal probability, uniformly spreading the samples. This can considerably reduce the time that is needed to explore most of the design space, although it does not ensure that the samples will always have similar performance uniformity-wise.

2.1.2. Machine learning for regression analysis

Once enough points have been sampled, the surrogate model can finally be trained. However, choosing a surrogate model to replace a high fidelity model (e.g. FEM) is not always straight-forward. Certain models are computationally efficient for simple linear relationships but make poor predictions when non-linearities are present, while more computationally expensive models may have improved accuracy for the same amount of training data in these cases.

Most surrogate models are part of the machine learning field. Essentially, machine learning is the ability of a computer program to learn from experiences when given enough data, and some metric to assess its performance. The rapid advances in machine learning have brought forth new types of predictors that offer accuracy and flexibility despite the complexity of the problem. However, since a machine learning program can only make predictions based on training data, the nature of the beast is such that it can never achieve a flawless accuracy. The degree of accuracy can be measured by *machine learning performance* metrics. These performance metrics assess the error that is measured between the prediction and the data corresponding to the true solution. Therefore, the performance that is achieved is also a function of the metric that is chosen to evaluate the learning process. More than one metric can be considered to improve the performance, but on the expense of computational time.

Four distinct machine learning categories exist, namely the *supervised machine learning*, *unsupervised machine learning*, *semi-supervised machine learning*, and *reinforcement learning* [13].

In supervised learning the training data have already been labelled by the user before being fed to the program. This means that supervised learning can be employed when making numerical predictions. For example, the house pricing at a specific location can be predicted by feeding labelled numerical data to the program, such as the number of bedrooms of the house, its size, its age, etc. Since the objective of this work is to utilise a surrogate model for regression analysis, this falls under the umbrella of supervised learning and therefore the rest of the categories are omitted.

2.1.3. Types of surrogate models

One of the most popular non-linear surrogate models used in aerospace engineering, is the Radial Basis Function (RBF) model [14]. The radial function ϕ relies only on the difference of the distance of the data point $\boldsymbol{x}$ from some centre point $\boldsymbol{c}$, evaluated at the Euclidean norm.

$$\phi(\|\boldsymbol{x}-\boldsymbol{c}^{(i)}\|_2) \tag{2.1}$$

The centre point is typically approximated as the average Euclidean distance between the data points [15]. The prediction of the model is then determined by

$$\hat{\boldsymbol{y}} = \sum_{i=1}^{n_t} w_r^{(i)} \phi(\|\boldsymbol{x}-\boldsymbol{c}^{(i)}\|_2) \tag{2.2}$$

where w_r^i are weights that serve as the radial basis function coefficients, and nt are the number of basis functions. One of the most typical basis functions is the Gaussian function:

$$\phi\left(\boldsymbol{x_i}-\boldsymbol{x_j}\right) = \exp\left(\frac{\|\boldsymbol{x_i}-\boldsymbol{x_j}\|_2{}^2}{{d_0}^2}\right) \tag{2.3}$$

where i and j are indices that go up to the number of basis functions and d_0 is a basis function scaling hyperparameter that needs to be properly defined in order to achieve accurate predictions. A more complete list of radial basis functions can be found in [16].

Another prevalent surrogate model is the Kriging model. This model is derived from the summation of basis functions and a selected stochastic process. The prediction for the widely used ordinary Kriging model is then made as

$$\hat{\boldsymbol{y}} = \beta_0 + Z(\boldsymbol{x}) \tag{2.4}$$

where β_0 is a deterministic constant, and $Z(x)$ represents the stochastic process with a mean value of zero and a covariance of

$$cov\left[Z\left(\boldsymbol{x}^{(i)}\right), Z\left(\boldsymbol{x}^{(j)}\right)\right] = \sigma^2 R\left(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}\right) \tag{2.5}$$

where R is the spatial correlation and σ is the standard deviation of the stochastic process that is selected [9]. Although there are various correlation models one could select, the Gaussian model is quite common for engineering purposes [16]:

$$R\left(\boldsymbol{x}^i, \boldsymbol{x}^j\right) = \prod_{l=1}^{n_t} \exp\left(-\left(\theta_l\left({x_l}^{(i)} - {x_l}^{(j)}\right)\right)^2\right) \tag{2.6}$$

where θ stands for the unknown Kriging hyperparameters (tunable settings that dictate the behaviour of the model) which are to be estimated [16]. The Kriging surrogate model requires a lot of training time if the dimension of the problem is high. The training time can be reduced by combining Kriging with a partial least squares method, in order to reduce the number of hyperparameters that need to be properly tuned. [9]. This is accomplished by projecting the problem onto a smaller space by a number of principal components. The obtained information from the projection is then passed to the Kriging correlation matrix so that the amount of hyperparameters is reduced accordingly. The correlation function in this instance is:

$$R\left(\boldsymbol{x}^i, \boldsymbol{x}^j\right) = \prod_{l=1}^{n_t} \prod_{k=1}^{h} \exp\left(-\theta_k\left({w_l}^{(k)} {x_l}^{(i)} - {w_l}^{(k)} {x_l}^{(j)}\right)\right)^2 \tag{2.7}$$

where k is an index that goes up to the number of principal components h.

Surrogate models are increasingly gaining popularity in the aerospace field by replacing higher order models such as FEM and computational fluid dynamics simulations for optimisation purposes, owing to the continuous advances in the field [16]. Much of this comes from the research into machine learning, driven by the relative ease of acquiring large amounts of data nowadays.

2.1.4. Artificial neural networks as surrogate models

Perhaps the most promising and prominent machine learning technique is that of an Artificial Neural Network (ANN). Such a network is very versatile and can be used for a variety of purposes, such as image and speech recognition, classification, and regression analysis among others [17–19]. Therefore, an ANN can act as a successful surrogate model of FEM by performing a regression analysis, given a sufficient amount of data to achieve a certain accuracy.

In mathematics and engineering a great deal of inspiration is drawn from biological processes. This is the case once more, with the development of ANNs, which imitate the nervous activity of working brains. The artificial neuron follows the same working principles as its biological inspiration.

The work that served as a landmark in the field, was the development of the Perceptron algorithm by F. Rosenblatt in 1958 [20]. The Perceptron considered a *linear threshold unit*, which accepted numerical inputs and produced numerical outputs. Each artificial neuron was connected with other neurons through numerical weights, forming the following linear relationship that calculates the weighted sum:

$$z = \boldsymbol{x}^T \boldsymbol{w} \tag{2.8}$$

where $\boldsymbol{x}$ is the input vector, $\boldsymbol{w}$ is the weight vector and z is the linear weighted sum output at the *output layer*. In the lingo of ANNs, a *layer* is a vector that contains certain information as the process is being executed. If all the neurons of one layer are connected with all the neurons of the previous layer, then this is called a *dense layer* [13]. However, the output layer is not complete yet, as this weighted sum is forwarded to a function, known as the *activation function*. Various activation functions can be used, and their role is very significant for the final output.

An improvement of the concept can be achieved by including a numerical bias that is added to the weighted sum. This bias vector, containing one term per artificial neuron, can increase the accuracy of the results when the output vector has components that on average do not pass through the origin of the function's coordinate system.The complete mathematical form of the Perceptron is then given by:

$$\hat{\boldsymbol{y}}_{\boldsymbol{w},\boldsymbol{b}}(\boldsymbol{x}) = \Phi\left(\boldsymbol{x}^T \boldsymbol{w} + \boldsymbol{b}\right) \tag{2.9}$$

where Φ is the activation function and $\hat{\boldsymbol{y}}$ is the output vector. The subscripts $\boldsymbol{w}$ and $\boldsymbol{b}$ of the output vector are the weight and bias vectors that need to be tuned in order to increase the accuracy of the results.

The Multi-Layer Perceptron

The main problem of the Perceptron was that it was unable to solve complex problems, since it could not handle non-linearities. Another of its deficiencies was that it could not solve relatively simple problems such as an Exclusive OR (XOR) problem [21]. Although, it turns out that this limitation can be overcome if multiple Perceptrons are stacked together. This led to the prevalent *Multi-Layer Perceptron* (MLP) architecture that finds general use in the present. An MLP consists of at least three different layers, namely the Input Layer, the Output Layer, and a Hidden Layer (HL) that is between the previous two. Networks with such an architecture are called feed-forward networks because every layer feeds results into the subsequent one, leading to the output vector. An example of a MLP with four inputs, one dense HL and one output is presented in Figure 2.1.

Considering the complexity and non-linearity of a problem, the number of HLs can increase. An empirical rule-of-thumb that has been developed is that "*the optimal size of the hidden layer is usually between the size of the input and size of the output layers*" [23]. Typically, if an ANN consists of more than three HLs, it is called a *Deep Neural Network*; although the definition of an actual deep network has become a bit murky since the inception of that definition is dated back in the 1990s, and the number of HLs that are used currently to solve complex problems is counted in the hundreds [24]. However, care should be taken to select the appropriate number of HLs, as a low number would lead to the ANN underfitting the data, while on the other hand a high HL number could result in overfitting. In a similar fashion, the number of artificial neurons is equally important.

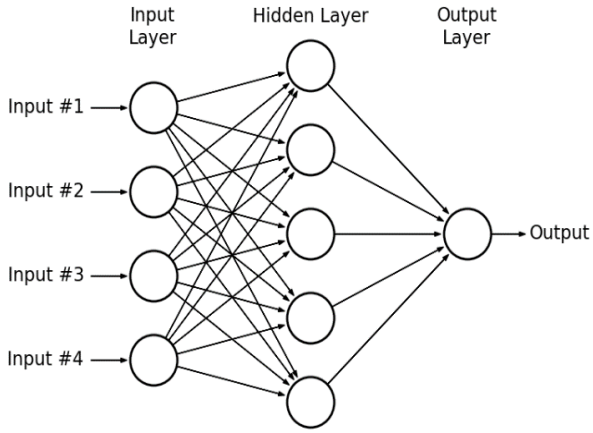


Figure 2.1: The architecture of a MLP with a dense HL between the input and output layers [22].

The Backpropagation algorithm

Adversely, the layer stacking of a MLP poses the significant problem of training the ANN efficiently [25]. A very efficient algorithm that was developed to perform this training function is the *backpropagation algorithm* [26]. Essentially, the algorithm firstly passes forward all the way from the Input Layer to the Output Layer, where it makes a prediction. The parameter tuning will be based on the difference of this prediction with the desired output, or in other words the error of the prediction. Now that the prediction has been made, the algorithm follows the same route in reverse. On its way back it will measure the error at every node of every HL, and finally tweak them to reduce that error.

The error can be measured via various well-known metrics, where the right selection can make a significant difference in the convergence rate of the hyperparameter tuning. For greater fine-tuning, more than one metric can be used if necessary. A function that calculates an error such as the one described here, is called a *cost function* (also known as a *loss function*).

The hyperparameter tweaking can be performed by the *gradient descent* algorithm, which is commonly utilised in mathematical optimisation problems (derivative-based algorithm as mentioned in Section 2.2.4). The gradient descent algorithm performs a partial derivation of the cost function with respect to the hyperparameter that is to be tweaked. The goal is to perform the chain rule derivation in order to nudge the optimisation algorithm towards the minimum of the cost function, by effectively finding the optimal values for the weights. Therefore, the calculated partial derivative will orient the optimisation algorithm towards a local minimum in the design space. Once the direction is known, it needs to be scaled towards the local minimum that has been found. This can be achieved by introducing a real-valued scalar *learning rate*.

As with most optimisation algorithms, the value of the learning rate is problem-dependent, and it may be that some trial-and-error attempts may be required by the user. If the learning rate is too small, it is easy to imagine that this might delay the convergence towards the minimum. On the contrary, a high learning rate might overshoot the valley of the design space, or even worse, result to the divergence of the algorithm.

In order to properly utilise these derivation procedures, the choice of the activation function is paramount. The most popular activation function has been the logistic sigmoid function, although it has fallen out of favour as of lately [26]. The function is defined in Equation 2.10.

$$\Phi(z) = \left(1 + e^{-z}\right)^{-1} \tag{2.10}$$

where $z \in \mathbb{R}$ can be any real number. This function is differentiable at its every point, which means that the backpropagation algorithm now functions as intended to optimise the hyperparameters.

2.1.5. Activation functions

Apart from the logistic sigmoid function, there are more functions that are being actively used, each one possessing its own advantages and disadvantages [27]. The choice of such a function is not always straightforward and it is problem-dependent.

An activation function needs to have a non-zero derivative in order for the backpropagation algorithm to function properly. However, certain function derivatives behave in a more desirable manner than others.

For instance, the values obtained from the gradient of the activation function can either explode or vanish depending on the inputs. This can result in either very slow convergence or complete divergence [28].

Another feature that requires consideration, is whether the activation function is bounded and the spread of its range. Plenty of the most popular activation functions are bounded and lie within the range of $[-1,1]$. The numerical stability of the backpropagation algorithm can be improved if the output of the activation function is bounded, although this may prove inefficient for the convergence of the hyperparameter optimisation process [28]. Even when bounded, there is also the so-called vanishing gradient problem that can arise. This occurs when the magnitude of the gradients reduces significantly close to its bounds. This usually leads to the first HLs being unaffected by the backpropagation algorithm, resulting in poor performance [13].

The Rectified Linear Unit (ReLU) function

The ReLU activation function has gained a lot of traction over the last years, making it one of the most popular functions for ANNs. The function will simply output a value of null if the input is any negative number, and returns the positive value back as a linear function. The function is still nonlinear however, rendering it an appealing choice for an activation function. It can be represented as

$$\Phi(z) = \begin{cases} z & z \geq 0 \\ 0 & otherwise \end{cases} \tag{2.11}$$

This means that the function contains one open bound, as $\mathbb{R} \in [-1,\infty]$. This function effectively tackles the vanishing gradient problem, since it has a null derivative for negative numbers and a derivative value of unity for positives ones. Another significant advantage that it possesses is that a sparse activation of neurons is possible, contrary to a dense one, which can lead to significant computational efficiency (although not always). This is possible because the artificial neurons do not have to fire a signal in an analog way.

However, the ReLU function has a its own drawbacks as well. It suffers from what is known as the *dying ReLUs*. As mentioned already, the gradient will always be null for negative values, which results in certain weight connections to be unaffected by the gradient descent backpropagation algorithm. In this case, the artificial neurons will not be responsive to the error as calculated by the algorithm, essentially dying. When the weighted sum of the neuron's input are all negative, it will only a null output, rendering it useless.

Leaky ReLU function

The problem of dying neurons can be tackled by a variation of ReLU, named *Leaky ReLU*. This function is able to provide a negative slope for negative inputs, and therefore the output will not be null this time. This is made possible via a scalar hyperparameter that defines how much the function will leak. In general, it is considered a superior version, and as B. Xu *et al.* [29] have shown, a comparison between the ReLU and the leaky ReLU function always favours the latter in terms of efficiency. As far as anything else is concerned, this function is equivalent to its predecessor. Other variants exist as well, such as the randomised leaky ReLU, and the parametric leaky ReLU, which are involved with altering or optimising the leaky hyperparameter in an efficient manner. These methods seem to perform equally as well as the leaky ReLU [29].

The Exponential Linear Unit (ELU) function

In 2015, a new activation function, named as the Exponential Linear Unit function, was developed by D. Clevert *et al.* [30], which managed to surpass all of the ReLU variants at that point in time in terms of training time and performance. Mathematically, it can be defined as shown in Equation 2.12.

$$\Phi(z) = \begin{cases} a\,(e^z - 1) & z < 0 \\ z & otherwise \end{cases} \tag{2.12}$$

where a is a non-zero real-valued hyperparameter. The main differences between the ELU and ReLU functions is that the former is able to output negative values, which takes care of the vanishing gradient issue and the dying of the artificial neurons. Unfortunately, the faster convergence rate that ELU function accomplishes is countered by the slower computation that it performs. Additionally, the fact that there is one more hyperparameter present that has to be configured, serves as another drawback.

In 2017, a variant of ELU was developed by G. Klambauer *et al.* [31], named Scaled Exponential Linear Unit (SELU), that achieved a self-normalisation of the ANN. This essentially means that the statistical mean

of the HL output will be null and the standard deviation equal to unity, completely eliminating both the vanishing and the numerical exploding complications. In practice, this results in greater performance of the ANN, especially if it is a deep neural network [13]. There are some requirements for this self-normalisation to occur however, such as a specific parameter initialisation, sequential network architectures (e.g. feed-forward networks), and all HLs need to be dense.

2.1.6. Cost functions

The choice of cost function is crucial in obtaining a network that exhibits good performance. An incorrect decision could even lead to complete divergence from the true function, as the gradient descent backpropagation algorithm requires cost functions that their gradient can be computed [28]. Additionally, the backpropagation algorithm can be performed only if the cost function is related just to the ANN output, as a dependence on other HLs would not work [32]. It is also important that the cost function is a scalar value, computed as an average cost from all the training instances, in order to calculate its derivative and pass it on to the gradient descent backpropagation algorithm [28].

Certain cost functions will now be presented, which are most frequently encountered in regression analysis.

Mean Squared Error (MSE)

The MSE cost function is probably the most prevalent function for regression analysis. It is defined as the sum of the squared distance between the target value and the prediction:

$$\frac{\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2}{n} \tag{2.13}$$

Because the difference is squared, larger difference values yield a higher error than smaller differences. Of course, this results in the stricter punishment of larger differences. The function is also generally preferred if the true value follows a normal distribution, adhering to the principle of maximum likelihood.

Mean Squared Logarithmic Error (MSLE)

This function is usually favoured when the true value has a large breadth of values and a minimal penalisation is to be enforced. In this case, the natural logarithm of both the true and predicted values is calculated first, and then the squared error. The function is defined by the following formula:

$$MSLE=\frac{\sum_{i=1}^{n}\left(log\left(y_i+1\right)-log\left(\hat{y}_i+1\right)\right)^2}{n} \tag{2.14}$$

Its behaviour is very similar to that of MSE for small values, but it corrects the erroneous behaviour of MSE for larger values.

Mean Absolute Error (MAE)

If the probability density function of the true value has extended tails, then the MAE function is an appropriate cost function to choose. Its calculation is simply the sum of the absolute difference between the true and predicted values.

$$MAE=\frac{\sum_{i=1}^{n}\left|y_i-\hat{y}_i\right|}{n} \tag{2.15}$$

However, if there are no outliers in the probability distribution, the MSE cost function may be considered instead.

2.1.7. Training a neural network

With all the important ANN tools and algorithms introduced, the network can now be trained to make predictions. The training process can prove difficult and may contain pitfalls, which lead to poor performance or unwanted behaviour from the ANN. Therefore, the techniques that are commonly employed to circumvent these problems will be discussed. Some of these techniques are not always applied, and their implementation should be considered on a case by case basis. Some of the techniques that have been developed

are used more frequently in Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN), which are quite common in applications such as speech and image recognition, respectively. Although, all of the aforementioned networks can be successfully employed for regression applications, the RNN and CNN types shine when there is a temporal component to the data. Thus, when it comes to regression, feed-forward networks are most commonly used [6, 33–36], and therefore the techniques mentioned will be limited to this relevant domain.

Dataset preparation

During training it is very common to withhold some of the data from the ANN, in order to assess its accuracy and avoid overfitting the data. A very popular range that yields good results, is to feed the network about 80-90% of the total available data, and use the rest for testing purposes [37]. This small set is known as a test set, and its components can be chosen randomly from the complete set. However, a purely stochastic approach can lead to a significant sampling bias if the dataset is small. This problem can be avoided if the data is categorised in different strata, where the test set will incorporate enough data from each stratum, such that the data is correctly representative. This is known as *stratified sampling*.

Hyperparameter initialisation methods

An important element that has not been discussed so far, is the initialisation of the weights of the ANN, after which they will change via the gradient descent backpropagation algorithm for instance. The simplest approach is probably to randomly generate numbers for the neuron connections, through a zero mean normal distribution $\mathcal{N}(0,1)$, although this introduces certain problems. In 2010, X. Glorot and Y. Bengio [38] showed that the very popular sigmoid function at the time, along with the aforementioned random weight initialisation, caused the output variance of each HL to be significantly higher than the corresponding input variance. In essence, there was a constant increase in the variance from the first HLs towards the last ones. In other words, this describes the vanishing gradient problem. The authors claimed that two conditions have to be met if this problem is to be dealt with. First, the output variance of each HL needs to be equal to that of its inputs, and secondly the gradient variance has to be the same during the forward and reverse pass of a HL. Since they concluded that it is not possible to meet both conditions for networks that have a dissimilar number of artificial neurons and inputs, they offered a compromise that refers to the weight initialisation. According to their method (aptly named Xavier initialisation), the weights are randomly chosen from a zero-mean Gaussian distribution and a variance of

$$\sigma^2 = 2\left(fan_{in} + fan_{out}\right)^{-1} \tag{2.16}$$

where fan_{in} stands for the inputs and fan_{out} for the neurons of the layer. Alternatively, a uniform distribution can be used instead:

$$r = \frac{\sqrt{3}\sqrt{2}}{\sqrt{fan_{in} + fan_{out}}} \tag{2.17}$$

As for the family of ReLU functions, a very popular weight initialisation method was proposed by He *et al.* [115], which addressed the rectifier nonlinearities. In this case, the variance simple changes to

$$\sigma^2 = 2\,{fan_{in}}^{-1} \tag{2.18}$$

The He initialisation can also work for the ELU activation function as well. However, the SELU function requires its own initialisation method in order to self-normalise, which is known as LeCun initialisation and draws samples from a truncated zero-centred Gaussian distribution, with a variance of:

$$\sigma^2 = \frac{1}{fan_{in}} \tag{2.19}$$

Batch normalisation

The Xavier and He initialisation methods are very effective at mitigating the vanishing or exploding gradient issues at the start of the training process. However, there is still a possibility that the problem may occur mid-process. Therefore, an additional method may be applied to alleviate the problem, known as *batch normalisation*.

During the ANN training the dataset is divided in many sections, called batches. The *batch size* determines how many instances of data each batch will contain, and it is considered a hyperparameter. The reason that the full dataset is not used instead, is that it can be very computationally expensive for large datasets.

In 2015, S. Ioffe and C. Szegedy [39] proposed that batch normalisation is yet another tool that helps with accelerating the training process of an ANN. This method performs operations on the inputs of each HL by centring them around the origin and normalising them. The result of the operation is then scaled and shifted by utilising two vector parameters respectively. These operations are only possible if the mean and variance of the inputs are calculated first. Thus, the algorithm formulae are presented in the order of their execution:

$$\begin{aligned} \mu_B &= \frac{\sum_{i=1}^{m_B} \boldsymbol{x}^{(i)}}{m_B} \\ {\sigma_B}^2 &= \frac{\sum_{i=1}^{m_B} \left(\boldsymbol{x}^{(i)} - \mu_B\right)^2}{m_B} \\ \hat{\boldsymbol{x}}^{(i)} &= \frac{\boldsymbol{x}^{(i)} - \mu_B}{\sqrt{{\sigma_B}^2 + \epsilon}} \\ \boldsymbol{z}^{(i)} &= \boldsymbol{\gamma} \otimes \hat{\boldsymbol{x}}^{(i)} + \boldsymbol{\beta} \end{aligned} \tag{2.20}$$

where the subscript B refers to the mini-batch, m_B is the number of instances of the mini-batch, $\boldsymbol{\mu}_B$ is the mean vector of the inputs for the mini-batch, $\boldsymbol{\sigma}_B$ is the standard deviation vector for the inputs of the mini-batch, $\hat{x}^{(i)}$ represents the origin centred and normalised inputs for the instance i, ϵ is a very small term that is added to prevent numerical instabilities (division by zero), $\boldsymbol{\gamma}$ is a vector consisting of one scale parameter for each input, $\otimes$ is the operator of element-wise multiplication, $\boldsymbol{\beta}$ is the vector of the output shift for the HL, and finally $\boldsymbol{z}^{(i)}$ is the result of the whole operation.

The authors of the paper found that this algorithm can significantly improve the training performance of the network. Moreover, they determined that higher learning rates can be chosen in the gradient descent backpropagation algorithm, further expediting the training process. On the other hand, additional parameters are introduced that need to be tweaked, making the process more difficult. Another consideration is that the added operations also slow down the prediction making of the ANN.

Early stopping

A common mistake when training an ANN is to allow it to overfit all of the given points. This will result in poor performance when the network will make a prediction on unseen data. A very efficient way to avoid this issue is regularisation. Although batch normalisation served a different purpose, it can also act as a means to regularise the network.

The training of ANNs usually requires a lot of repetitions of going through the whole dataset. Every one of these repetitions is called an *epoch* in neural network lingo. In every epoch, the data is fed to the ANN in a different pattern, in hopes of achieving a greater generalisation upon seeing "new" input. An epoch is not to be confused with an iteration, which is the number of batches that the complete dataset is divided into. Generally, when the model is trained for many epochs it will also begin learning the statistical noise that is inherent in the data. But enough epochs should be chosen such that the model is capable of generalising. This effectively renders the epoch number as a hyperparameter to be selected by the user [40]. One approach would be to train multiple models on a different epoch size and measure their performance. Of course, training multiple complete models is quite inefficient and computationally expensive.

While the network is being trained, part of the evaluation is performed after each epoch on a portion of the dataset that has been kept aside. This set is known as the validation set during training and it should not be mistaken with the training set that is also withheld from the complete dataset. It is common that the validation set is about 20-30% of the whole dataset, but the decision should be made based on how much data is available for the network. It is considered good practice to measure the error of the dataset, as well as for the validation set, with the chosen cost function. The error, and thus performance, is measured at the end of every epoch.

Having defined what an epoch is, the *early stopping* method can now be discussed, which is arguably one of the best regularisation methods [27]. If at the end of an epoch the error increases with respect to the previous epoch, then the training can be terminated. A great indication of overfitting is the increase of the validation set error after it has been stable and small in previous epochs. Therefore, it is quite common to use

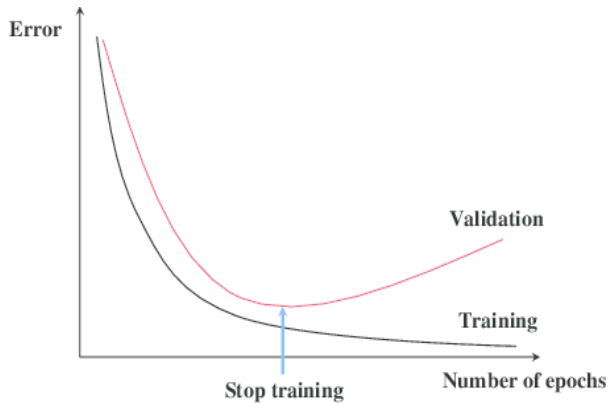


Figure 2.2: The principle of early stopping when the validation set's error increases after a certain epoch [41].

the error of the validation set as the metric to determine when the training should stop. An example of both errors, showcasing the point that overfitting is observed, is presented in Figure 2.2.

However, it is possible that the validation set will decrease again at later epochs, obtaining an even lower error. Thus, a decision has to be made as to how much slack or patience the network should exhibit. The early stopping patience is a hyperparameter that determines how many epochs are allowed after stagnation or an increase in error has been measured. Selection of a large patience value will make the network training slower but may result in better performance in the end.

Network Optimisers

Up to this point, various ways to improve performance and increase the training efficiency have been discussed. What has not yet been touched upon are the alternatives to the gradient descent backpropagation algorithm, which can lead to further improvements. The gradient descent algorithm has been widely popular mainly due to its simplicity and small number of hyperparameters. In the following paragraphs, other popular and more modern network optimisers will be discussed in terms of historical appearance.

Momentum algorithm

As already mentioned, the gradient descent algorithm uses a learning rate hyperparameter that dictates the scale of the step towards the minimum. A more intuitive approach based on momentum was proposed by B. Polyak in 1964 [42], named *momentum optimisation*. In this approach, the design space is related more to real-world physics, where a sharp drop towards a minimum would accelerate the particle travelling through it. Thus, if the gradient is a large value, the step would increase as well. Additionally, all of the previous gradients are stored and they are subtracted from the momentum vector. In order to restrict the acceleration from reaching excessive values, a scalar momentum hyperparameter is introduced. The algorithm is then differentiated from the gradient descent algorithm, as follows:

$$\begin{aligned} \boldsymbol{m} &\leftarrow \beta\, \boldsymbol{m} - a \nabla_{\boldsymbol{\theta}} C(\boldsymbol{\theta}) \\ \boldsymbol{\theta} &\leftarrow \boldsymbol{\theta} + \boldsymbol{m} \end{aligned} \tag{2.21}$$

where $\nabla_{\boldsymbol{\theta}}$ is the gradient with respect to $\boldsymbol{\theta}$, the new momentum hyperparameter is β and $\boldsymbol{m}$ is the momentum vector. The momentum algorithm is able to travel faster across flat regions of the design space, as long as it has already accumulated enough momentum. This can significantly increase the efficiency of the algorithm with respect to gradient descent. The algorithm is prone to overshooting valleys however, owed to the momentum that it has built up. This may lead in constant back and forth in the valley, until it reaches the bottom. Yet another disadvantage is that the user has to tune another hyperparameter, namely the momentum hyperparameter β.

Nesterov Accelerated Gradient

A simple variation of the momentum algorithm that led to a greater training efficiency was proposed by Y. Nesterov [43] in 1983, called *Nesterov Accelerated Gradient*. In his paper Y. Nesterov proposed that the gradient

that prescribes the momentum should not utilise the gradients of previous positions, but rather calculate the slope ahead. The algorithm is then defined in quite a similar way to the momentum algorithm, except for the difference in the cost function:

$$\begin{aligned} \boldsymbol{m} &\leftarrow \beta\,\boldsymbol{m} - a\,\nabla_{\boldsymbol{\theta}}\,C(\boldsymbol{\theta} + \boldsymbol{\beta}\,\boldsymbol{m}) \\ \boldsymbol{\theta} &\leftarrow \boldsymbol{\theta} + \boldsymbol{m} \end{aligned} \tag{2.22}$$

This updated version of the momentum algorithm usually works better than the initial algorithm, since the direction of the algorithm is mostly correct, thus the calculation of the gradient towards that direction leads to a training speed up.

Adaptive Gradient Descent (AdaGrad)

Moving away from the concept of momentum and back to vanilla gradient descent, J. Duchi *et al.* [44] proposed an adaptive learning rate for the gradient descent algorithm as it moves in the design space. The algorithm requires two steps and its computations are slightly more expensive. On the upside, the training is usually more efficient, owing to the algorithm's robustness, especially for quadratic problems. Additionally, the learning rate's initial user input is not as crucial anymore.

Essentially, the first step of the algorithm collects the square of the cost function's gradient and adds it to the previous value of the collection vector. In the second step the gradient is scaled down, yielding an adaptive learning rate.

$$\begin{aligned} \boldsymbol{s} &\leftarrow \boldsymbol{s} + \nabla_{\boldsymbol{\theta}}\,C(\boldsymbol{\theta}) \otimes \nabla_{\boldsymbol{\theta}}\,C(\boldsymbol{\theta}) \\ \boldsymbol{\theta} &\leftarrow \boldsymbol{\theta} - a\,\nabla_{\boldsymbol{\theta}}\,C(\boldsymbol{\theta}) \oslash \sqrt{\boldsymbol{s} + \varepsilon} \end{aligned} \tag{2.23}$$

Here s represents the collection vector, ε is a very small value that is used to avoid numerical explosion, and $\oslash$ is the matrix element-wise division operation.

Unfortunately, the AdaGrad algorithm has the tendency of stopping before reaching the global optimum [13]. This happens because the learning rate is reduced significantly too soon and the algorithm cannot explore the design space anymore.

RMSProp algorithm

During one of their lectures, G. Hinton *et al.* [45] proposed a fix to the aforementioned AdaGrad problem of stopping too early. Their algorithm, called RMSProp, collects the gradients only from the recent positions of the algorithm, eliminating the excessive slowing of the AdaGrad algorithm in that manner. This is achieved by introducing a *decay scalar* hyperparameter β in the first step of the AdaGrad algorithm:

$$\begin{aligned} \boldsymbol{s} &\leftarrow \beta\,\boldsymbol{s} + (1-\beta)\,\nabla_{\boldsymbol{\theta}}\,C(\boldsymbol{\theta}) \otimes \nabla_{\boldsymbol{\theta}}\,C(\boldsymbol{\theta}) \\ \boldsymbol{\theta} &\leftarrow \boldsymbol{\theta} - a\,\nabla_{\boldsymbol{\theta}}\,C(\boldsymbol{\theta}) \oslash \sqrt{\boldsymbol{s} + \varepsilon} \end{aligned} \tag{2.24}$$

Despite the additional hyperparameter, this algorithm is preferred over AdaGrad as a result of its better performance.

Adaptive moment estimation (Adam)

In 2014, D. Kingma and J. Ba [46] proposed the *adaptive moment estimation* algorithm, which borrowed ideas from RMSProp and momentum optimisation. Specifically, it utilised the hyperparameters that change the momentum and the scale of the square of the gradient. This algorithm is more involved than the previous ones discussed so far, including more steps and more hyperparameters. The construction of the algorithm does not borrow equations from its inspirations one to one, but rather makes small changes to make this combination work.

$$\begin{aligned} \boldsymbol{m} &\leftarrow \beta_1\,\boldsymbol{m} - (1-\beta_1)\,\nabla_{\boldsymbol{\theta}}\,C(\boldsymbol{\theta}) \\ \boldsymbol{s} &\leftarrow \beta_2\,\boldsymbol{s} + (1-\beta_2)\,\nabla_{\boldsymbol{\theta}}\,C(\boldsymbol{\theta}) \otimes \nabla_{\boldsymbol{\theta}}\,C(\boldsymbol{\theta}) \\ \hat{\boldsymbol{m}} &\leftarrow \frac{\boldsymbol{m}}{1-{\beta_1}^t} \\ \hat{\boldsymbol{s}} &\leftarrow \frac{\boldsymbol{s}}{1-{\beta_2}^t} \\ \boldsymbol{\theta} &\leftarrow \boldsymbol{\theta} + a\,\hat{\boldsymbol{s}} \oslash \sqrt{\hat{\boldsymbol{m}} + \varepsilon} \end{aligned} \tag{2.25}$$

The parameters β_1 and β_2 control the rate of decaying of average gradients and squared average past gradients respectively, and t stands for the number of repetitions. The algorithm tries to increase the values of $\boldsymbol{m}$ and $\boldsymbol{s}$, as they are biased towards the origin at the start of the algorithm. Although the algorithm has three hyperparameters, the flexibility of the adaptive learning rate renders this hyperparameter less sensitive to the choice of its initial value.

Nesterov adaptive moment estimation (Nadam)

In 2016 T. Dozat [47] combined the Adam algorithm with the Nesterov momentum optimisation algorithm and reported better performance than the Adam algorithm all around. The algorithm is slightly more expensive but based on the report, the performance boost is worth it. No new hyperparameters are required for Nadam. The Nesterov implementation into the Adam algorithm was performed as such:

$$\begin{aligned}
&\boldsymbol{m} \leftarrow \beta_1 \boldsymbol{m} - (1-\beta_1)\nabla_{\boldsymbol{\theta}} C(\boldsymbol{\theta}) \\
&\boldsymbol{s} \leftarrow \beta_2 \boldsymbol{s} + (1-\beta_2)\nabla_{\boldsymbol{\theta}} C(\boldsymbol{\theta}) \otimes \nabla_{\boldsymbol{\theta}} C(\boldsymbol{\theta}) \\
&\hat{\boldsymbol{m}} \leftarrow \frac{\boldsymbol{m}}{1-\prod_{i=1}^{t} {\beta_i}^t} \\
&\hat{\boldsymbol{s}} \leftarrow \frac{\boldsymbol{s}}{1-{\beta_2}^t} \\
&\bar{\boldsymbol{m}}_t \leftarrow \frac{\nabla_{\boldsymbol{\theta}} C(\boldsymbol{\theta})\left(1-\beta_t\right)}{1-\prod_{i=1}^{t}\beta_i} + \beta_{t+1}\hat{\boldsymbol{m}} \\
&\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + a\,\hat{\boldsymbol{s}} \oslash \sqrt{\hat{\boldsymbol{m}} + \varepsilon}
\end{aligned} \tag{2.26}$$

It is interesting to note that for some cases T. Dozat found that RMSProp can outperform the Nadam algorithm, but not Adam. This highlights that there is no best algorithm, and that the optimiser is case-dependent.

2.2. Optimisation algorithms

2.2.1. Introduction to the concept of optimisation algorithms

When seeking to optimise a structure, an accurate and fast prediction is only useful if paired with an algorithm that will perform said structural optimisation. Since the field of optimisation is of immense importance in quantitative disciplines, numerous optimisation techniques and algorithms have been developed that one can implement. Unfortunately, there is no clear winner among them, as per the famous *no free lunch theorem*, stating that if one does not make any assumptions about the data, then there is no reason to prefer one model over the other. Therefore, some algorithms perform better in certain problems and worse in others [48].

2.2.2. Single-objective optimisation

In most optimisation problems, the aim is to minimise a certain function, called the objective function. If only one objective function is considered, then the optimisation problem is called a single-objective optimisation problem. In real life problems, most of the solutions to such a problem would be unattainable, due to limiting constraints. Therefore, as an example, a very simple minimisation problem that is constrained to only positive numbers for the input vector may be

$$\begin{aligned}
&\underset{\theta}{minimise}\ f(\boldsymbol{x}) \\
&subject\ to\ \boldsymbol{x} > 0
\end{aligned} \tag{2.27}$$

There are numerous algorithms that can find a solution to the problem of minimising such an objective function, with varying degrees of computational efficiency. A very important point to make here is that a solution can have many local minima in its design space but only one global minimum. Thus, when performing an optimisation it is crucial to understand if a local minimum or a global one has been found by the algorithm.

2.2.3. Multi-objective optimisation

In reality, multiple items need to be considered in engineering optimisation, such as cost, time, performance, etc. Of course, this means that different objective functions will have to be included in the optimisation. Such a problem is then aptly called a multi-objective optimisation problem. Unfortunately, the nature of the beast is such that it is no longer clear which solution is best, as found by an optimisation algorithm [49]. Therefore, once the optimisation has been performed, a variety of possible candidate solutions is found, requiring a decision-maker to evaluate them and choose the one considered best among them. There are a few methods that can help the decision-maker in assessing the candidate solutions, with most of them incorporating certain weights to highlight the best solution [49].

It is considered instructive to first take a step back, in order to better grasp the problem. At this point, it is important to introduce the notion of Pareto optimality. A solution is said to be Pareto optimal if one of the objectives can only be improved in the expense of another objective. In other words, a better solution with respect to one objective cannot be obtained without sacrificing the optimality of one or more of the other objectives. A Pareto optimal solution will then lie on a Pareto frontier, which can be thought of as including the best possible solutions. All the other solutions away from this frontier are said to be dominated by the Pareto optimal solutions, and are therefore sub-optimal. An illustrated example, where two objectives are considered, can be seen in Figure 2.3. The curve in this case is the Pareto frontier and the green marks on it are the Pareto optimal solutions, while all the other solutions are dominated by them.

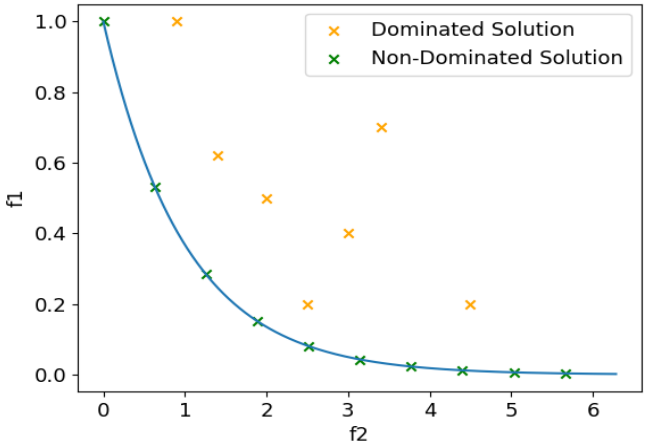


Figure 2.3: Optimal solutions lying on the Pareto front.

2.2.4. Types of optimisation algorithms

Some algorithms are better at finding the global minimum than others that can get stuck at a local minimum more easily. A brief, yet representative list of the ways that optimisation algorithms can be categorised is:

- *Derivative-based algorithms.* It is probable that these algorithms find the most use in most engineering fields. The derivative of the function (or gradient when non-scalar values are involved) is utilised to point the optimisation algorithm towards a direction where a minimum valley is present. For that reason these algorithms are also commonly called descent algorithms. Once they find the direction they are to follow, they take a step towards that valley. Care should be taken in selecting an appropriate value for the step size, since a high value can lead to the algorithm overshooting the bottom of the valley. However, there are methods that utilise an adaptive step size to overcome this problem. Furthermore, derivative-based algorithms can be divided by the order of their derivative. In practice, first and second order derivatives are used. A second-order derivative is computationally more expensive, but it is also able to gain information on both the direction and step size.

- *Direct method algorithms.* These algorithms are able to find a minimum without taking a derivative of the function. Typically, they search in different directions iteratively. More sophisticated types perform computations regarding the step size, direction and increased likelihood to approach the design space that contains the global minimum.

- *Stochastic method algorithms.* In this case, the randomness of the algorithms is employed. Stochastic methods can also be combined with other types of algorithms, and they seem to improve the convergence to a minimum in most cases. By drawing from a selected probability distribution, the design space can be explored in a more efficient manner. Once sufficient exploration has been done and the regions containing minima are identified, a faster convergence towards a solution is executed.

- *Population method algorithms.* In the majority of cases, these algorithms are inspired by processes that are found in nature. Many different names are also assigned to them, such as evolutionary optimisation algorithms and metaheuristic algorithms. A heuristic algorithm is one that sacrifices accuracy and does not solve an exact optimisation problem, instead favouring an approximation is more computationally efficient. In a population method, many points from the design space are sampled, resulting in a fast exploration. This circumvents the problem of getting stuck in a local minimum to a certain degree, and thus they can approach the global minimum in an approximate manner. The points, commonly called individuals, may be selected from one of the sampling methods that have already been mentioned. The biggest class of population algorithms are perhaps the Evolutionary Algorithms (EAs). These algorithms update the individuals with each passing iteration (or generation), until an individual that minimises the function (or a vector of individuals for a multi-variate function) is found [51, 52].

Population method algorithms that are inspired by nature, such as genetic algorithms, have gained traction in the last two decades, although their inception is dated back to the mid 1960s [53], with the development of Evolutionary Strategy (ES) algorithm. Since then, many variations have spawned from the initially proposed ES algorithm, seeking to improve its efficiency, similarly drawing inspiration from processes observed in nature. Their popularity is mainly attributed to their ability to effectively ignore local minima and asymptotically converge towards the global optimum. Evolutionary algorithms were used extensively to carry out the work that was performed in this thesis, and thus, their working principles are further expanded upon in this section.

The research on EAs is vast, owing to the research that has been conducted over decades. As stated before, since the inception of the GA, many other nature-inspired variations have spawned, hoping to improve upon the initial concept. However, this proliferation of nature-inspired variations has been criticised [54], and it is thought that these variations fail to bring anything novel to the table, and they only contribute in creating more nature-inspired analogies. Additionally, it would be impractical to include a complete list of all the research as well as all the variations of the initially proposed GA. Therefore, only the most relevant algorithms towards the completion of the objectives set in this work will be considered.

2.2.5. Genetic algorithm

The prominence of GAs can be attributed to their flexibility and possible application to a wide range of problems [55]. Of course, this is reflective of the concept from which they were devised, since natural processes have evolved to be very efficient under various types of environments.

In biological evolution, those individuals who are the fittest among a population are more likely to pass on their genes to the subsequent generation. In 1975, a variation of the ES algorithm was proposed, termed a Genetic Algorithm (GA) [56]. This variation proved much more popular than its predecessor, where the main difference lies in the selection and evolution of the population. In ES the parents are selected first, and then through breeding their children are produced, whereas a GA only gradually selects the parents and produces children until enough of them have been created [52]. In the context of mathematical optimisation, the concept of fitness can be thought of as the inverse of the result of an objective function in a minimisation problem. In other words, while the algorithm seeks to minimise the objective function of a minimisation problem, this also means that it seeks the individuals with the highest fitness to pass on their genes. Essentially, these genes are the variables $\boldsymbol{x}$ of an objective function $f(\boldsymbol{x})$. In design space, the individual is a point that is referred to as a chromosome or a genotype [57]. The fittest individuals then pass on their genes to the subsequent generation, by first undergoing two genetic operations, namely the crossover and mutation operation.

Chromosome Initialisation

In most applications, the chromosomes are real or integer parameters and hence they belong in $\mathbb{R}^n$ or $\mathbb{Z}^n$ space respectively. To begin the optimisation procedure, an initial population is required, with a certain

combination of genes $\boldsymbol{x}$. The generation of these points can be performed by any of the methods that were discussed in Section 2.1.1 (e.g. random sampling, LHS, etc.).

Selection operation

Although, the general idea is to select the best individuals with the highest fitness to produce more children, there are many approaches when it comes to this selection process. Each process possesses its own advantages and shortcomings, and it should be chosen on the unique characteristics of the given problem to be solved [58].

Perhaps the most unimaginative, yet simple, process is that of *random selection*. In this case, the individuals are selected randomly from the population of the current generation to produce children. An effective implementation of this selection operator should use a permutation scheme to avoid the repetitive selection of individuals.

An approach that is less stochastic, is the *truncation selection*. In this selective method, the parents are chosen out of the n best parents of the population. This is rather simple to implement and can prove to be efficient [59].

Roulette wheel selection, is another selection operation, where more randomness than the truncation selection is involved in choosing the parents. First, the fitness of all the individuals is compared and ranked to the rest of the population. Then, a probability of selection is related to the fitness values, where the fittest individuals have a higher likelihood of being selected. However, it is considered that this method introduces too much stochastic noise to be preferred over other methods. A more recent version based on stochastic acceptance improves on the initial concept, by accepting a random individual i, with a probability of f_i/f_M, where f_i is the fitness of the individual and f_M is the maximum fitness found in the current generation [60].

One of the most popular selection operators is the *tournament selection*. This process involves a tournament between certain individuals to decide who will pass on their genes. Out of n randomly selected individuals, a tournament is held, or in most cases many tournaments within the same generation, where the fittest individual effectively kills the rest of the participants. This process allows for some of the weaker individuals to live, granted they are only paired with even weaker individuals in their own tournament [61]. A tournament has certain characteristics, such as size and *selection pressure*. This pressure is the likelihood of the individual to participate in a certain tournament, and it has been shown that it is an important parameter in order to achieve faster convergence [58].

Furthermore, by maintaining only the best individuals that have been encountered in generation n for the next generation $n+1$ is termed an elitist selection scheme and it has been found to generally improve the performance of the GA, since the best solutions found so far are not lost [62].

Crossover operation

The crossover operator dictates how the genetic material of the parents will be combined to form the child [63]. Unlike in nature, the child can inherit genes from multiple parents and not necessarily from just two. There are various methods to perform the crossover, with ongoing research to improve upon existing methods, and therefore some of the most popular ones will be presented.

Quite possibly the simplest approach and the standard for binary problems (also possible to be applied for real-valued problems), is the *k-point-crossover* operation. Here, the child's genes are divided in k sections, where each section has genes from a specific parent. For instance, in single-point-crossover there are two sections divided by the crossover point, where the first one has the genes from parent A, and the other one inherits from parent B. Any number of crossover points can be selected to form the wanted sections. A popular variation of the point-crossover is the *uniform-crossover* operation [63]. This crossover randomly selects each point with a probability of 0.5, from either parent. Alternatively, a *half-uniform-crossover* can be performed, where the differences between the parents will be identified first and then take half of that from the other parent [64, 65]. A representation of the uniformity of this operation can be seen in Figure 2.4.

Yet another popular method is the Simulated Binary Crossover (SBX) [66]. This crossover can be applied to both real-valued and integer-valued optimisation problems. As its name aptly implies, the SBX method simulates the operation of the single-point-crossover on binary strings. The process via which the children inherit their characteristics is stochastic in nature. Firstly, the ratio of the absolute difference in the values of the children over that of the parents is calculated:

$$\beta = \left| \frac{x^{(2,t+1)} - x^{(1,t+1)}}{x^{(2,t)} - x^{(1,t)}} \right| \tag{2.28}$$

Figure 2.4: A representation of the genes of a child, as inherited by the parents via a uniform-crossover operation [51].

where t is the number of the generation and x shows the individual. A random number is then generated, between the interval [0,1]. Consequently, a value for β can be found from a selected probability density function, such that the area under the function is equal to the value that was randomly generated. The probability density function in this case is:

$$P(\beta) = \begin{cases} 1/2\left(\eta_c + 1\right)\beta_i{}^{\eta_c} & \beta_i \leq 1 \\ 1/2\,\frac{\eta_c+1}{\beta_i{}^{\eta_c+2}} & otherwise \end{cases} \tag{2.29}$$

Notice that a parameter η_c has been introduced, and it is referred to as the distribution index. It is recommended that it is given the value of 2.0 for single-objective optimisation problems [66]. An expanded version of this crossover operation has been developed, where the distribution index is self-adaptive as the design space is being searched for both single-objective optimisation and multi-objective optimisation problems as well [67].

Mutation operation

The mutation operation is carried out after the crossover operation has been performed. This operation allows for the random mutation of the genes that the child has inherited from its parents. The reasoning behind this operation is to allow for a better exploration of the design space. For instance, a random mutation can escape a local minimum and include more genetic variability into the population. Of course, mutating the genes of the offspring very frequently will considerably slow down the convergence towards a minimum. Therefore, in accordance with mutation in nature, the probability of the operator should be low (typically less than 2% [57]).

A point of concern regarding the mutation operator is the feasibility of the mutated variable in constrained problems. If certain regions of the design space are off limits due to the constraints of the optimisation problem, then it should not be possible for the mutation to produce points that lie within these regions. In these cases, a bias can be applied to direct the search [68].

Typically two different approaches are followed to perform the mutation. For binary problems, the chromosome has a certain probability of being "flipped", resulting in the only other possible value. Therefore, if a chromosome has k bits, it is common to choose a mutation rate with a value equal to $1/k$, which corresponds to one mutation for every child chromosome on average. For problems in $\mathbb{R}$ space, a zero-mean Gaussian noise function is typically added to the offspring, although a bit-flip approach could also be implemented [51]. Therefore, the new offspring are:

$$\boldsymbol{x}' = \boldsymbol{x} + \sigma\,\mathcal{N}(0,1) \tag{2.30}$$

where σ represents the mutation rate in this case. Different distributions can be selected however, such as the one that is used for the SBX operation.

2.2.6. Differential evolution algorithm

A different and more recent heuristic approach to that of the GA, is the Differential Evolution (DE) algorithm [69]. The primary advantage of this algorithm is the efficiency by which it converges to the global extremum [70]. The building blocks of the DE algorithm are similar to that of the GA, albeit a higher order of complexity is involved as the DE algorithm is more sophisticated.

The algorithm is capable of mutating each individual in the population by utilising other individuals of the same population, resulting in the improvement of the individual [71]. Essentially, the algorithm can be considered as an adaptive mutation algorithm, where the mutation rate is higher when the characteristics of the population are spread out throughout the design space but it decreases as a solution converges towards an extremum.

The working principle of the DE algorithm is vector addition and subtraction to generate new children within the population. At first, three individuals are chosen at random within the population, $\boldsymbol{x_{r1}}$, $\boldsymbol{x_{r2}}$, and $\boldsymbol{x_{r3}}$. Then a child is generated according to the following formula:

$$\boldsymbol{v} = \boldsymbol{x}_{r1} + F(\boldsymbol{x}_{r2} - \boldsymbol{x}_{r3}) \tag{2.31}$$

where F is a differential weight. In that way, a mutated child away from the other three points is generated, successfully exploring the design space in an efficient manner. The child can then be crossed-over with each member of the population, and after a new group of offspring has been produced, a comparison between the children and the parents can be performed with the fittest replacing the other individuals. The mutation process is shown in Figure 2.5.

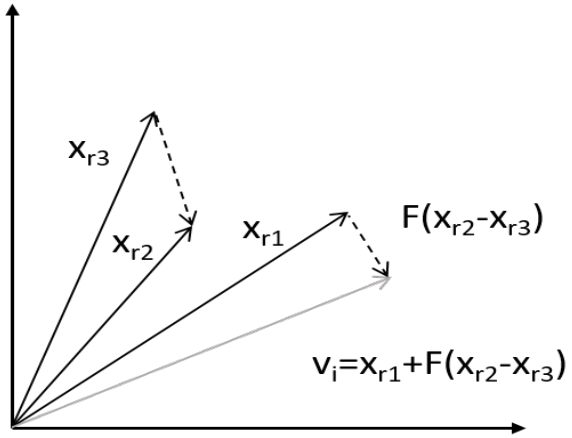


Figure 2.5: The production of a new child from the initial vectors, following the operations of vector addition and subtraction. Recreated from: [57].

The typical cross-over operation that is performed in a DE algorithm produces a trial vector [69]:

$$u_{ji,G+1} = \begin{cases} v_{ji,G+1} & if\left(randb\left(j\right) \leq CR\right) \ \ or \ \ j \equiv rnbr\left(i\right) \\ x_{ji,G} & if\left(randb\left(j\right) > CR\right) \ \ or \ \ j \neq rnbr\left(i\right) \end{cases} \tag{2.32}$$

where G stands for the generation index, j is an index for the individuals within a population $\in [\![1,n]\!]$, $rnbr\left(i\right)$, $randb\left(j\right)$ is the j^{th} calculation of a random number following a uniform distribution $\in [0,1]$, CR is a user-defined crossover constant $\in [0,1]$, $rnbr\left(i\right)$ is a randomly selected index $\in [\![1,n]\!]$ which makes sure that $u_{i,G+1}$ inherits at least one feature from $v_{i,G+1}$. The selection is then made based on whether the trial vector $u_{ji,G+1}$ as a lower cost function than the vector $x_{ji,G}$, where the $x_{ji,G+1}$ replaces the latter, otherwise the vector $x_{ji,G}$ is maintained.

A significant improvement to the algorithm has been made by Elsayed *et al* [72] by introducing a variation of the chosen operator, which produces new CR and F values for each individual. This effectively renders the algorithm adaptive to the changes across generations by altering the user-defined hyperparameters.

2.2.7. Non-dominated sorting genetic algorithm II

When a complex problem that includes many variables that need to be optimised simultaneously is present, a multi-objective optimisation algorithm is necessary. All the aforementioned EAs were only concerned with single-objective optimisation problems, although alternative versions that hold the same principle concept have been used for multi-objective optimisation problems. In practice, non-dominated sorting is used in order to move a population of solutions towards the Pareto front with each passing generation. The first algorithm to utilise this concept is known as the Non-dominated Sorting Genetic Algorithm (NSGA) [73]. However, this algorithm is hardly ever used in practice, since it suffers from low computational efficiency, does not incorporate an elitist scheme, and there is a further need to specify a sharing parameter between the solutions. An improved version that tackled these problems served as an update to the previous algorithm, with the name NSGA II [74]. The algorithm has been found to be so successful and popular that it now frequently serves as the benchmark against which novel algorithms are compared to [75].

Non-domination ranking essentially performs an evaluation of all the solutions in the design space and they are assigned a level based on their proximity to the Pareto-front. Since the Pareto front contains all the non-dominated solutions of the entire domain, they are said to be non-dominated individuals of level/rank 1. Therefore, as the level increases, the solutions are further away from the optimum Pareto front.

Another important aspect of a multi-objective optimisation algorithm is the capability to produce solutions that are spread around in the design space. Thus, as they move towards the Pareto front they provide for a lot of coverage. This is crucial for the subsequent decision-making of which solution should be chosen as optimal for the given problem. One can imagine that a lot of solutions that are clustered in a small region of the design space, does not allow for much flexibility. The NSGA II algorithm was developed with this in mind, by incorporating the scheme of a crowding distance. This crowding distance utilises the Manhattan distance, otherwise known as the L_1 norm or the taxicab distance, which is defined as:

$$\sum_{i=1}^{n} |p_i - q_i| \tag{2.33}$$

where $\boldsymbol{p}$ and $\boldsymbol{q}$ are vectors. Therefore, the distance that is computed between the solutions is not that of the closest points. Instead, the population is grouped with respective to each of the considered objectives. The fitness difference of every point is then summed up between two neighbours. The first and the last solution in this sorting are assigned to an infinite crowding distance and are thus immediately selected. The aim here is to select neighbouring points with a high crowding distance. The idea of this sorting is demonstrated by an example in Figure 2.6.

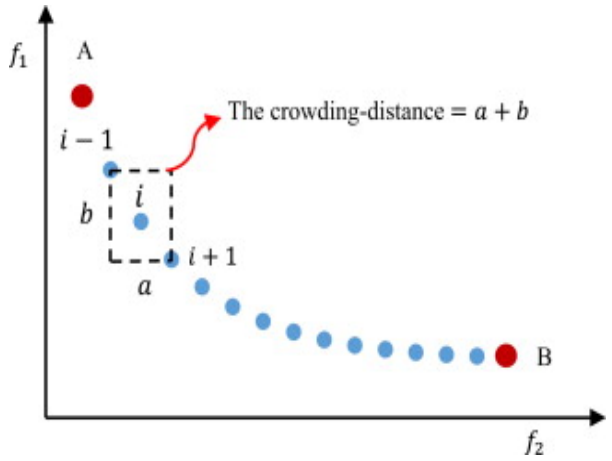


Figure 2.6: The crowding distance between various points lying on a solution front, utilising the L_1 norm in the NSGA II algorithm [76].

Finally, the NSGA II algorithm applies a selection pressure by employing a binary tournament mating selection. Thus, the individuals that are found are first compared by their Pareto level and then by their crowding distance. With the majority of engineering problems involving multiple objective functions, the NSGA II algorithm has found countless applications. Although, close to two decades have passed since its inception, it is still the algorithm that is predominantly used in most multi-objective optimisation problems, due to its simplicity with respect to more novel but complicated algorithms [77–80].

2.2.8. Non-dominated sorting genetic algorithm III

A more recent algorithm that hopes to improve on NSGA II, is aptly named NSGA III [81, 82]. The algorithm was developed with many-objective (defined as more than 4 objectives) optimisation in mind. There are certain problems inherent to such a high-order optimisation problems, such as the expensive computation of the crowding distance in high-order design space, the larger fraction of the population that is non-dominated which does not leave room for the generation of many new solutions in a generation, leading to inefficiency.

The NSGA III aims to overcome such problems by altering the way the selection operator functions. The idea in this algorithm is to promote the diversity among individuals by providing a number of reference points. These points have to be well-spread in order to cover most of the design space. After the initial points

have been provided as parameters, they are updated in an adaptive manner. This operation completely replaces the crowding distance approach that is followed in NSGA II.

If little is known of the problem that is to be optimised, the selection of the appropriate reference points becomes quite difficult. Rather, a different technique can be employed, by utilising the method developed by Das and Dennis [83], called the weighted-sum approach, which is capable of generating reference points on a normalised hyper-plane, which is uniformly aimed towards all of the objective axes and contains an intercept of one on each of the axes.

The algorithm first tries to find solutions around the reference direction that contains the least ones so far, and then assigns them to that direction. In the case where the reference direction does not have any solutions yet, then the solution having the smallest perpendicular distance in the normalised design space will be passed on to the next generation. The normalisation of the population is adaptive, where the vertices of the hyper-plane that correspond to the extreme values of the population (ideal point) for every objective function are found and then subtracted by the objective function which they represent, effectively translating the ideal vector containing all the ideal points to a zero vector.

An example between the selection and grouping of individuals between the NSGA II and NSGA III algorithms is presented in Figure 2.7.

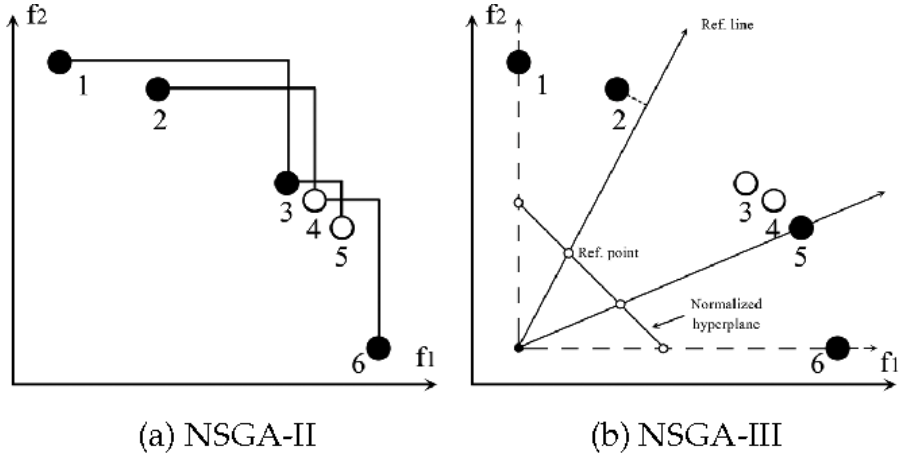


Figure 2.7: An example of the difference in the selection of solutions, between the NSGA II and NSGA III algorithms, for two objective functions [84].

The algorithm serves as an efficient computing approach to various problems, without having to tweak many parameters [85]. Although it was found that the advantages over algorithms with properly tweaked parameters is minuscule, the main benefit of the algorithm lies in its robustness and convenience.

A variant of the NSGA III, called the Unified NSGA II (UNSGA III), has shown that further increases in performance are possible for mono-objective and multi-objective problems by introducing tournament pressure [84].

2.2.9. Constraint handling

So far, the inclusion of factors that constraint the optimal solution were not considered. A constraint can significantly alter the optimal solution by reducing the feasible design space. Since constraints are almost guaranteed in most engineering problems, an efficient way of dealing with them is necessary. Thus, a lot of research has been involved in constraint handling over many decades. Therefore, some simple approaches will be introduced for reference, along with the constraint handling methods that were utilised in the current work. The reader is encouraged to read the work of M. Montes and C. Coello [86] for an extensive historical review of constraint handling in nature-inspired optimisation algorithms.

In multi-objective optimisation algorithms, the constraints can be handled in a simple manner if one considers the violation of each solution with respect to the imposed constraint as another objective function that needs to be minimised [87]. In such an approach, the individuals can be ranked depending on the degree of the violation they are performing. A more refined method considers the constraints as objective functions themselves, where the population is subsequently subdivided in sub-populations each representing the constraints [88]. The primary drawback of these methods is when a high number of constraints is present, where the search cannot be directed towards the optimum in an efficient manner, due to possible lack of pressure.

When considering single-objective optimisation algorithms a popular way to handle constraints is through penalty functions. The most extreme version of a penalty function is the *death penalty*, but it is also the simplest to implement. Essentially, it performs a check on whether a constraint is violated after the crossover and mutation operators. If the solution does not lie in the feasibility region, it is discarded and the process continues until a feasible solution is found. Albeit simple, it is easy to see why this approach would take a long time to converge as it does not use any metric to direct it towards the feasible region.

By abandoning the concept of the death penalty, and moving towards different penalty functions, a greater efficiency can be achieved. The penalty function will deteriorate the fitness of a solution if it has violated any constraints. This deterioration is controlled by the degree of the violation:

$$f(\boldsymbol{x})' = f(\boldsymbol{x}) + ag(\boldsymbol{x}) \tag{2.34}$$

where $g(\boldsymbol{x})$ represents the constraint function and a is the penalty factor. Since it is highly likely that the optimal solution will be at the border with an infeasible region for most optimisation problems, then if the initial solution happened to be in the infeasible region, its deterioration will also move the search towards the optimum. Unfortunately, the value of the penalty factor is problem-dependent, but an adaptive factor can also be computed if it is included in the genetic optimisation process [89].

Another popular approach to constraint handling is the repair operator. Such an operator can act as a projection of the infeasible solution to feasible design space. There are various ways this can be performed however. A simple approach would be to direct the infeasible solution towards the closest solution in the feasible design space through a direct line. However, in order for this to work, one can imagine that a requirement is that the solution space is continuous. Another important requirement is that knowledge of the design space can be obtained, as otherwise the repair operator cannot be adapted. Thus, the equation of the constraint must be known. Since it is possible to search within the infeasible space with that approach, more solutions can be considered in contrast to an approach such as the death penalty method [90].

A more sophisticated technique than the ones already mentioned, is one that can self-adapt the fitness formulation, such as the one developed by R. Farmani and J. Wright [91]. Here, the constraint violations are represented by an infeasibility measure that penalises the infeasible solutions in two stages. In this approach, infeasible solutions with a small violation are kept if they have a low objective function value. The reasoning behind this is that this can be more efficient in problems that consider a lot of constraints, where the solutions may lie on the constraint bounds. Of course, the solutions that are not close to the constraint bounds are seen as definitely non-optimal and they are penalised.

The infeasibility metric can be calculated by the sum of the normalised constraint violation values. Firstly, the worse an infeasible solution is the more it is penalised, where the infeasible solution is penalised by twice the value of the best's solution objective function, and the rest are penalised exponentially according to their infeasibility. This method is advantageous because there is no parameter tweaking and the global optimum can be identified from within the infeasibility region.

2.2.10. Performance metrics

Finally, it is important to have certain metrics that measure the performance of the optimisation algorithm (e.g. convergence). There are many problems in the optimisation field that are used for benchmarking and comparison purposes for new algorithms, and their Pareto front is well-known. Therefore, certain metrics that measure the efficiency of an algorithm to converge towards that Pareto front have been suggested, although when the optimal solution is not known beforehand, the *hypervolume* metric (or S-metric) [92] is of interest.

In multi-objective optimisation problems, the hypervolume between a non-dominated front and a reference point is measured, to serve as a metric of that front. If the problem is 3-Dimensional (3-D) it can simply be called as the volume, and for a 2-Dimensional problem (2-D) an area. This metric will compute a scalar hypervolume, from all the solutions to a user-defined reference point in the design space. Essentially, the solutions and the reference point will be the vertices of this computed hypervolume. In that way, different solutions or fronts can be compared to the same reference point, and whichever hypervolume has the highest value indicates the best set of solutions. It is very important that the reference point is selected such that a hypervolume can be constructed from all the vertices. One can imagine that if the reference point lies among the solutions, it would be very difficult or impossible to create such a geometrical hypervolume.

2.3. Composite structures

Structural optimisation is a field that incorporates all of the ideas introduced so far. The more complicated the structure, the more important it is to apply an efficient approach in terms of computational time and accuracy. Therefore, the necessity for sophisticated optimisation approaches is critical in structures made from composite materials. Composites have an inherent complexity that renders their analysis more difficult and time consuming to perform. Of course this is rather unfortunate, since composite structures can realise significant weight savings, which are always coveted in the aerospace sector. Hereafter, all the structures mentioned are considered to be made of composite materials.

2.3.1. Variable stiffness composite structures

As already mentioned, a composite laminate typically allows a constant angle orientation for each of its individual plies. However, greater design flexibility and performance can be achieved by spatially varying these ply angles to construct a variable stiffness (VS) composite, instead of a constant stiffness (CS) composite. In the following paragraphs, the manufacturing methods of VS structures will be explained, along with the inherent manufacturing constraints that affect their design. Furthermore, the various cases where these structures have been employed, will be discussed, as well as their modelling technique. Finally, the optimisation approaches that have been followed, in addition to any surrogate models that were utilised in conjunction, will be mentioned.

2.3.2. Manufacturing of VS composites

At the time of writing, there are two popular manufacturing methods for VS composites:

1. Automated Fibre Placement (AFP)
2. Continuous Tow Shearing (CTS)

The latter method is essentially a modification of AFP [93]. A recent work by Bittrich *et al.* [94] also recommended the Tailored Fibre Placement (TFP) method, as a suitable manufacturing option, which is also a variation of the AFP process. The TFP method will not be included in the discussion however, since most of the work concerning VS structures is performed with the AFP and CTS processes.

They all possess their advantages and disadvantages, which impose certain restrictions that are relevant to mathematical optimisation. Other considerations, such as the capital cost of the machines, deposition rates, the ease of use, etc., will not be included in the coming discussion.

Automated Fibre Placement method

The most widely used method for VS manufacture and for modelling, is the AFP manufacturing method [95, 96]. In this method, a band of tows, known as a course, can be placed and steered on a complex shaped surface by a machine. A tow is a bundle of many fibres that creates a packet, to improve handling and placement time efficiency. Some common tow widths are 3.175 mm, 6.35 mm, and 12.7 mm [97]. There is a limit on the number of tows that can be within a course, which is typically about 32 tows for small tow widths. The machines that are employed, are computer controlled and they commonly have a robotic arm with 6 DoF and an additional axis of rotation for a mandrel, if the surface is to be rotated [98]. An example of an AFP machine in action, is shown in Figure 2.8.

The course is pulled towards the part surface and is heated in the process for increased tackiness. Pressure is then applied by a roller to better conform the course and have it adhere onto the surface. The machine is also able to control the feed rate of individual tows, which allows for the option of steering the fibres. The tows can be cut by the machine head and restarted again.

The courses follow a reference curve when laid down, which is the centre line of the course. There are two different ways that can be employed for the placement of the courses, namely the parallel and shifted methods [100]. The parallel method is capable of producing constant thickness composites by having the fibre paths lie at a constant distance from the reference curve. However, it is not used in practice due to the limits that it imposes on the allowed curvature of the steered tows, since excessive curvature may lead to the fibres wrinkling. This manufacturing constraint is one of the most important factors that hinders the optimality of VS structures, and it is simply known as the curvature constraint.

The shift method finds wider acceptance because it does not restrict the design space as much as the parallel method. In the shift method, all the tows remain parallel to the reference curve. For instance, if

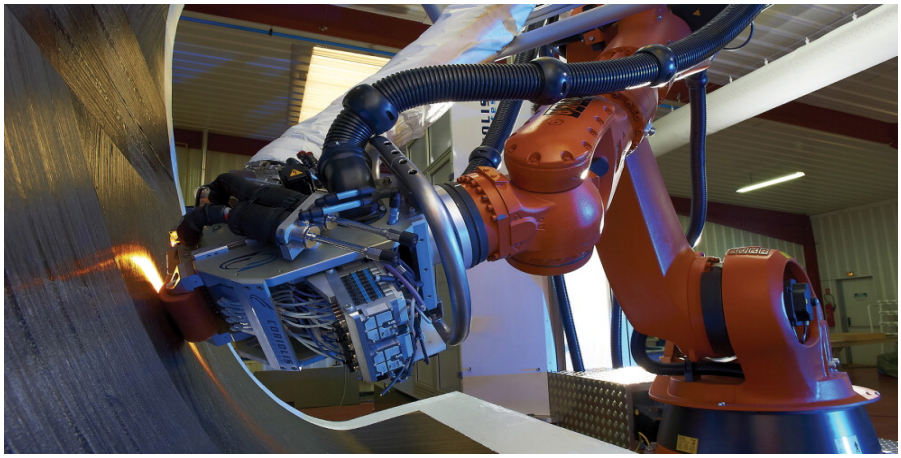


Figure 2.8: An AFP machine laying down layers on a concave mould [99].

the angle variation is only a function of one axis in Cartesian coordinates, then the course will be shifted perpendicular to that axis, as shown in Figure 2.9. Typically, the centre-point of the structure is selected to act as a reference for the shift. When the fibres are steered, the finite width of the courses gives rise to a different problem, where the steered courses cannot be placed completely parallel to one another. This is also true for CS composites if the geometry of the part is complex. One approach to remedy this is by varying the individual tow width, which can only be performed in discrete steps [101].

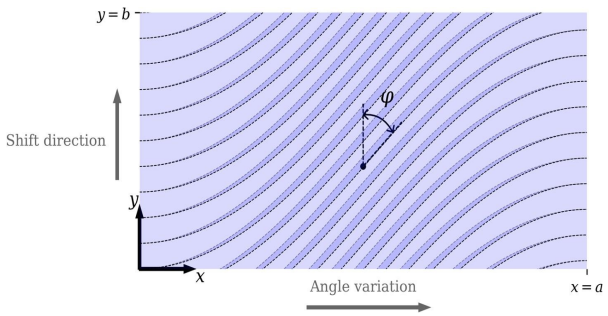


Figure 2.9: The selection of the shift direction with respect to the angle variation for a VS plate [102].

This method then leads to certain regions having overlaps or gaps, as presented in Figure 2.10. Both of them are considered defects and they are undesirable in the structure. The overlap of the tows leads to increased local thickness, while gaps are typically filled with resin at the time of curing, resulting in lower elastic properties.

There are different strategies that take these defects into account by utilising the option of cutting and restarting the tows. One strategy, known as the tow-drop method, is to not allow any overlaps to occur, by terminating the tows immediately as they intersect the boundary of another course. However, this would create wedge-shaped gap areas in-between, with engineering properties similar to the composite's matrix. Another strategy is to only allow overlaps by cutting the tows once both vertices of the tow have crossed the boundary. This will result in the unwanted effects of a bumpy surface and increased weight. Conventionally, a coverage metric is used to describe the manufacturing strategy, as described by B. Tatting and Z. Gürdal [104]. For instance, a 0% coverage does not allow overlaps, while a 100% coverage strategy does not allow for any gaps to exist. A compromise between the two is also possible, as presented in Figure 2.11.

The amount of gaps and overlaps will be determined by how much the courses are shifted to cover the whole structural geometry. By keeping the course width w_{tow} constant, the effective width w_e of the course will vary along the shift direction, as can be seen in Figure 2.12b. The effective course width was approximated by a simple expression by Blom *et al.* [105], and is defined as:

$$w_e \approx \frac{w_{tow}}{\sin(\phi)} \tag{2.35}$$

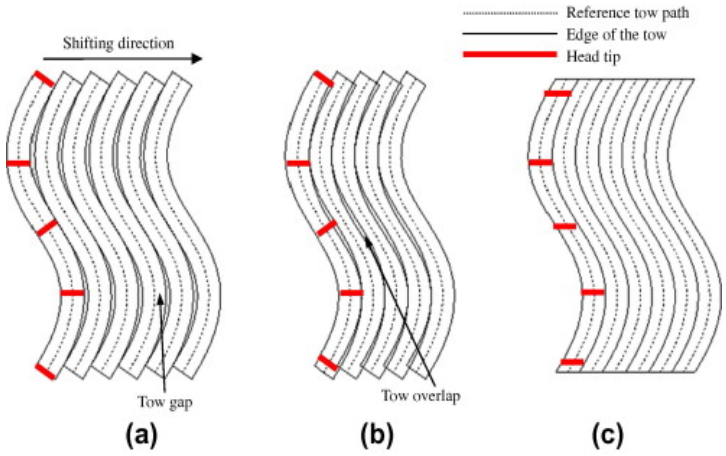


Figure 2.10: A depiction of the possible gaps/overlaps of the placed courses via the shift method [103].

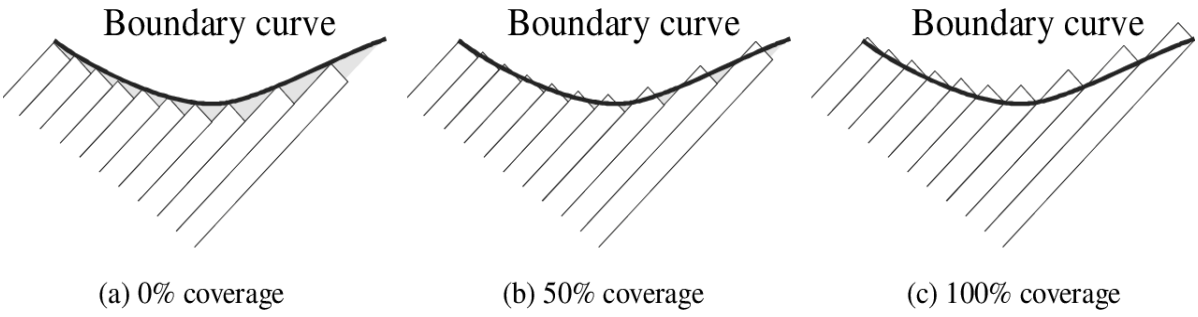


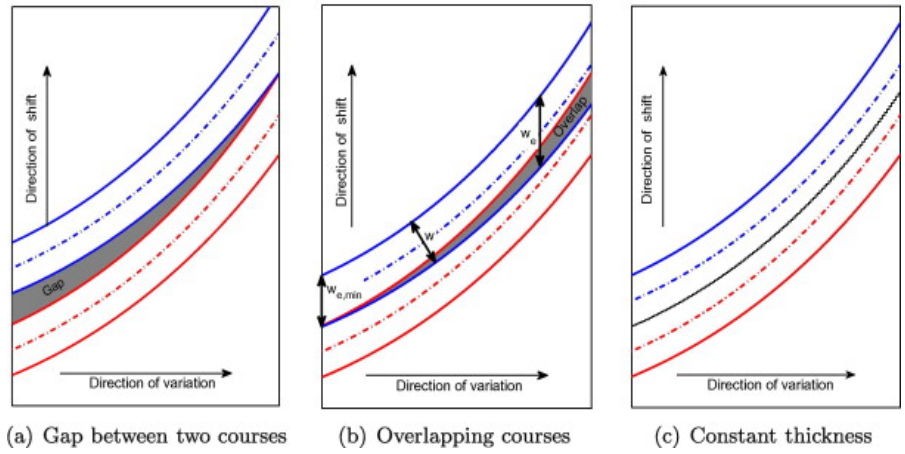
Figure 2.11: The coverage metric of the shifting method, depicted with three different strategies [104].

where ϕ is the variable fibre angle along a direction. For instance, the gaps (Figure 2.12a) can be eliminated from the structure by selecting the appropriate shift value. On the other hand, if the minimisation of overlaps is desirable, then the shift value can be chosen to be the minimum effective course width. In Figure 2.12c, the defects are eliminated altogether by having the machine head cut the tows and then restart the tow feed [105].

A technique that can help spread out these defects, is called staggering. By spreading out the defects, it is ensured that a specific region will not be negatively affected too much by a build-up of defects. Ply staggering moves the plies along a chosen direction as presented in Figure 2.13, and while the defects are still present in the plane of the ply, there is no build-up along the thickness of the laminate.

Gaps and overlaps are the most common defect types and their presence results in additional defects, such as stress concentration regions and fibre angle deviation and waviness at the transition zone between the defect and the pristine laminate. Of course, many defects that are unrelated to gaps and overlaps can also occur with the AFP process. The works of G. Lozano *et al.* [93] and F. Heinecke and C. Willberg [101] provide reviews of the various defects that might take place and the subsequent manufacturing limitations that they impose, although they are typically not taken into account in design as much as the gap/overlap defects [107]. Some of these possible defects are:

- Fibre angle deviation at the course boundaries with respect to the reference curve.
- Tow misalignment due to inaccuracy of the machine head.
- Machine head collisions with complex geometries.
- Fibre separation on concave surfaces if the tow does not adhere well enough.
- The tows might get twisted as the AFP machine head lays them down, resulting in gaps.
- Contamination of the laminate by the fuzz that is created through the friction between the tow edges and the machinery that comes in contact with the edges.



(a) Gap between two courses (b) Overlapping courses (c) Constant thickness

Figure 2.12: An example of the gap/overlap defects that can occur via the shift method [105].

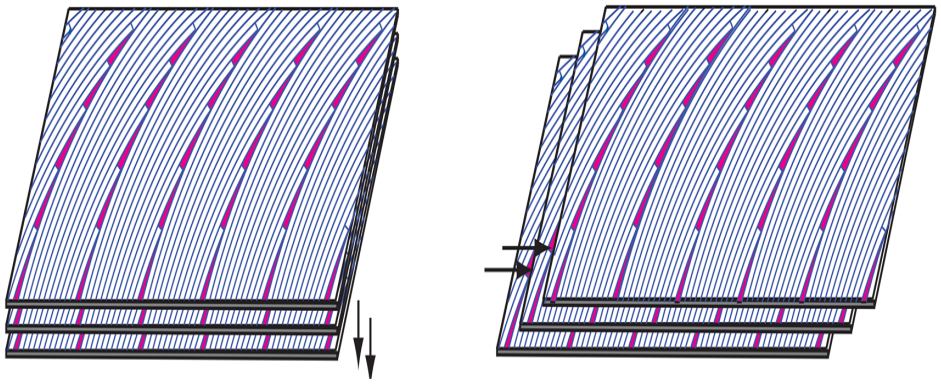


Figure 2.13: A presentation of the ply staggering technique that spreads out the defects within the laminate [106].

Continuous Tow Shearing method

The CTS manufacturing method seeks to improve the robustness of the AFP method by eliminating the gaps and overlaps that are inherent in the latter approach. In this technique, tows are sheared in their plane to vary their angle, instead of bending them as is done with AFP [108]. This is made possible by utilising shear compliant semi-impregnated tows instead of stiffer pre-impregnated slit tapes that are commonly used in the AFP process [103]. This tow shearing results in tow paths that align exactly with the reference curve, avoiding the formulation of gaps or overlaps. Another advantage that is reported by Kim *et al.* [109] is that this method allows for an increased course curvature (30 mm), further relaxing the manufacturing curvature constraint.

The components of a CTS machine head are shown in Figure 2.14. At the head's tip, the compaction shoe and the pinch device make up the tow guide roller and the gripping shoe [103]. The semi-impregnated tows are driven through a small gap and laid onto the mould surface, where tows are pressured by the gripping shoe. One end of the tow is restricted by the gripping shoe and the other end moves along the shifting direction, allowing for its in-plane shear deformation.

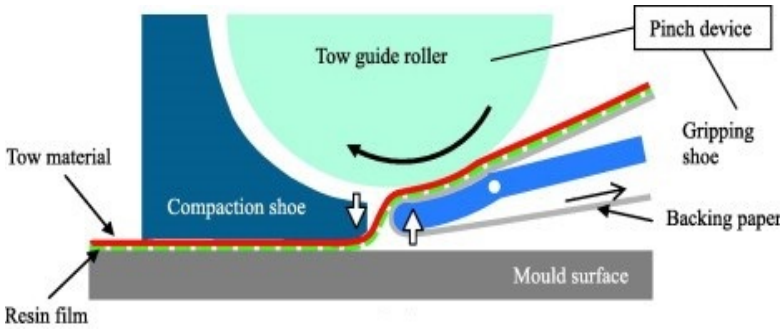


Figure 2.14: The machine head of the CTS method, placing tows onto a mould surface [103].

The fact that no gaps or overlaps occur with the CTS approach certainly makes it a great alternative to AFP, but the CTS method is associated with a different type of defect which is inherent to the process. When the tow is sheared with an angle θ, the initial tow width w_0 is reduced perpendicular to the tow direction, resulting in a local thickness increase due to the fibres being loaded on top of each other, as shown in Figure 2.15.

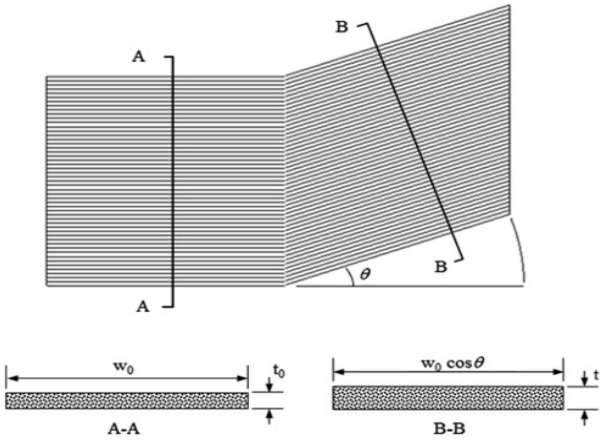


Figure 2.15: Thickness increase of a laminate due to tow shearing [108].

Since the volume of the tow is preserved, the increased thickness t can be calculated simply by considering the shearing angle θ and the initial thickness t_0:

This process can be controlled to yield variable thickness structures, which can be optimised accordingly. However, an additional consideration is that a thickness variation is sometimes undesirable, as in aerospace applications where a smooth outer surface is often necessary. This is also true for the AFP method if an overlap technique is employed, since tow overlap build-up can lead to unsymmetrical structures with the

outer surface being bumpy and a smooth inner mould surface [104]. Thus, a trade-off should be performed to determine which process is more appropriate.

2.3.3. Modelling techniques for VS structures

In order to categorise the various existing modelling techniques for VS structures, the concise classification by S. Ijsselmuiden [110] will be used. The modelling techniques then boil down to three distinct categories:

1. Functional fibre path representation
2. Direct stiffness modelling
3. Discrete stiffness representation

The different modelling methods will be introduced alongside certain historical papers that marked the beginning or the foundation for said modelling technique.

Functional fibre path representation

A significant amount of work has been devoted to explicitly modelling the tow trajectory on a surface. In that way, it is ensured that the structural continuity will not be disrupted and that the design can indeed be manufactured subject to the constraints that were imposed. Once the path is generated for each ply, they can be individually constructed either by parallel placement or with the shift method. The number of design variables can be reduced with this technique, although the degree of the reduction depends on the equation that is used to formulate the tow trajectory.

In 1993, Z. Gürdal and R. Olmedo [1] introduced the concept of the functional fibre path for VS structures by linearly varying the fibre angle of a plate, and reporting a significant increase of the buckling load with respect to CS plates. In another collaboration of the first author in 2002, B. Tatting and Z. Gürdal [98] introduced a popular notation for linearly varying tows, which was adopted and still used in the field. The equation that allows the fibre path to vary between two points in a Cartesian coordinate system x-y is defined by

$$\theta\left(x'\right) = \phi + T_0 + \frac{\left(T_1 - T_0\right)\left|x'\right|}{d} \tag{2.36}$$

where ϕ is the angle of rotation of the fibre path, T_0 and T_1 are the fibre angles defined at the starting and finish points of the trajectory respectively, and d is the distance between these points. A representation of a trajectory generated by this formulation is given in Figure 2.16.

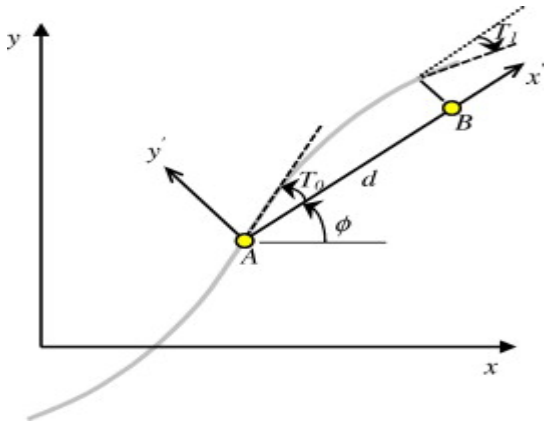


Figure 2.16: A depiction of a linearly varying tow from point A to point B [111].

The notation that was proposed for such a linear variation has the compact form of $\phi < T_0|T_1 >$. The linear variation of the tow trajectory is still frequently used, but other non-linear formulations of the functional have achieved greater improvements by further expanding the design space.

The field of functional fibre representation has been quite active with exploring different mathematical formulations that can define the path trajectory. In 1995, a Non-Uniform Rational B-spline (NURBS) formulation was proposed by Nagendra *et al.* [112], where a discrete set of control points (CPs) is selected such that

the path can vary non-linearly between them. These curves are commonly found in graphical software due to their great flexibility. NURBS curves require the definition of their order, weighted CPs and a knot vector. The knot vector essentially determines what the effect of the CPs will have on the curve. While this allows for a great amount of freedom for the curve, the selection of the parameters needed to construct a NUBRS curve is not a straight-forward task.

Other less flexible curves have also been used with success ever since, in which superior VS designs were found with respect to CS designs. This may signify that the manufacturing constraints that are often imposed, restrict the design space to the degree that the difference between simpler and more sophisticated trajectory functionals is not pronounced as much.

Some of the polynomials that have been employed in the literature so far are Legendre polynomials [113], Splines [114], and Lagrange polynomials [102, 115, 116]. More specifically, the Lagrange polynomial as proposed by Z. Wu *et al.* [116] seems to be a popular choice for the trajectory generation due to the allowed curve flexibility with only a small number of CPs. The trajectory can be varied in two perpendicular directions by pre-selecting a finite amount of CPs to form a $M \times N$ grid, as shown in Figure 2.17. The polynomial permits a non-linear variation by means of a double series function of $(M-1)^{th}$ by $(N-1)^{th}$ order polynomials. The generated angle variation is then defined by polynomial coefficients at the CPs, where the values in-between are interpolated. The authors mention that the CPs need not be uniformly spaced, however a good selection of their spatial coordinates can accelerate convergence in terms of optimisation.

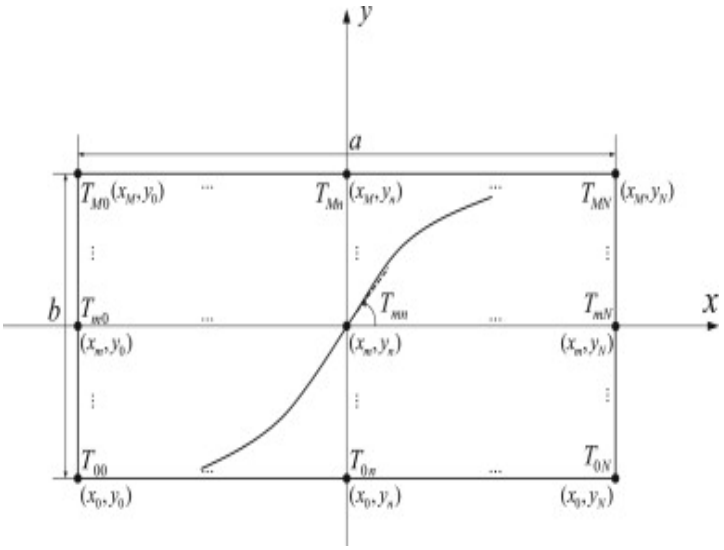


Figure 2.17: Non-linear variation of the fibre trajectory based on a Lagrange polynomial [116].

The equation formula requires the number of CPs M and N, the angle coefficients at these points T_{mn}, and the coordinates $x-y$ along which the variation θ is allowed to vary.

$$\theta(x,y) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} T_{mn} \prod_{m \neq i} \frac{x - x_i}{x_m - x_i} \prod_{n \neq j} \frac{y - y_j}{y_n - y_j} \tag{2.37}$$

Alternatively, the fibre variation can be along one direction instead. Although this limits the available design space, manufacturing via the shifting method is only viable through this way since the shifting occurs perpendicular to the axis of variation. Equation 2.37 is then simplified as follows:

$$\theta(x) = \sum_{m=0}^{M-1} T_{mF} \prod_{m \neq i} \frac{x - x_i}{x_m - x_i} \tag{2.38}$$

where M is the number of the CPs, T_m are the selected angles at these CPs and x represents the coordinates that follow the axis along which the fibre angle varies.

A different approach to generating designs would be to directly consider the curvature of the trajectory. Thus, in order to ensure that the curvature constraint is enforced, Blom *et al.* [97] proposed a constant curvature trajectory defined as,

$$\sin(\phi) = \sin(T_0) + \kappa |x| \tag{2.39}$$

where ϕ is the variable fibre angle, T_0 is the initial fibre angle, κ is the constant curvature and x is the spatial coordinate. This formula simplifies the design process by taking care of the most important manufacturing constraint, although the feasible design space is limited considerably.

In their work, Lopes *et al.* [117] considered the response of the strength and buckling load of a VS panel with a cut-out. More specifically, they applied a set of First Ply Failure (FPF) criteria to check whether any of the laminate's plies have failed in pre-buckling or post-buckling. The fibres were allowed to vary linearly as per Equation 2.36. They performed FE calculations discretely for every T_0 and T_1 in the range of $0-90^\circ$ with a step of 5°. Every element was assigned its own unique layup based at the centroid of the element. They report improvements over conventional CS laminates, by directing the load away from the plate's centre, towards the edges which are supported.

The first work that considered VS cylinders was that of B. Tatting [118] in 1998. The author sought to improve the buckling load by linearly steering the fibre along the cylinder's axis of rotation, when it is subjected to axial compression, pressure and torsion. Only a marginal improvement was reported over the traditional CS counterpart however. An additional load case of pure bending was examined, with the fibre varying circumferentially that time. Substantial improvements were found for this case, with the author concluding that circumferential stiffness tailoring is more suitable for the cases where the load also varies along the circumference (e.g. bending and shear).

A buckling comparison between axial and circumferential linear fibre variation of a cylinder was performed by S. Güldü and A. Kayran [119]. In their study, they applied different load cases, namely compressive axial, bending and torsional loading. In contrast to B. Tatting, they report that the cylinder with axial fibre variation performed better in all considered load cases. Their suspicion was that a linear variation between the two ends of the cylinder did not produce enough design variables to yield a satisfying solution. To remedy this, they sectioned the cylinder in different axial strips, where the fibres could linearly vary independently within these sections. In that way, they were able to better capture the capabilities of circumferential steering and improve the initial results.

A numerical and experimental comparison between a CS and VS cylinder was performed by E. Labans and C. Bisagni [2]. The authors conducted a buckling and free-vibrations analysis to determine whether VS cylinders can provide superior results and what type of numerical model can accurately capture their structural response. Experimentally, the cylinders were subjected to uni-axial compressive loading until the point of buckling and free-vibration experiments were conducted subsequently. The inherent manufacturing imperfections of the cylinders were measured and used as input in the numerical model. The stiffness variation was handled by sectioning the cylinder longitudinally and assigning a constant layup to each section. The buckling test showed that the CS cylinder buckled at a higher load compared to the VS counter-part, but at a lower shortening value due to compression. However, the authors note that the VS cylinder was quite insensitive to the initial geometric imperfections in contrast to the CS cylinder, owing to the variation in stiffness that may attract load away from the areas suffering from imperfections. The vibration test yielded similar results for the two cylinders. Finally it was reported that the rather simplified numerical model was able to capture the structural response quite well, with larger errors up to 15% being reported only for the higher mode shapes for both cylinders and analyses.

M. Rouhi *et al.* [120] performed a comparison between a Quasi-Isotropic (QI) CS cylinder and a VS cylinder. A bending loading case was considered and the fibre steering varied circumferentially in a linear manner. The cylinder was divided into equally spaced axial strips and each strip was assigned a fibre orientation. In order to increase computational efficiency, the fibre orientations are initially assigned to a small amount of points that span a large section of the cylinder each and the smaller strips between these points acquired a value by the linear interpolation of these points. Although a significant increase in buckling load was found over the base CS cylinder, the reported increase from experimental testing was less than the numerically predicted value. However, they state that this can be attributed to the tow waviness that was exhibited due to fibre steering.

Another pure bending-induced loading case of a VS cylinder was considered by Blom *et al.* [105]. The goal of the work was to tailor the stiffness properties of the cylinder, while evaluating the buckling and strength response of the structure. A constant fibre curvature was utilised to circumvent the manufacturing constraint and increase computational efficiency. Once more, a circumferential variation was preferred, with the cylinder sectioned along its length, where each section had an individual fibre orientation. The comparison with a CS cylinder found that a 17% improvement in buckling is possible, granted it is to occur before material failure, which was an additional constraint.

Direct stiffness modelling

In direct stiffness modelling, the design variables are the components of the stiffness matrix of the engineering theory that is employed to model the laminate (Classical Lamination Theory (CLT), First Shear Deformation Theory (FSDT), etc.). The process involves additional complexity and restrictions for physically realistic designs, but the payoff is the reduced number of design variables, since the number of plies is disassociated with the model.

The approach that is most popular is the direct stiffness parametrisation via the *lamination parameters*. These parameters allow for the direct tweaking of the stiffness matrix components to find the optimal design [121]. Once the matrix components are determined, further post-processing is required to determine what are the actual fibre paths which lead to such parameters. In total, the design parameters are reduced down to a value of 12 if the popular CLT is applied in the model. In addition, J. Grenestedt [122] proved that a reformulation of a design problem in terms of the lamination parameters, yields a feasible domain with a closed convex surface, where efficient gradient-based optimisation algorithms can be utilised to converge to the global minimum.

For the sake of brevity, the lamination parameters will be introduced without the derivation of CLT. The interested reader is directed to one of the many books that introduce the subject in detail [123, 124]. The stiffness of a plate is dictated by the so called $\boldsymbol{ABD}$ matrix in CLT, where $\boldsymbol{A}$, $\boldsymbol{D}$, and $\boldsymbol{B}$ are sub-matrices that govern the in-plane, out-of-plane and coupling behaviour between the two, respectively. The theory allows for the existence of invariant parameters U_i that depend on the material and not the fibre orientation of the ply. By using lamination parameters, the $\boldsymbol{ABD}$ matrix can be written as follows [125]:

$$\begin{aligned}
\boldsymbol{A} &= h\,(\Gamma_0 + \Gamma_1 V_{1\boldsymbol{A}} + \Gamma_2 V_{2\boldsymbol{A}} + \Gamma_3 V_{3\boldsymbol{A}} + \Gamma_4 V_{4\boldsymbol{A}}) \\
\boldsymbol{B} &= \frac{h^2}{4}\,(\Gamma_0 + \Gamma_1 V_{1\boldsymbol{B}} + \Gamma_2 V_{2\boldsymbol{B}} + \Gamma_3 V_{3\boldsymbol{B}} + \Gamma_4 V_{4\boldsymbol{B}}) \\
\boldsymbol{D} &= \frac{h^3}{12}\,(\Gamma_0 + \Gamma_1 V_{1\boldsymbol{D}} + \Gamma_2 V_{2\boldsymbol{D}} + \Gamma_3 V_{3\boldsymbol{D}} + \Gamma_4 V_{4\boldsymbol{D}})
\end{aligned} \tag{2.40}$$

The laminate thickness is given by h, the matrices Γ_i contain the invariants U_i and the lamination parameters for each sub-matrix are given by $V_{i\boldsymbol{A}}$, $V_{i\boldsymbol{B}}$, $V_{i\boldsymbol{D}}$. The invariants are contained within the matrix Γ_i, as such:

$$\begin{aligned}
\boldsymbol{\Gamma}_0 &= \begin{bmatrix} U_1 & U_4 & 0 \\ U_4 & U_1 & 0 \\ 0 & 0 & U_5 \end{bmatrix}, \boldsymbol{\Gamma}_1 = \begin{bmatrix} U_2 & 0 & 0 \\ 0 & -U_2 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \boldsymbol{\Gamma}_2 = \frac{1}{2}\begin{bmatrix} 0 & 0 & U_2 \\ 0 & 0 & U_2 \\ U_2 & U_2 & 0 \end{bmatrix}, \\
\boldsymbol{\Gamma}_3 &= \begin{bmatrix} U_3 & -U_3 & 0 \\ -U_3 & U_3 & 0 \\ 0 & 0 & -U_3 \end{bmatrix}, \boldsymbol{\Gamma}_4 = \begin{bmatrix} 0 & 0 & U_3 \\ 0 & 0 & -U_3 \\ U_3 & -U_3 & 0 \end{bmatrix}
\end{aligned} \tag{2.41}$$

The lamination parameters are obtained by the integration of the fibre orientation of each ply through the thickness

$$\begin{aligned}
(V_{1\boldsymbol{A}}, V_{2\boldsymbol{A}}, V_{3\boldsymbol{A}}, V_{4\boldsymbol{A}}) &= \int_{-0.5}^{0.5} (\cos(2\theta), \sin(2\theta), \cos(4\theta), \sin(4\theta))\,\mathrm{d}\bar{z} \\
(V_{1\boldsymbol{B}}, V_{2\boldsymbol{B}}, V_{3\boldsymbol{B}}, V_{4\boldsymbol{B}}) &= 4\int_{-0.5}^{0.5} \bar{z}\,(\cos(2\theta), \sin(2\theta), \cos(4\theta), \sin(4\theta))\,\mathrm{d}\bar{z} \\
(V_{1\boldsymbol{D}}, V_{2\boldsymbol{D}}, V_{3\boldsymbol{D}}, V_{4\boldsymbol{D}}) &= 12\int_{-0.5}^{0.5} \bar{z}^2\,(\cos(2\theta), \sin(2\theta), \cos(4\theta), \sin(4\theta))\,\mathrm{d}\bar{z}
\end{aligned} \tag{2.42}$$

where $\bar{z}$ is a normalised through the thickness variable. The final important step is to solve the inverse problem of finding the laminate stacking sequence that corresponds to these variables after the optimisation of the lamination parameters.

The first authors that utilised the method to optimise a laminate were Hammer *et al.* [126] in 1997. They used the CLT as their basis and formulated the problem in the case where orthotropic laminates are considered. Ever since, many authors have chosen this modelling approach on the grounds of computational efficiency and robustness [121, 127, 128].

Discrete stiffness representation

This modelling representation was the first one to be used for VS structure modelling, and is perhaps the simplest. This is owed to the straight-forward way that the variable angle orientations are handled. The structure is discretised in many finitely small regions, so that each region will be assigned its own stiffness. The stiffness of each region is a function of the fibre angle of each ply at that location, as depicted in Figure 2.18. Naturally, this technique typically employs the FEM to perform this discretisation and arrive at a solution. Care must be taken to ensure that manufacturing constraints (e.g. curvature constraint) are not violated due to the discretisation.

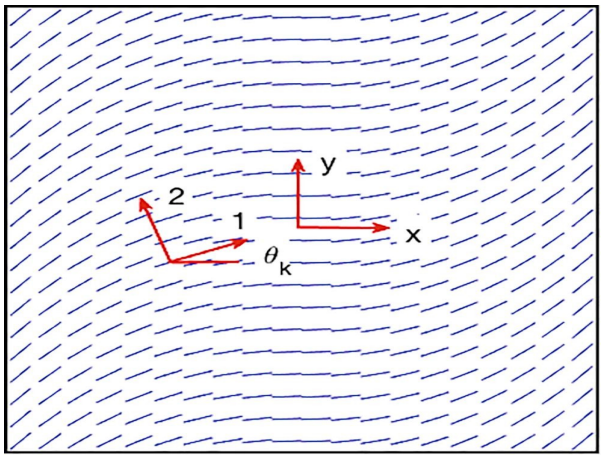


Figure 2.18: A VS plate that has discrete variable fibre angles in its plane [129].

In 1991, M. Hyer and R. Charette [3] discretised a plate containing a hole with a small amount of elements, in hopes of optimising the fibre path around the hole via tow steering. They utilised symmetry and only modelled one quarter of the plate, as shown in Figure 2.19.

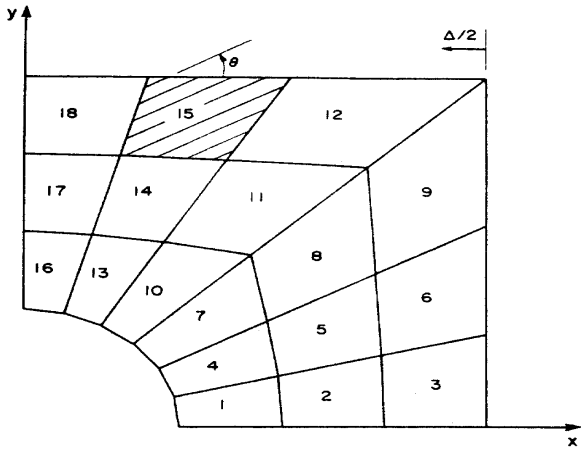


Figure 2.19: The discretized quarter of a plate with a central hole from the work of M. Hyer and R. Charette [3].

Each element was assigned a unique stiffness based on a laminate of the type $[\pm 45/\theta_6]_S$. Thus, only one variable fibre angle θ was tied to each element, amounting to only 18 design variables. Nevertheless, they reported significant increases in both buckling load and strength of the plate, piquing the interest in the field of VS structures.

2.3.4. Models that account for manufacturing defects

The majority of the models so far only take the curvature constraint into account. While, there are many defects that can occur in the manufacturing process, the features that are primarily considered are the gaps or overlaps and the width of the tows [93, 107]. Some of the other possible defects have a more stochastic

nature and may also be difficult to incorporate into a model. The first author that examined the effect of tow width and overlaps in VS plates, was P. Langley [130]. In his work, the plate was discretised and the elements were assigned a fibre angle after calculating the distance from the element centroid to the tow reference path, using the Golden Section search algorithm. A check was performed whether that distance was less than half of the tow width in order to determine if an overlap is present and then increase the element's thickness.

The effect of gaps resulting from tow drops, on the strength stiffness of a VS laminate was studied by Blom *et al.* [97], using the FEM. The tow width and the number of plies were both variables in this work. They reported a reduction of the laminate's strength with increasing tow width, since this leads to an increase in gap area. The authors also successfully utilised ply staggering to minimise the collocation of defects through the thickness. The gap areas were considered to be filled with resin and their mechanical properties were assumed to be these of the matrix. This necessitates the use of a fine mesh such that these small gaps can be appropriately captured.

In their work, K. Fayazbakhsh *et al.* [131] introduce a method to model gaps/overlaps that does not require such a fine mesh, yielding increased computational efficiency. They proposed a *Defect Layer* method which is able to calculate the gap/overlap area percentage of individual elements within the discretised mesh. In other words, the gap/overlap of each ply occupies a specific area of each element, which is taken into account, and the results from each ply are averaged to produce the properties of the element. The authors considered two cases to showcase their method, namely a 0% coverage and a 100% coverage strategy. It is reminded that these strategies correspond to no overlaps and no gaps respectively. If only gaps are present, the elastic properties of the element are reduced according to the percentage of gaps within the element, as shown in Figure 2.20. For the case of overlaps, an increase in thickness that is proportional to overlap area is performed.

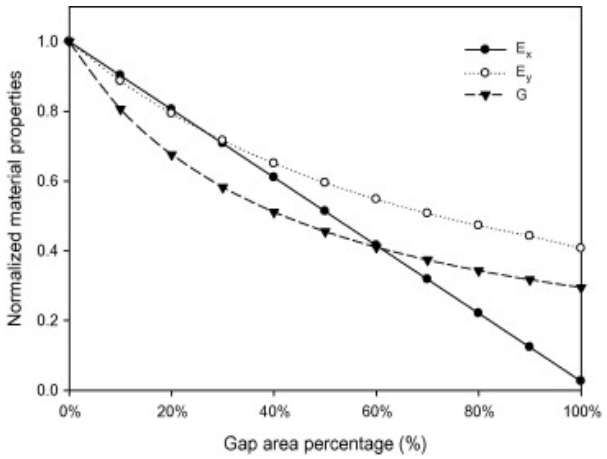


Figure 2.20: The reduction in normalized elastic properties of the element with respect to the gap area percentage, as considered by the authors of the Defect Layer method [131].

In the same work, a comparison was made between a defect-free model and the two aforementioned coverage strategies. They report a decrease in laminate strength by following a gap strategy with respect to a defect-free model, while an overlap strategy leads to better strength and buckling load. Additionally, it was found that gaps have a higher negative effect on the in-plane stiffness and that overlaps have a higher positive effect on the buckling load, although the additional mass due to the overlaps was not accounted for.

The Defect Layer method has been adopted by other works ever since, where the numerical modelling of VS structures is involved, due to its accuracy and computational efficiency [132–134]. A later work by the same group of authors [135] investigated the effect of the number of tows and the tow width with respect to the gap/overlap area. They concluded that an increase of the number of tows contained within a single course reduces the area affected by defects, but it also reduces the number of designs that can be manufactured since the curvature constraint must be satisfied over the whole course. Additionally, they found that an increase in tow width has the opposite effect and the defects cover a larger area.

V. Mishra *et al.* [136], proposed a different homogenisation approach to further improve the computational efficiency of calculating gaps in the model. In their approach the tow width is not explicitly modelled, and the amount of gaps is measured by analytically calculating the gradient of the distribution of the tow angle within a ply. The authors call this approach, the *Smearing method.* However before applying the method, a correlation between the defects and the ply properties is performed through a homogenisation process.

The properties of every element can then be determined. A comparison between the Defect Layer method and the Smearing method shows that the latter approach leads to a significant reduction in computational time by an order of magnitude. However, the drawbacks of this method are that it is more intricate, and only serves as an approximation. The authors report stiffness results which are predicted within 5% accuracy on the conservative side, while similar accuracy is observed for the prediction of buckling with the results being non-conservative this time.

2.3.5. Optimisation of VS structures

All of the modelling techniques mentioned thus far, have been used with success to generate VS designs. They all possess their benefits and shortcomings but they can all generate viable structures that can be optimised. In contrast to CS structures, VS designs have a higher number of design variables, which renders the optimisation problem more computationally expensive.

This higher order dimensionality usually prohibits the use of gradient-based optimisation algorithms. The time that is required to compute the gradients increases exponentially with an increasing number of design variables, thus a constriction of the variables is often necessary for the sake of efficiency. This narrows the design space and may lead to a sub-optimal solution.

The high number of design variables also typically leads to many local minima within the design space. As with gradient-based methods, deterministic direct search approaches may easily get stuck in a local minimum. Furthermore, the deterministic direct search methods are computationally inefficient when a large number of design variables is involved. For these reasons, they are rather unpopular in the optimisation of VS structures, as reported by H. Ghiasi *et al.* [137], in their review paper of the optimisation algorithms utilised for VS structures.

The same paper states that population based methods find wider acceptance, since it is possible to approximate the global minimum, albeit in a less computationally efficient manner [137]. X. Legrand *et al.* [138] considered that each element can be assigned each own fibre angle as a design variable. They used a GA to optimise the fibre variation in order to increase the strength of bolted joints and bolted lugs. Although they were successful in doing so, only laminates with a single ply were considered.

An increase in the buckling load of a laminate containing a central hole, was sought in the work of J. Huang and R. Haftka [139]. This time, the laminate consisted of many plies, where the variation of the fibres around the hole was represented by a piece-wise bilinear interpolation function for each ply. The far-field distribution was fixed at an angle of 0°. A combination of a GA with a conjugate gradient method was used to improve efficiency and pinpoint the global minimum. The goal of the GA was to by-pass the local minima, and once the valley containing the global minimum is located, the conjugate gradient method is employed. Even with the design variables being active only around the area of the hole, they reported a significant increase in the buckling load of the laminate.

Similarly, a combination of a gradient-based method with an EA was adopted by D. Peeters *et al.* [140], in order to increase the buckling load of a laminate. In their work, they employed both a fibre variation and a variable thickness design. Initially, a gradient-based optimisation is performed based on the lamination parameters technique, where the thickness and stiffness matrix are obtained. The mapping back to continuous fibre angle distribution is then performed by an EA that is intertwined with a gradient-based algorithm. The addition of thickness variation to the fibre steering, resulted in a weight decrease of 15% without reducing the buckling load.

A multi-step approach was also used by M. Rouhi *et al.* [141], where they optimised a VS cylinder to increase its buckling capabilities in a bending loading case. In order to increase the computational efficiency of the optimisation, they employed a surrogate model to replace the expensive FEM simulations. More specifically, they performed a comparison between different basis functions for the RBF surrogate model, to ascertain which function yields the best results. The best basis function was then chosen to act as the objective function, after being trained from sample points that were generated via the LHS method. A GA was then utilised to optimise the discretised fibre angles, where convergence was checked for. If the GA did not converge, then the bounds of the LHS tightened around the current optimum found by the GA, and the retraining of the RBF model would commence until convergence. The authors then conclude that a buckling load improvement of up to 25% is possible against a CS design.

A comparative study on the topic of surrogate models for the optimisation of VS structures was conducted by M. Nik *et al.* [142]. Four different surrogate models were compared, namely Polynomial Regression, RBF, Kriging, and Support Vector Regression. They based the comparison on different metrics, such as the coefficient of determination (R^2), relative average absolute error, and relative maximum absolute error. For

benchmarking purposes, they optimised both a VS laminated plate and a VS cylinder. The angle variation of the structures was then optimised by a GA. They reported that Polynomial Regression performed worst for all considered metrics, while Kriging and RBF performed the best and were the most robust.

The buckling load of a VS cylinder was optimised with the use of a particle swarm optimisation algorithm in the work of S. Pitton *et al.* [36]. A novel periodic reference path definition was proposed to decrease the design variables without heavily restricting the design space. A neural network acted as a surrogate model to replace the FEM analyses, by sampling points via the LHS method. The dataset generation was performed multiple times in order to avoid any bias introduced by the LHS approach. A buckling load improvement of 4% was found, although the initially imposed curvature constraint needed to be relaxed to achieve this value.

The particle swarm optimisation algorithm was also preferred by S. Güldü and A. Kayran [119] in their work of optimising a VS cylinder to obtain a higher buckling load. The effect of the increased thickness due to the tows overlapping was also taken into account in their model. The coordinates of each element centroid were determined and a calculation is performed, whether the shifted tow distance from the element centroid is less than half of the tow width for any shifted path. If yes, then the element is assigned that thickness value. A comparison between axially and circumferentially stiffness varying cylinders was performed under various loading conditions. The fibre path varied linearly between two points on the cylinder, in all cases. The comparison showed that the axial stiffness variation provided the best results in all the considered loading cases, with the circumferential variation even performing similar to the benchmark CS cylinder in one of them. The authors suspected that the lack of additional points (only two were initially considered) for the linear variation, led to such poor performance. In a subsequent optimisation analysis, they concluded that an increase in intermediate points, where the fibre path can linearly vary between each of them, leads to a substantial performance increase for the circumferential fibre variation.

In their aforementioned work, M. Nik *et al.* [135] obtained the Pareto front of a VS plate for the competing designs of maximum stiffness and maximum buckling load, using the NSGA II algorithm. They also compared the Pareto fronts of plates with a complete gap strategy (0% coverage) and a complete overlap strategy (100% coverage). The authors reported that a complete gap strategy moves the defect-free Pareto front towards lower values for both stiffness and buckling load, while the opposite is true for the case of the complete overlap strategy. They conclude that the difference between the two fronts is maximised for laminate designs which result in maximum buckling load improvement, and it is minimised for designs that offer the highest stiffness. This is owed to the low angle variation when the stiffness is maximised, since only a small amount of defects is produced in that case, in contrast to the higher angle variation seen in optimal buckling designs.

A different type of multi-objective optimisation was performed by T. Guimarães *et al.* [143], in which the buckling load and flutter speed of a VS laminated plate were optimised in conjunction. The authors utilised a DE algorithm that was able to handle multiple objectives, where the optimisation was performed multiple times to address the stochastic nature of the algorithm. The fibre path was modelled by means of Lagrange polynomials which necessitate the use of CPs within the coordinate system of the plate. Different optimisation analyses were conducted where the number of CPs, or in other words the order of the polynomial, were varied. These analyses were performed in order to ascertain to what degree the optimisation results can change with the increasing DoF of the fibre path. The authors found that significant improvements were possible with the continuous increase of the degree of the polynomial. Since flutter and buckling are competing with each other, a decision needs to be made as to what is the optimal design, as chosen from the Pareto front.

In one of their papers in a series of VS cylinder optimisations, M. Rouhi *et al.* [144] performed a multi-objective optimisation on a cylinder subjected to bending. The optimal configuration for a VS bending-induced cylinder is to increase the in-plane stiffness of the tensile side in order to attract load away from the compressive region, which leads to increased buckling loads. However, if the sign of the applied moment is reversed, the previously optimised configuration becomes sub-optimal. Therefore, the goal of this work was to increase the buckling load by considering both sides of the cylinder in optimisation as two different objective functions. As with their previously mentioned work, the authors used a RBF surrogate model and a GA within an iterative scheme, where the bounds of the sampled points were tightened as they trained the surrogate model until convergence was achieved. When compared against a CS cylinder, an improvement of 25% was reported for the VS cylinder.

2.4. Concluding remarks

The main goal of the chapter was to identify and present the literature that is relevant to the optimisation of a variable stiffness cylinder, while also acquainting the reader with the necessary technical background

regarding surrogate modelling. To that end, this chapter was primarily concerned with the presentation of three different topics:

1. **Artificial Neural Network**. The concept of neural networks for engineering purposes was given. The focus was on feed-forward multi-layer perceptron networks and their capabilities to increase computational efficiency through approximated predictions. Advanced topics concerning their training, performance, and optimisation were also included in the discussion.

2. **Population-based optimisation algorithms**. Popular algorithms that considered a single or multiple objectives were introduced, along with the metrics to evaluate them. Various approaches in handling the optimisation constraints were also examined.

3. **Variable stiffness composite structures**. The relatively novel field of variable stiffness composite structures was presented. The methods of their manufacture and numerical modelling were discussed, along with the inherent constraints and defects. The literature work on the performance boost over traditional structures was studied, and the means of their optimisation and related computational efficiency were also considered.

The knowledge gained from this literature study better illuminates the methods and techniques that have already been employed, as well as their advantages and drawbacks exhibited in their results. Therefore, this information serves as a stepping stone and as a guide for the work conducted in this thesis.

As shown, there are many methods that can be exploited to optimise the variable stiffness cylinder, with some methods being more successful than others. In the vast majority of cases, a single objective was considered, usually being the buckling load of the cylinder. Furthermore, improvements in performance were found by allowing for the fibre trajectory to vary linearly within a finite number of sections. The implementation of a curvature constraint was always a consideration in the optimisation problem, although the inclusion of defects in the models was quite rare. The incorporation of the FEM to perform the calculations was highly favoured, with a small amount of endeavours utilising surrogate models to replace the expensive FEM models. Out of all of these works, only a single one considered the application of neural networks for the optimisation problem [36], although the performance boost was not as significant as other related efforts have reported.

The multitude of options that are available to perform a structural optimisation is great and it is frequently the case that not all of the state-of-the-art approaches are being considered. Additionally, as with most optimisation problems, a concurrent optimisation of multiple objectives is often required in practice. Although these objectives may be in the form of a constraint when given the actual problem formulation, it is nevertheless insightful to obtain the Pareto front for various objectives, to get a better grasp on the behaviour of the objectives within the design space. The examples of multi-objective optimisation for variable stiffness structures that were found in literature were scarce, opening up to the further exploration of the interaction between various objectives and the simultaneous improvements that can be obtained from these structures.

Another gap found in the considered literature, is the absence of a comparison study between the performance of neural networks and other surrogate models to take the place of the FEM. While in the work of M. Nik *et al.* [142], a comparison between various surrogate models for variable stiffness structures was performed, neural networks were not included. Considering that neural networks can achieve superior results for very complicated problems, it begs the question whether an increase in the model complexity (e.g. high order polynomials to dictate the fibre path) can better be predicted by a network able to comprehend this complexity.

The lack of methods which are able to model the defects that are inherent to the manufacturing approaches of variable stiffness structures is perhaps attributed to the small number of studies that take them into account. At this point in time, two methods seem to be the most promising, the Defect Layer method [131] and the Smearing method [136]. The advantage of the latter is its computational efficiency at the expense of accuracy and simplicity. So far, no work has been carried out on constructing a surrogate model for the Defect Layer method to reduce its computational expense and to investigate any potential accuracy reductions.

3

Methodology

The methodology that was followed in the undertaking of the current work will be laid out in this section, as well as the reasoning behind the approaches that were taken.

3.1. Preliminaries and chapter layout

A brief summary of the decisions taken on the execution of the thesis are reported in this section, as well as the rationale behind the layout of the chapter.

The main goal of this thesis was to investigate the results of a multi-objective optimisation of a VS cylinder by employing surrogate modelling techniques to reduce the computational time required. To that end, 5 objectives were optimised:

- Buckling load
- Free vibrations frequency
- Axial stiffness
- Lateral stiffness
- Strength in axial compression

These objectives are some of the most frequent design drivers for cylinders in aerospace, and it was thus determined that they are sufficient for the comparison between VS and CS cylinders. Although in practice the mass is generally minimised with various constraints on the structural properties, this study is concerned with the structural capabilities that can be obtained by the different cylinders. Therefore, the aim was to obtain the Pareto front in the objective space, and then to perform a comparison based on the findings.

To perform the optimisation, an optimisation algorithm has to be selected first. Although various algorithms might perform the optimisation successfully, the degree of success and the time required for completion are dependent on the algorithm selection. Since this optimisation problem contains 5 objectives, it is defined as a many-optimisation problem, as discussed in Section 2.2.8. Therefore, the discussed UNSGA III algorithm overcomes the inefficiency of other algorithms when it comes to many objectives by searching the design space more efficiently. For that reason, that the algorithm was chosen to be employed for the optimisation problem.

The primary surrogate model that was used to perform the predictions during the optimisation process was a feed-forward MLP neural network, which is the used most commonly for regression. Other types of neural networks, such as the RNN and CNN types, perform best when the data is of the temporal kind [18, 19, 145]. Therefore, the increased complexity, training and prediction time of these neural networks does not ensure a greater prediction accuracy for the regression tasks that were considered. However, literature suggested that the RBF and Kriging surrogate models find greater usage in the field of aerospace engineering. Thus, in order to avoid a myopic approach, these commonly used surrogate models were also considered and meaningful comparisons were performed to ascertain whether ANNs indeed exhibit superior performance

for the tasks at hand. These comparisons are performed after all the data has already been obtained and they are not part of the optimisation process.

Finding the neural network architecture that yields the optimal accuracy is particularly difficult due to the many hyperparameters that need to be tuned. Therefore, it was decided to use an optimisation algorithm to aid with the tweaking of the hyperparameters. This optimisation problem was defined with a single objective, being the minimisation of the gradient descent metric. The very popular DE algorithm was utilised for this task, owing to its superior performance on many benchmark optimisation problems over its counter-parts.

Multiple VS modelling techniques were discussed in 2.3.3. All of them have certain benefits and drawbacks and they are all viable candidates as to how the VS structure can be modelled. It is re-iterated that this modelling concerns how the variability of the fibre angles is generated and whether additional features such as gaps or overlaps are incorporated in the model. Typically, this is the first step in any VS related work and it is crucial as it lays the foundations for the rest of the project. In this thesis, it was decided that a continuous functional fibre representation should be utilised to generate this angle variation. The geometry would subsequently be discretised in many elements in order to perform a FE analysis. In this way, it is ensured that the model will be accurate and that the imposed curvature manufacturing constraint pertinent to tow steering is met, provided that the discretisation is fine enough.

The manufacturing method of choice is the AFP shifting method, as it is the most popular in literature and finds widespread use in industry, although the CTS method seems to be gaining steam fast. It was also decided that the full coverage strategy (only overlaps) would be investigated, as literature shows that it yields mechanical properties that are superior with respect to the gap method, although most authors do not take into account the additional mass due to the overlaps. Of course, each method has other advantages and disadvantages which were discussed in Section 2.3.4, and generally speaking a good design should weigh them in.

Furthermore, two different manufacturing techniques were identified in Section 2.3.2 regarding the AFP shifting method, namely the axial and circumferential angle variations. Their advantages and disadvantages have already been presented, although it is not yet clear which approach yields superior results for specific structural objectives. For that reason, both of them were investigated in separate optimisations in order to draw further conclusions.

The presented methodology on optimisation is firstly performed on cylinders without cut-outs, hereby mentioned as pristine for the sake of brevity, and then to ones with cut-outs. Initially, a CS cylinder is optimised so that it can act as a benchmark to compare the VS cylinder against. The improvement of the CS cylinder over a QI design is also noted, so that the extent to which the CS cylinder has been optimised can be quantified. Subsequently, different optimisations of VS cylinders following either the axial or the circumferential angle variation are executed, in order to make a comparison between the two angle variations and the CS cylinder. Lastly, the same methodology is applied to a cylinder containing cut-outs to identify the improvements found in that case.

Additionally, certain Python packages were extensively used such as *Tensorflow* [146] for the training of neural networks, *smt* [9] for the construction of the RBF and Kriging models, and *pymoo* [147] for the optimisation process. The results that compare computational times were generated on a 4-core system with Intel® Core® i7-6700HQ CPU @ 2.60GHz, 8 GB RAM, and NVIDIA GeForce GTX 960M GPU @ 1.1 GHz and 4 GB VRAM.

With these preliminaries in mind, the structure of the chapter is now presented:

1. In Section 3.2 the construction of the cylinder that follows a functional representation is given. The steps needed to develop the complete form of the cylinder from an initial curve are explained.

2. Section 3.3 deals with the approach of how the cylinder geometry is discretised to accurately capture the stiffness properties and partition them into elements to be used in FEM analyses.

3. The methodology concerning the surrogate model predictions of the variable stiffness properties is discussed in 3.4, where a suitable approach to reduce the time required to obtain the element properties is offered.

4. The manner in which the optimisation problem is tackled is given in Section 3.5. The subjects concerning the optimisation such as design variables and hyperparameters, are presented in the section.

5. The finite element modelling is discussed in Section 3.6, where the solver, element type and boundary conditions are elaborated upon.

3.2. Construction of the cylinder from a functional curve representation

As it was explained in Section 2.3.2, the AFP machine lays multiple bundles of fibres, which are called tows, at the same time. These tows form what is defined as a course. Therefore, the cylinder can be constructed from multiple courses that are shifted perpendicular to one another until the complete cylinder mould is filled.

When it comes to the modelling aspect, the first step in the development of the structural properties of the cylinder is to generate a curve that defines the angle variation. This curve can be thought of as a single fibre. This reference curve can then be off-set to form the parallel boundaries of a course. In other words, the reference curve lies at the centre of the course and two off-set curves can be generated from it to define the edges of the course. Multiple courses can then be created by a shifting operation with respect to the axial or circumferential directions, in order to fill the space of the cylinder. A block diagram of the aforementioned is given in Figure 3.1.

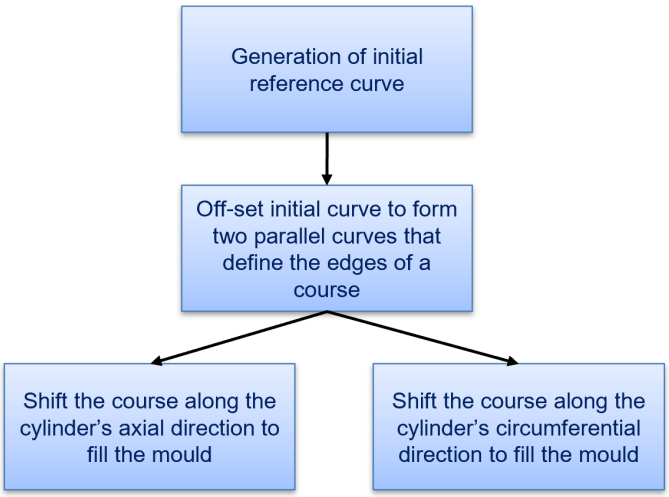


Figure 3.1: A block diagram that describes the modelling process to fill a cylinder from an initial reference curve.

3.2.1. Cylinder geometry and properties

The dimensions of the cylinder and the material properties are defined in this section. The geometry of the cylinder was selected to be identical to that of the work of S. Pitton *et al.* [36], where an ANN acted as a surrogate model, so a comparison can be drawn. For the cut-out cylinder, two holes of different dimensions were considered. The centre of the cut-outs aligned with the middle of the cylinder's length and they were placed on opposite sides. The cut-outs had the reverse major and minor axes dimensions and therefore they are defined by the same two values. The purpose regarding the inclusion of two different cut-outs was to capture how the angle variation will conform to their distinctive shape, since certain cut-out geometries affect specific structural objectives in a different manner. Furthermore, the strength of the material system AS4/8552 was taken from [148], where the A-basis values were deemed most appropriate for design. The data of the cylinder and its properties can be seen in Table 3.1.

3.2.2. Modelling technique

Having defined the geometry and material properties of the cylinder, the next step is to model the spatial angle variation. Arguably, the functional representation that yields the greatest design space is a trajectory generated by NURBS curves. This assures that the optimal solution will be a set of the possible curves that can be generated via this approach. On the other hand, the design space cannot be searched thoroughly within a practical time frame. For that reason, the Lagrange polynomial that was presented in Equation 2.38 was used instead, as it allows for an efficient alternative of a sufficiently large design space that requires significantly less variable inputs. Equation 2.38 is repeated below for clarity:

$$\theta(x) = \sum_{m=0}^{M-1} T_{mF} \prod_{m \neq i} \frac{x - x_i}{x_m - x_i} \tag{3.1}$$

Table 3.1: Material properties of *AS4/8552*

Geometry of the cylinder					
Cylinder Length [mm]	Cylinder Radius [mm]	Ply Thickness [mm]	No. of Plies [-]	Tow Width [mm]	No. of Tows [-]
705	300	0.181	8	3.175	32
Elastic Properties					
E_{11} [GPa]	E_{22} [GPa]	E_{12} [GPa]	E_{13} [GPa]	E_{23} [GPa]	v_{12} [GPa]
141	10.3	4.5	4.5	2.25	0.3
A-Basis Material Strength					
ε_{1T}	ε_{1C}	ε_{2T}	ε_{2C}	ε_{12}	
0.0155	0.001	0.004679	0.0217	0.015028	
Geometry of the Elliptical Cut-outs					
Major Axis of First Ellipse [mm]			Major Axis of Second Ellipse [mm]		
200			300		
Density [kg/m^3]					
1580					

where M is the number of the CPs, T_m are the selected angles at these CPs and x represents the coordinates that follow the axis along which the fibre angle varies. In this case, the angle coefficients at the CPs, as well as their spatial position, can comprise the design space. The inclusion of the latter will considerably enlarge the design space however, and since it is likely that the angle coefficients are more important for the considered optimisation problems, it was decided that the CPs will be linearly spaced from one another and will not be a part of the design space.

As it has already been mentioned, the AFP shifting method requires that the course trajectory varies along an axis and is shifted perpendicular to it, in order to cover the whole area. For the case of a cylinder, the axial and circumferential variation, hereby also abbreviated as AV and CV respectively, are depicted in Figure 3.2. It is noted that the cylinders are not shown to scale.

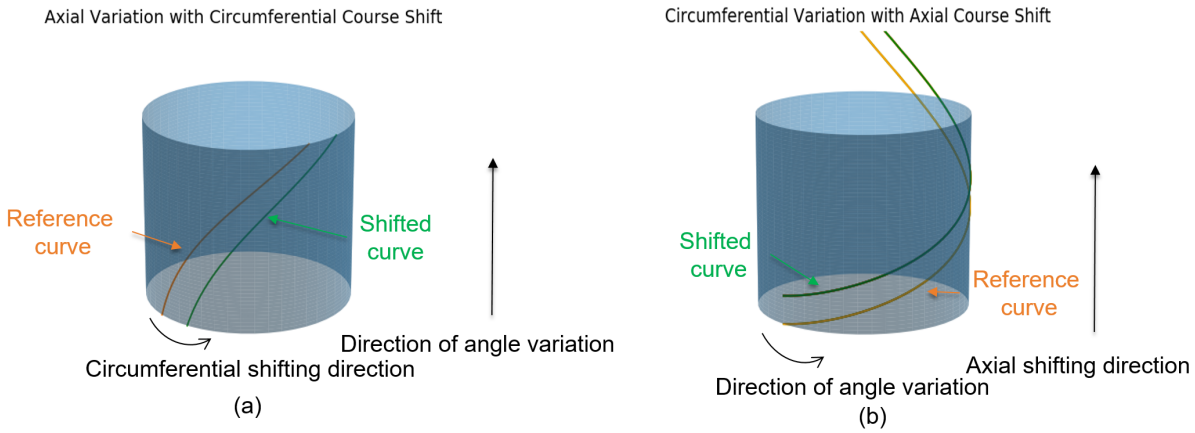


Figure 3.2: A representation of the (a) axial variation and (b) circumferential variation.

Since the curve of the course needs to be formed explicitly, its position can be determined by manipulating equation 2.38, which provides the variation of the angle. By unrolling the cylinder to form a 2-D plane, let the axis of the cylinder be z and the circumference denoted by s, then the coordinates of the curve can be obtained by the trigonometric formula:

$$\frac{\mathrm{d}z}{\mathrm{d}s} = \tan(\theta) \Rightarrow \int \frac{1}{\tan(\theta)} \mathrm{d}z = \int \mathrm{d}s \tag{3.2}$$

At this point the equation will be solved for the smallest amount of CPs, which is two, and their positions lie at the bottom and top of the cylinder respectively. For the case of axial variation, $z \in [0, L]$, with L being the

height of the cylinder. Equation 3.2 can then be integrated to obtain the unknown values of s. By substituting θ as a function of z from 2.38:

$$\begin{aligned}&\frac{\mathrm{d}z}{\mathrm{d}s} = \tan\left(\theta\left(z\right)\right) \Rightarrow \int \frac{1}{\tan(\theta(z))}\mathrm{d}z = \int \mathrm{d}s \Rightarrow \\ &\frac{\ln\left(\tan\left(\frac{zT_2}{L} - \frac{(z-L)T_1}{L}\right)\right)}{\frac{T_2}{L} - \frac{T_1}{L}} - \frac{\ln(1-\tan(\frac{zT_2}{L} - \frac{z-LT_1}{L})^2)}{\frac{2T_2}{L} - \frac{2T_1}{L}} + C_0 = s\end{aligned} \tag{3.3}$$

where C_0 is the coefficient of integration. Forcing the initial condition of $s = 0, @z = 0$, the coefficient becomes:

$$C_0 = -\frac{L\left(2\ln\left(\tan\left(T_1\right)\right) - \ln\left(1+\left(\tan\left(T_1\right)\right)^2\right)\right)}{2\left(T_2 - T_1\right)} \tag{3.4}$$

Substituting the coefficient C_0 back to the previous expression, the circumferential position s can be determined:

$$s = \frac{L}{2\,T_1 - 2\,T_2}\left(-2\ln\left(\tan\left(\frac{T_1 L - \left(T_1 - T_2\right)z}{L}\right)\right) + \ln\left(1+\left(\tan\left(\frac{T_1 L - \left(T_1 - T_2\right)z}{L}\right)\right)^2\right) + 2\ln\left(\tan\left(T_1\right)\right) - \ln\left(1+\left(\tan\left(T_1\right)\right)^2\right)\right) \tag{3.5}$$

In a similar fashion, the variable z can be calculated in the case of circumferential variation, by assuming that two CPs exist at the start and at $2\pi r$, where r is the radius of the cylinder, and forcing the same initial condition:

$$\begin{aligned}&\int \tan\left(\theta\left(s\right)\right)\,\mathrm{d}s = \int \mathrm{d}z \Rightarrow \frac{\ln(1+\tan(\frac{sT_2+(2\pi r-s)T_1}{2\pi r})^2)}{\frac{T_2-T_1}{\pi r}} + C_0 = z \Rightarrow \\ &C_0 = -\frac{\ln\left(1+\left(\tan\left(T_1\right)\right)^2\right)\pi\, r}{T_2 - T_1} \Rightarrow \\ &z = \frac{\pi\, r}{T_1 - T_2}\left(\ln\left(1+\left(\tan\left(T_1\right)\right)^2\right) - \ln\left(1+\left(\tan\left(\frac{2\,\pi\, r\, T_1 - s\left(-T_2+T_1\right)}{2\pi\, r}\right)\right)^2\right)\right)\end{aligned} \tag{3.6}$$

Thus, the curve can be constructed explicitly for two CPs for both cases of angle variation. Unfortunately, no real solution can be obtained for a higher number of CPs, and the integration has to be performed numerically for these cases. One more limitation exists for the axial angle variation, where the angle coefficients T_m need to have the same mathematical sign ($\pm$), as otherwise the solutions are completely imaginary or the curve will self-intersect. This problem does not present itself for the circumferential variation, although as it will be explained later on, the curves might still self-intersect.

Depending on the case of angle variation, the positions of the CPs should be unique and along perpendicular axes, as shown in Figure 3.3. The CPs are shown as lines, since the angles calculated by the equation are equal along the span of the lines. In Equations 3.3-3.6 it was assumed that the starting positions are the same at $s = z = 0$, however the ending position is at $z = L$ and $s = 2\pi r$ for the axial and circumferential angle variation respectively. Different positions could have been selected, e.g. midspan values at $L/2$ and πr to then utilise symmetry, but the former positions were preferred because they enlarge the design space. Since the termination point of only one direction is defined for a given angle variation case, Equations 3.3-3.6 dictate the perpendicular ending position, which could lead to many windings around the cylinder for the case of AV, and unnecessary overlap at the same location for the CV. These designs are most probably sub-optimal with respect to mass, and it was expected that they will vanish during the multi-objective optimisation scheme.

Having generated the central reference curve, the curves that bound the course can be generated next. This can be accomplished in multiple ways, but the simplest approach is to off-set the reference curve to form a family of parallel curves. This is possible by utilising Equation 3.7 to off-set the reference curve by half the course's width value:

$$\begin{aligned}s_d &= s + \frac{w_{course}\, z'(t)}{2\sqrt{s'(t)^2 + z'(t)^2}} \\ z_d &= z + \frac{w_{course}\, s'(t)}{2\sqrt{s'(t)^2 + z'(t)^2}}\end{aligned} \tag{3.7}$$

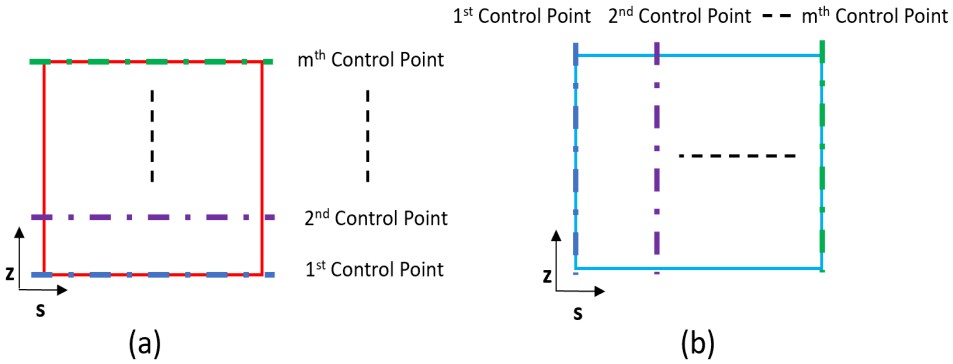


Figure 3.3: The distribution of CPs on the unrolled cylinder for the case of (a) axial angle variation and (b) circumferential angle variation.

The variable w_{course} stands for the width of the course, while $z'(t)$ and $s'(t)$ are the derivatives of the parametrically given $z(t)$ and $s(t)$, where t is the parameter. Therefore, the course can be defined by the two outer curves that bound it.

Since the coverage method adopted in this work only allows for overlaps, the course needs to be shifted by the appropriate value that eliminates gaps and minimises overlaps. This value is essentially the effective course width, which can be approximated as such for the cases of AV and CV respectively:

$$\begin{aligned} w_e &\approx \frac{w_{course}}{\max(\sin(\theta(z)))} \\ w_e &\approx \frac{w_{course}}{\max(\cos(\theta(s)))} \end{aligned} \tag{3.8}$$

These equations are approximations due to the variation of the angle, leading to the boundary curves of the course having different angles along a given axis, and thus their accuracy can differ significantly for certain course trajectories. Therefore, a simple distance minimisation between the parametrically defined courses was used, so the courses touch without any unnecessary overlaps.

3.2.3. Design viability

Although the equations that have been discussed so far can fill the cylinder with courses, not all of the designs that can be modelled are viable. Some of these problems come in the form of either real life constraints and some are inherent to the mathematical depiction of the model. These limitations are considered in this section.

Curvature constraint

A manufacturing constraint regarding the curvature of the tows should be imposed to the optimisation problem, as the optimised designs would be unrealistic otherwise. The determination of the number of tows in a course and the tow width affect the value of the manufacturing constraint, which limits tow curvature. The vast majority of studies use a specific pair of values for the aforementioned variables, due to greater curve freedom and flexibility they provide, and for that reason, the same values were selected in this work, yielding the curvature constraint of $1.5748\ 1/m$. The calculation of the curvatures of the fibres are calculated by taking the gradient of their individual angle values. However, the curvature has significant values only along the axis of variation, and thus it suffices to take the derivative of the angle with respect to that orientation. In all cases the derivatives were calculated analytically, although it is not recommended for CPs that are higher than five, since it is a long and tedious process and numerical methods should be preferred instead.

Whether the manufacturing constraint is satisfied directly depends on the number of CPs that will be chosen for a given design, since an increase in CPs will result in an enlargement of the design space. To better illustrate the importance of CP number selection, a random curve that varies along the cylinder's axis is presented in Figure 3.4.

It is evident that the curve cannot satisfy the manufacturing constraint easily as the flexibility of the curve increases. Although more viable designs should be possible with a big CP number, the computational time that is needed to search for them might prove to be exceptionally high. Therefore, a compromise needs to

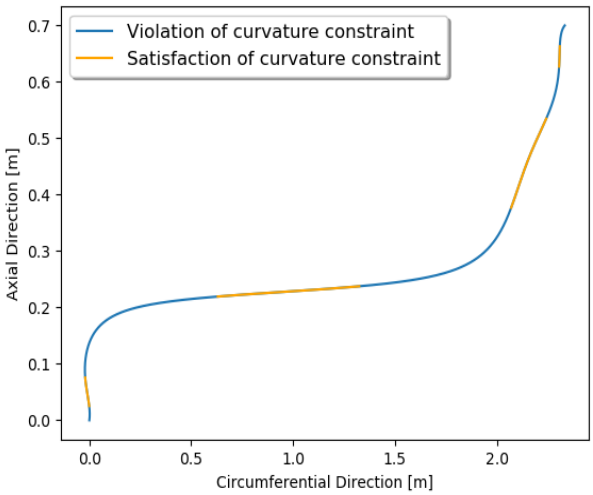


Figure 3.4: A randomly generated curve showcasing the likely violation of a curvature limit for a flexible curve.

be made with respect to the flexibility of the curve and the expected optimal designs. Thus, the design space was sampled via the LHS method, in order to generate sufficient samples to determine which number of CPs should be used in the ensuing optimisation. The construction of the LHS was performed by the minimisation of the maximum correlation coefficient, in order to ensure that the sampling is representative. In Figure 3.5, the amount of acceptable designs are presented as a function of the number of CPs. It is clear that the design space is enlarged significantly for small increases in the number of CPs, and there is a steep decrease in viable designs for the chosen sampling amount. For the given cylinder geometry, it seems that the larger design space of circumferentially varying angles is able to provide the same amount of viable solutions as the AV when it comes to the manufacturing constraint. By considering the full design space, it can be seen that the curves level-off at 5 CPs, where the viable designs are effectively null. Therefore, it was decided that only CPs with a value of up to 4 would be considered in the rest of the study.

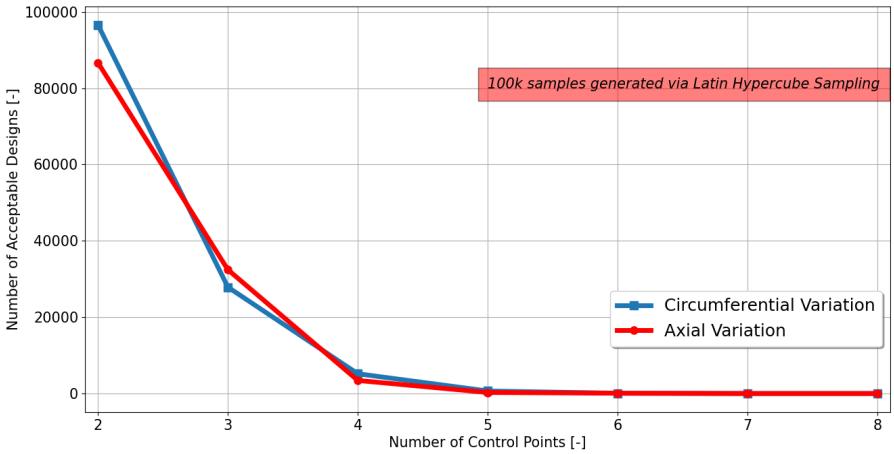


Figure 3.5: Designs that satisfy the manufacturing constraint as a function of the number of CPs.

Course creation

Not all reference curves lead to viable courses, as some parallel curves can self-intersect due to excessive curvature. Such an occurrence is shown in Figure 3.6, where both boundary curves self-intersect. In most cases the self-intersection of the curves is a sub-set of the violation of the manufacturing curvature constraint, and designs that satisfy that constraint are likely to produce viable courses.

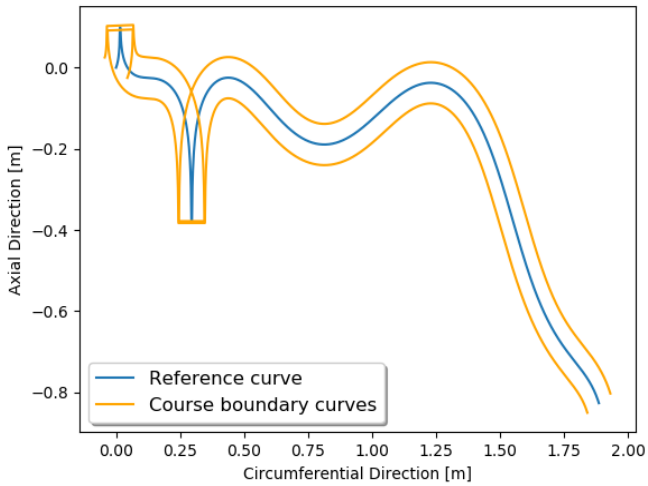


Figure 3.6: An example of a self-intersecting curve in a course due to parallel off-setting.

Cylinder area coverage

The way the reference curve is set-up is to start and terminate at well defined locations that bound the geometry of the cylinder. However, when a course is created by off-setting the reference curve, the boundary curves that are normal to the reference curve do not necessarily start and terminate at the correct locations, which may lead to the creation of gaps that need to be eliminated. This can best be seen in Figure 3.7, where there are visible gaps between two courses for the case of cylinder axial angle variation.

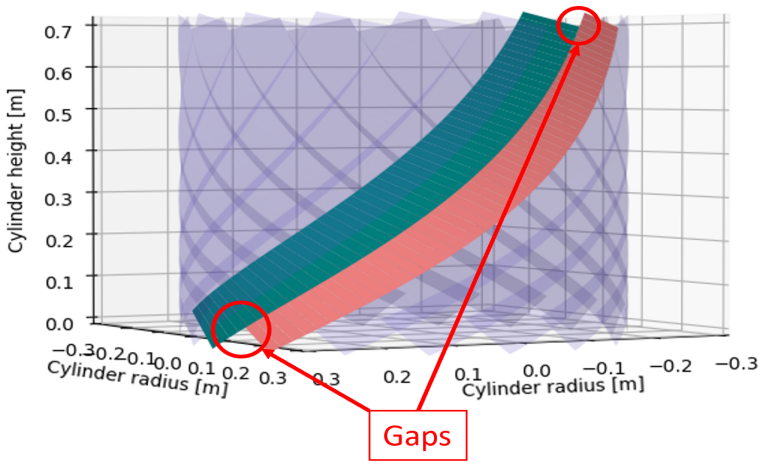


Figure 3.7: Problematic gaps between courses due to the parallel off-set of the reference curve of the course.

Another point of consideration is related only to the case of an axial angle variation of the cylinder. As the courses are shifted circumferentially and the final course fills the cylinder, some additional overlap may be present, which can be an unwanted feature. This problem occurs because the number of shifts of the initial course is an integer number and the final shift can cover more area than is needed. In Figure 3.7 the circumferentially covered cylinder is shown, where the darker colours indicate a higher thickness due to overlaps. It can be seen that in this case the final course results in considerable overlap. This can be avoided if the tows of the course are trimmed to have a variable width, which is possible for some AFP machines. However, since the scope of this work was to model and account for the overlap defects and their distribution, the width of the courses was kept constant, as otherwise the same trimming procedure could be employed between the rest of the courses, which would defeat the purpose.

Another coverage problem might arise only for the case of circumferential angle variation, where the first and final courses can have significantly different orientations when they meet, resulting in the creation of

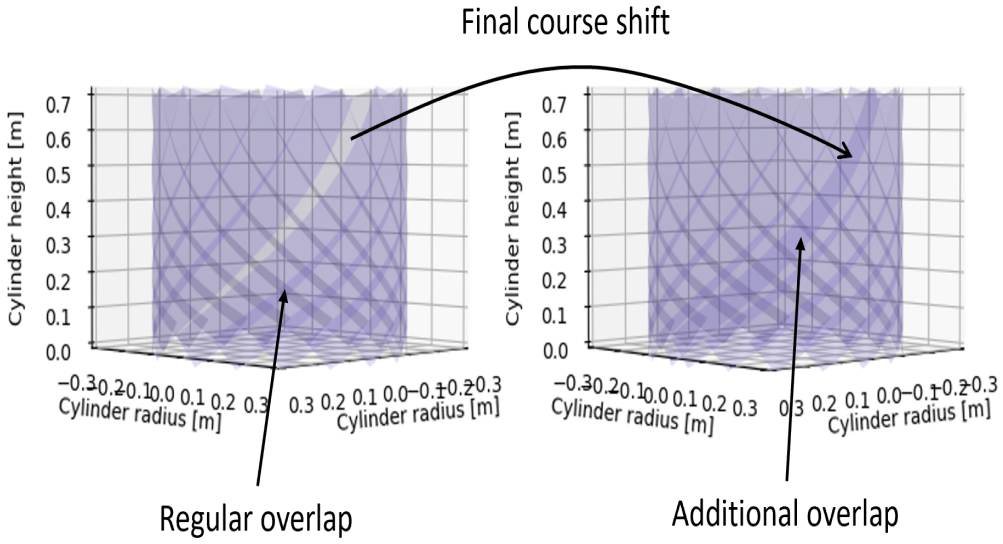


Figure 3.8: An illustration of the additional overlap that can occur with the last course that is needed to cover the cylinder.

gaps. Such an occurrence is presented in Figure 3.9. On the other hand this orientation miss-match could lead to additional overlaps instead of gaps for certain designs.

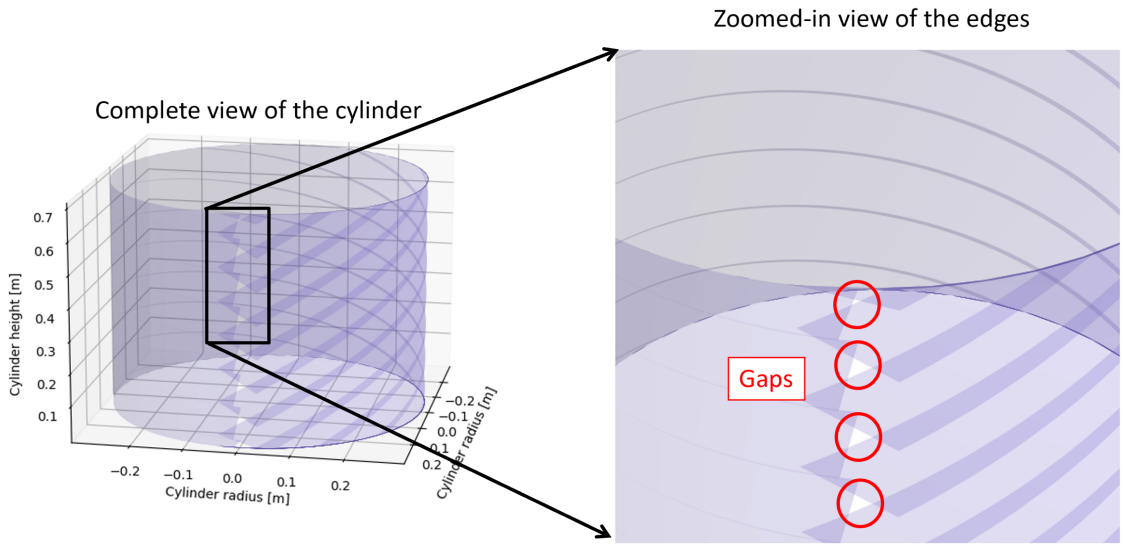


Figure 3.9: Presentation of a case where gaps are present due to the first and final courses having different angles where they touch.

Another problem related only to the circumferential angle variation is that additional overlaps can also occur if the number of shifts is unnecessarily high. To showcase this, let n be the proper number of shifts that leads to full coverage without any additional overlap. In Figure 3.10 a course is shifted upwards in order to cover the unrolled cylinder. The determination of the upward shifts is not a straight-forward matter if additional overlap is to be avoided. The problem becomes more complicated if gaps are also present and additional shifts might fill them. Therefore, the solutions to these problems cannot be found analytically.

Since the coverage problems that have been presented are dependent on the case of the angle variation, two different numerical schemes were developed to mathematically optimise the courses to cover the cylinder without adding unnecessary weight. The optimisation is accomplished by finding the optimal combination of extending the course to fill the gaps and determining the right number of courses. The course extension was a linear extrapolation for the sake of simplicity. The optimisation schemes for the AV and CV are presented in Equation 3.9, respectively.

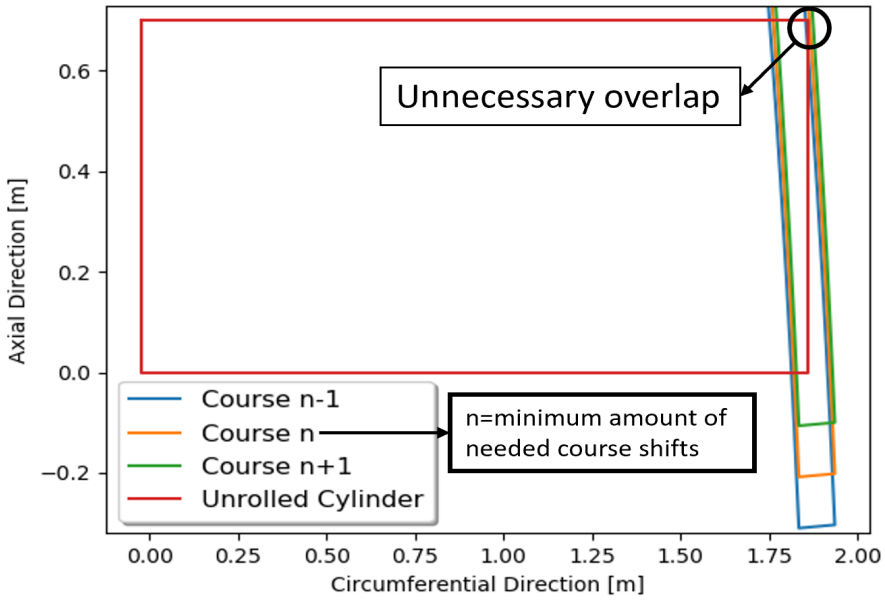


Figure 3.10: Unnecessary overlap in the case of circumferential angle variation.

$$\begin{aligned} &\min\left(n+e_b+e_t\right), \quad \min\left(n_b+n_t+e_b\right) \\ &s.t. A_c - A_n \leq \epsilon, \qquad s.t. A_c - A_n \leq \epsilon \end{aligned} \tag{3.9}$$

For the case of axial angle variation, n stands for the number of courses, e_b is the extension of the courses downwards, e_t is the extension of the courses upwards and for the case of circumferential angle variation, n_b is the number of downward shifts, n_t is the number of upward shifts and e_b is the course extension downwards. They are both subject to the same constraint where A_c is the analytically calculated area of the cylinder that is completely filled and without additional overlap, A_n is the area that is found iteratively in the optimisation, and ϵ is a small numerical error. This constraint ensures that the cylinder will be completely filled while the objective function is minimising the variables that lead to additional overlap.

The optimisation algorithm that was used was the Hooke-Jeeves algorithm, which is a direct search algorithm [149]. The algorithm evaluates the expression $f(\boldsymbol{x} \pm \alpha \boldsymbol{e^i})$ in i^{th} dimesional space by searching the objective space in small steps α for different basis vectors e^i , where the directions resemble a compass. Since the computations at each iteration are quite cheap, sufficient iterations around the objective space are used to guarantee that the global minimum is found. The algorithm was chosen for its simplicity and good convergence for low-dimensional problems, as the optimisation problem is not complicated and convergence to the solution is achieved relatively fast. The difficulty of the optimisation problem lies in the calculation of the current area in each iteration, as the area of each course and the subsequent overlaps cannot be calculated analytically, and for that reason it had to be performed numerically.

3.3. Stiffness variation of the cylinder

In order to capture how the stiffness varies throughout the cylinder, the varying thickness and angle fields need to be calculated first. To that end, the cylinder was discretised to a number of elements in order to assign these two properties to each element. This pre-processing step leads to the element properties being communicated to the software that performs the FEM analyses at a later stage. The right determination of the number of elements is crucial in order to achieve convergence in the FE analysis and in the correct calculation of the fields. However, a problem arises in FE analysis when each design is different and requires a different amount of elements for convergence. For that reason, it was decided to achieve convergence for a baseline quasi-isotropic (QI) CS design ($[0/90/45/-45]_S$) and then apply a safety factor of 1.1 in both the axial and circumferential direction in order to be on the conservative side. Another convergence consideration is whether the fields are accurate, and once again, each design requires a different amount of elements to ensure that convergence has been achieved. Therefore, two representative axial angle variation designs with different degrees of curvature are investigated to study the effect of the number of elements on the angle and

thickness fields. The two designs were selected to have a very low and high curvature respectively, in order to perform a comparison near the extremes, as shown in Figure 3.11.

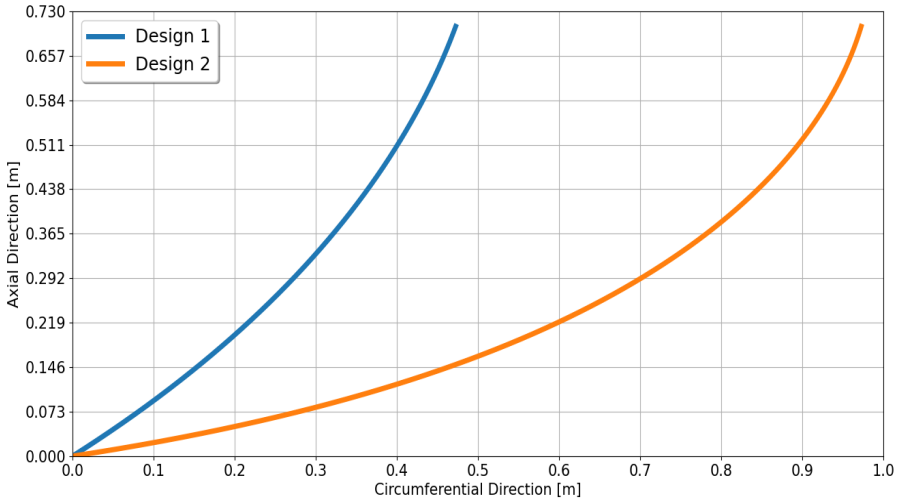


Figure 3.11: Two design curves with different degree of curvature for a convergence investigation.

The convergence performed in the FE analysis for the QI CS design yielded a greater number of elements in the circumferential direction than in the axial, mainly driven by the accurate calculation of the first buckling mode, and therefore a similar ratio of circumferential to axial elements was used in the investigation, whenever applicable. In order to ascertain whether convergence has been achieved, the statistical variance of the field values for each element was calculated, since a constant variance would mean that the field has been captured correctly and a further increase in the number of elements does not yield further improvements. The normalised variance versus the number of elements is presented in Figure 3.12 in log-log scale, where it is observed that the angle field needs only a small amount of elements for its convergence, while the thickness field exhibits a knee with a smaller upwards slope afterwards. It seems that the convergence of the thickness field requires a very high number of elements, which would make the optimisation process completely impractical. Therefore, a cut-off point of 3500 elements was selected, which satisfies the angle field and the convergence related to the FEM analyses, as is shown in Figure 3.12. This plot is also representative of the circumferential angle variation, as the results are very similar in that case as well.

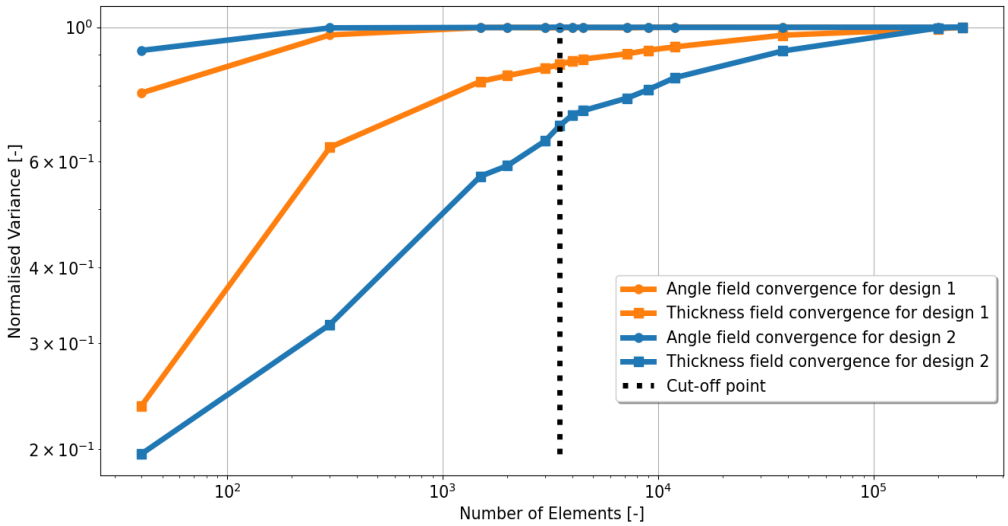


Figure 3.12: Convergence of the angle and thickness fields for selected designs.

3.3.1. Thickness variation

The variation of the fields can also be visualised for the selected amount of elements. The numbering of the elements was performed row by row, from bottom to top, as is shown in Figure 3.13, where n denotes the number of columns and k the number of rows.

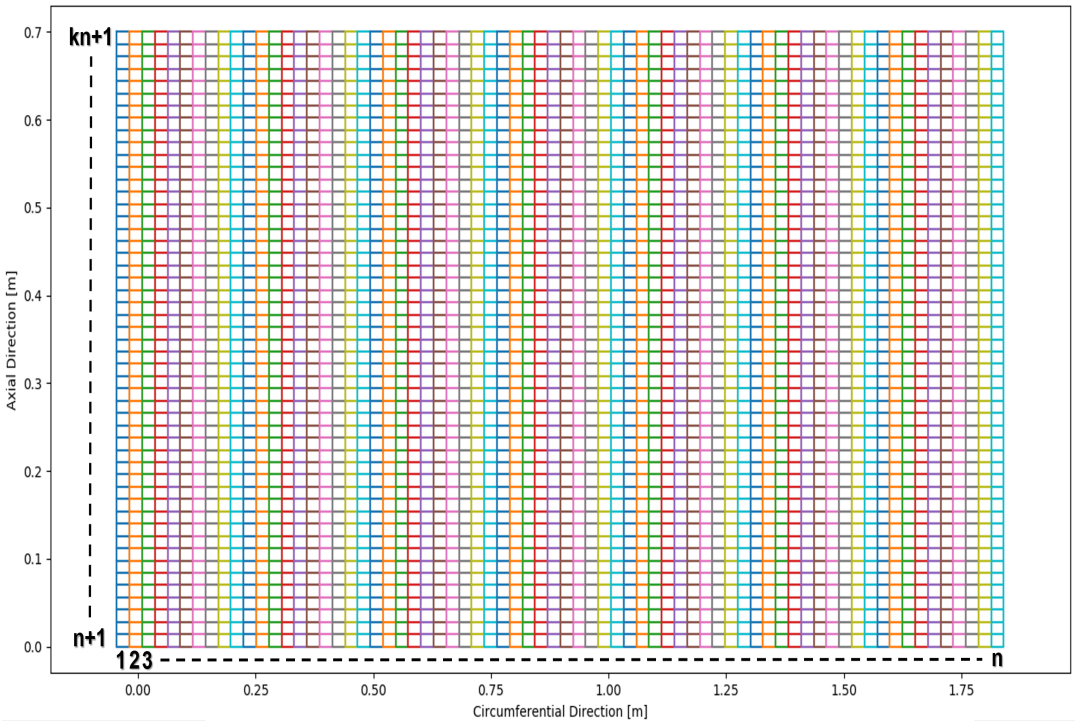


Figure 3.13: The numbering of the elements that discretise the cylinder.

As an example, Figure 3.14 presents the courses of an axial angle variation cylinder that has been completely filled and optimised for gaps and overlaps. It can be seen that there is significant overlap at 30% of its height due to its small angle at this position, while the top has substantially less overlaps because of the respective high angle.

The thickness distribution for each element of this example is shown in Figure 3.15, where the number of overlaps for each element is displayed. It can be seen that the results seem quite noisy and that certain bands have similar values. These bands are essentially the elements on the same row of the mesh, where it is expected that in the case of axial angle variation the thickness variation along the cylinder's circumference must be minimal. The same general trend as what was discussed for Figure 3.14 is observed, where the highest number of overlaps occurs close to 30% of the cylinder's height and the lowest value of no overlaps is in the range of the final elements, which represents the top of the cylinder.

So far, the discussion was concerned with only a single composite ply. When it comes to a laminate, the method of staggering the plies was used in order to avoid excessive thickness build-ups in certain locations. Since an effective ply staggering spreads out the thickness uniformly, it was decided to superimpose the thickness values of each ply for each corresponding element and minimise the average variance between the elements. In that way, once the variance has been minimised it is ensured that the plies have been uniformly distributed. The minimum staggering distance is the distance between two elements of the same row. Thus, the design variables are the staggering distances of each ply, in terms of discrete element distance jumps. For this simple optimisation problem, the same Hooke-Jeeves algorithm as for the cylinder coverage was employed. The unconstrained optimisation problem is given as:

$$\min \left(\frac{\sum_{i=1}^{n} (t_i - \bar{t})^2}{n-1} \right) \tag{3.10}$$

where n is the number of elements, t_i is the laminate thickness of the i^{th} element and $\bar{t}$ is the mean

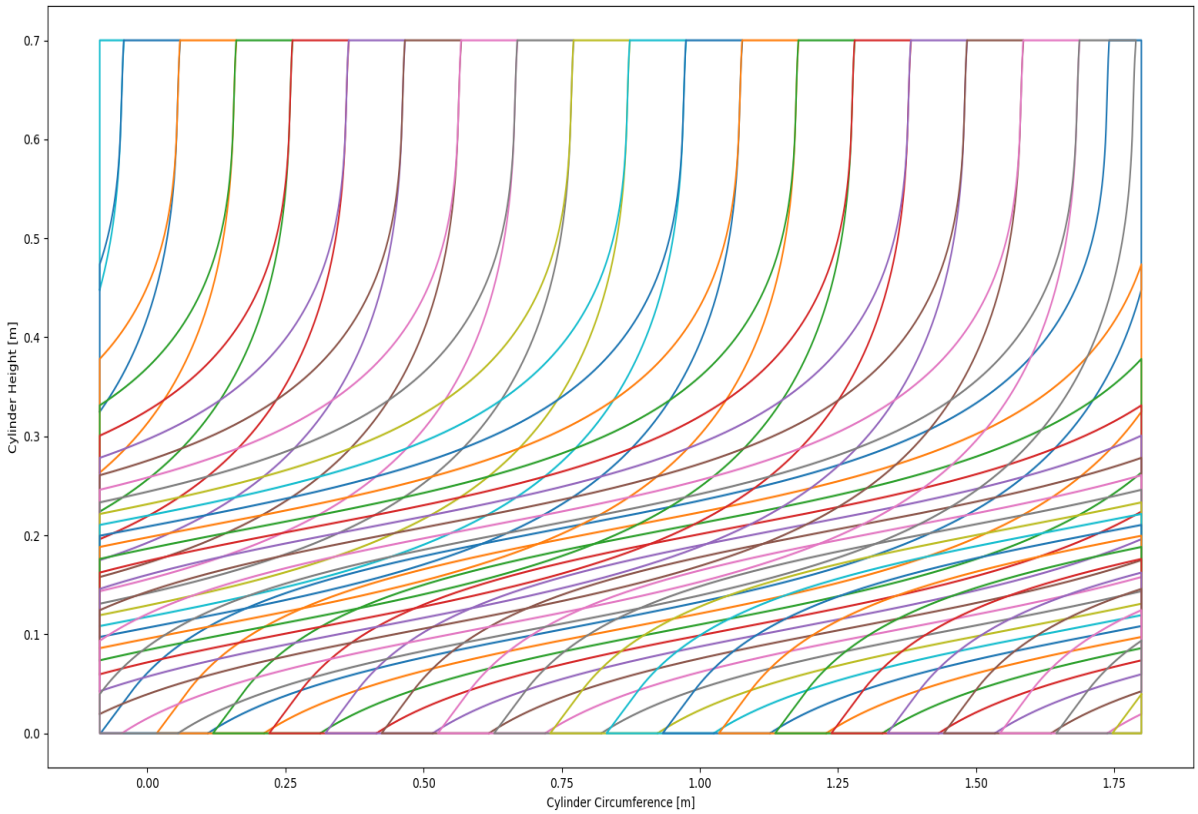


Figure 3.14: An example of the courses of an axial angle variation cylinder.

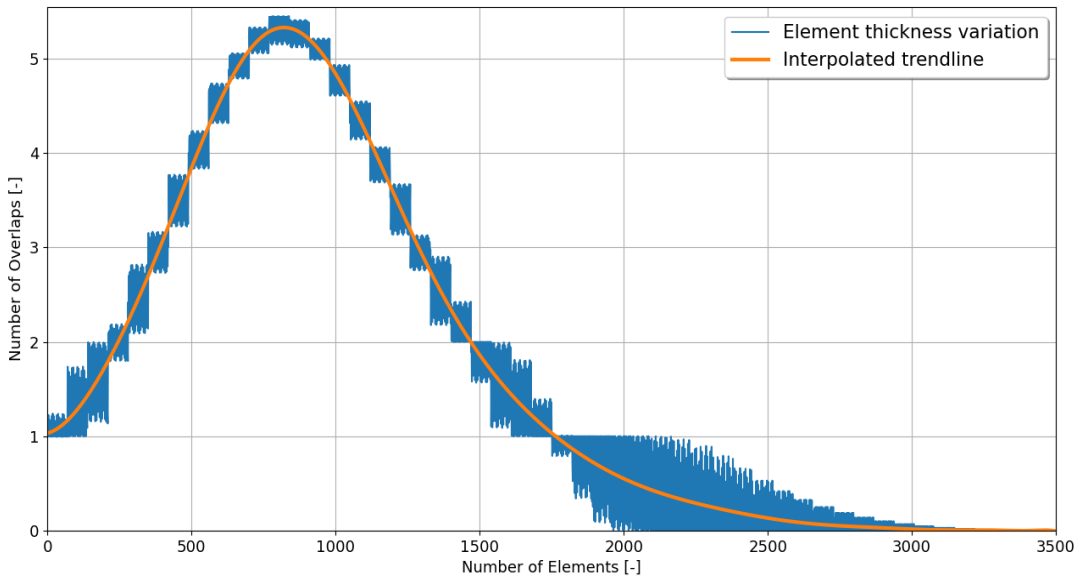


Figure 3.15: Thickness variation of every element of a selected example design.

laminate thickness.

3.3.2. Angle variation

In literature, the angle variation is most frequently calculated as being constant perpendicular to the axis of variation. In other words, rows or columns of elements can be homogenised to have the same values. This simplifies the model and is very efficient computationally-wise. In reality, the angle varies throughout a single course as depicted in Figure 3.16, rendering the homogenisation approach an approximation. A more accurate approach was favoured in order to capture this variation within the courses. Thus, the course was discretised by multiple curves, other than the reference and boundary curves, as shown in Figure 3.16.

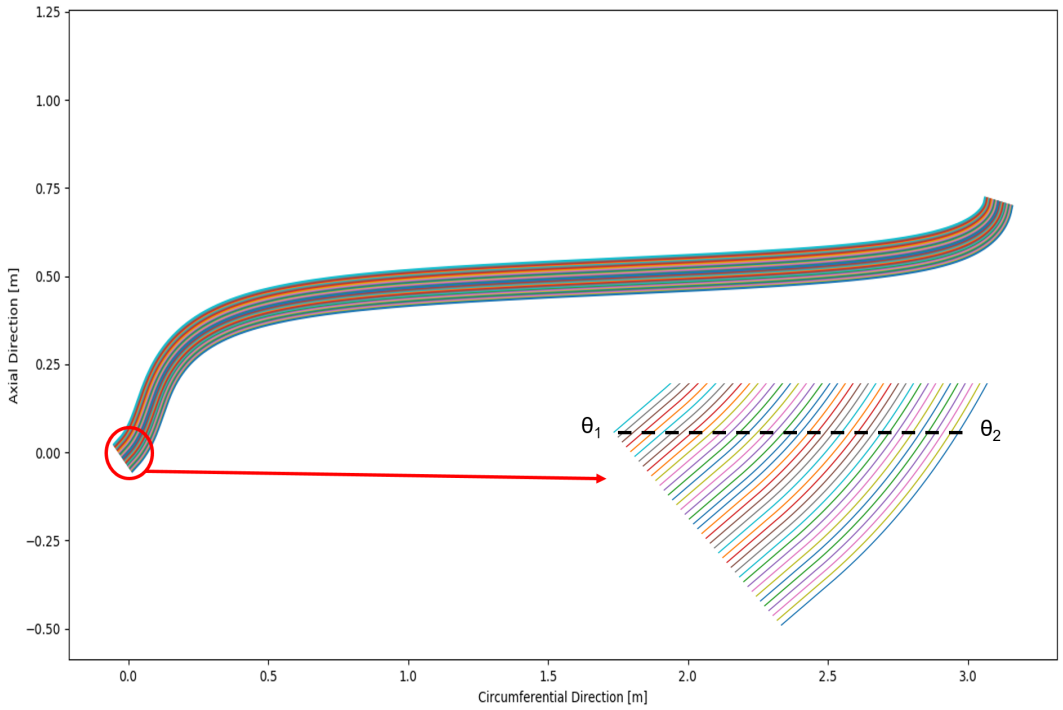


Figure 3.16: Angle variation within a single course.

Naturally, the number of curves dictates the accuracy of the calculation, which is dependent on the design. In order to reduce the required computational time, 80 curves was deemed an appropriate amount for each course, based on the number of elements that has already been chosen.

In the current method, an angle value is assigned to an element by considering a given curve's proximity to the element's centroid but also by factoring in the amount of common area between the element and the course that contains the curve. In other words, a weighted average approach was utilised, which is given as:

$$\theta_e = \frac{\sum_{i=1}^{n} \theta_i w_c}{\sum_{i=1}^{n} w_c} \tag{3.11}$$

where θ_e is the element angle, n is the number of curves, θ_i is the angle of the i^{th} curve and w_c is the weight of the course calculated as a fraction of common area with the element. It is noted that the angle values that were presented earlier for the convergence study of the angle field were calculated by Equation 3.11.

The ongoing example of the cylinder is shown discretised by course curves in Figure 3.17. It is observed that the number of curves seems sufficient and results in a fine discretisation for the examined case.

The angle variation for each element can then be plotted as seen in Figure 3.18, where the homogenised approach is also included for comparison. It can be identified that the homogenised approach follows the weighted average approach and it almost acts as an average for some rows of elements. Although this homogenisation is not as accurate, it is significantly cheaper computationally-wise and can serve as a good

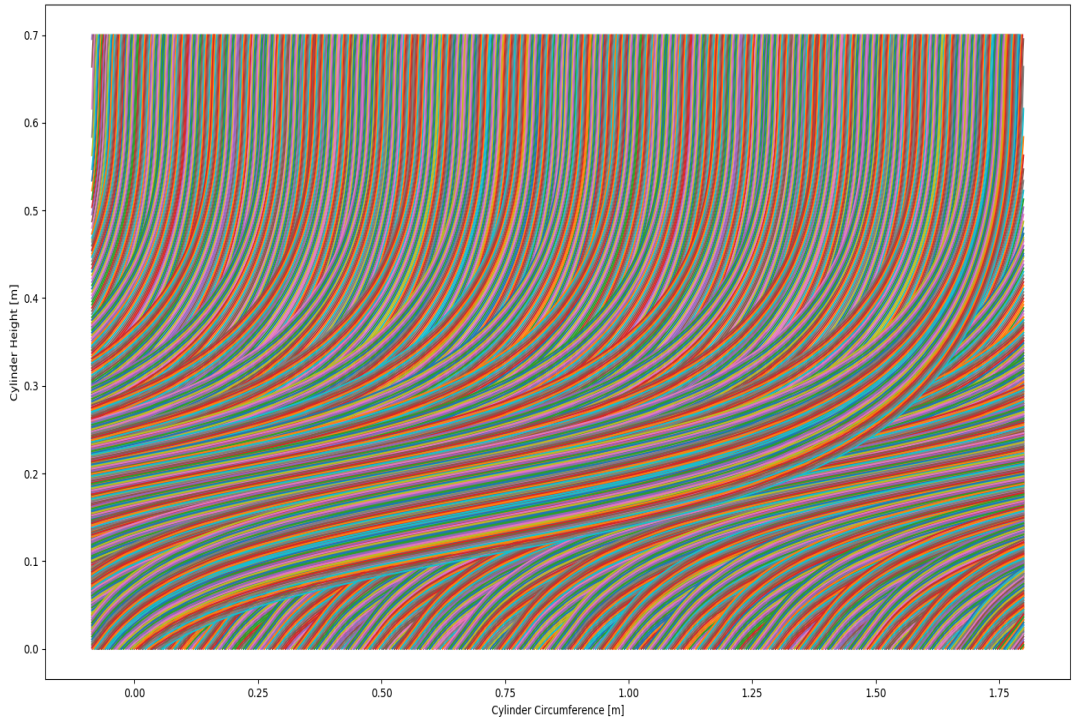


Figure 3.17: A cylinder discretised by course curves.

alternative. However, its accuracy can vary greatly between designs and it should be considered whether there is little angle variation within the course preferably.

3.4. Set-up of the artificial neural network architecture for element property prediction

The angle and thickness fields are element properties that require time-consuming computations and their approximation with a surrogate model is desirable instead. An ANN can act as the surrogate model in this case as the element properties can be predicted in a regression analysis. Thus, the angle coefficients at the CPs can serve as the input to the network in an ideal case, with the output being the element properties. Although one network can be trained to predict both element properties, in this work separate ones were trained to focus on each one.

The ANN should be able to make accurate predictions, so the element properties can then be passed on to the FEM software. Unfortunately, as the accuracy of the prediction increases, so does the amount of training data that is required. For that reason a compromise between accuracy and number of input data needs to be made, in order for the optimisation to finish in a realistic time frame. Of course, since the aforementioned values cannot be known beforehand, an arbitrary decision for the sufficient accuracy was made, where a value of 85% was selected to be calculated on the validation data. Although there is no single metric that can define exactly what accuracy means, in this case the Coefficient of Determination, or otherwise the statistical R^2, was chosen to represent the accuracy because of its popularity in literature and its resemblance to the concept of general accuracy that goes up to 100%.

Since the element properties can be quite noisy, the ANN is likely to overfit the data. Therefore, many of the techniques that were discussed in Section 2.1.7 were also employed in this case, in order to force the network to generalise more. To begin with, the input data was divided into training and validation data, with values of 80% and 20% respectively. The data was also shuffled and stratified, so it is ensured that the input will be representative. The data were also standardised accordingly by removing the mean and scaling the variance to unity, so a zero-mean Gaussian distribution is obtained. In that way, a bias towards certain features is avoided, as well as other problems like possible divergence of the training algorithm. Where appropriate, the technique of batch normalisation was also employed in order to improve the stability and performance of the network by centring the input of the HLs around the origin and normalising them.

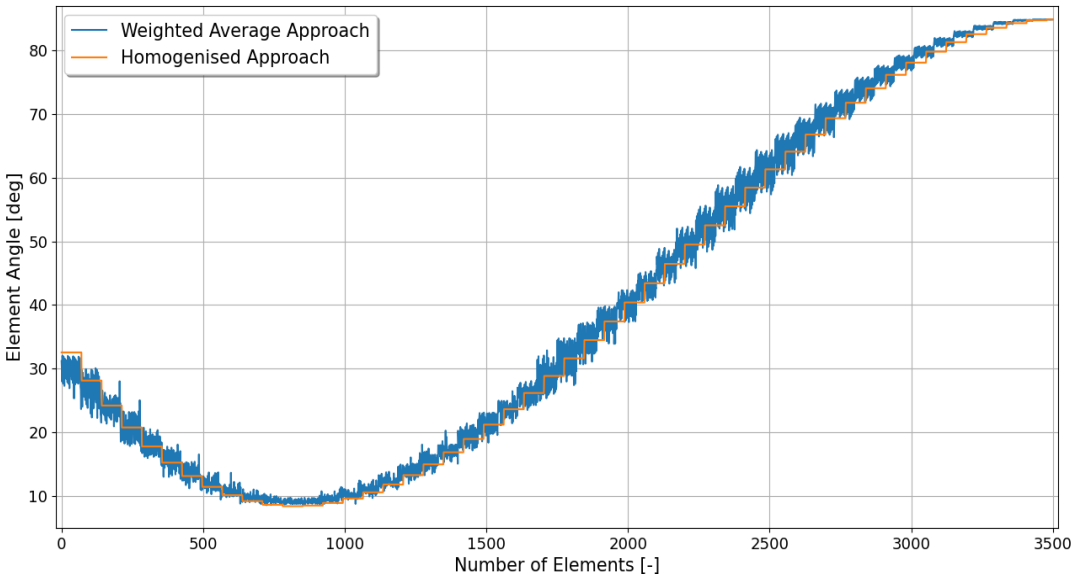


Figure 3.18: Angle variation of each element for a selected example design.

When it comes to hyperparameter initialisation, such as the weights and biases, it is only relevant if a gradient descent backpropagation algorithm is utilised, since a population-based approach will create the first generation according to the sampling technique that was selected. In this work, the gradient descent backpropagation algorithm was favoured, since it expedites the process significantly if its own hyperparameters are fine-tuned.

The selection of the initialisation method is generally performed based on the activation function which is utilised in the ANN model. The most novel activation functions are the ELU and SELU at the time of writing, which go well together with the He and LeCun initialisations respectively. The SELU function tends to outperform all the others, but it requires more HLs, which hinders the speed of the model's architecture optimisation. Both of them were included as possible hyperparameters to determine the optimal ANN architecture.

An early stopping method was also utilised, where the model is reverted back to the point that it exhibited the lowest MSE value on the validation data. A stopping patience value of 100 was also used, so that the full epoch set would not be examined, yielding a speed boost in the training process. For that reason, the number of epochs was chosen to be very high.

The Nadam optimiser was selected to perform the backprogation of the neural network, since it is regarded as one of the most state-of-the-art training algorithms in the literature with superior performance results with respect to other optimisers. Its many hyperparameters and longer computation time are its only drawbacks, but the performance gains should outweigh them.

The network contains various hyperparameters that have to be properly adjusted to increase the model's accuracy. In addition, the architecture of the ANN that maximises performance is also unknown. Thus, an optimisation scheme can provide the variables that yield the best results. The preferred metric for the gradient descent algorithm is the MSE in literature, and although various regression metrics have their own advantages and disadvantages and a combination of them could be used to accelerate convergence, the increase in computational time by calculating more than one metric retards the optimisation process. Therefore, the objective to be minimised is the summation of the MSE metrics for the training and validation data.

The design variables include the real-valued hyperparameters of the Nadam algorithm, such as β_1, β_2 and its learning rate, as presented in Equation 2.26. The integer-valued number of neurons and HLs were also varied, although for regression problems the number of HLs that perform best tends to be small, and thus its upper bound value was 5 layers.

As a summary, the hyperparameters that were optimised and the ones that remained fixed can be seen in tabular form in Table 3.2 and Table 3.3, respectively.

The hyperparameters of the ANN were optimised by a DE algorithm, owing to its great efficiency in finding global extrema. The implementation of the algorithm is the one proposed in [72], where the algorithm's hyperparameters CR and F are adaptively changing from variations of the chosen operator. The variants for

Table 3.2: Summary of the ANN hyperparameters to be optimised.

Hyperparameter	Minimum Boundary	Maximum Boundary
No. of neurons per hidden layer	2	1000
Hidden layers	1	5
Activation function	ELU	SELU
Learning rate	0.0001	0.01
β_1	0.9	0.999
β_2	0.9	0.999
Batch size	16	64

Table 3.3: Summary of the ANN hyperparameters that are fixed in the optimisation.

Hyperparameter	Fixed Parameter
Epochs	1000
Early stopping epochs	50
Early stopping minimum change to qualify as an improvement	0.02
Initialisation	He normal/LeCun normal
Bias	On/Off
Loss function	Mean Squared Error
Output activation function	Linear unit activation function
Optimiser	Nadam

the mutation and crossover operations are identical to the ones discussed in 2.2.6, where one individual in the population is randomly selected following a uniform distribution and may undergo a binomial crossover depending on the value of its cost function.

The mathematical optimisation of the problem has the form:

$$\min(MSE_t + MSE_v) \tag{3.12}$$

where MSE_t is the mean squared error of the training data, MSE_v is the mean squared error of the validation data and n is the total number of neurons in the network. Once the model is optimised, its R^2 is calculated on the validation data to ascertain whether it meets the 85% accuracy criterion. If it does, then it will carry on making predictions that are passed on to the FEM model. The training commences every 1000 individuals from the generations of the main optimisation algorithm that is described in Section 3.5.

Although, it is very convenient to only provide the angle coefficients as input, additional information is likely to improve accuracy at the expense of the time required to generate this information. For instance, the Hooke-Jeeves optimisation used for the coverage of the cylinder provides the course extension and the number of shifts needed. Even though the optimisation is relatively fast, this category of data takes the longest to obtain. In contrast to these features, other information that could improve the performance of the ANN can be obtained almost immediately from the evaluation of the angle coefficients, such as the complete angle variation of the curves, the starting and ending positions of the course, and the maximum and minimum axial and circumferential widths of the course. It is also reminded that the shifting distance of a course is equal to its effective course width. In order to determine which approach should be preferred, 1000 samples were generated via the LHS method for 2 CPs in the case of axial angle variation. Then three different models were trained, one with the angle coefficients as input, the second with the cheaply computed additional information also included, and the last with all the aforementioned features.

The accuracy of the ANN model was also compared to two other popular surrogate models for regression, namely the RBF [14] and the Kriging with partial least squares [9] models, which were presented in Section 2.1.3. This comparison also serves as a validation for the accuracy of the ANN. More specifically, the RBF model used a Gaussian basis function and the Kriging model utilised a Gaussian correlation function.

The real-valued hyperparameters of the RBF and Kriging models were found with a simple GA optimisation algorithm with the same operations and optimisation hyperparameters as in the algorithm chosen for the multi-objective optimisation problem, which is discussed in Section 3.5.2. For the case of Kriging, the GA was combined with the efficient derivative-based optimisation of constrained optimization by linear approximation [150], where the GA would optimise the initial position of the gradient search algorithm along with the number of principal components. For clarity, it is restated that the RBF hyperparameter is a basis

Table 3.4: Comparison between different surrogate models, having only the CPs as input.

Thickness							
Metric	ANN	RBF	Kriging	Metric	ANN	RBF	Kriging
Normalised MSE	0.972636688	1	0.98767507	MSE	1.30922E-08	1.34605E-08	1.32946E-08
Normalised MAE	0.996263066	1	0.999150493	MAE	8.7985E-05	8.8315E-05	0.00008824
Normalised R^2	-0.447650885	-1	-0.815629619	R^2	-0.018717921	-0.041813658	-0.034104458
Angle							
Metric	ANN	RBF	Kriging	Metric	ANN	RBF	Kriging
Normalised MSE	0.979166991	1	0.993037625	MSE	235.9936508	241.0147126	239.3366776
Normalised MAE	0.995325587	1	0.997104314	MAE	12.24873664	12.30626119	12.27062612
Normalised R^2	-0.12285531	-1	-0.951230581	R^2	-0.002557827	-0.020819829	-0.019804458

Table 3.5: Comparison between different surrogate models, having immediately available information as input.

Thickness							
Metric	ANN	RBF	Kriging	Metric	ANN	RBF	Kriging
Normalised MSE	0.442640401	1	0.871958122	MSE	2.18705E-09	4.94092E-09	4.30828E-09
Normalised MAE	0.859654168	1	0.758306354	MAE	3.46379E-05	4.02929E-05	3.05543E-05
Normalised R^2	1	0.353989766	0.885485857	R^2	0.666646998	0.235986215	0.590306489
Angle							
Metric	ANN	RBF	Kriging	Metric	ANN	RBF	Kriging
Normalised MSE	0.046063352	1	0.099403901	MSE	0.589207243	12.79123687	1.271498847
Normalised MAE	0.412076756	1	0.451206049	MAE	0.60051025	1.457277657	0.657532494
Normalised R^2	1	0.94281626	0.995782093	R^2	0.996375076	0.939398622	0.992172458

function scalar d_0 defined in Equation 2.3, and the Kriging hyperparameters θ_k are a number of weights that are reduced during the projection onto a smaller space which was shown in Equation 2.7. It is worth noting that this process is considerably faster than the optimisation of the ANN architecture.

The case where only the CPs were provided as inputs, was examined first. The comparison was based on three popular metrics, in particular the MSE, MAE, and the statistical R^2. The results are presented in Table 3.4, with the optimal results having a green colour and the worst ones being red.

It is clear that the performance for all models is nowhere near satisfactory, with very large MSE and MAE values, and even negative values for R^2. The advantage of having only the CPs as inputs is the swift training of all the models, but this is not warranted at such an expense in accuracy. The next comparison is the one with additional inputs that can be generated just by having the angle coefficients (e.g. angle variation, course width, etc.). This additional information can be acquired almost immediately with insignificant computation time, although the total number of inputs is orders of magnitude higher. The amount of inputs depends on the discretisation of the curve, and for this study the input number was equal to 305. The comparison with the additional data is presented in Table 3.5.

It is observed that the picture is completely different now, with the overall accuracy being significantly better. It can be said that the prediction of the angle variation is satisfactory in all models, although the thickness is not predicted as accurately, with an R^2 difference of about 33% in the best case. The superior accuracy for the angle variation was to be expected, as the weighted average method that calculates the element angles does not deviate too much from the angle variation of the initially obtained curve which serves as an input. On the other hand, the element thickness assignment is a more complicated process, which is not as periodic along a row/column of elements as the angle variation. It is speculated that by increasing the number of el-

Table 3.6: Comparison between different surrogate models, having computationally expensive information as input.

Thickness							
Metric	ANN	RBF	Kriging	Metric	ANN	RBF	Kriging
Normalised MSE	0.329827963	1	0.71729582	MSE	1.63545E-09	4.95849E-09	3.5567E-09
Normalised MAE	0.713515251	1	0.599083418	MAE	2.86343E-05	4.01313E-05	2.4042E-05
Normalised R^2	1	0.316627507	0.905073042	R^2	0.754383141	0.238858454	0.682771844
Angle							
Metric	ANN	RBF	Kriging	Metric	ANN	RBF	Kriging
Normalised MSE	0.034716873	1	0.101700564	MSE	0.447271302	12.88339845	1.310248892
Normalised MAE	0.354906806	1	0.461647757	MAE	0.508770735	1.433533326	0.661787444
Normalised R^2	1	0.942500311	0.994754326	R^2	0.997248507	0.939907028	0.992017266

ements, the thickness assignment to each element would become clearer, although this would considerably increase the computational time needed.

The ANN provided the best accuracy across each metric, apart from the MAE metric in the thickness variation prediction, where the Kriging model was superior. In general, the Kriging model had very similar results to the ANN and it was the second to best model. The Gaussian RBF proved to be the least reliable, especially in the case of the thickness variation.

In the final comparison of the surrogate models, the computationally expensive information of the number of course shifts and course extension, as obtained by the Hooke-Jeeves optimisation, was included as input. This amounts to a total of 308 inputs for this study. The results are shown in Table 3.6.

The general trend is similar as in the intermediate case, although the accuracy has improved all around. However, the prediction for the thickness variation is still not optimal, but this is also subject to the number of samples that were used to train the models. Seeing as the improvement in accuracy from the latest additional data is not significant, it was decided to carry on with intermediate scheme, which does not include the computationally expensive inputs.

3.5. Cylinder optimisation

In this section, the set-up for the optimisation of the cylinder for both the CS and VS methods is explained. The design variables of the problem are defined, as well as the optimisation algorithm that was chosen to solve it. The approach to handling the hyperparameters, convergence and termination criteria, and incorporation of the surrogate model, among others, are detailed in the coming paragraphs.

3.5.1. Optimisation set-up for the constant stiffness cylinder

In order to make a fair comparison between the CS and VS cylinders, the former cylinder should be optimised first to serve as a comparison reference. The type of the investigated problem is considered as a many-objective optimisation problem, as it contains more than 4 objectives. Conventional multi-objective optimisation algorithms struggle to provide diverse optimal solutions for high-dimensional problems such as this. Therefore, the novel UNSGA III algorithm was considered as the best choice in this work due to its advantages in higher dimensions, as explained in Section 2.2.8. Since the reference directions for the algorithm are not known beforehand, the Das-Dennis method was used to sample the reference directions in a uniform manner throughout the design space.

In the interest of time, it was decided that the layup would be forced to be symmetric, reducing the number of design variables by half, which amounts to 4 different angle design variables, one for each ply. In case the ANN cannot achieve a satisfactory accuracy, a total allowable number of samples was set as an upper limit, in order to allow the surrogate model to carry out the optimisation with the accuracy that has obtained from the samples gathered up to that point. This decision was made for the sake of the project's completion within the allocated time frame.

The UNSGA III algorithm has its own hyperparameters that need to be selected and finding their optimal

values can prove difficult. Therefore, the selection of these values was based on what is reported to be successful in most cases in literature. For the crossover operation, the SBX method was used, since it applies in both real-valued and integer-valued variables, and it is popular in literature. The crossover probability from two parents was set to 0.5 and the η_c distribution index was given a value of 10.0. As for the mutation operation, a polynomial mutation similar to the crossover was chosen, with a value of 20.0 for η_c. The initial sampling was performed via the LHS method and the selection process utilised a tournament selection. Additionally, a termination criterion that accounts for changes in both the design and objective space was implemented, where the optimisation stops if the average change is less than 0.5° or less than 1°, respectively, in a window of the last 30 generations. The maximum number of allowed generations was selected based on a practical time frame as not to exceed the allocated time for the completion of this thesis.

The optimisation problem for the CS cylinder is then defined as:

$$\min(\lambda, \omega_n, K_a, K_l, P_{FPF}) \tag{3.13}$$

where λ is the critical buckling eigenvalue of the 1^{st} buckling mode, ω_n is the natural circular frequency of the 1^{st} free vibrations mode, K_a is the axial stiffness, K_l is the lateral stiffness, P_{FPF} is the FPF load according to the maximum strain criterion.

3.5.2. Optimisation set-up for the variable stiffness cylinder

The VS cylinder optimisation utilised the UNSGA III algorithm just as in the case of the CS cylinder, considering that the number of objectives does not change between the two. Specifically for the instance of VS, two different optimisation runs are required for the axial and circumferential angle variations, and their respective subsets. The subsets for each type are the number of CPs that formulate the tow trajectory. However, this would mean that 12 unique optimisation runs would have to be performed in total, for the pristine cylinder and the one containing cut-outs, which was deemed infeasible time-wise. Therefore, a compromise was made, where the same amount of optimisation runs would be performed, although some in a different manner. More specifically, some optimisations would firstly run based on FEM analyses and afterwards on ancillary surrogate models once sufficient data has been obtained to train them.

The compromise for the sake of time is made by having the same surrogate model run a different optimisation case entirely from start to finish. For example, an optimisation for the case of axial angle variation with 2 CPs is performed and a surrogate model is trained to carry on with this optimisation. The same model would then be utilised to run an entire optimisation for the cases of 3 and 4 CPs. The drawback of this approach is that the surrogate model is expected to be very precise and resemble the FE analyses for the aforementioned cases, without necessarily having been trained on representative data. Therefore, some samples from the other sub-cases with a different number of CPs were obtained via the LHS method, amounting to 10% of the total allowable number of samples from the primary source.

When choosing the sub-case that should provide the data for the surrogate model, the right choice is not obvious. For instance, if the data is obtained from a sub-case with 2 CPs, the drawback of the data not being representative for the rest is increased due to the limited design space, although the viable designs that do not self-intersect and do not violate the curvature constraint are increased for a fixed amount of samples. On the other hand, the opposite is true for a sub-case with 4 CPs. Therefore, it was arbitrarily decided that the pristine cylinder would acquire its data from 2 CPs and the cylinder containing cut-outs would obtain the data from a sub-case of 4 CPs, so that a comparison can be performed in the end to determine the best approach. In that way, by examining the results from both AV and CV, a higher certainty decision on the winner of the comparison can be made.

The design variables differ between the AV and CV, in order to account for the fact that the AV case can only have angle coefficients of similar mathematical sign. Thus, the design variables are:

- $x_{i*j} \in (0, 90)$ for the axial angle variation, plus an additional $x_j \in [\![0, 1]\!]$.
- $x_{i*j} \in (-90, 90)$ for the circumferential angle variation.

In this instance, x stands for the individual of the population, i is the number of CPs and j is the number of symmetric plies in the laminate. The additional integer set for the axial angle variation case determines whether the angle coefficients will be negative (null value) or positive (unity value). This scheme is better illustrated in Figure 3.19, where each coloured bracket indicates the set of angle coefficients for a ply.

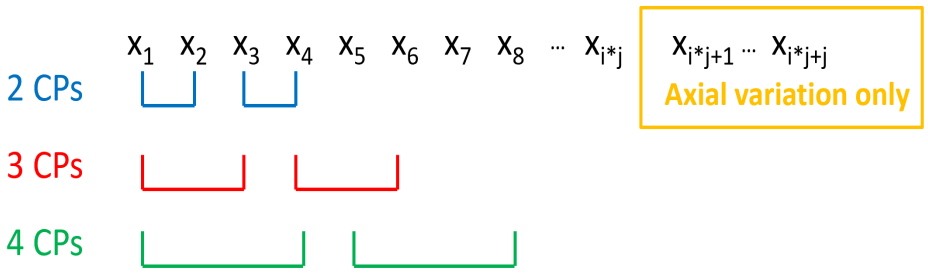


Figure 3.19: Design variables in the optimisation of the VS cylinder.

The same hyperparameter set-up that was described for the CS case regarding the UNSGA III was also implemented for the VS cylinder. The curvature of the course was also constrained, with respect to the manufacturing curvature limit. It was also decided to limit the mass of the cylinder, as the mass is one of the most important objective in practical design, and the CS and VS designs can be compared on more equal grounds. For that reason, the VS mass was not allowed to surpass that of the CS cylinder for a relative change of more than 50%.

The optimisation is then defined as:

$$\begin{aligned} &\min\left(\lambda, \omega_n, K_a, K_l, P_{FPF}, M\right) \\ &s.t.\left(\kappa_{max} \leq \kappa_{limit}, M \leq 1.5\, M_{CS}\right) \end{aligned} \tag{3.14}$$

where M and M_{CS} are the structural masses of the VS and CS cylinder respectively, κ_{max} is the maximum curvature of a course, with κ_{limit} being the manufacturing curvature limit.

3.5.3. Surrogate modelling for the finite element method

The surrogate model that was used for the FEM analyses is very similar to the ones used for predicting the element properties. Different networks were trained for each objective in order to optimise the performance of each network to the respective objective, by using data from each analysis. The training of the surrogate model commences every 400 collected samples and if all objectives have an accuracy above 85% for two consecutive times, then no further FEM analyses are run and the networks will be making predictions instead. If this condition is not met by the 7^{th} training iteration, the networks will carry out the optimisation anyway, in the interest of time. The input for the network is the element properties of each ply of the laminate, which quadruples the amount of inputs with regards to the element properties surrogate model. The thickness variation input is first altered by staggering the plies, since the staggering operation is a very fast process and drastically improves the performance of the ANN.

An ANN model was also used for prediction making for the case of the CS cylinder, in the same manner as with the VS cylinder. The only difference was the input of the network, as the angles of each ply of the laminate were only used, which amounts to a difference in the thousands with respect to the VS model input.

The optimisation scheme was selected to be the DE algorithm for this surrogate model as well. The same hyperparameters were used, since there are no clear distinctions between the two surrogate models.

3.6. Finite element model

The FEM software that was used in this work is *Abaqus FEA 2019* by *Dassault Systemes*. The software provides a variety of solvers, eigenvalue extraction algorithms and meshing techniques which are necessary for the calculation of the optimisation objectives for the cylinder.

3.6.1. Analysis procedure

Buckling analysis

The buckling performance of the cylinder is assessed via a linear buckling eigen-analysis. Generally speaking, structures with stiff responses are better approximated by such an analysis or, at best, it provides an upper bound. Cylinders are very imperfection-sensitive however, so eigen-analyses do not necessarily predict the critical load or the corresponding buckling mode without including the initial geometric imperfections. Eigen-analyses can still be used in the context of a comparison and they serve as an excellent metric to measure the buckling response of the structure. However, it is noted that the VS cylinders tend to be less imperfection-sensitive which is considered a significant advantage, but this aspect was not considered in the comparison between the VS and CS designs, since only the critical eigenvalues of the first modes were compared. The eigenvalue buckling problem can be formulated as such:

$$\left(K_0^{NM} + \lambda_i K_\Delta^{NM}\right) u_i^M = 0 \tag{3.15}$$

where N and M are the DoF of the entire model, i is the number of the buckling mode, K_0^{NM} is the stiffness matrix, K_Δ^{NM} is the stress stiffness matrix, λ_i is the eigenvalue of the i^{th} mode, and u_i is the eigenvector of the i^{th} mode. The stress stiffness matrix is calculated via a perturbation method, with an incremental load Q^N. If no pre-loading is introduced in the structure, the critical buckling load corresponds to the product of the i^{th} eigenvalue with the incremental load, $\lambda_i Q^N$ [151].

Two different methods are available for buckling eigenvalue extraction in Abaqus, the *subspace* and *Lanczos* methods. The workings of these methods are not explained here, as the algorithms are relatively long, and their working principles can be found in numerical analysis textbooks or the Abaqus theory manual [152, 153]. The software user manual recommends the subspace method if less than 20 eigenvalues are requested, as it is likely to be the faster extraction process. For that reason, the subspace method was chosen in this work. The default number of subspace vectors and iterations as calculated by Abaqus for the given cylinder design was selected. In case a higher amount was needed for certain designs, the values were increased in an incremental manner until convergence.

Free vibrations analysis

In typical design, the first mode of vibration is the most critical one and drives the design. Therefore, only the natural frequencies were of concern for the conducted comparison. The natural frequencies of the various designs can be obtained via an eigen-analysis. In the case where no damping is present, the eigenvalue problem is given as:

$$\left(K^{NM} - \omega^2 M^{NM}\right) \phi_i^M = 0 \tag{3.16}$$

where M^{NM} is the mass matrix, K^{NM} is the stiffness matrix, N and M are the model's DoF, ω is the circular natural frequency, and ϕ is the vibration mode. There are three different eigenvalue extraction processes available in Abaqus, namely the subspace, Lanczos, and *Automatic multi-level sub-structuring* methods. The latter approach is best suited for very large models, where sub-structuring is favoured if the response of a specific region is required. In this case, the Lanczos eigensolver was selected, as it is recommended by the Abaqus user manual on the grounds of it being the most robust. The recommended values for the block size and the maximum number of block steps for the first frequency mode were also used, which are 1 and 80, respectively.

Stiffness analysis

When it comes to stiffness requirements, constraints on axial, lateral, and torsional stiffness typically need to be met. Closed cylinders exhibit a high torsional stiffness and hence this is not driving the design. For that reason, the axial and lateral stiffnesses are the optimisation objectives. The axial stiffness is seen as the simplest objective to optimise, opposed to the lateral stiffness which has both bending and shear stiffness contributions. The axial stiffness can be simply determined by the ratio of an axial load to the axial deformation P_z/u_z, and the lateral stiffness of a certain radial direction is similarly obtained by dividing the radial load by the corresponding radial deformation P_ϕ/u_ϕ. Thus, the deformations were obtained by separate linear static analyses.

First Ply Failure analysis

Last but not least, the strength of the cylinder is another important objective that usually drives the design. Depending on the philosophy of the design, different approaches can be taken when accounting for the structure's strength. When it comes to composite structures, failure is not easily determined accurately, and 3-D ply or laminate criteria might be more precise. In this case, a conservative FPF approach is followed as it will provide the quickest and simplest form of strength comparison. The non-interactive strain criterion was selected for that purpose. For the sake of simplicity, other effects that influence the strength (e.g. stacking sequence), are not taken into account in the form of design constraints, like the 10% rule, which states that at least 10% of the fibres should be oriented towards each of the four principal directions: 0°, 45°, −45°, 90°, so that unforeseen secondary load cases can be withstood [123]. The FPF load has a better accuracy for pristine laminates, and as such the introduction of cut-outs is expected to increase its inaccuracy. Nevertheless, the optimisation of this objective serves as a good indicator to perform a strength comparison. The strain criterion is given as follows:

$$\max\left(\frac{\varepsilon_1}{\varepsilon_{1T}}, \left|\frac{\varepsilon_1}{\varepsilon_{1C}}\right|, \frac{\varepsilon_2}{\varepsilon_{2T}}, \left|\frac{\varepsilon_2}{\varepsilon_{2C}}\right|, \frac{\varepsilon_{12}}{\gamma_{12}}\right) \tag{3.17}$$

where ε_1 and ε_2 are perpendicular in-plane strains along the directions of 1 and 2, ε_{12} is the shear strain, T and C refer to the critical tensile and compressive strains, and γ_{12} is the critical shear strain. These failure indices signify failure if their ratio equals unity. The calculation of these indices is performed by Abaqus for the sake of expediency, only for the case of CS cylinders. Subsequently, the maximum index is scaled by the reciprocal of the initially applied unit load to obtain the FPF load of the structure. For the case of VS cylinders, the principal strains of the top-most ply are transformed to a global coordinate system according to Equation 3.18, where it is assumed that each ply has identical strains due to strain compatibility between the plies.

$$\begin{aligned}
\varepsilon'_x &= \frac{\varepsilon_x + \varepsilon_y}{2} - \frac{\varepsilon_x - \varepsilon_y}{2}\cos(2\theta) - \frac{\varepsilon_{xy}}{2}\sin(2\theta) \\
\varepsilon'_y &= \frac{\varepsilon_x + \varepsilon_y}{2} + \frac{\varepsilon_x - \varepsilon_y}{2}\cos(2\theta) + \frac{\varepsilon_{xy}}{2}\sin(2\theta) \\
\varepsilon'_{xy} &= \left|\varepsilon_x - \varepsilon_y\right|\sin(2\theta) - \varepsilon_{xy}\sin(2\theta)
\end{aligned} \tag{3.18}$$

The reason for this discrepancy is that the VS model is built up by many laminates, since each element has a laminate of its own. Abaqus readily provides FPF results for the top-most and bottom-most plies but further results for the plies in-between would require as many analyses as there are elements, rendering the approach impractical.

3.6.2. Element type

The primary criterion for the selection of the element type was to achieve convergence with as few elements as possible, since the pre-processing of the element properties can become very computationally expensive with an increasing number of elements. To that end, the quadratic doubly-curved rectangular Abaqus element S8R was chosen for every analysis case. The element contains 8 nodes having 6 DoF each, employs reduced integration, utilises the shear-flexible Mindlin theory,and calculates strains which are approximate to those of the Koiters-Sanders shell theory. A limitation of this element is that it requires a regular mesh to achieve a fast convergence, as shear transverse locking can occur otherwise, which is a point of concern for the irregular mesh of the cut-out cylinder. Figure 3.20 presents representative meshes of the cylinder with and without cut-outs.

3.6.3. Boundary conditions

The model of the cylinder has a coupling constraint at its ends, coupling the nodes of the circumference of each side with a reference point, respectively. This coupling was enforced so the cross-section of the cylinder would remain circular, simplifying the analysis.

The boundary conditions were thus imposed at the reference points. The top side of the cylinder was fixed, meaning that all DoF movement was restricted. On the other hand, the cylinder was free to move on the bottom side, where the loading was applied. The free vibrations analysis was load-free. In all other cases, the load had a magnitude of unity.

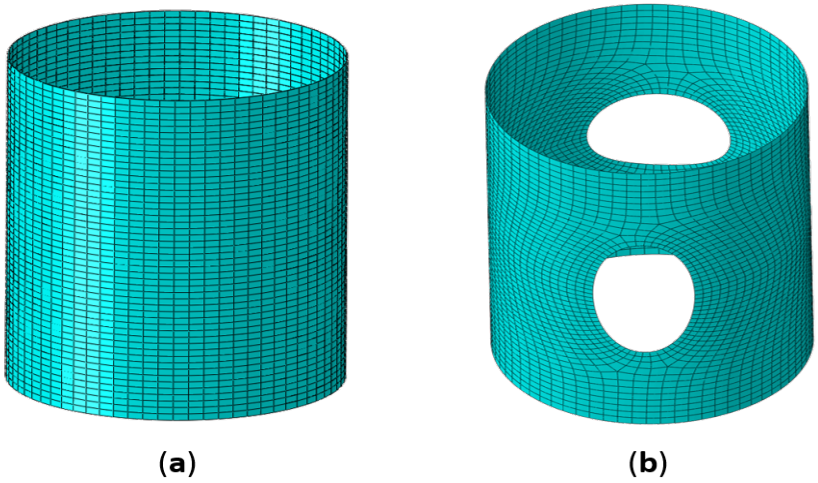


Figure 3.20: Representative cylinder mesh for the (a) pristine and (b) cut-out cylinders.

In the analysis cases of buckling, axial stiffness and axial strength determination, an axial compressive load was applied at the bottom reference point. Although the cylinder may be buckling-critical or strength-critical in other directions, the axial direction is often-times the crucial one in aerospace design.

The lateral stiffness is determined by applying a lateral load along an arbitrary radial direction for the pristine cylinder. This approach can yield the best lateral stiffness design for the case of a pristine CS cylinder due to symmetry, although the loading direction affects the results in any other case. If the design that provides the best lateral stiffness all around the circumference is desired, then multiple lateral load analyses could be run in a clock-wise manner, with the best lateral stiffness design being determined by the highest mean value of these analyses. However, that would require significant time and is deemed infeasible within the time frame of this study. Therefore, the random direction of the lateral load will yield optimal designs only if the structure is loaded along this specific axis. In the case of a cylinder with cut-outs, the lateral load is pointed in-between the two cut-outs that are present on the cylinder, as not to bias the results.

The boundary conditions of the cylinder are shown in Figure 3.21 for the pristine case, although the same conditions apply for the cut-out cylinder.

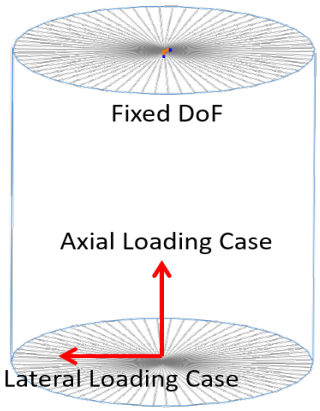


Figure 3.21: Boundary conditions of cylinder model.

3.7. Process flowchart

A simplified flowchart depicting the processes required to perform the optimisation of a VS cylinder is shown in Figure 3.22. The flowchart was condensed to convey the general ideas behind the implemented methodology in a concise manner but it should not be followed to the letter if the results are to be reproduced.

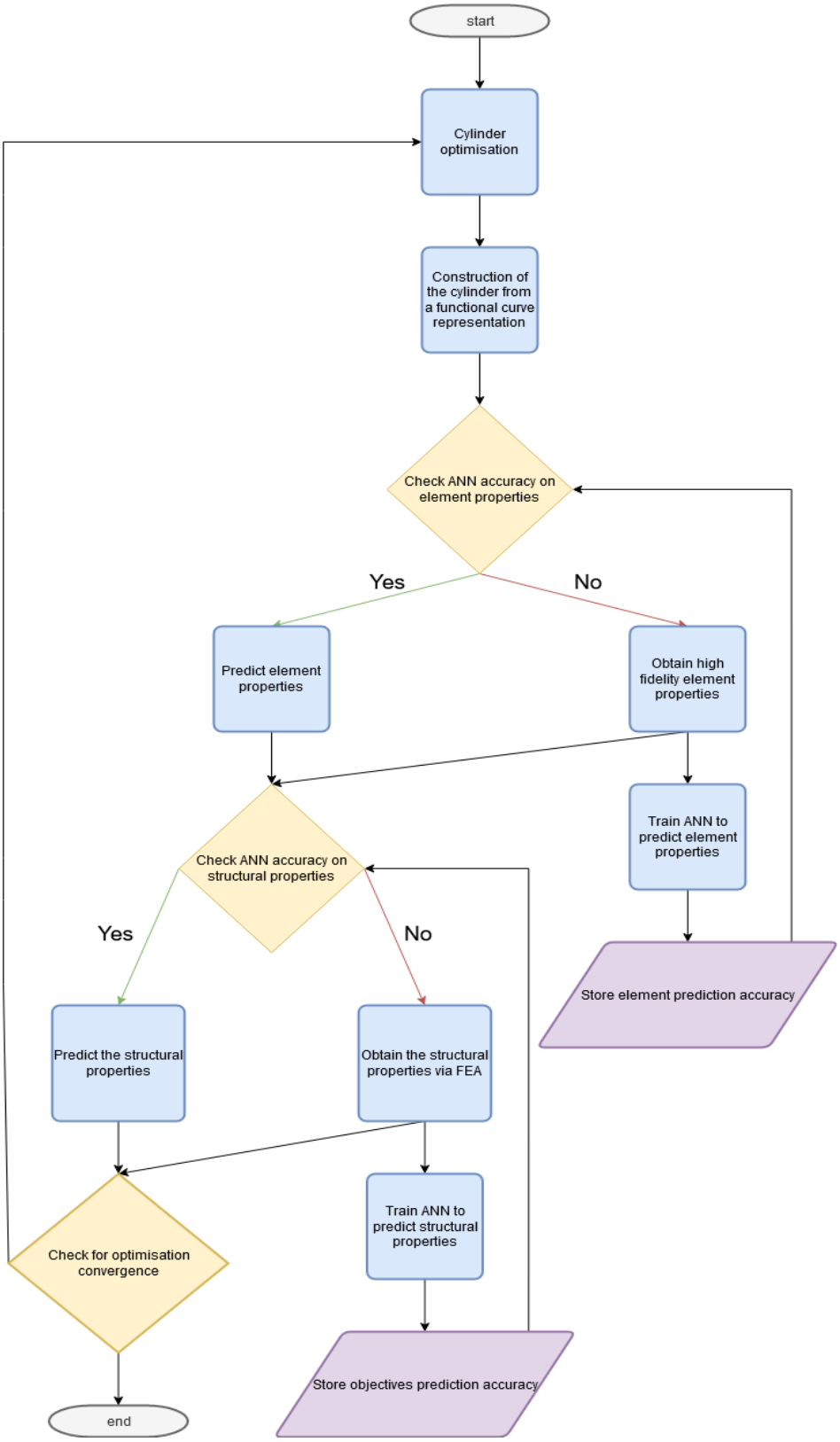


Figure 3.22: A flowchart of the processes needed for the cylinder optimisation.

4

Optimisation results

The optimisation results for the pristine cylinder and the cylinder with cut-outs are presented in this section. The outcomes of the comparison between the CS and VS cylinders are also discussed subsequently.

4.1. Pristine cylinder optimisation

The goal of the optimisation is to explore the objective space and identify the Pareto optimal solutions. In view of the fact that the high dimensionality of the problem would require a considerate amount of time to determine the complete Pareto front, the optimisation was terminated based on the satisfaction of the termination criterion mentioned in Section 3.5.2.

4.1.1. Constant stiffness cylinder

The CS cylinder optimisation was performed first to act as the baseline design. It is mentioned again that the convergence of the optimisation can be determined by a convergence metric, such as the hypervolume calculation in multi-objective optimisation problems. Therefore, an arbitrary reference point that allows for the creation of a geometrical hypervolume was selected. This leads to an arbitrary hypervolume scalar number, which by itself does not provide much information. The actual change of the hypervolume with each generation of the optimisation algorithm is of interest instead. This change is observed in Figure 4.1, where the normalised hypervolume (in blue) is plotted against the corresponding generation of the optimisation algorithm for the true data, as well as the ones predicted by the ANN. It is observed that the algorithm did not converge as is evident by the reduction of the hypervolume after having peaked in earlier generations. Therefore, the ANN was trained after a set number of generations to perform predictions and continue the optimisation process, as was mentioned in Section 3.5.3, in case convergence is not achieved.

The normalised values of the optimisation objectives can also be seen in 4.1 to provide additional information as to how they affect the hypervolume. Most of the objectives obtain good values in the first generations, however a reduction is observed later on for the majority. It is noted that the FPF load objective attained greater values only when the rest of the objectives were reduced.

In order to properly assess the accuracy of the surrogate model, all of the predictions would have to be compared to data generated by FEM. These analyses would be too time consuming to perform however, which defeats the goal of surrogate modelling. Attention on the extremities of the design space was given instead, where analyses were run only for the best results of each objective, since they are more likely to exhibit a greater inaccuracy. The results are given in Table 4.1, where the predictions seem to be in excellent agreement with the actual FEM analyses.

The results of the QI laminate $[0/90/45/-45]_S$ were also included as a reference to showcase the relative difference between the optimised extremities. Most objective values had a significant increase with respect to the QI laminate, after certain optimised designs were found by the optimisation algorithm. The most notable increases are observed for the axial stiffness (61.5%) and FPF load (77.9%) objectives, which require the fibres to be oriented towards the axis of the cylinder. On the contrary, the frequency of the free vibrations increased only by 3.2% for the corresponding optimised design. Lastly, the designs for the buckling and lateral stiffness objectives achieved intermediate values of 13.9% and 17.8%, respectively.

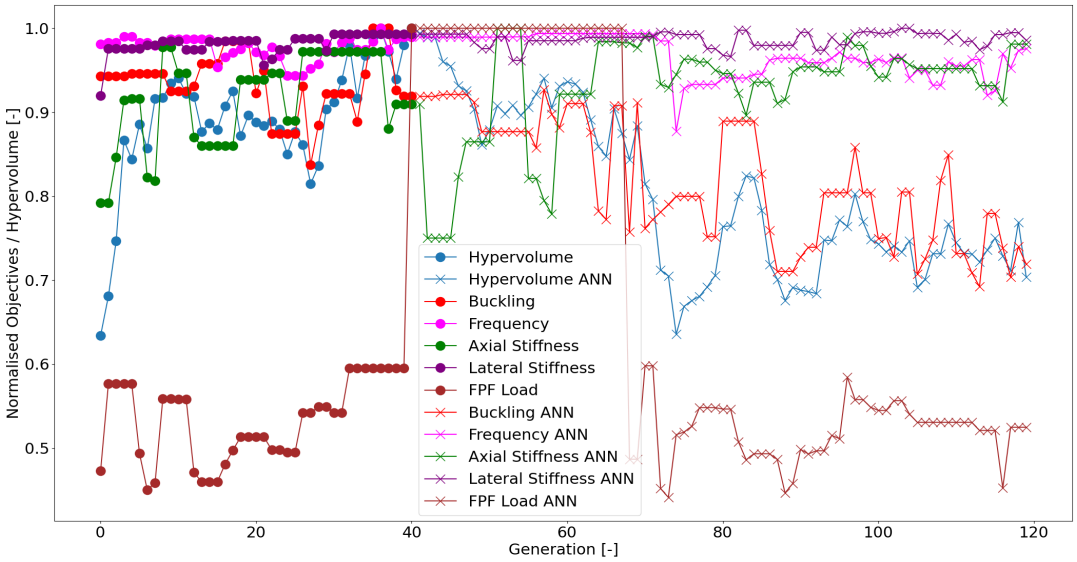


Figure 4.1: Hypervolume convergence of the CS cylinder.

Table 4.1: Comparison between predictions and analyses values for the best results of each objective.

Optimal values	QI reference	Optimised FEA Designs	Relative difference between QI and optimised designs [%]	ANN prediction	Relative difference between FEM and ANN [%]
Buckling [kN]	352.68	410	13.9	409	0.21
Frequency [Hz]	335	346	3.2	346	0
Axial Stiffness [MN/m]	210	546	61.5	544	0.48
Lateral Stiffness [MN/m]	24	29	17.8	29	0
First Ply Failure [kN]	131	592	77.9	592	0

The optimal architecture of the ANNs varied for each objective as is seen in Figure 4.2. The figure shows the inputs to the network, the number and type of layers and the corresponding amount of weights and biases. Regarding the bias vectors, they were not computed at all when batch normalisation layers were present as they would not affect the accuracy and would only increase the computational time required.

It is observed that in all cases but one, the SELU activation function with its self-normalising properties was preferred. It is noted that the batch normalisation layers do not constitute an additional HL with respect to the upper bound of HLs given in the optimisation algorithm. Another remark not shown in the figure, is the lower learning rate that was required for the Nadam algorithm to converge in the case of the SELU layers. The resulting architectures are consistent with the literature that the SELU activation function with LeCun initialisation and a higher number of HLs generally outperforms ELU with batch normalisation layers and He initialisation. It is surprising however that the lateral stiffness objective preferred a relatively deep network with the latter architecture. It is also not clear what the connection between the complexity of the prediction of an objective and its ANN architecture is. Of course, these architectures could change given more optimisation time.

A representative graph of the ANN training process is depicted in Figure 4.3, where the epochs are plotted against the training and validation MSE metric. A decaying learning rate was also implemented in order to speed up the training by measuring the validation error every 15 steps and reduce the learning rate by a factor of 0.45 when the error stops dropping. These hyperparameters were found to work best for most cases via a parametric study and hence they were selected as fixed values. The plot shows that both the training and validation loss metrics are minimised by the Nadam algorithm with increasing epoch values. No overfitting is observed and thus, no regularisation was necessitated.

The ANN was also compared with the RBF and Kriging surrogate models and the results are presented in Table 4.2. The comparison shows that all of the models performed exceptionally well at predicting each optimisation objective, boasting R^2 values above 90%. It is interesting that all of the models performed best in at least one objective. It seems that the objectives that had a higher accuracy for all the surrogate models, yielded either the RBF or Kriging models as their best performers, while the ANN model was superior for the

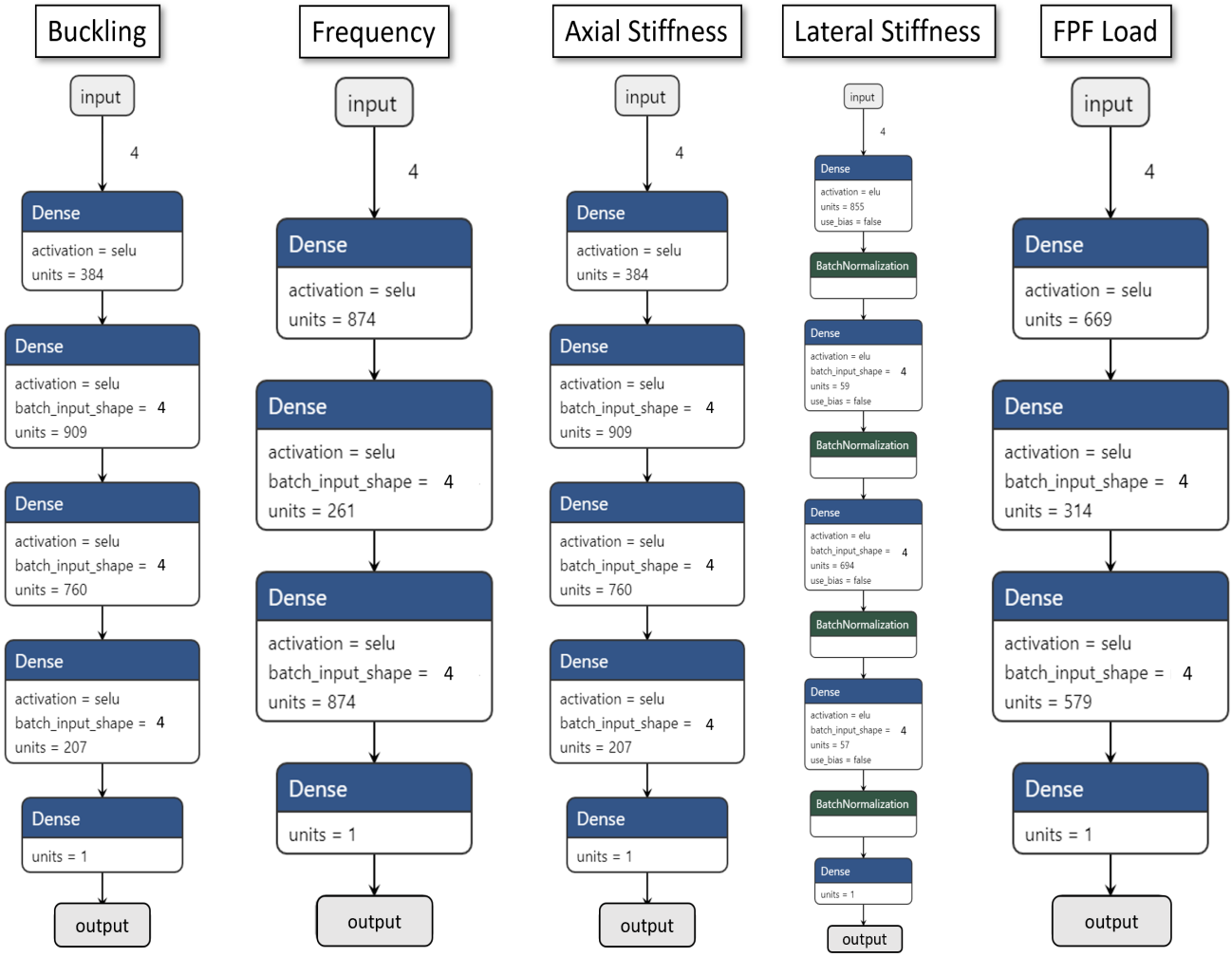


Figure 4.2: Neural network optimised architectures for each objective.

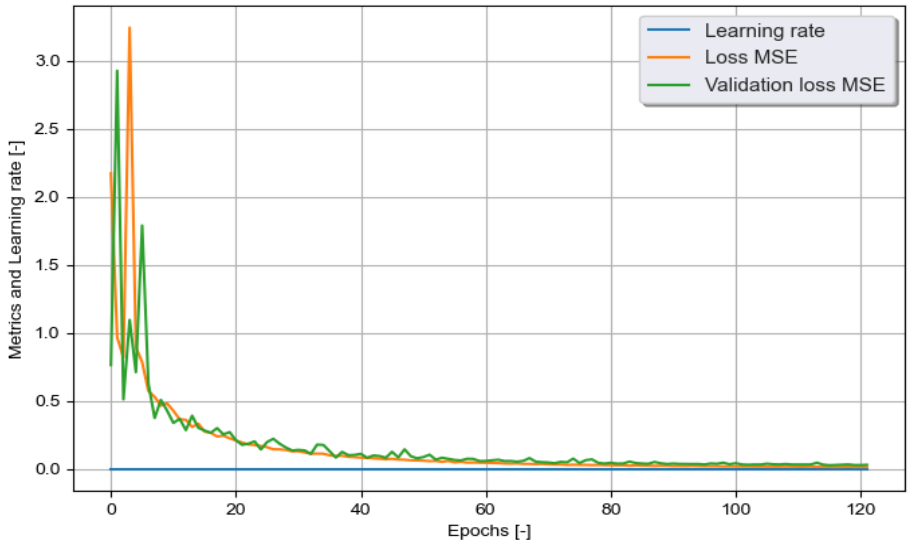


Figure 4.3: Representative minimisation of the ANN cost function.

Table 4.2: Comparison between different surrogate models on the accuracy of the predictions of the objectives.

Buckling							
Metric	ANN	RBF	Kriging	Metric	ANN	RBF	Kriging
Normalised MSE	0.57	1.00	0.66	MSE	1.12E+08	1.98E+08	1.31E+08
Normalised MAE	0.93	1.00	0.76	MAE	7513.38	8096.00	6157.01
Normalised R^2	1.00	0.96	0.99	R^2	0.95	0.91	0.94
Frequency							
Metric	ANN	RBF	Kriging	Metric	ANN	RBF	Kriging
Normalised MSE	0.69	0.69	1.00	MSE	15.67	15.52	22.60
Normalised MAE	1.00	0.72	0.78	MAE	2.75	1.98	2.14
Normalised R^2	1.00	1.00	0.99	R^2	0.98	0.98	0.98
Axial Stiffness							
Metric	ANN	RBF	Kriging	Metric	ANN	RBF	Kriging
Normalised MSE	1.00	0.73	0.19	MSE	1.16E+12	8.47E+11	2.25E+11
Normalised MAE	1.00	0.54	0.20	MAE	7.72E+05	4.16E+05	1.53E+05
Normalised R^2	0.97	0.98	1.00	R^2	0.97	0.98	0.99
Lateral Stiffness							
Metric	ANN	RBF	Kriging	Metric	ANN	RBF	Kriging
Normalised MSE	1.00	0.97	0.80	MSE	3.14E+14	3.04E+14	2.52E+14
Normalised MAE	1.00	0.79	0.67	MAE	1.40E+07	1.10E+07	9.39E+06
Normalised R^2	0.99	0.99	1.00	R^2	0.95	0.95	0.96
FPF load							
Metric	ANN	RBF	Kriging	Metric	ANN	RBF	Kriging
Normalised MSE	0.31	0.69	1.00	MSE	1.35E+08	3.02E+08	4.36E+08
Normalised MAE	0.58	0.69	1.00	MAE	7294.15	8676.71	12637.84
Normalised R^2	1.00	0.98	0.97	R^2	0.99	0.97	0.95

objectives that were more difficult to predict.

Given the fact that the RBF and Kriging models require significantly less time to train for such a small number of inputs (one input per symmetric ply, 4 in total), they are certainly more than viable candidate surrogate models for this task. It is possible that a specific ANN architecture could outperform them but that would likely require a significant amount of time spent on architecture optimisation to ascertain. A comparison between the computational times required to train the surrogate models and for prediction-making can be seen in Figure 4.4.

Two different ANN architectures were chosen to represent the extremes (shallower and deeper nets), chosen from the architectures found in Figure 4.2. Enough iterations were performed to gather representative data, with the error bars being the standard deviation found in the data. It is noted that the training time is displayed in a logarithmic scale. It is clear that the RBF model achieves the fastest training with respect to the other surrogate models by a significant margin. On the other hand, the Kriging model took the longest to train, while the ANNs had a similar speed. When it comes to the prediction time, the RBF performs the best once again, with the Kriging model as a close second. The ANNs take the longest to perform a prediction, with the deeper net performing the worst. Lastly, it is mentioned that the comparison is not entirely fair, since the CPU was utilised for the RBF and Kriging models and the GPU for the neural networks. Nevertheless, the utilisation of the GPU for neural networks has been made effortless with most implementations of the process found in various software, and this is the reason why the comparison was performed in such a way.

A very important aspect of surrogate modelling is the time which can be saved by making approximations. Thus, the total time that is required for the optimisation to complete with and without the use of surrogate models is shown in Table 4.3. It can be observed that the buckling objective requires the most amount of time with respect to the other objectives. It is also noticed that the difference between the time needed to complete the optimisation with the utilisation of surrogate models is many orders of magnitude smaller than the optimisation with only the FEM analyses. The total amount of time that is saved is therefore 8.6 days. The time that was taken to obtain the data was included in this calculation, however the training time of the neural networks was not.

When it comes to the structural objectives optimisation, the optimisation algorithm identified the results that are likely to lie on the Pareto front. These results are all considered good and non-dominated, requiring a decision based on a case by case basis as to which solution should be preferred. Nevertheless, special

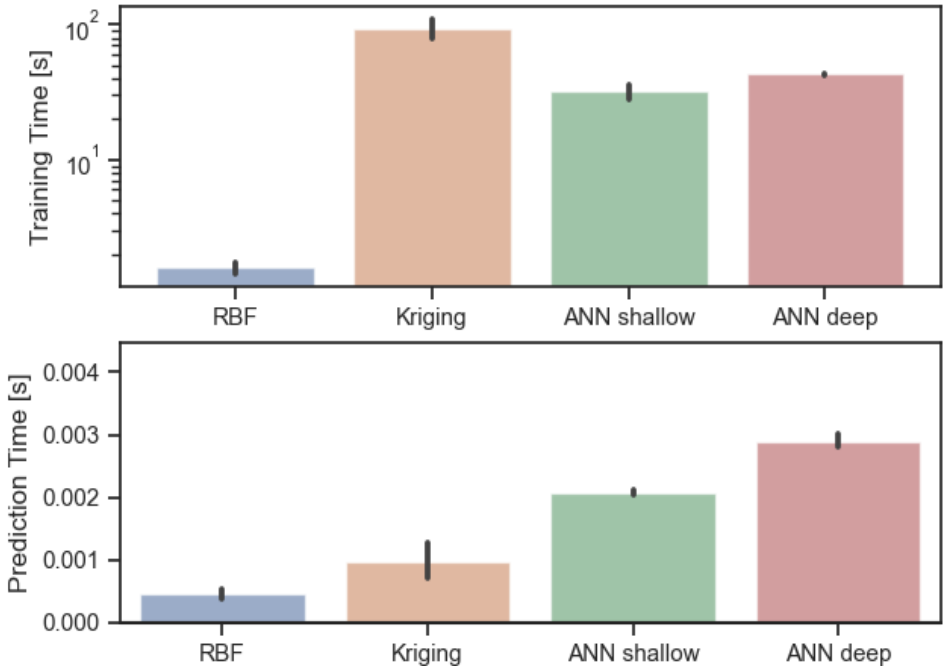


Figure 4.4: Comparison of the computational times for the training and value prediction between different surrogate models.

Table 4.3: A comparison of the total time required for the optimisation to complete with and without the use of surrogate models in the case of constant stiffness cylinders.

Structural Objective	Average FEA Time [s]	Optimisation Time Required w/o Surrogate Model [s]	Optimisation Time Required with Surrogate Model [s]	Time Saved [days]
Buckling	65.0	311808	14.4	3.6
Frequency	28.5	136704	14.4	1.6
Axial Stiffness/FPF Load	32.5	156096	14.4	1.8
Lateral Stiffness	28.5	136656	14.4	1.6
Total	154.4	741264	57.6	8.6

attention is given to the results that contain the highest value for an objective, and are called hereafter as optimal designs.

In regards to the relationship between the objectives, a very insightful way to visualise the results is to plot each objective with one another, as shown in Figure 4.5.

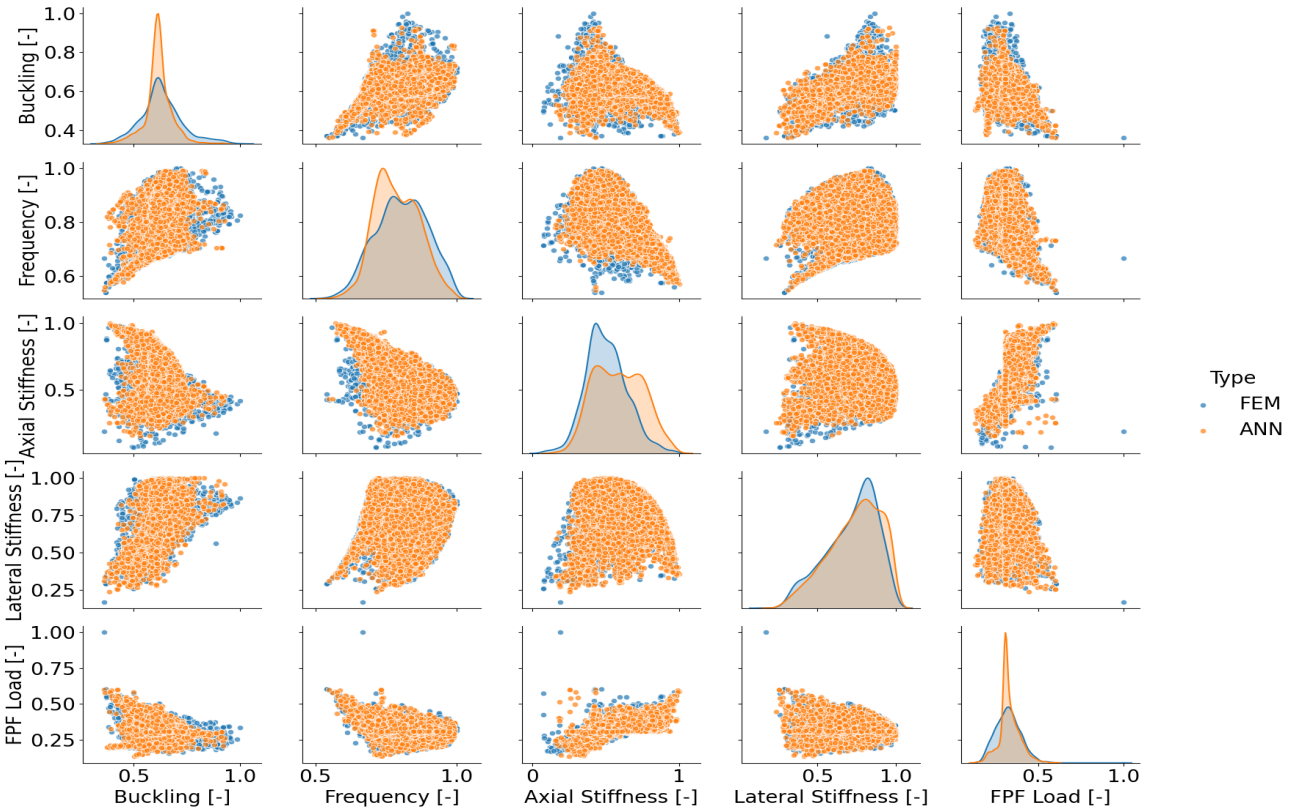


Figure 4.5: Scatter plot of pairs of objectives.

The diagonal represents the density estimation of the objective with respect to the others. The FEM results and ANN predictions are distinguished in different colours in order to inspect them separately. It can be seen that the ANN predictions closely follow the patterns from the FEM results in a qualitative sense. Any differences in the density plots can primarily be attributed to the different stages of the optimisation, since the surrogate model was deployed further on in the optimisation process. It is also mentioned that certain results are obscured due to overlap between the FEM and ANN bubbles in the graph.

Upon further examination, the results are rather intuitive as certain relationships can be established. For instance, the axial stiffness correlates well with the FPF load but there is a distinct negative correlation with the other objectives. However, the best FPF load is also one of the worst axial stiffness values. A similar trend is displayed between lateral stiffness and buckling, although not as extreme, where the correlation is positive for the most part, but is harder to determine. Perhaps the most difficult correlation to distinguish qualitatively is the frequency-lateral stiffness relationship.

The degree of correlation can be quantified by a non-linear method such as Spearman's correlation, which can be seen in Figure 4.6.

The figure suggests that the strongest positive correlation is between axial stiffness and the FPF load, while the most negative is that between axial stiffness and frequency. Weak correlations are established between frequency and lateral stiffness, and axial and lateral stiffness.

It is difficult to obtain an insightful depiction of the objective space, due to its high dimensionality. One possible 2-D representation is to have multiple axes and the solutions depicted by points in this space. The positions of these points is then determined by the objective values of the point-solution. For instance, a solution with only a very good buckling value would lie on the respective buckling axis. If the solution contained good values for other objectives as well, then the point would be shifted accordingly, even if its best value is that of buckling. Therefore, the concept resembles that of the tug-of-war rope game between different axes.

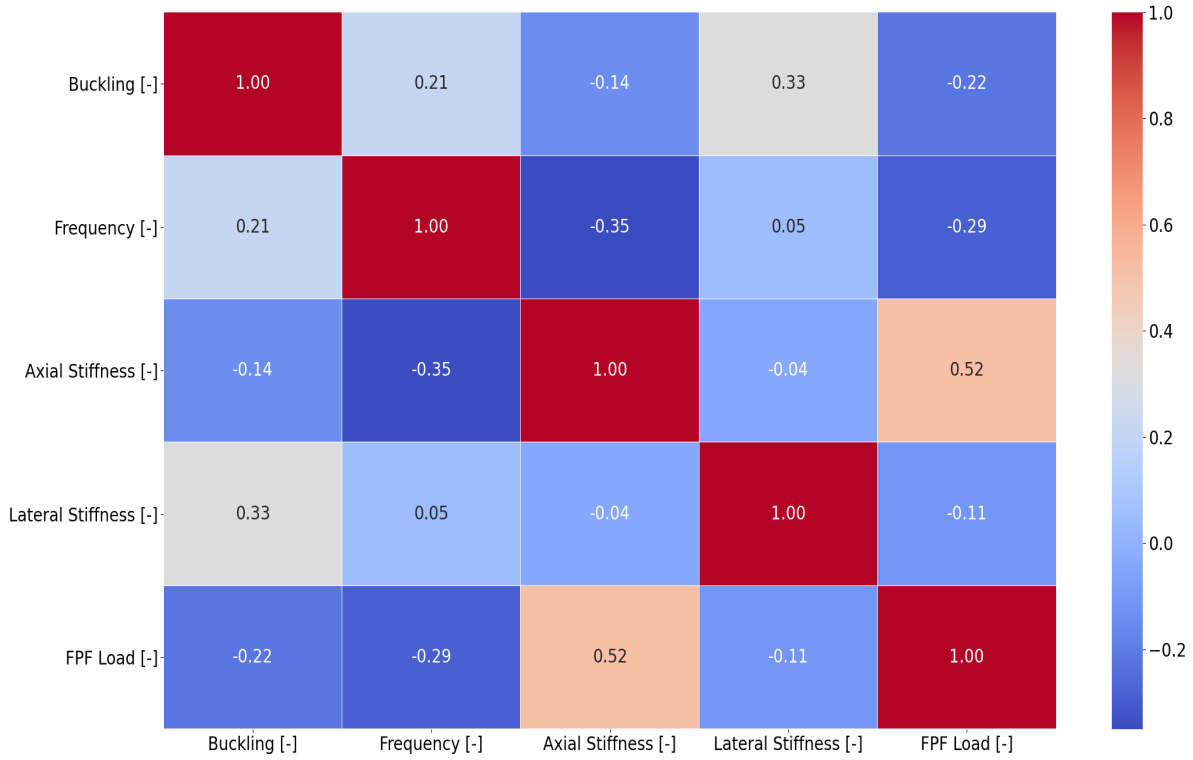


Figure 4.6: Spearman correlation between objectives.

This is presented in Figure 4.7. The optimal solutions have been given a colour to distinguish them.

Unfortunately, such a plot does not indicate whether one solution contains a higher value for a given objective with respect to other solutions, since it is relative. Nevertheless, the spread of the solutions can still be visualised. For instance, most points lie close to the frequency and axial stiffness axes, indicating that most solutions have relatively high values for these objectives. Outliers are also visible, such as the solution containing the optimal FPF load, which suffers in all other objectives. This can also be seen with the buckling objective, although it is not as pronounced.

This relationship can be examined in deeper depth by means of a petal diagram for the optimal solutions, as shown in Figure 4.8.

In this case every optimal solution is given by circles having different slices for each objective. The more the slice is filled, the higher the value for this specific objective is. For example, the optimal buckling solution contains a filled buckling slice, which means that it has the best buckling value out of all the solutions, while it also contains very low values for the axial stiffness and the FPF load. Interestingly enough, the optimal buckling and optimal frequency solutions seem very similar. Furthermore, the maximisation of either the axial stiffness or the FPF load would result in the other structural properties being very poor.

It is intuitive to study the eigenmodes from the buckling and frequency analyses to investigate the differences between different designs. The 1^{st} modes for the optimal cases of buckling and free vibrations frequency can be seen in Figure 4.9 (a) and (b), along with the modes of the QI laminate, respectively in (c) and (d), in order to provide a reference. The buckling modes of the two designs have a high number of circumferential half-waves, with the optimised design having them at a lower angle with respect to the longitudinal axis of the cylinder. On the other hand, it seems that the modes of the natural frequency are identical for the different designs, both exhibiting less half-waves along the circumference in comparison to the buckling modes.

Naturally, the design space is also of great interest. One method of reducing the dimensionality down to 2-D, is by having parallel coordinates represent the angle of each ply. Such a plot is displayed in Figure 4.10, where the optimal solutions have been given their own colour. Only 4 out of the 8 plies are shown due to symmetry. Thus, the 4 plies have a specific angle value that is displayed in the figure. In order to avoid any confusion, the line that connects the points between the consecutive plies does not signify a certain fibre movement. In other words, the graph may be interpreted as a point cloud, although lines were preferred as

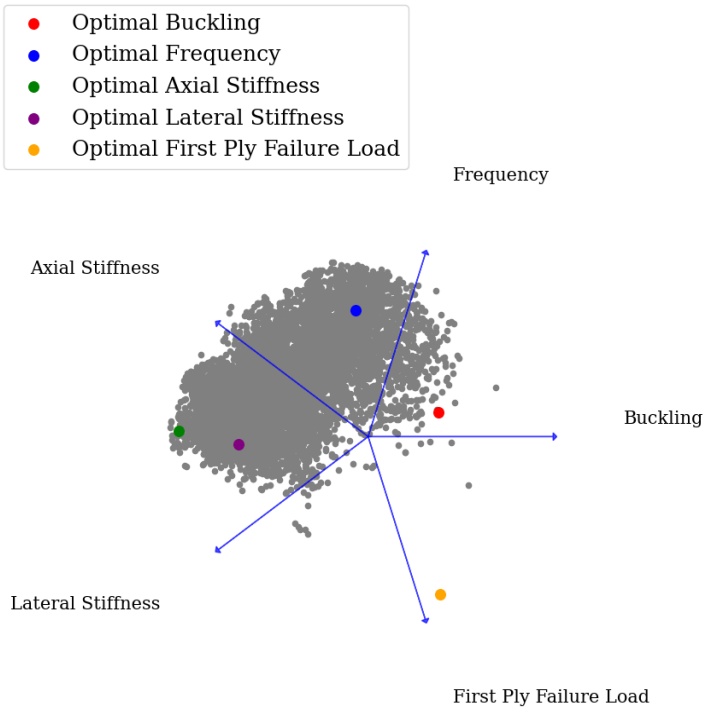


Figure 4.7: 2-D representation of the objective space.

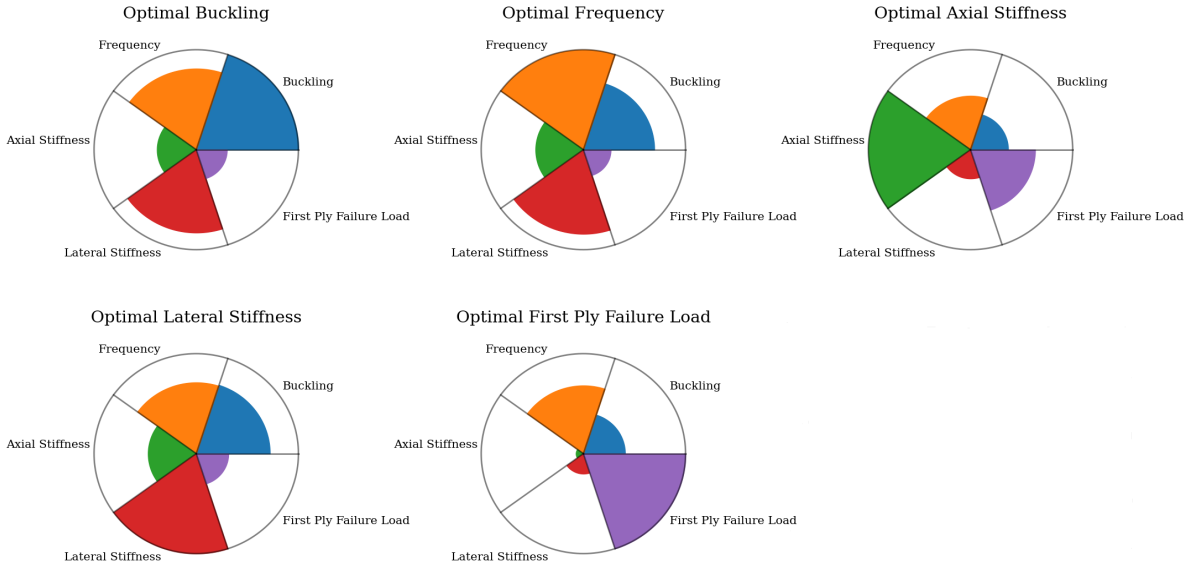


Figure 4.8: Petal plots of the optimal solutions.

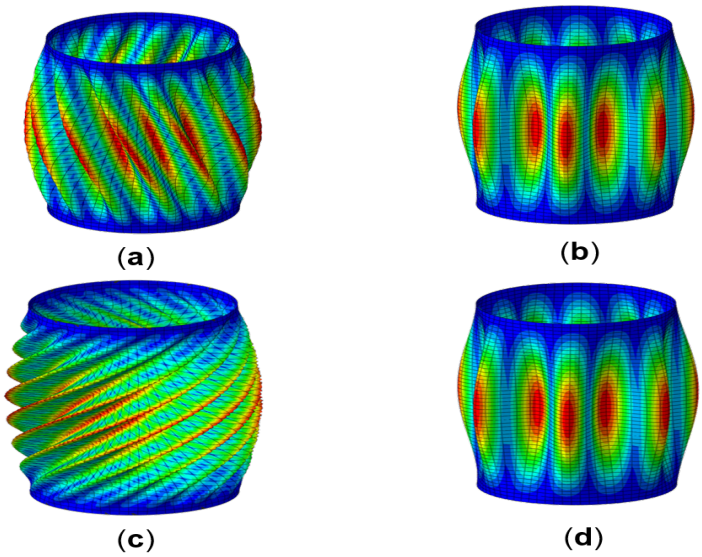


Figure 4.9: Presentation of the first modes for the objectives of (a) buckling and (b) free vibrations frequency for the optimised cylinders and of (c) buckling and (d) free vibrations frequency for the QI cylinders.

they aid in recognising the patterns in the design space.

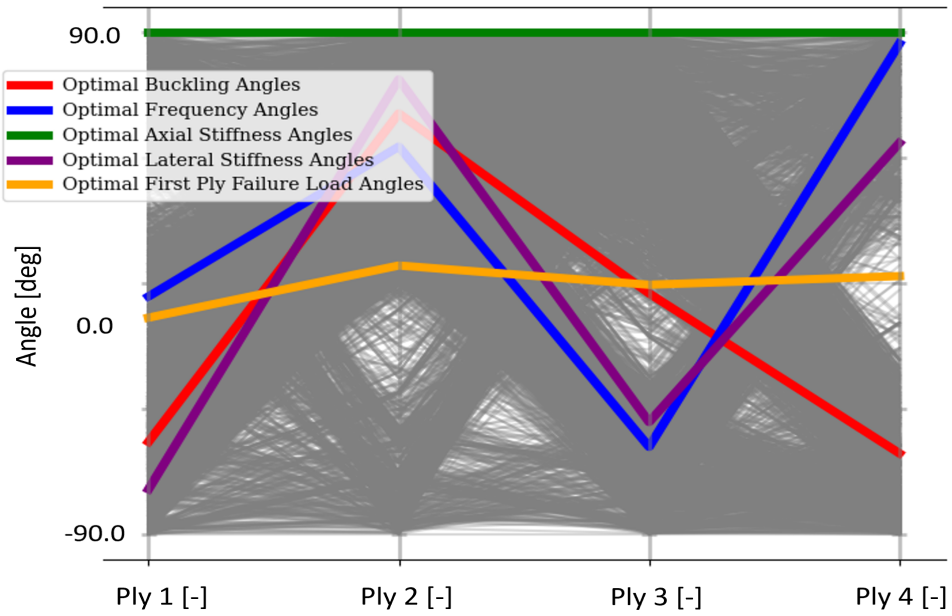


Figure 4.10: Parallel coordinate plot of the design space.

The vertical axis of the plot indicates the angle values of each ply and is in the range of $[-90°,90°]$ degrees, with each tick being a change of 10°, while the horizontal axis shows the numbering of the plies. It is reminded that at 0° the fibres follow the circumference and at 90° they go across the longitudinal axis of the cylinder. The information obtained from the design space is that most solutions tend to avoid the region around 0° and mainly favour positive angles over slightly negative ones. The optimal designs are also given in tabular form in Table 4.4.

The design for the optimal axial stiffness is unsurprising, since all of the plies are oriented towards the application of the mechanical load. On the other hand, the optimal FPF load design is more interesting since all of its plies are very close to 0°, but it is distinctively better than an all-0° design. When it comes to the design of the lateral stiffness, the optimisation algorithm yielded a result that is balanced and has the fibres

Table 4.4: The ply angles of the optimal designs.

Objective	1st Ply Angle [deg]	2nd Ply Angle [deg]	3rd Ply Angle [deg]	4th Ply Angle [deg]
Optimal Buckling Angles	-56.9	60.8	-3.7	-61.0
Optimal Frequency Angles	-4.8	48.9	-58.7	85.8
Optimal Axial Stiffness Angles	90.0	90.0	90.0	90.0
Optimal Lateral Stiffness Angles	-73.9	73.0	-49.8	50.2
Optimal First Ply Failure Angles	-12.4	6.3	-0.6	2.5

aligned more towards the loading direction. The designs for the optimal buckling and frequency do not seem to favour a balanced laminate. Of course it should be kept in mind that the designs that generate optimal solutions for each objective are not unique and a very different laminate could yield very similar results.

4.1.2. Variable stiffness cylinder

Having established the design and objective spaces for the CS pristine cylinder, the VS pristine cylinder was optimised next. The designs that provided the highest value for each objective for the CS cylinder were injected in the first generation of the subsequent VS cylinder optimisations, in order to facilitate the search around these locations in the design space. The two different angle variation cases with a different amount of CPs are presented in this section. These various instances are then compared with each other, as well as with the CS case. To reiterate, the examined VS cases are 6 in total:

- Axial Angle Variation 2,3,4 CPs
- Circumferential Angle Variation 2,3,4 CPs

Axial Variation - 2 Control Points

The AV case with 2 CPs is presented first. It is reminded that the FEM results were obtained with 2 CPs for the pristine cylinder. The surrogate models were then trained primarily on these samples and performed the optimisation for the other cases entirely on their own.

In every optimisation, it is important to investigate whether the optimisation algorithm has achieved convergence with respect to a convergence criterion. In this instance, as the hypervolume that defines the objective space converges to a constant value, it can be said that convergence has not been achieved. Since the termination criterion was based on a fixed amount of generations, it is not guaranteed that the algorithm has converged for each VS case. The normalised values of the hypervolume and for each objective in this case, are presented in Figure 4.11 against the number of generations, showing that the algorithm has indeed not converged.

It should be noted that the injected designs from the CS optimisation yield the highest values at the start of the process and they are reduced in subsequent generations. On the other hand, the hypervolume has peaks and valleys, indicating that there is constant change and convergence has not been achieved. Nevertheless, the optimisation was terminated as has already been mentioned. A distinction is also made between the as-obtained FEM results and the ANN predictions after the surrogate models had been trained.

The relation between each objective is examined next. The differences between the FEM results and the ANN predictions are also of interest here, in order to ascertain that the predictions behave as expected. This is shown in Figure 4.12. Once again, the diagonal represents the density estimation of each objective in relation to the others. Along with the objectives, the mass is also included in the plot to show the mass distribution across the various designs and to highlight which mass values yield superior results. The ANN predictions mostly overlap with the FEM results, indicating that the accuracy is sufficiently good. It is re-iterated that any potential differences in the density plots could also be due to the different designs of the later stages of the optimisation, upon which the surrogate model had to perform the predictions.

Unlike the CS optimisation case, certain constraints had to be met in the VS cylinder optimisation. Therefore, it is necessary to visualise which designs meet the imposed constraints, as illustrated in Figure 4.13. The infeasible designs are the ones that are deemed too heavy and/or exhibit too high a tow curvature, as per equation 3.14. It is shown that most of the designs that are feasible also happen to exhibit higher values for each objective. In addition, the average mass was quite close to the base value of the CS cylinder. Since a high tow curvature correlates with heavier designs, it seems that the results with high values are better represented by smaller values for tow curvature and mass. The union of feasible and infeasible designs in the mass density

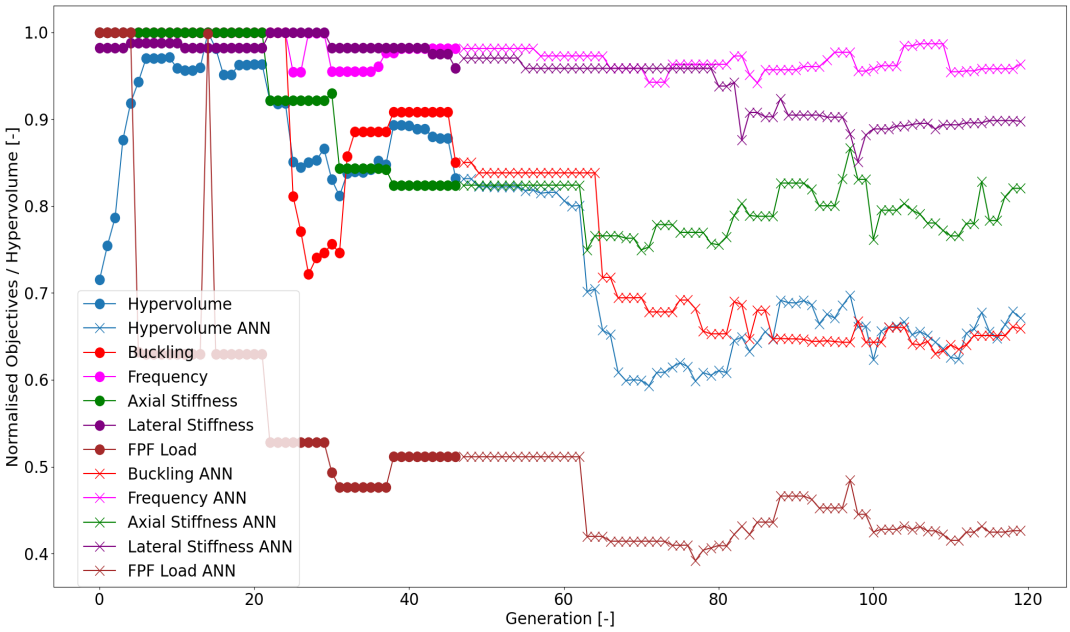


Figure 4.11: Generational objective change for the AV case with 2 CPs.

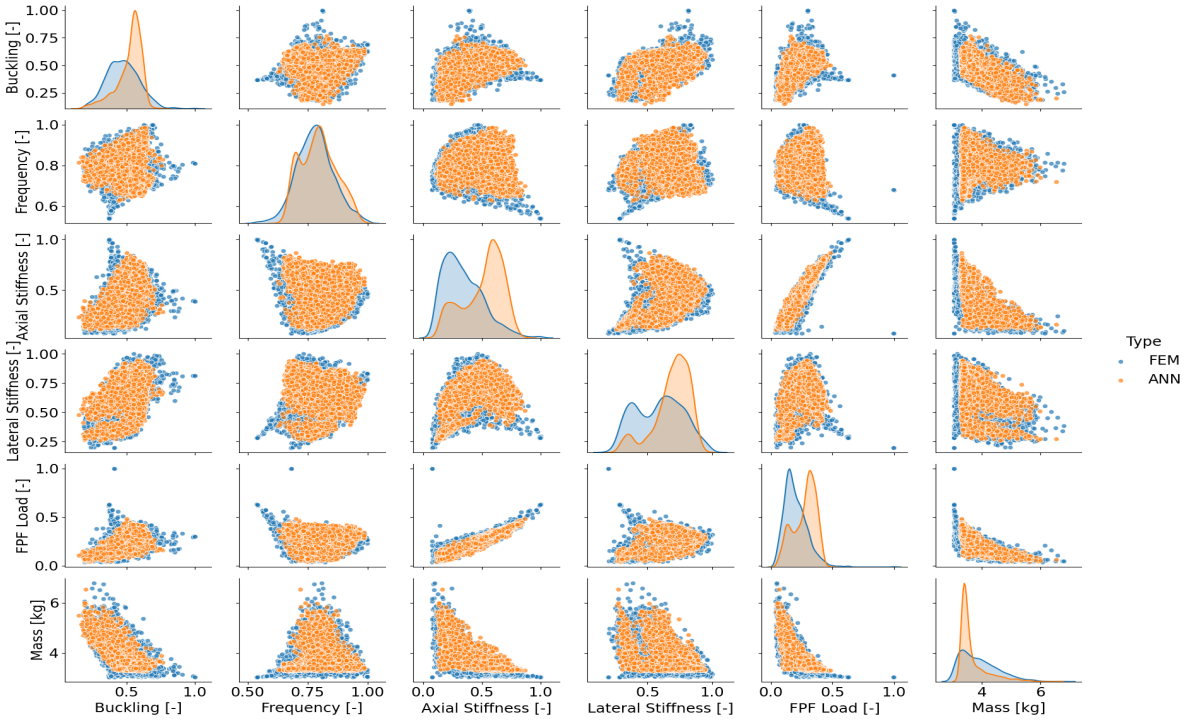


Figure 4.12: Objective relationship for both FEM and ANN results in the AV case with 2 CPs.

plot signifies the infeasibility due to the tow curvature. The rest of the discussion and results will be limited to the feasible designs.

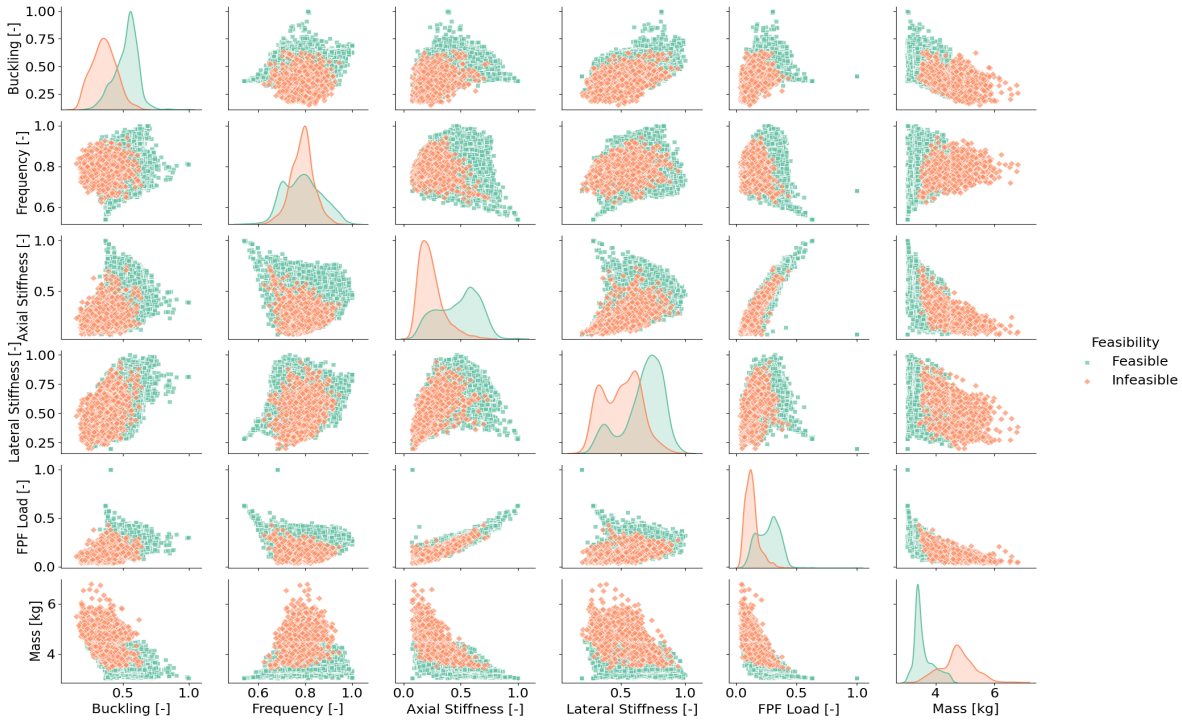


Figure 4.13: Design feasibility with respect to the imposed constraints for the AV case with 2 CPs.

The correlation between objectives can be quantified with Spearman's correlation method, as presented in Figure 4.14, where the mass is also included. Although, the correlation was rather intuitive in the CS instance, it is not the case any more. It is observed that the positive correlation between objectives is more pronounced. What is more, objectives with high correlation in the CS instance, such as buckling and frequency, now exhibit the lowest positive correlation instead. Another interesting aspect is the mass, which correlates negatively with all objectives but the frequency. This means that most objectives suffer by an increase in mass due to overlaps.

The objective space can be visualised via a star coordinate plot like in the CS case. This is depicted in Figure 4.15, with the solutions with the highest values for each objective in different colours. It is evident that most of the solutions obtained high axial and lateral stiffnesses, as well as frequency values. This is similar to the CS case, which shows a similar objective space.

Another important aspect is to identify the solutions that result in higher values than the CS instance for each objective. In this case, the objective space contained only one solution that managed to do that, which is for the lateral stiffness objective. The solution was very similar to the CS one with only a mild stiffness variation. The petal plot that was used for the CS cylinder can be employed to showcase the similarities of the two solutions in the objective space, as shown in Figure 4.16. It can be seen that for a very minor increase in lateral stiffness there is a higher reduction in buckling. The increase in lateral stiffness is quantified as a 5.5% against a 4% increase in mass but when standardised with respect to mass the relative increase in relation to the CS one is 2%.

The design space can be investigated by a parallel coordinate plot, as seen in Figure 4.17. The angle variation within one ply is thus shown for each design. It can be seen that the optimal lateral stiffness design is the only one with variation within a ply among the so called optimal designs. In general, the designs that tend to have an angle close to 0° were avoided by the algorithm, favouring higher angles. Since only the feasible designs have been presented, this behaviour can be explained by the interpretation that designs close to 0° tend to yield a higher structural mass. This is very reasonable considering that for the axial angle variation case a circumferential course shift will likely lead to more overlap with shallow angles.

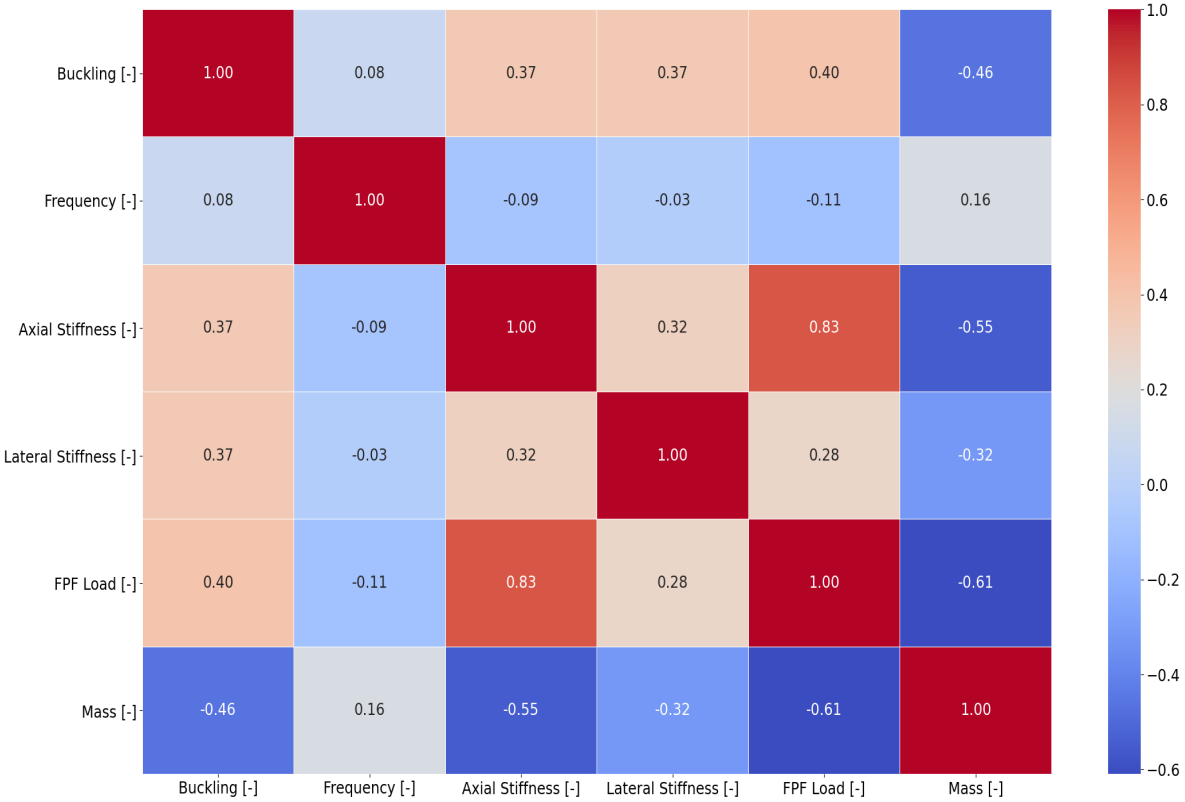


Figure 4.14: Correlation heatmap between objectives for the AV case with 2 CPs.

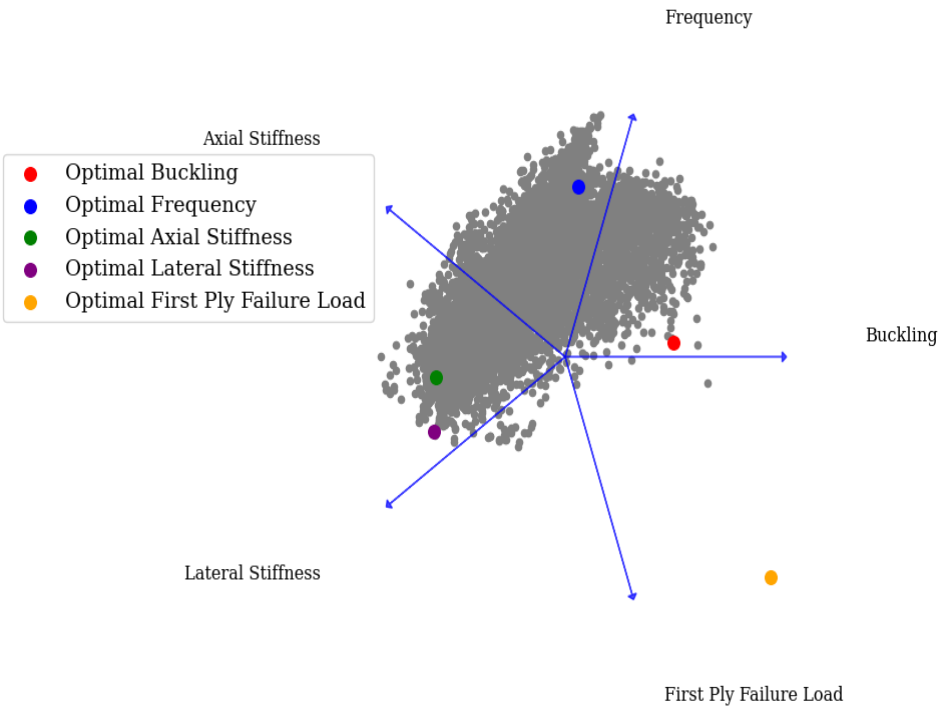


Figure 4.15: 2-D representation of the objective space for the AV case with 2 CPs.

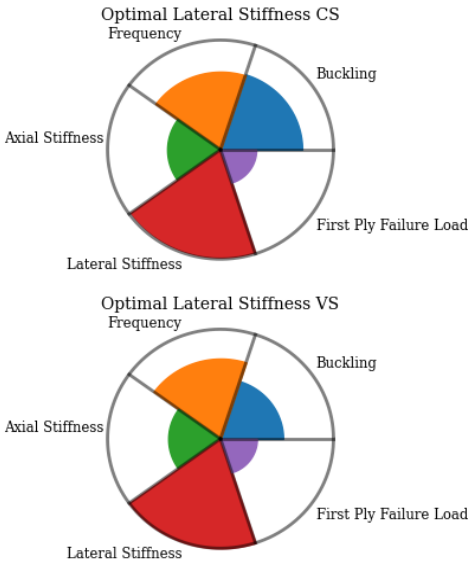


Figure 4.16: Petal plot of the optimal solution for the lateral stiffness objective in the AV case with 2 CPs.

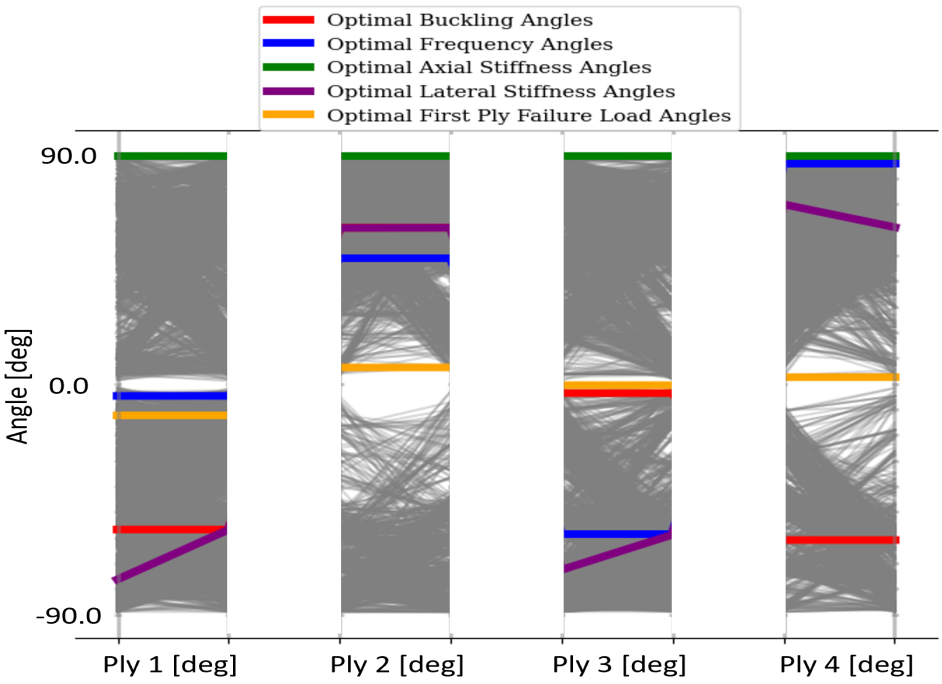


Figure 4.17: Parallel coordinate plot of the design space for the AV case with 2 CPs.

Circumferential Variation - 2 Control Points

The CV case for 2 CPs is presented next. The hypervolume seems like it has not converged to a constant value, as shown in Figure 4.18. Once more the hypervolume rapidly increases in the first generations, then decreases in the intermediate generations until it reverts to a positive increase towards the end. The ANN results follow the same paths after taking over, indicating that the predictions are accurate over a generation.

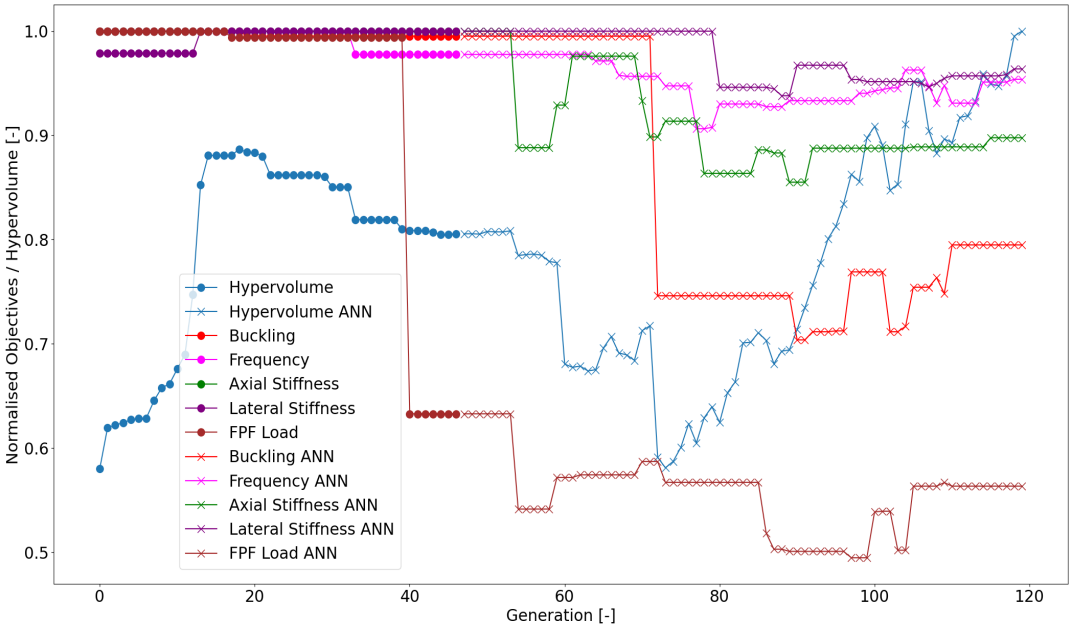


Figure 4.18: Generational objective change for the CV case with 2 CPs.

The distribution of the FEM and ANN results for each objective can be seen in the pair plot in Figure 4.19. It is observed that the average mass and the distribution peak of the solutions are higher than in the AV case with 2 CPS. Additionally, the ANN predictions tend to be closer with one another in clusters, while there are certain outliers for the FEM results.

The feasibility of the designs is presented in Figure 4.20. The primary difference between the AV and CV case with 2 CPs is that the driver for the infeasibility is the mass and not the tow curvature. This can be seen by the clear separation of the feasible and infeasible clusters along the mass axes and the smaller overlap in the mass density plot. The designs yielding the best values lie in the feasible region, although this is not as pronounced as in the AV case with 2 CPs.

This time the correlation is mainly positive between most of the objectives, as is shown in 4.21. The bigger difference is that frequency has a positive correlation with the rest of the objectives and that it now correlates negatively with mass. Thus, it seems that overlaps do not contribute positively at all in this case.

In the objective space the solutions are very similar to the previous instances. Figure 4.22 shows that more solutions are drawn to the buckling axis, although the difference is not very distinct.

Once more, the only optimal solution that surpassed the CS case was for the lateral stiffness objective. The design resulted in a marginal increase as is shown in the petal plot in Figure 4.23. It is also observed that the frequency increased at the expense of the buckling value. The relative increase in lateral stiffness is 20.0% against an 18% increase in mass, which amounts to 2% increase with respect to the CS case when standardised in relation to the mass.

Figure 4.24 depicts the design space for this case. It appears that the optimisation algorithm did not explore the design space as thoroughly in this case and instead focused on similar designs. For example, there was a complete lack of negative values for the second ply. The optimal lateral stiffness design exhibits a significant variation in its first ply to obtain this improved solution, although two of its plies have a CS.

Axial Variation - 3 and 4 Control Points

The optimisation objectives for the rest of the CPs were obtained solely by the ANN surrogate models as predictions. The results were quite similar for both 3 and 4 CPs, and therefore they will be presented together for the sake of brevity.

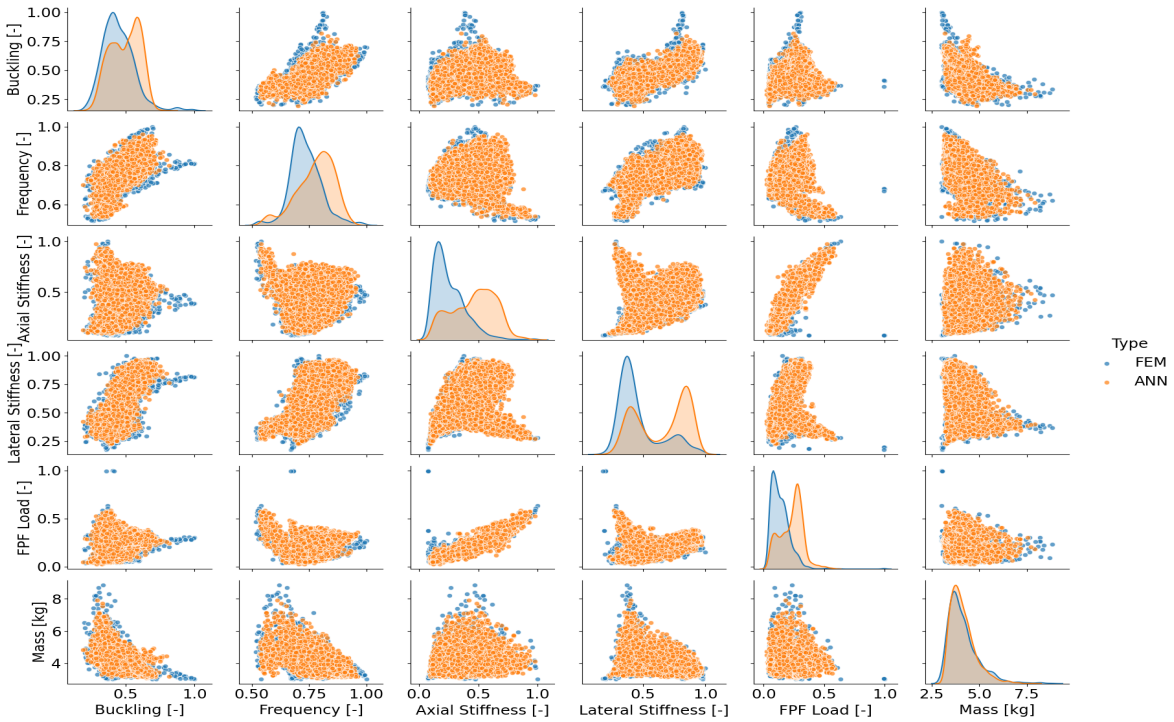


Figure 4.19: Objective relationship for both FEM and ANN results in the CV case with 2 CPs.

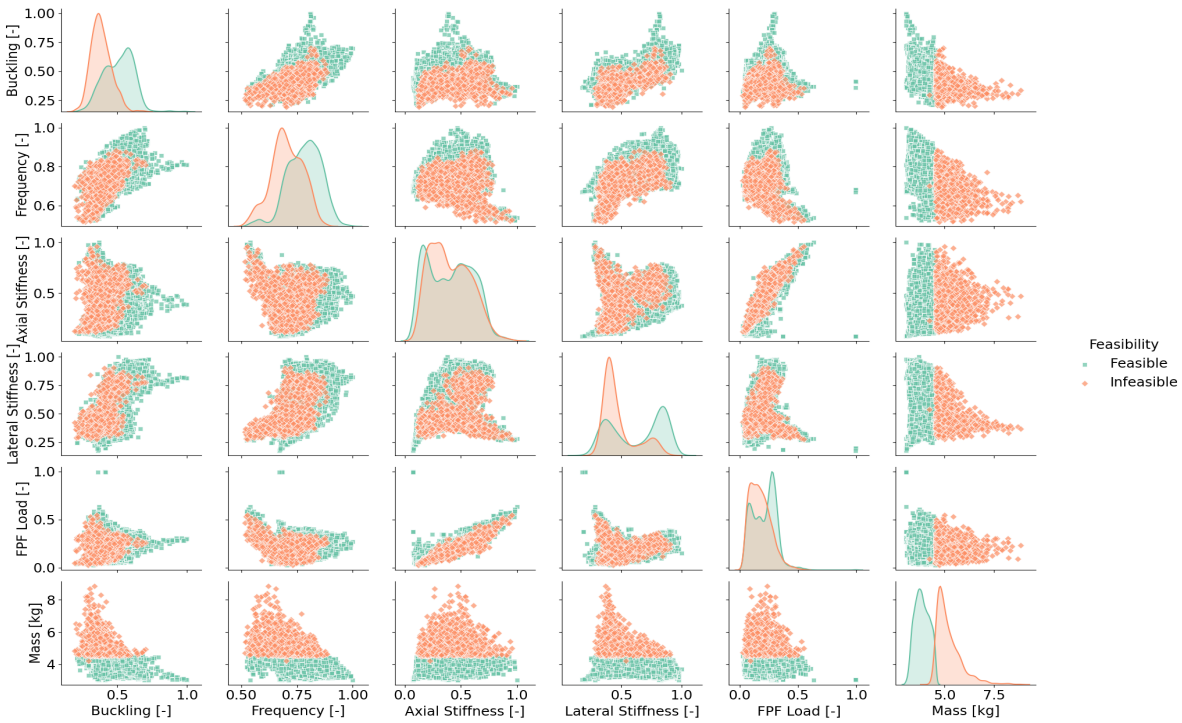


Figure 4.20: Design feasibility with respect to the imposed constraints for the CV case with 2 CPs.

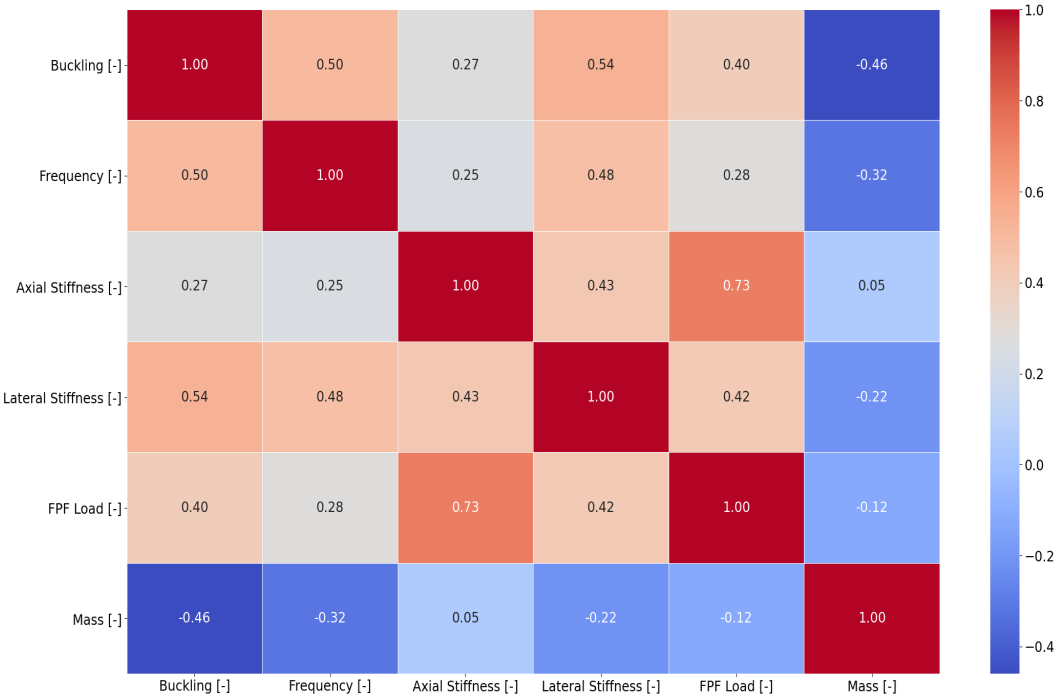


Figure 4.21: Correlation heatmap between objectives for the CV case with 2 CPs.

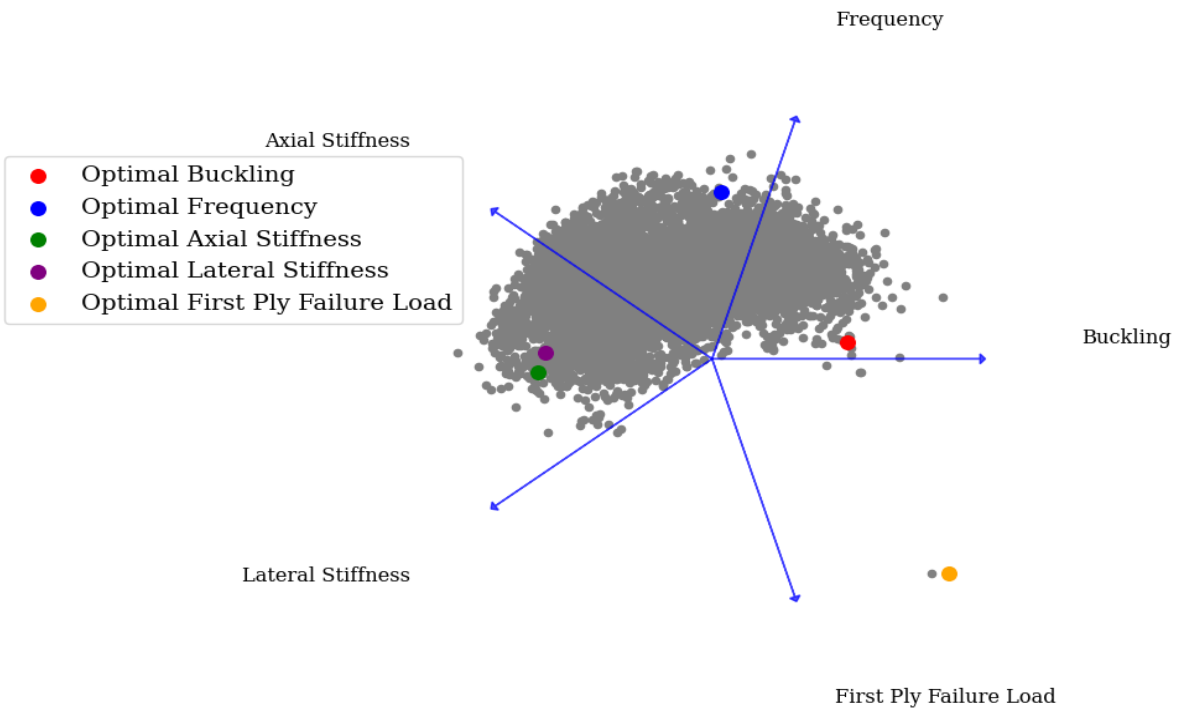


Figure 4.22: 2-D representation of the objective space for the CV case with 2 CPs.

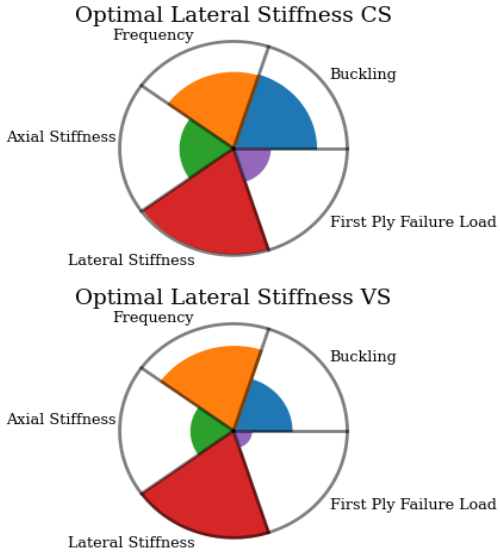


Figure 4.23: Petal plot of the optimal solution for the lateral stiffness objective in the CV case with 2 CPs.

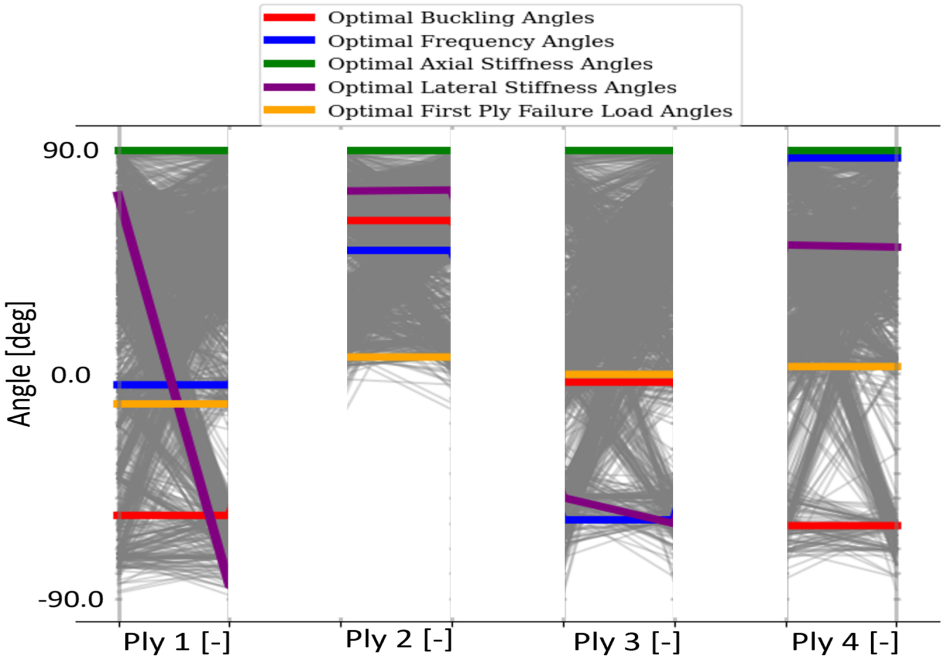


Figure 4.24: Parallel coordinate plot of the design space for the CV case with 2 CPs.

In Figure 4.25 (a) and (b), the convergence plots of the multi-objective optimisation algorithm can be found for the 3 and 4 CPs cases, respectively. From the hypervolume results it seems that in both cases the inter-generational change is not very significant towards the end, although it can be said that convergence has not been achieved in the objective space. It is observed that the FPF load and the axial stiffness results experience the steepest drop after the first generations and remain in that state. On the other hand, the lateral stiffness appears to have consistently high values throughout the generations.

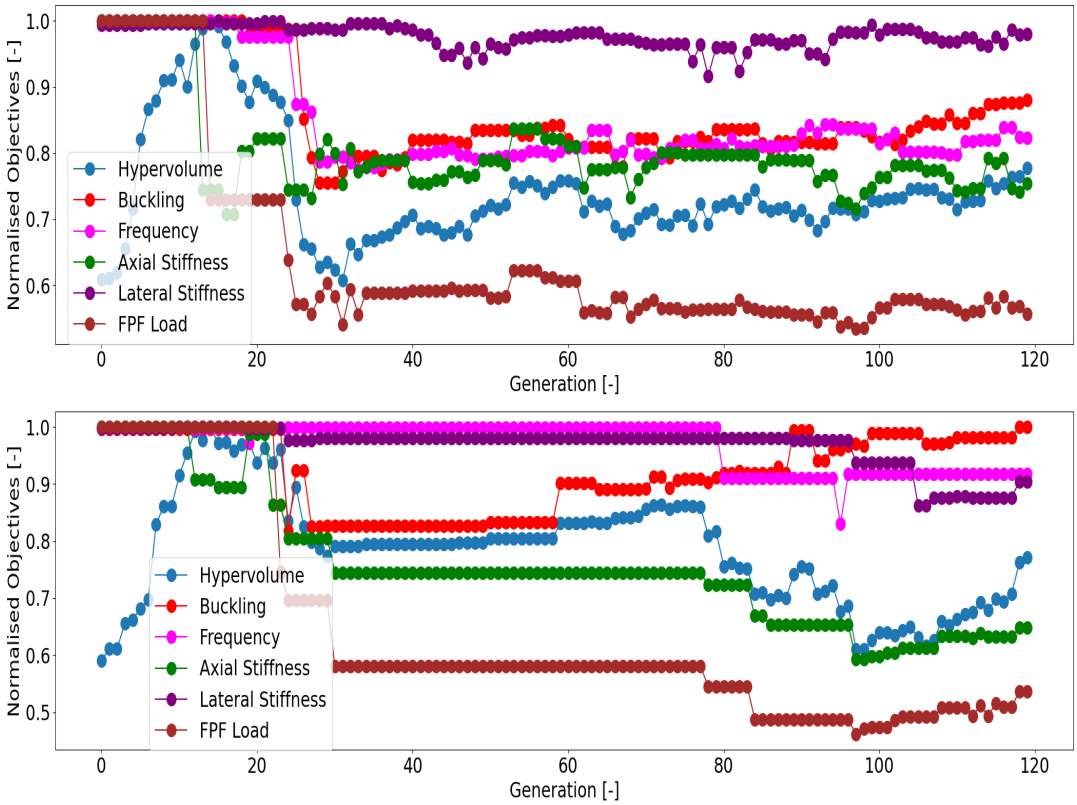


Figure 4.25: Generational objective change for the AV case with 3 CPs (top) and 4 CPs (bottom).

The feasible designs for 3 and 4 CPs are presented in Figure 4.26 (a) and (b), respectively. This time around, the feasible designs in relation to the structural mass are primarily found in a very narrow band. This band also contains the best results for each objective, highlighting that increased structural mass reduces the objective efficiency when you standardise the results. It is evident that most of the design infeasibility is attributed to the high mass due to overlaps. A significant difference between the two cases is witnessed when it comes to the feasible designs, as very few of them can be identified in the cases of 4 CPs. This was to be expected and is in line with the exploratory search of Figure 3.5 in Section 3.2.3. It can also be said that the trends of the various objective relationships very closely match in both cases, as well as the distribution of feasible and infeasible designs in the objective space region.

The objective relationships are quantified and are displayed in Figure 4.27 (a) and (b) for 3 and 4 CPs. The numbers match the visual trend that was exhibited by having very similar values. More specifically, the mean standard deviation for all objectives is 13%, while the highest standard deviation is displayed by the buckling objective at 21%. The primary difference between the two cases is the relationship between the buckling and lateral stiffness objectives, where the correlation is positive for 3 CPs but negative for 4 CPs. It seems that the latter case is the odd one out, considering that this objective correlation is the same for the case of 2 CPs, and such a negative correlation is of course detrimental to the design.

The tug-of-war between objectives is shown in Figure 4.28 for (a) 3 and (b) 4 CPs. In both cases the solutions are distinguished by their good frequency and lateral stiffness values in a relative manner. A minor difference between the two cases is the higher number of designs that exhibit good buckling load designs for 4 CPs. The solutions that have been identified to have the best values for each category and are defined as optimal, are the same in both cases and lie in the same spots in the objective space. These designs happen to

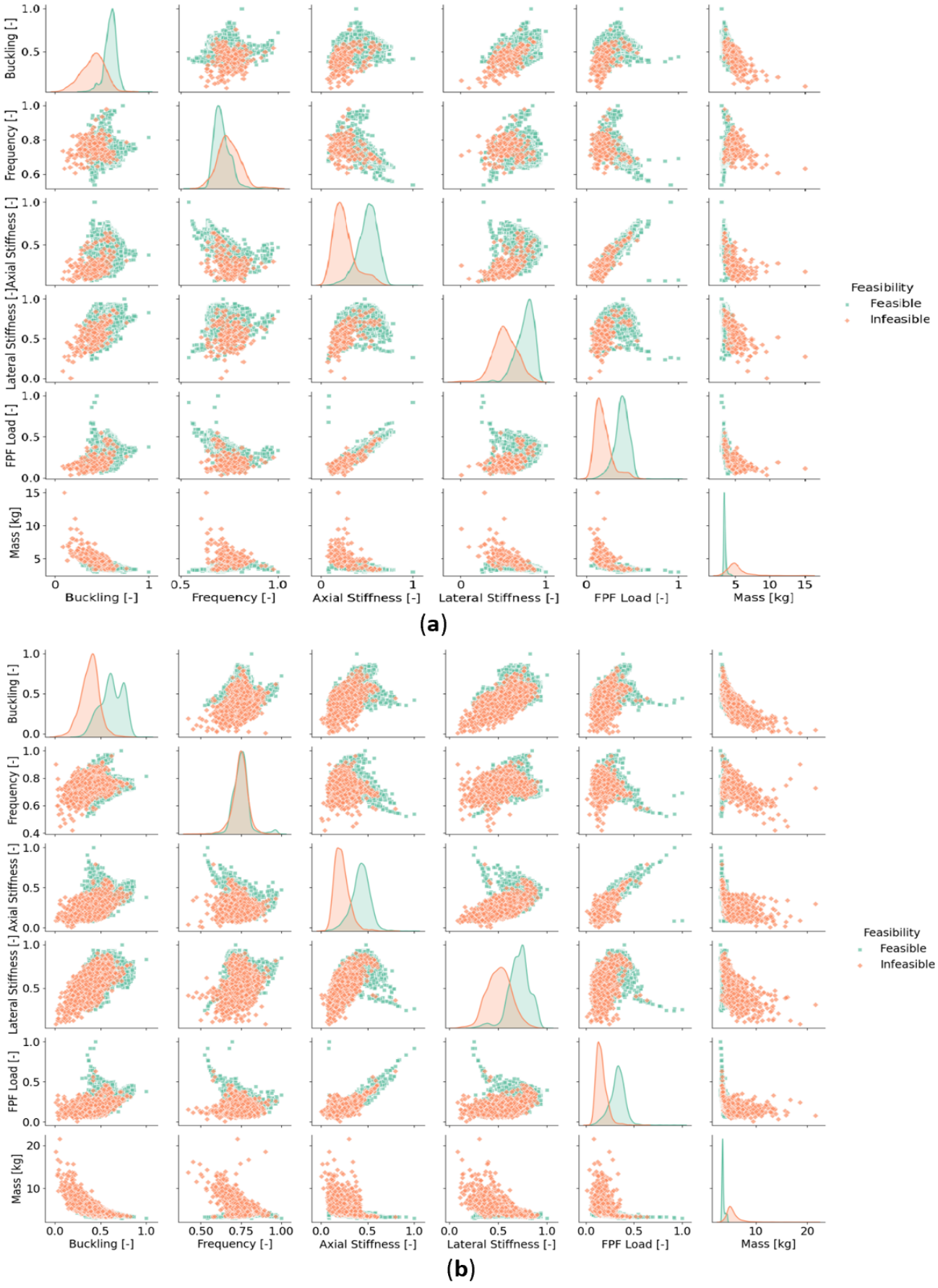


Figure 4.26: Design feasibility with respect to the imposed constraints for the AV case with a) 3 and b) 4 CPs.

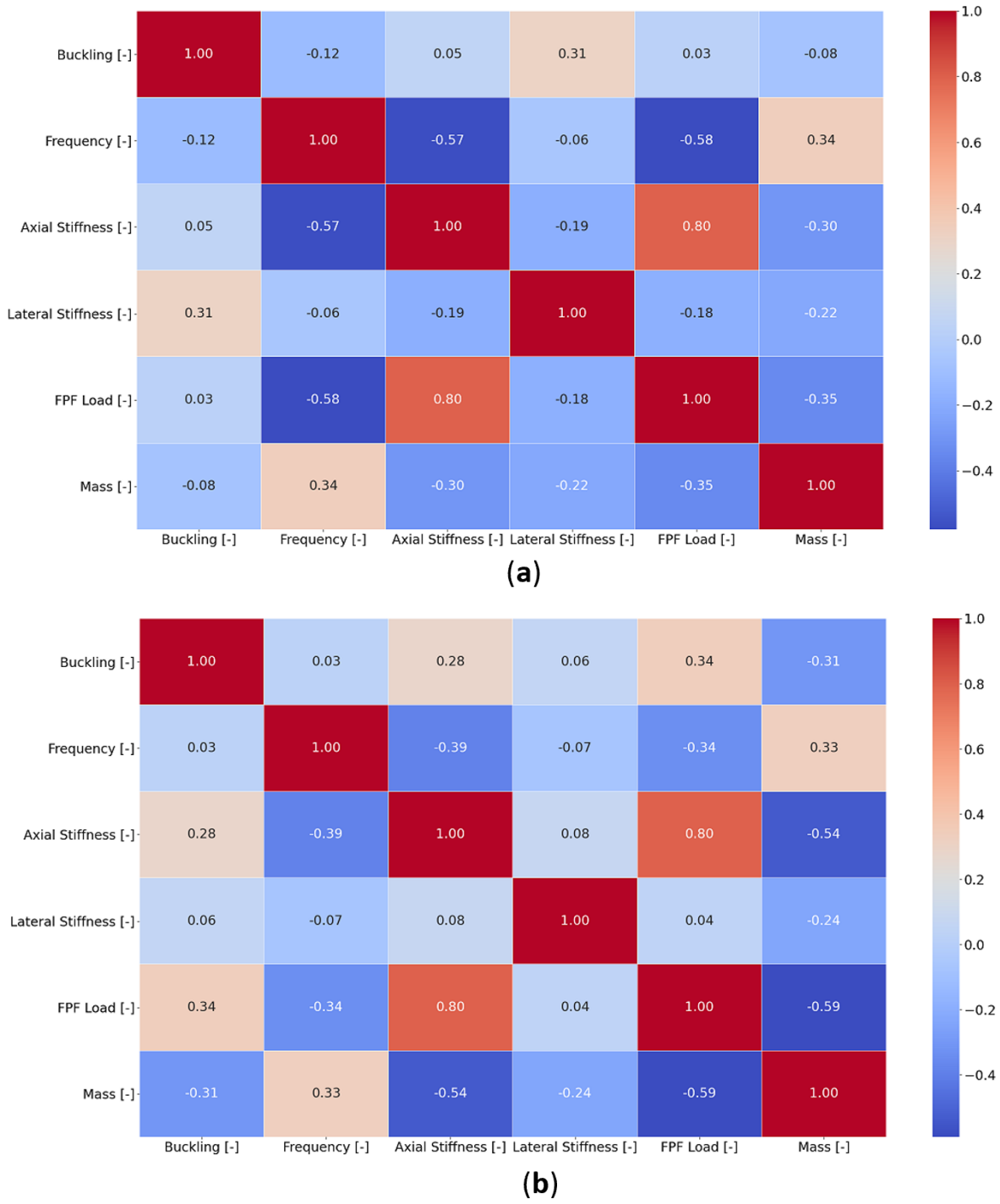


Figure 4.27: Correlation heatmap between objectives for the AV case with (a) 3 and (b) 4 CPs.

be the inputs of the corresponding optimal CS designs. Therefore, unlike the 2 CPs case, no design was found which yields higher values for any of the objectives.

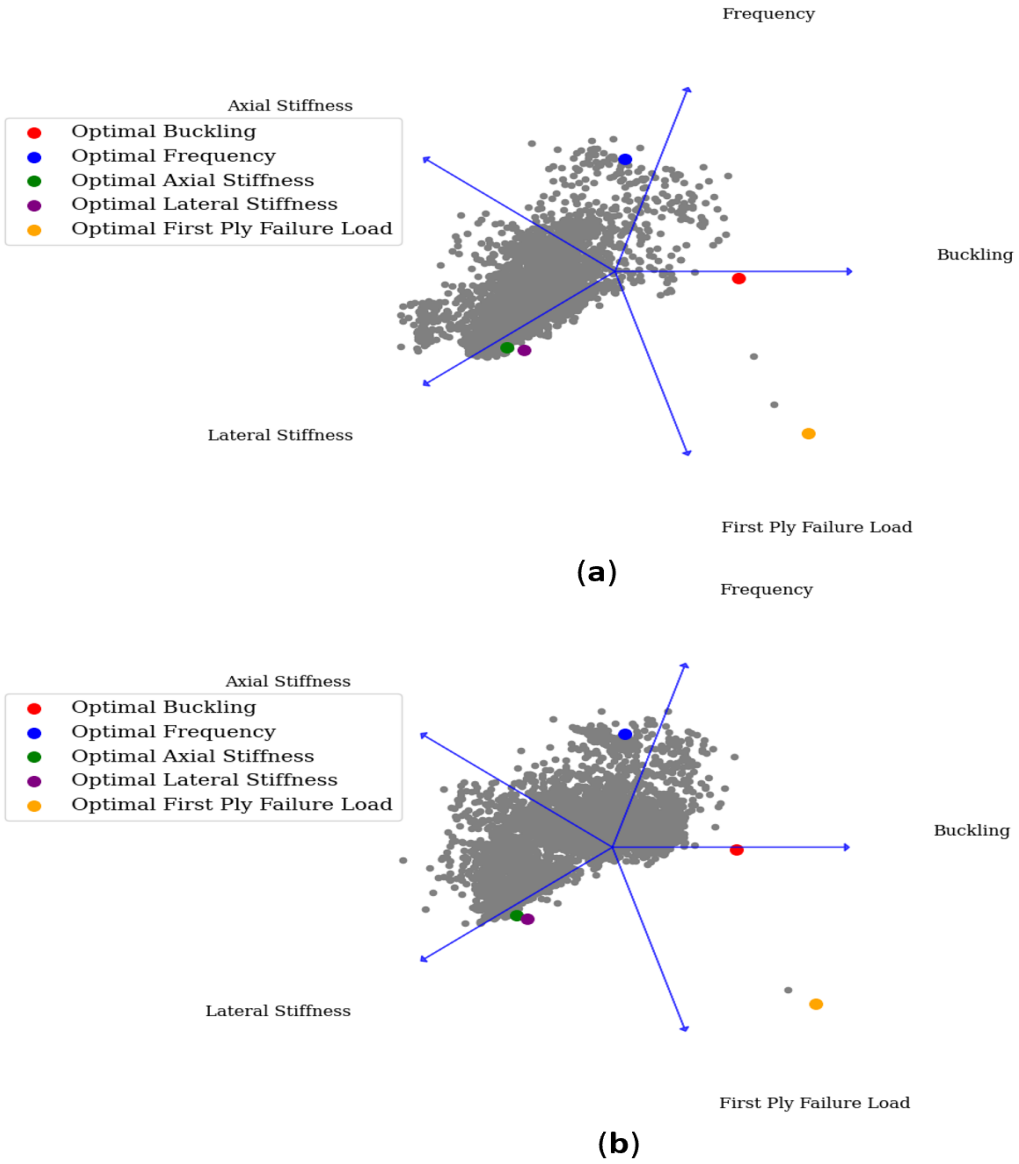


Figure 4.28: 2-D representation of the objective space for the AV case with (a) 3 and (b) 4 CPs.

The design space for the cases of 3 and 4 CPs is presented in Figure 4.29 (a) and (b), respectively. They both showcase similar trends on whether certain regions are densely or sparsely explored. The 1^{st} and 3^{rd} plies favour high negative angles, while the 2^{nd} and 4^{th} favour the opposite. It is reminded that in the case of 2 CPs, only an avoidance around the 0° region was witnessed. This also points to the fact that these designs tend to result in more overlap that decreases the structural efficiency when the mass is standardised. Of course it must be noted that these results are contained in the set of feasible designs and the infeasible designs could have potentially filled the empty space that is observed.

Circumferential Variation - 3 and 4 Control Points

The optimisation convergence plots for the CV cases of 3 and 4 CPs can be seen in Figure 4.30 (a) and (b), respectively. The lack of change in the hypervolume for the case of 4 CPs indicates a stagnation of the results in the objective space and that the multi-objective optimisation algorithm has converged. It was decided to manually bypass the implemented termination criterion to continue the optimisation for this case, as the surrogate model would perform the bulk of the optimisation process efficiently. It is also noteworthy that for

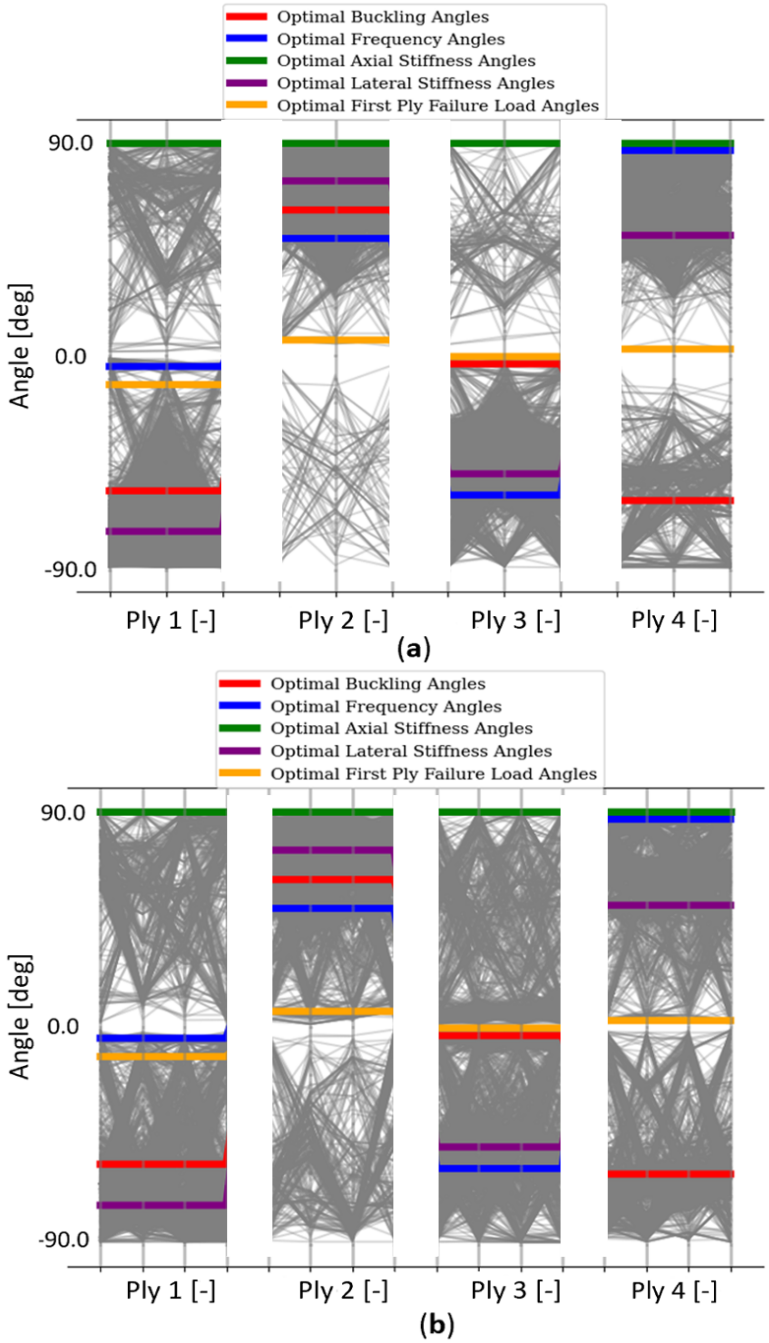


Figure 4.29: Parallel coordinate plot of the design space for the AV case with (a) 3 and (b) 4 CPs.

the same case, the FPF load objective is not the worst performer but instead it is the buckling load, which has not been observed before thus far. Apart from this, the rest of the objectives seem to consistently provide good results within a generation.

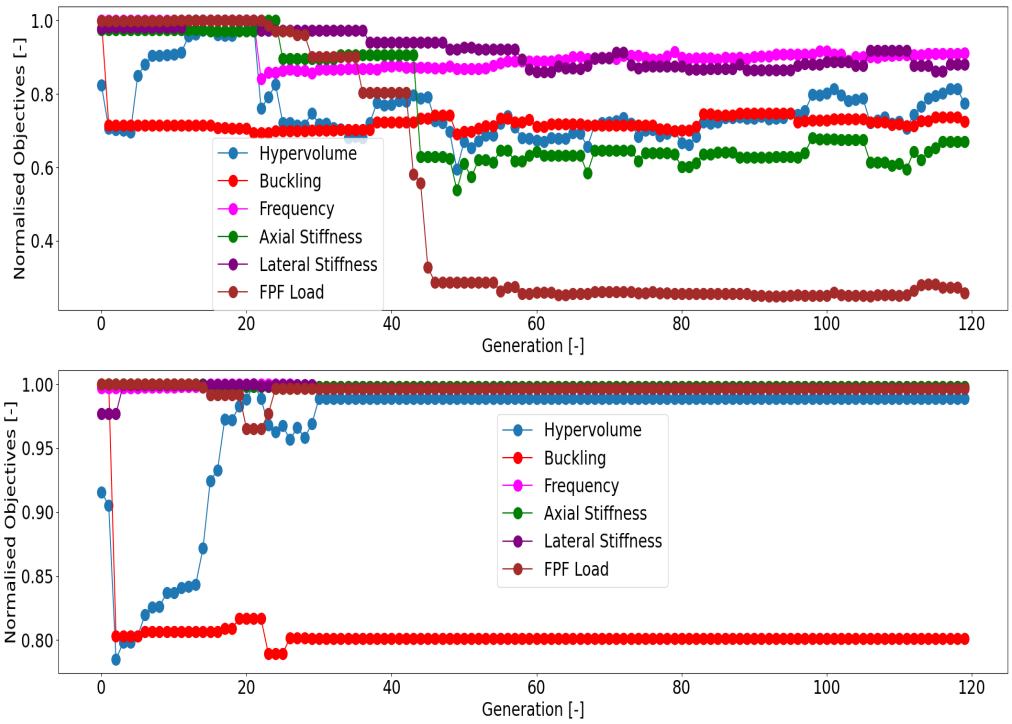


Figure 4.30: Generational objective change for the CV case with (a) 3 CPs and (b) 4 CPs.

An inspection of the feasible designs for 3 and 4 CPs was performed and shown in Figure 4.31 (a) and (b), respectively. Unlike the case with 2 CPs, there is greater overlap between the feasible and infeasible density plots regarding the mass, which is to be expected since the curvature can attain higher values with the tows becoming more flexible with the additional CPs, and therefore also drives a significant part of the infeasible designs. Additionally, the best results for each objective tend to be in the infeasible set, opposite of what was witnessed for the case of 2 CPs. Once more, the case with 3 CPs contains a greater number of feasible designs that the one of 4 CPs.

The correlation between the optimised objectives is presented in Figure 4.32 (a) and (b) for 3 and 4 CPs, respectively. The positive correlation between objectives has been reduced in both cases with respect to the 2 CPs case, although this is not as pronounced with the buckling objective. The buckling objective also exhibits the highest standard deviation between the 3 and 4 CPs case of 16%, with the mean standard deviation being 13%. It is also observed that the positive correlation between the axial stiffness that has been standardised and the structural mass also increased substantially with each additional CP, while the rest of the objectives correlate negatively with mass.

In Figure 4.33 (a) and (b) the objective space can be seen, for 3 and 4 CPs, respectively. A significant difference between these two cases and that of 2 CPs is that there is a far greater number of results with high relative FPF load values. Thus, it seems that the results are better spread out in this tug-of-war between objectives. The identified optimal results are the ones stemming from the CS optimisation, and therefore there was no objective with greater values than these designs, unlike the 2 CPs case.

The designs for 3 and 4 CPs are shown in Figure 4.34 (a) and (b), respectively. This time, the feasible design space that was explored was primarily around the optimal designs, as identified by the CS optimisation, in contrast to the case with 2 CPs. It is also clear that the angle variation within each ply is moderate, as other designs not presented here have most likely violated the curvature constraint or have yielded too high a mass.

Surrogate model performance

With the optimisation objectives having been presented, it is important to check the accuracy of the ANN that acted as the surrogate model on the collected dataset samples. Firstly, the ANN architectures that per-

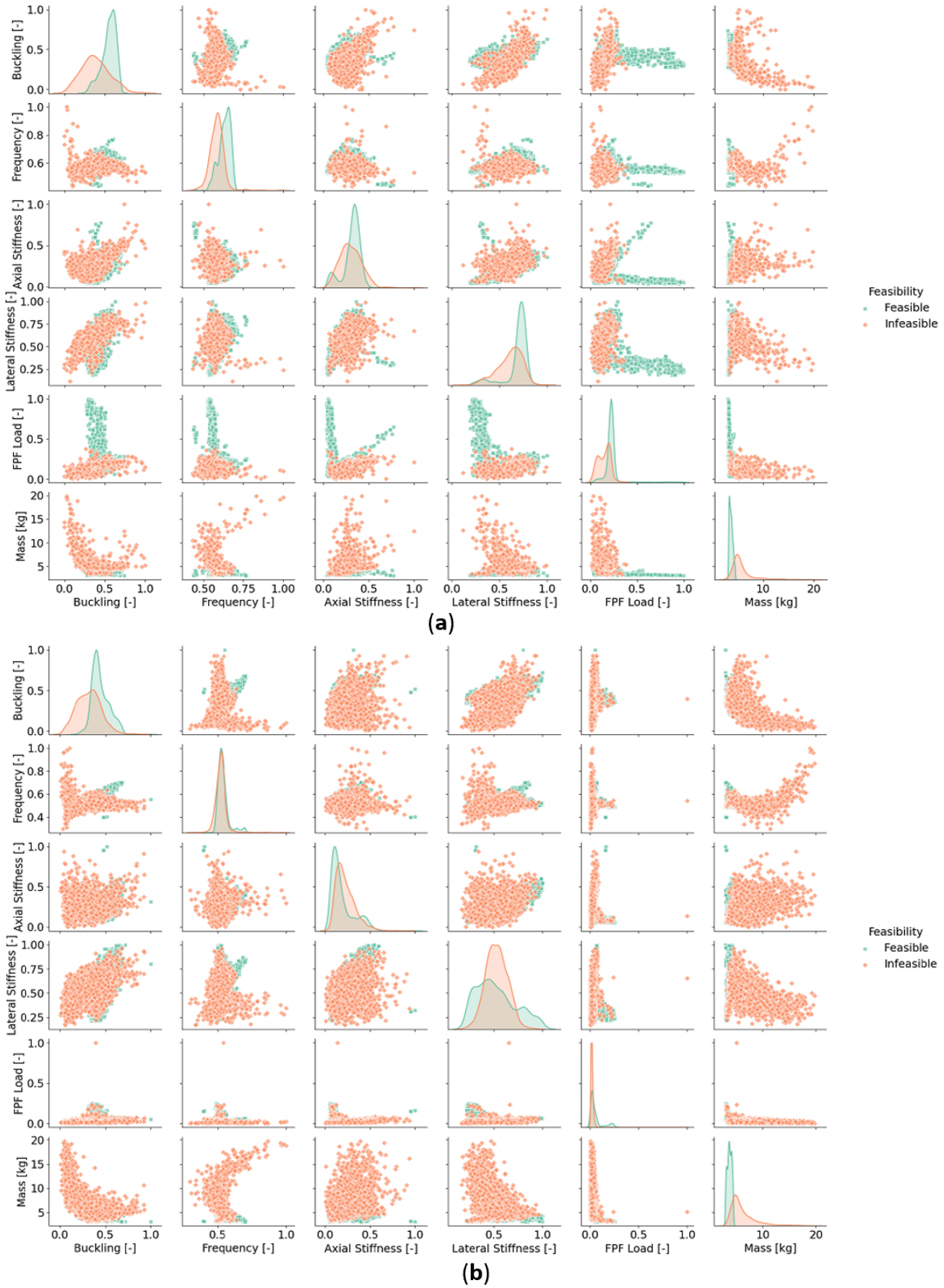


Figure 4.31: Design feasibility with respect to the imposed constraints for the CV case with a) 3 and b) 4 CPs.

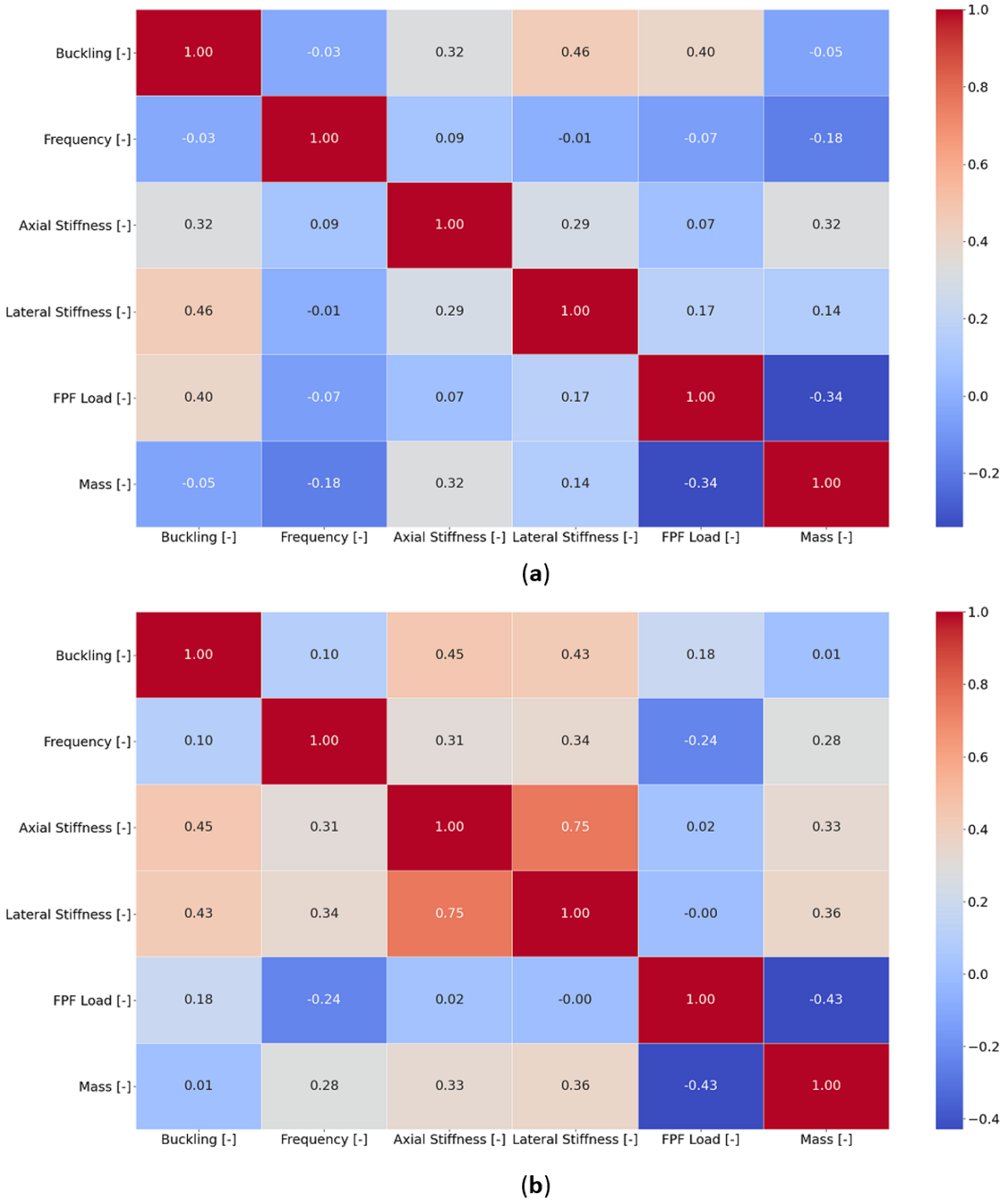


Figure 4.32: Correlation heatmap between objectives for the CV case with (a) 3 and (b) 4 CPs.

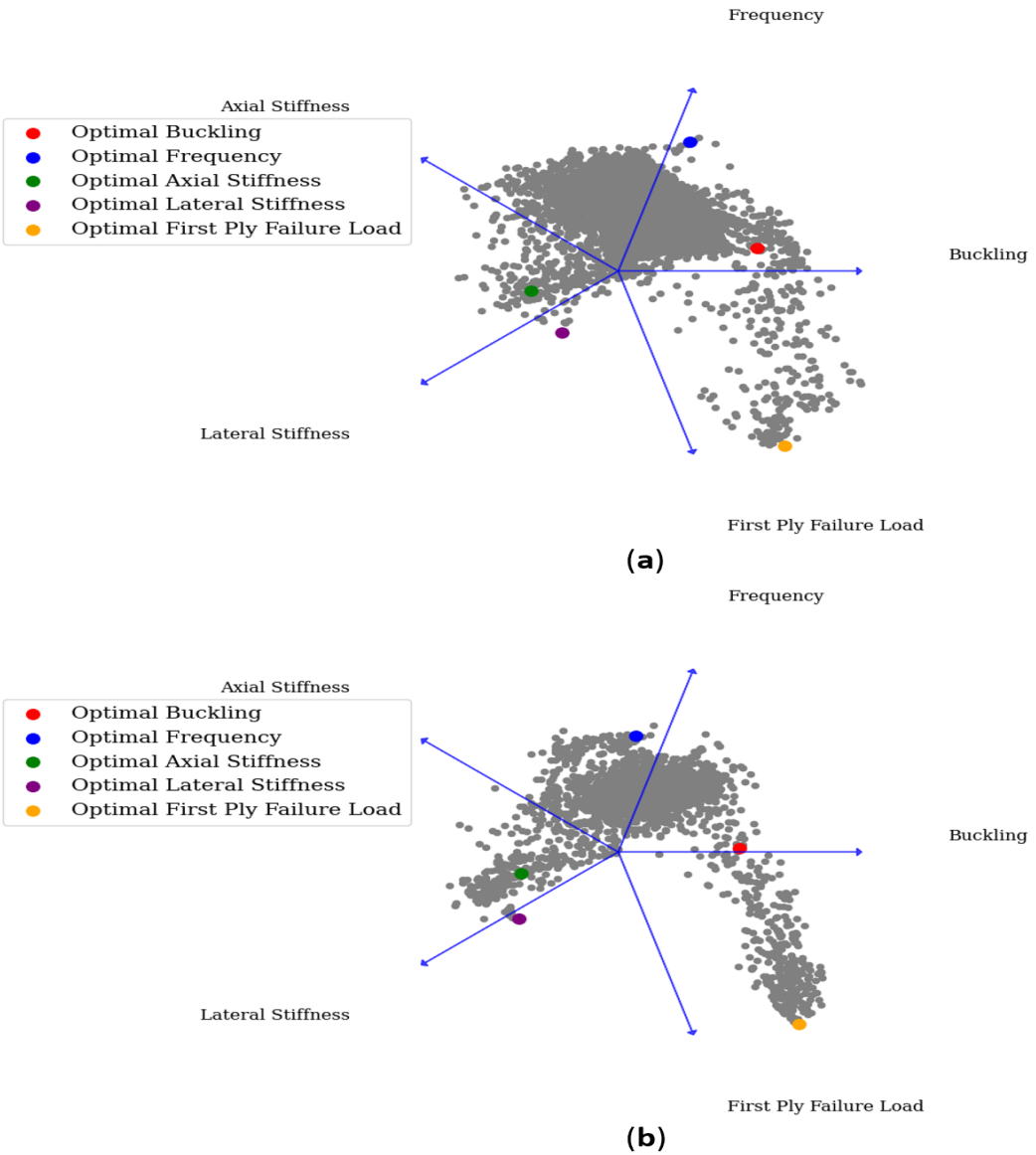


Figure 4.33: 2-D representation of the objective space for the CV case with (a) 3 and (b) 4 CPs.

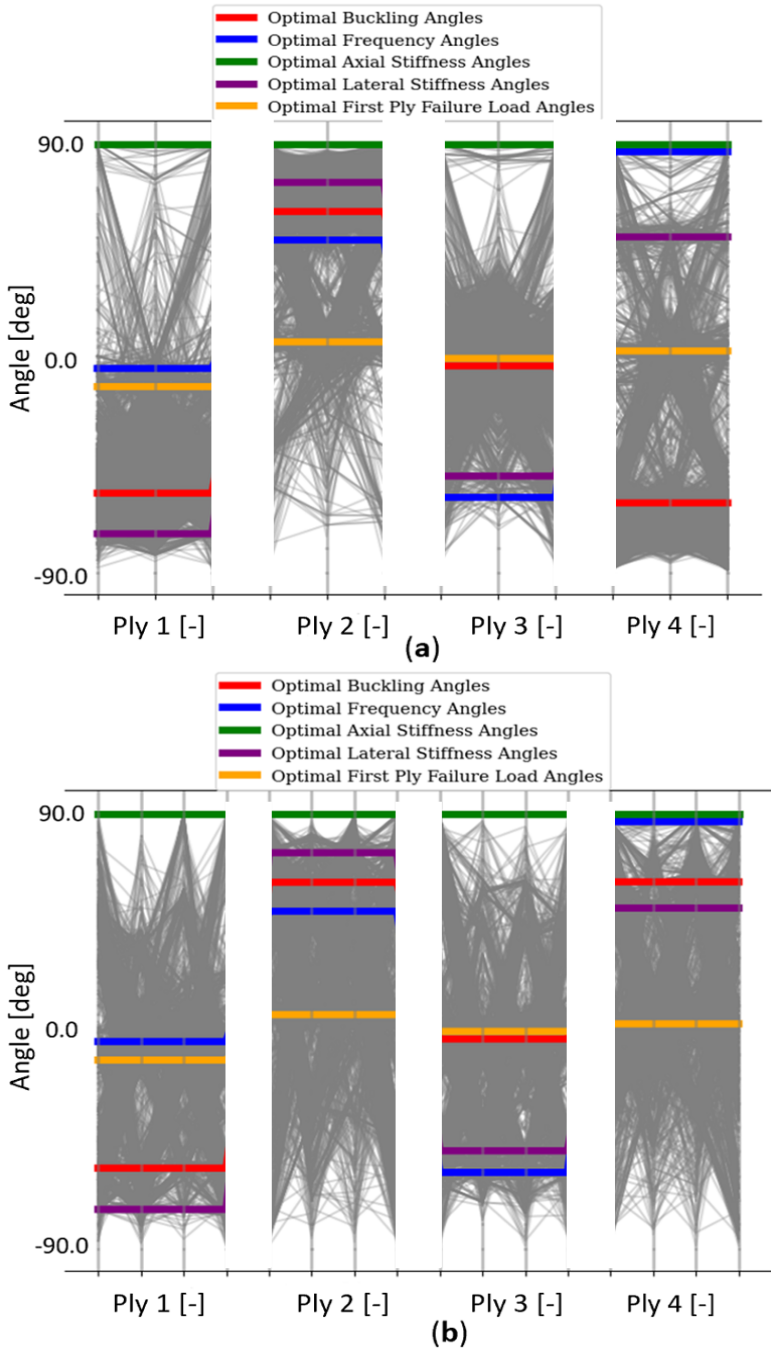


Figure 4.34: Parallel coordinate plot of the design space for the CV case with (a) 3 and (b) 4 CPs.

formed the predictions for the element properties is shown in Figure 4.35. The depth of each architecture is unsurprising, considering that it was very easy for the network to obtain a high accuracy for the angle predictions of each element, and therefore a shallow network was the optimal, while the opposite is true for the thickness predictions. In both cases the SELU architecture variant was deemed the most optimal.

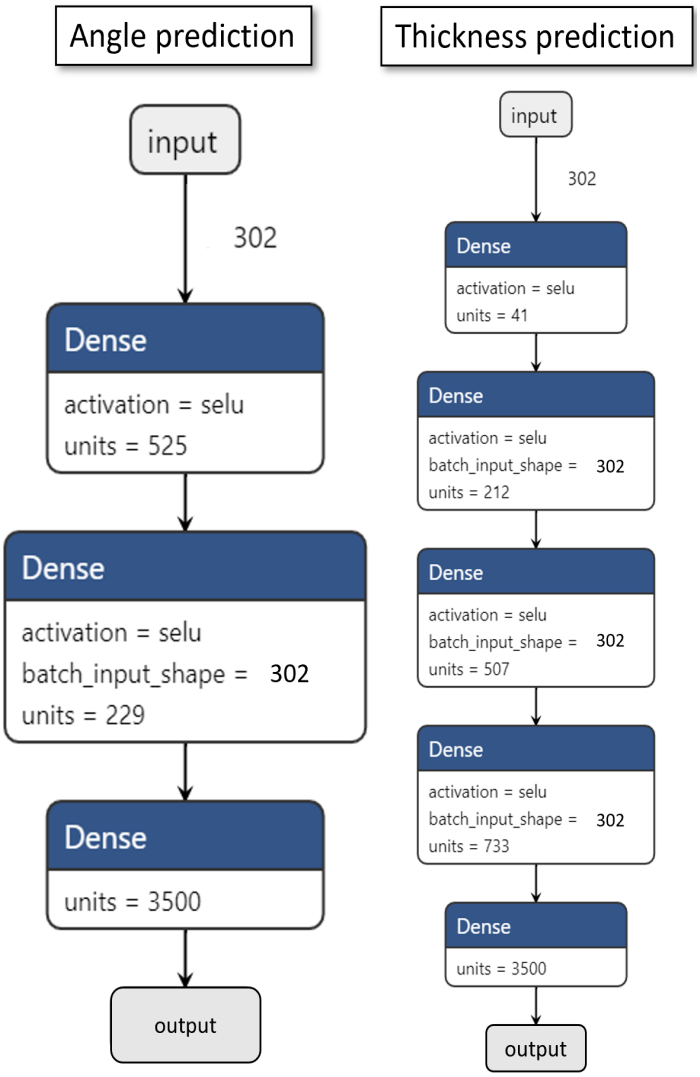


Figure 4.35: Neural network optimised architectures for the angle and thickness predictions of each element.

A comparison between different surrogate models was also carried out, which can be seen in Table 4.5. Although the difference is not significant in the case of the angle prediction, the picture is completely different for the thickness prediction of the elements. A possible cause for this vast difference between the models can be the poor performance behaviour for the RBF and Kriging models when they are required to make predictions for high-dimensional problems. This problem can be resolved by obtaining an excessive amount of data to compensate for this shortcoming [154]. Thus, the ANN is clearly superior in terms of accuracy for the element properties predictions, where a large input vector is present.

As with the CS case, the optimal architecture of the ANNs for each objective can be depicted as shown in Figure 4.36. This time, the ELU activation function with batch normalisation layers and the He initialisation method were preferred for most objectives, opposite to the CS instance. However, the prediction accuracy values for competing SELU architectures were very similar, although not shown here. Additionally, an architecture with 5 Dense HLs was preferred for the first time, for the FPF load objective. It was observed that during training, more than 4 HLs typically resulted in worse accuracy but that was not the case for this objective.

A representative ANN training is presented in Figure 4.37, similarly to the CS case. Once again, the Nadam

Table 4.5: Comparison between different surrogate models on the accuracy of the predictions of the element properties.

Angle							
Metric	ANN	RBF	Kriging	Metric	ANN	RBF	Kriging
Normalised MSE	0.55	1.00	0.79	MSE	0.8	1.45	1.15
Normalised MAE	0.53	1.00	0.71	MAE	0.63	1.18	0.84
Normalised R^2	1.00	0.92	0.98	R^2	0.99	0.92	0.98
Thickness							
Metric	ANN	RBF	Kriging	Metric	ANN	RBF	Kriging
Normalised MSE	0.08	0.99	1.00	MSE	9.30E-09	1.11E-07	1.11E-07
Normalised MAE	0.39	0.93	1.00	MAE	3.60E-05	8.52E-05	9.10E-05
Normalised R^2	1.00	0.22	0.2	R^2	0.93	0.21	0.18

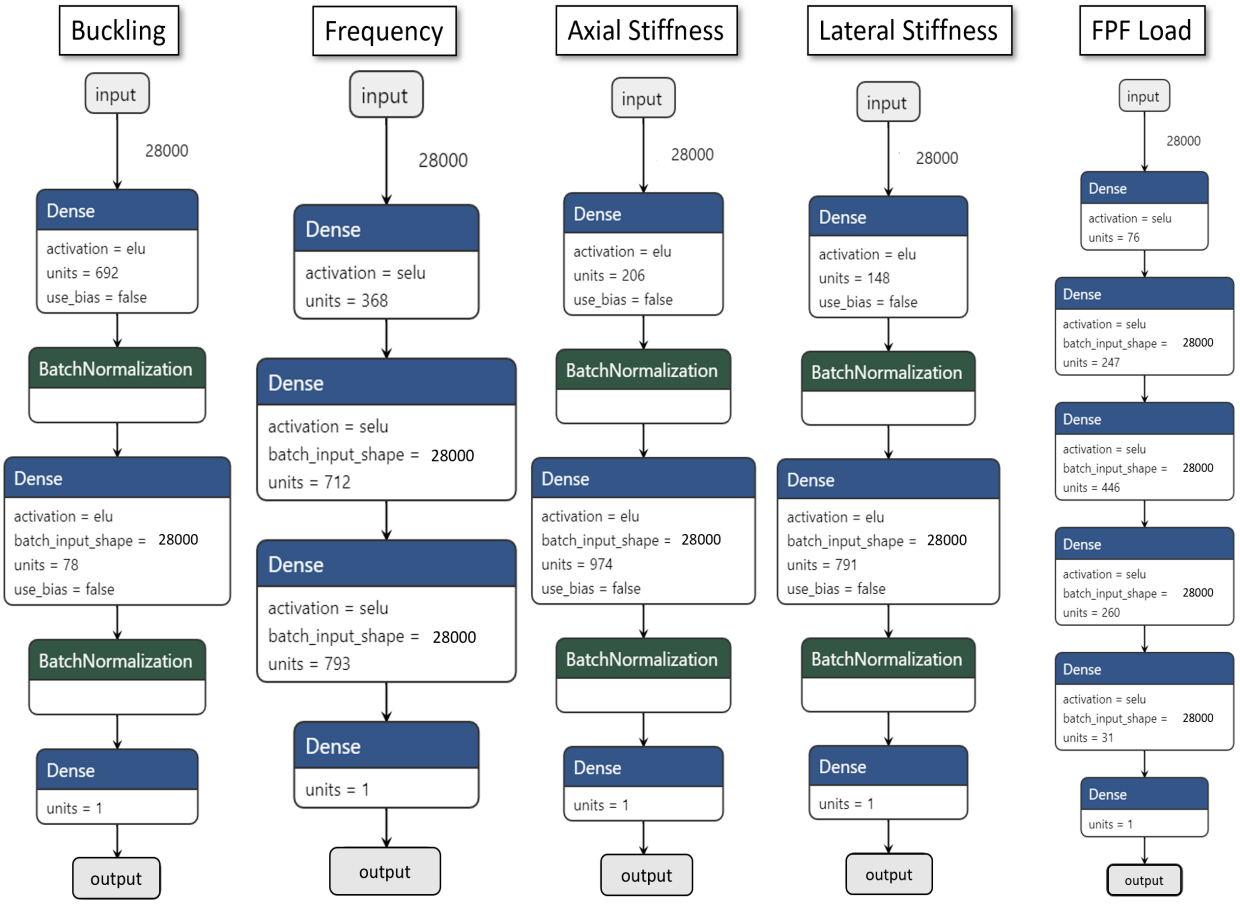


Figure 4.36: Neural network optimised architectures for each objective.

algorithm is able to minimise the MSE cost function smoothly for both the training and validation data. The absence of overfitting suggests that regularisation would likely increase the error if implemented. It is also noted that the algorithm does not require a lot of epochs to converge and therefore, it is promptly terminated by the early stopping criterion.

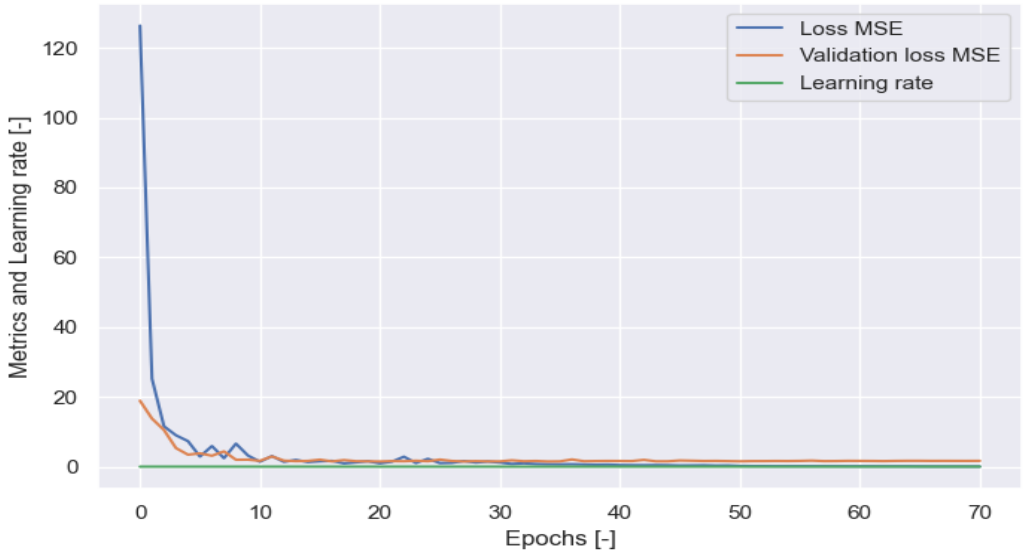


Figure 4.37: Representative minimisation of the ANN cost function for the VS case.

As with the element properties predictions, so too with the optimisation objectives, the ANN outperforms all other surrogate models, presented in Table 4.6. This can likely be attributed to the high-dimensionality of the problem, as was explained in the element properties comparison. The Kriging model is distinctly better than the RBF model and it seems that it can perform predictions much closer to the ANN values. It can be seen that the ANN achieves the highest accuracy for each objective, although it does not always surpass the 90% mark as with the CS case. The accuracy in both the CS and VS cases follow similar patterns however, with the buckling and FPF load objectives exhibiting the lowest values, highlighting their complexity.

The computational times to train the surrogate models and for them to make predictions are compared in Figure 4.38. The shallower and deeper network architectures were the ones depicted in Figure 4.36. The training and prediction times are both given in a logarithmic scale. The training times are quite different this time around, as compared to the CS results. The increase in dimensionality did not affect the ANNs in a meaningful way, although this can be attributed to a large extent to the GPU carrying out the computations, as there is a better scaling with more variables and data. Regarding the training time, the RBF model is similar to the deep network, while the Kriging model was the slowest one by many orders of magnitude. On the other hand, the Kriging model could output the fastest predictions and the RBF model was the worst performer in this category. In all cases, the deeper ANN required more training time and was slower at making predictions than its shallower counter-part.

A similar barplot is given in Figure 4.39 for the training and prediction times of the element properties. This time the inputs of the surrogate models were in the hundreds and not in the thousands as with the prediction of the objectives. The training times are very similar in both cases, although the aforementioned trends are less pronounced in this case. The only difference that can be observed is that the neural networks achieved the fastest prediction times by surpassing the Kriging model, which was the best in the previous instance.

The total optimisation time that was saved by utilising the ANNs can also be calculated and it is presented in Table 4.7. When compared to the time that was saved for the optimisation of the CS cylinder, a significant change in the total days required to complete the optimisation, with and without surrogate models, is noticed. More specifically, the CS optimisation saved 8.6 days , while a saving of 145.8 days was achieved in the VS case. The reason behind this difference can be attributed to the large amount of time that is needed to obtain the element properties in the proposed manner. This highlights the fact that a less accurate but faster

Table 4.6: Comparison between different surrogate models on the accuracy of the predictions of the objectives for the VS cylinder case.

Buckling							
Metric	ANN	RBF	Kriging	Metric	ANN	RBF	Kriging
Normalised MSE	0.19	1.00	0.56	MSE	2.54E+09	1.29E+10	7.21E+09
Normalised MAE	0.45	1.00	0.71	MAE	32970.57	73108.15	51933.87
Normalised R^2	1.00	0.02	0.55	R^2	0.80	0.01	0.44
Frequency							
Metric	ANN	RBF	Kriging	Metric	ANN	RBF	Kriging
Normalised MSE	0.10	1.00	0.45	MSE	69.99	668.43	300.26
Normalised MAE	0.28	1.00	0.61	MAE	5.70	20.45	12.56
Normalised R^2	1.00	0.00	0.62	R^2	0.90	0.00	0.55
Axial Stiffness							
Metric	ANN	RBF	Kriging	Metric	ANN	RBF	Kriging
Normalised MSE	0.08	1.00	0.35	MSE	1.16E+15	1.45E+16	5.13E+15
Normalised MAE	0.24	1.00	0.48	MAE	2.15E+07	8.77E+07	4.21E+07
Normalised R^2	1.00	0.00	0.70	R^2	0.92	0.00	0.64
Lateral Stiffness							
Metric	ANN	RBF	Kriging	Metric	ANN	RBF	Kriging
Normalised MSE	0.11	1.00	0.17	MSE	4.64E+12	4.20E+13	6.97E+12
Normalised MAE	0.26	1.00	0.34	MAE	1.40E+06	5.34E+06	1.82E+06
Normalised R^2	1.00	0.00	0.94	R^2	0.89	0.00	0.83
FPF load							
Metric	ANN	RBF	Kriging	Metric	ANN	RBF	Kriging
Normalised MSE	0.16	1.00	0.59	MSE	7.06E+08	4.43E+09	2.62E+09
Normalised MAE	0.27	1.00	0.68	MAE	12859.89	47017.07	31948.57
Normalised R^2	1.00	0.00	0.48	R^2	0.84	0.00	0.41

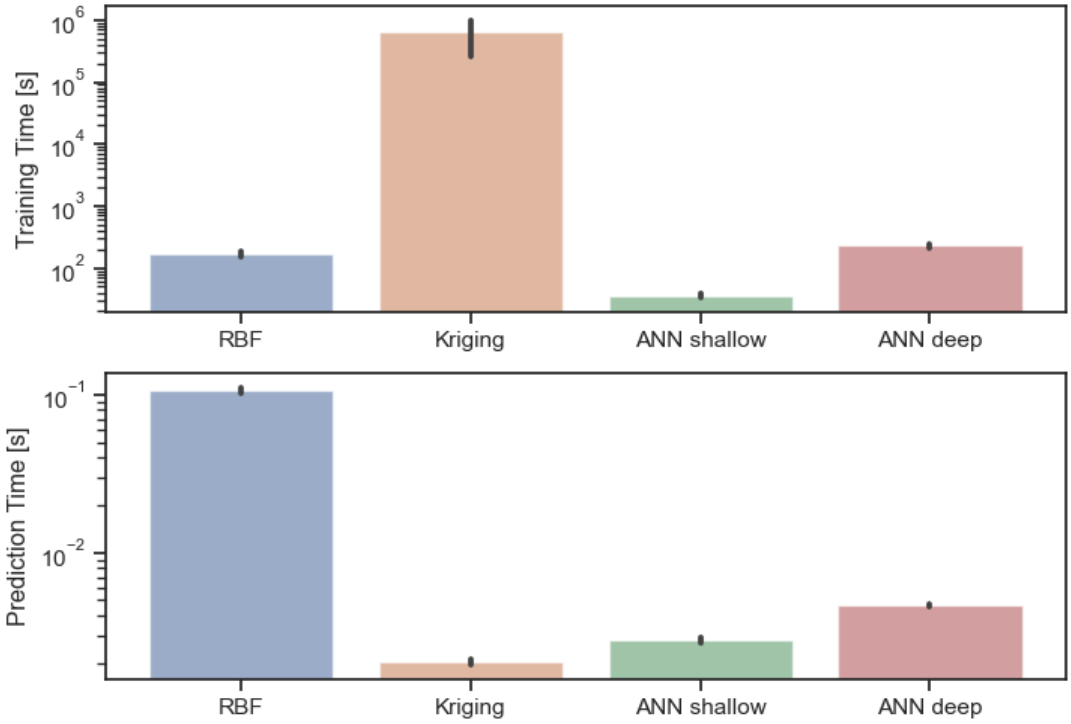


Figure 4.38: Computational times for the training and value prediction between different surrogate models for the optimisation objectives.

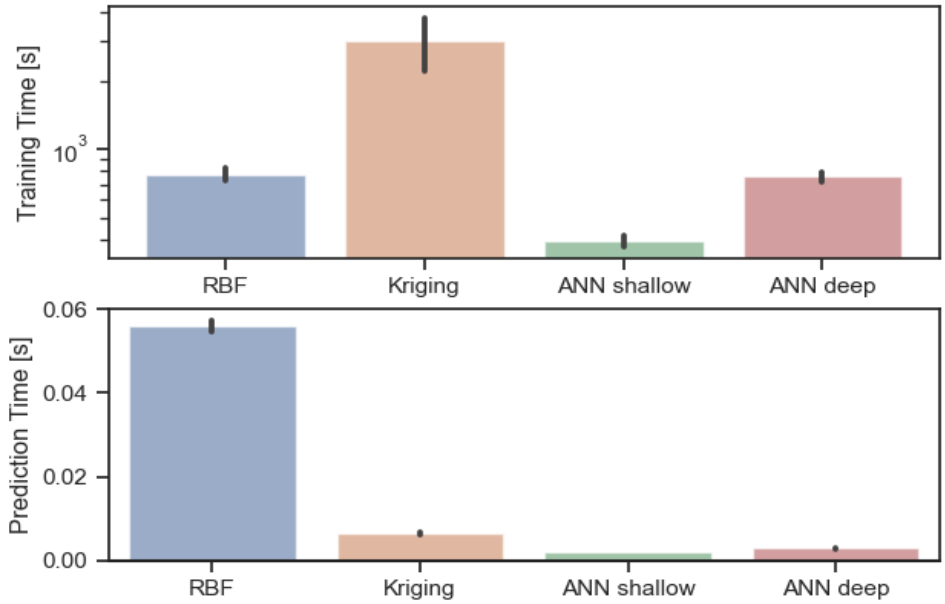


Figure 4.39: Computational times for the training and value prediction between different surrogate models for the element properties.

Table 4.7: A comparison of the total time required for the optimisation to complete with and without the use of surrogate models in the case of variable stiffness cylinders.

Structural Objective	Average Time to Obtain Element Properties [s]	Average Time to Predict Element Properties [s]	Average FEA Time [s]	Optimisation Time Required w/o Surrogate Models [days]	Optimisation Time Required with Surrogate Models [days]	Time Saved [days]
Buckling	-	-	65.0	41.9	4.0	37.9
Frequency	-	-	28.5	39.8	4.0	35.9
Axial Stiffness/FPF Load	-	-	32.5	40.1	4.0	36.1
Lateral Stiffness	-	-	28.5	39.8	4.0	35.9
Total	688.7	71.1	154.4	161.6	15.8	145.8

approach of calculating the element properties might be a desirable alternative if surrogate models are not to be implemented.

Angle variation model comparison

In order to identify the advantages and disadvantages of each angle variation approach, a comparison between them was performed, including the CS results. To that end, a weighted-sum scalar decomposition with equal weights for each objective was carried out, for each angle variation method, in order to compare their decomposed scalar values. The lowest scalar values were deemed to be of no interest, and thus only the top 100 all-around designs were selected for the comparison. These results were all viable designs and their objective values were normalised with respect to each objective category. Once more, the additional mass due to any overlaps was taken into account by standardising the results. This comparison can be seen at the top left of Figure 4.40 (blue chart). Additionally, each individual objective was investigated separately to obtain further insight.

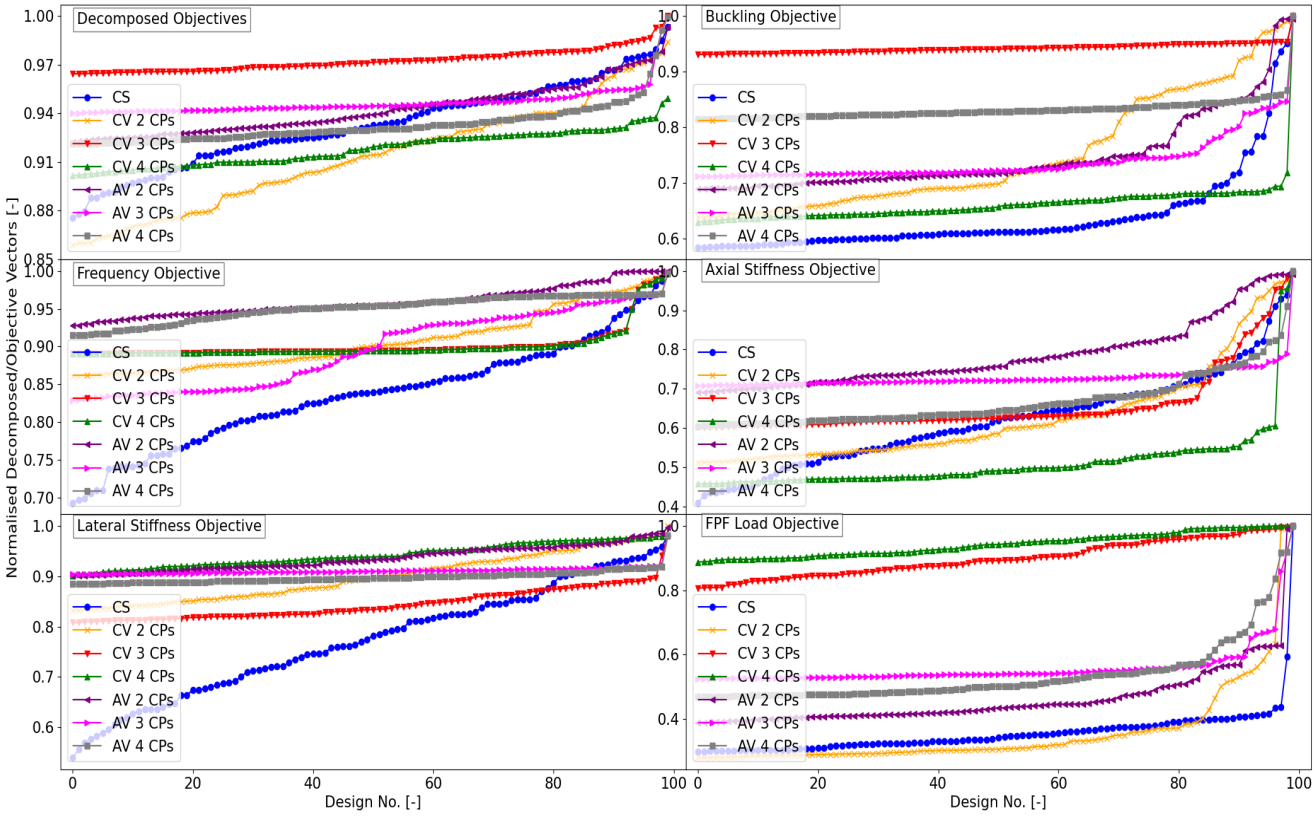


Figure 4.40: Comparison between angle variation models.

The decomposed values indicate that the CV with 3 CPs contains the best all-around design, while the AV with 4 CPs was a close second, although it only rose up towards the final designs. The CS designs exhibit

greater variation for each design, with its best design coming in 3^{rd}, although its designs performed poorly for the most part. It seems that an increase in the amount of CPs has a positive effect for the case of AV, but the same cannot be said for the CV, where it exhibits the worst performance out of all models for the case of 4 CPs.

When it comes to the specific objectives of the selected designs, it is observed that the various angle variation schemes perform better for certain objectives. For instance, the AV designs are able to attain superior frequency, axial and lateral stiffness values for most of the top 100 designs that were considered in their respective optimisations, although they were able to surpass the CS optimal design only for the latter case. The CV designs did not perform especially well for any case. The CS case exhibited the greatest variability overall for the considered top designs but it was not as pronounced for the buckling and FPF load objectives.

4.2. Cut-out cylinder optimisation

The cylinder with cut-outs is presented in this section. The layout is similar to the pristine cylinder since the objectives of interest are identical in both cases.

4.2.1. Constant stiffness cylinder

The baseline optimisation of the CS cylinder optimisation is given first. The convergence of the optimisation algorithm can be seen in Figure 4.41, where the algorithm did not converge, as it is evident from the hypervolume calculation. Additionally, surrogate model predictions were utilised to advance the optimisation, even though they resulted in a reduction of the hypervolume.

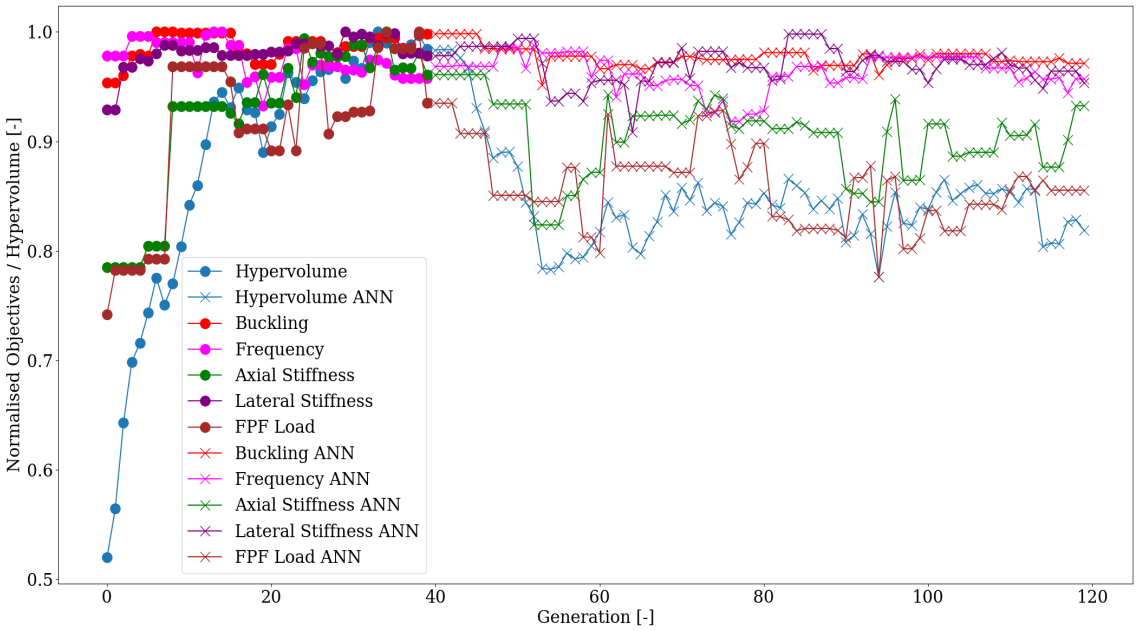


Figure 4.41: Hypervolume convergence of the CS cut-out cylinder.

In Figure 4.42, it can be seen that the ANN has greater overlap in its objective predictions, but this could also be attributed to the generated design space by the optimisation algorithm for the later generations in question. Additionally, the ANN predictions did not manage to surpass the FEM results.

The correlation between the objectives shows a good positive correlation between buckling and lateral stiffness, and axial stiffness and FPF load. The most significant difference between the CS designs of the cut-out and pristine cylinders is that in this case, the frequency exhibits a negative correlation with respect to buckling.

The objective space is visualised in Figure 4.44, where it can be seen that there is a good balance between the relative values of the different objectives. The axial stiffness and FPF load optimal designs are quite close, further highlighting the high correlation between the two objectives. Apart from the optimal FPF load objective, the rest of the objectives do not occupy extreme positions, indicating that these designs are versatile towards other objectives as well.

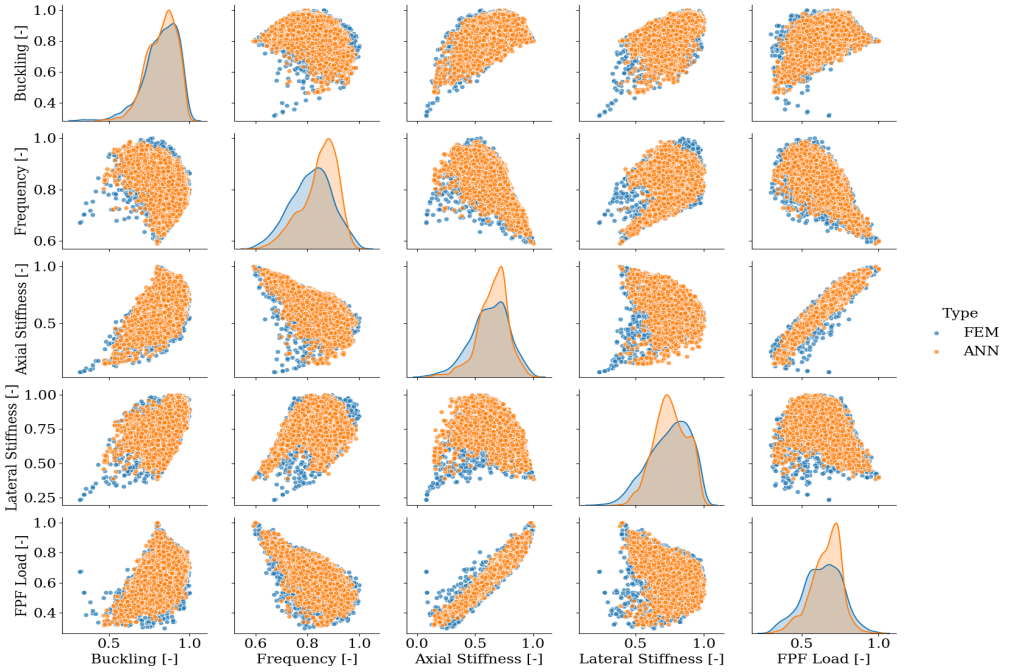


Figure 4.42: Objective relationship for both FEM and ANN results in the CS case of the cut-out cylinder.

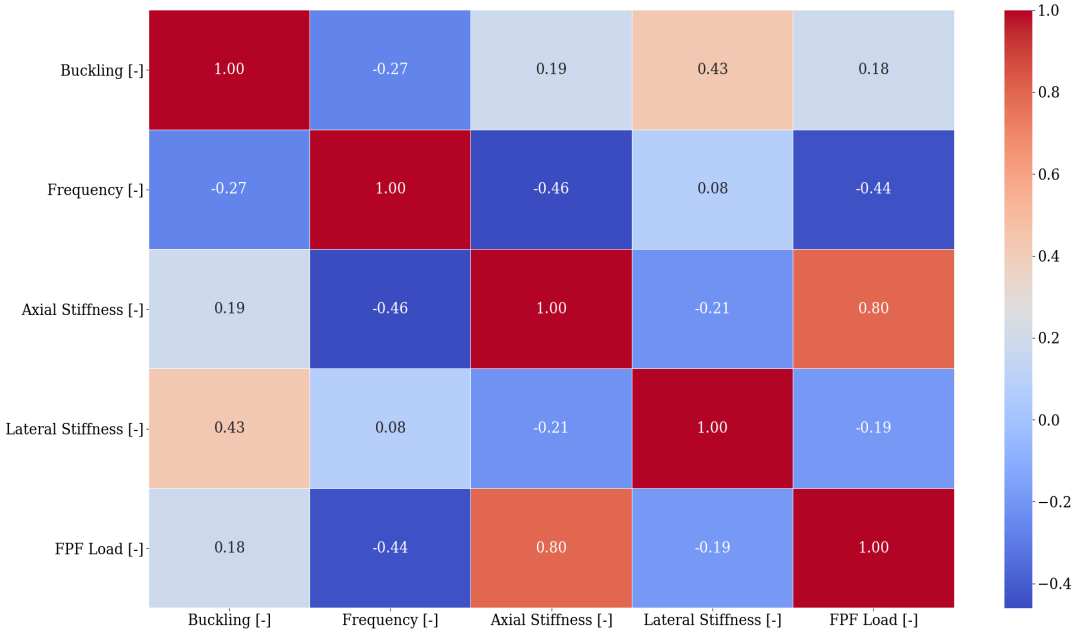


Figure 4.43: Correlation heatmap between objectives for the CS case of the cut-out cylinder.

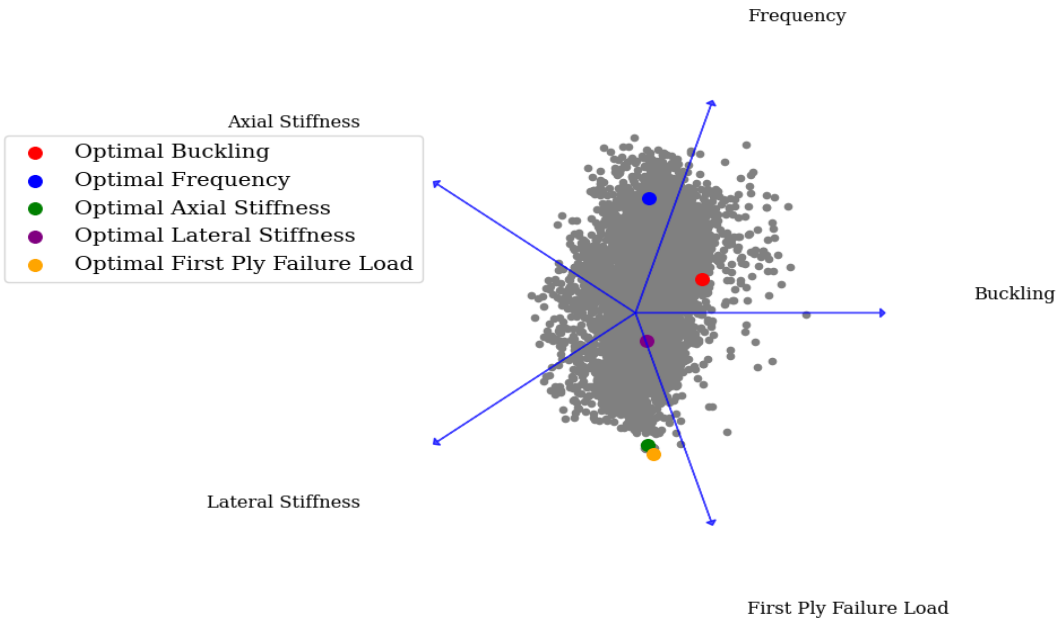


Figure 4.44: 2-D representation of the objective space for the CS case of the cut-out cylinder.

The relative objective values of the most interesting designs are investigated more clearly in Figure 4.45. It is observed that the axial stiffness and FPF loads are almost identical in that regard, while the rest of the designs exhibit similar values for these two objectives and possess equally high values for the remainder ones.

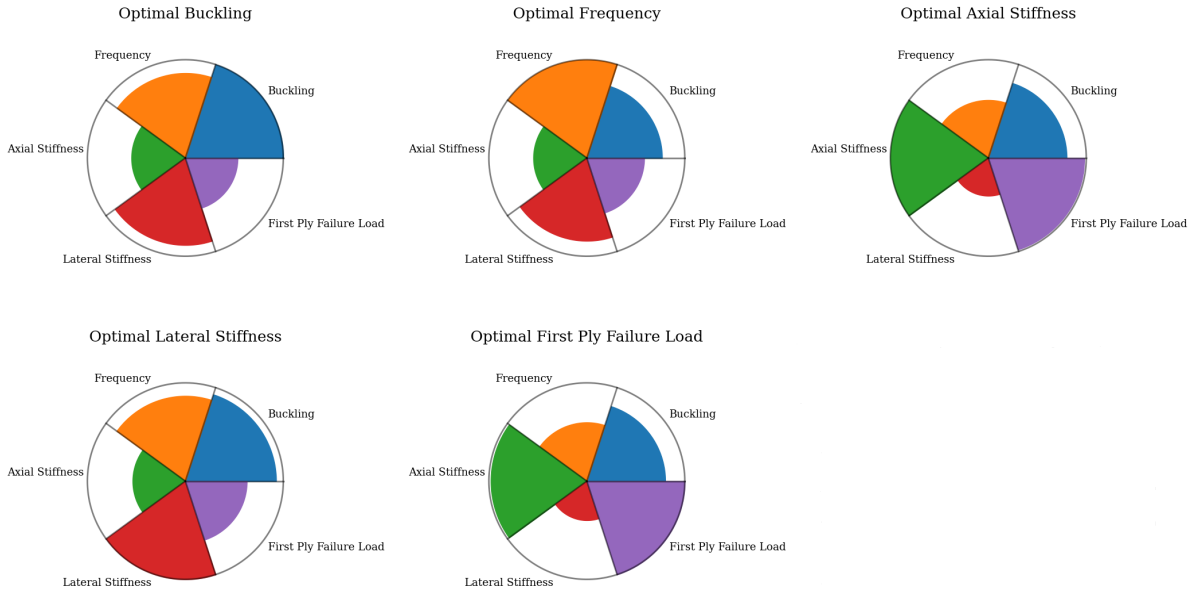


Figure 4.45: Petal plots of the optimal solutions for the CS case of the cut-out cylinder.

A visual comparison of the 1^{st} modes for the optimal designs for the buckling and free vibrations frequency objectives is shown in Figure 4.46 (a) and (b), as well as the modes of the QI laminate, respectively in (c) and (d). As can be noticed, the buckling modes are local in both cases, with the QI mode buckling around both cut-outs. In contrast, the optimal buckling design buckles only around the axially-elongated cut-out, albeit in a more pronounced way. On the other hand, the frequency modes are more similar with the highest displacement originating from the axially-elongated cut-out, although the optimal frequency design exhibits a longer displacement field, which stops at the circumferentially-elongated ellipse.

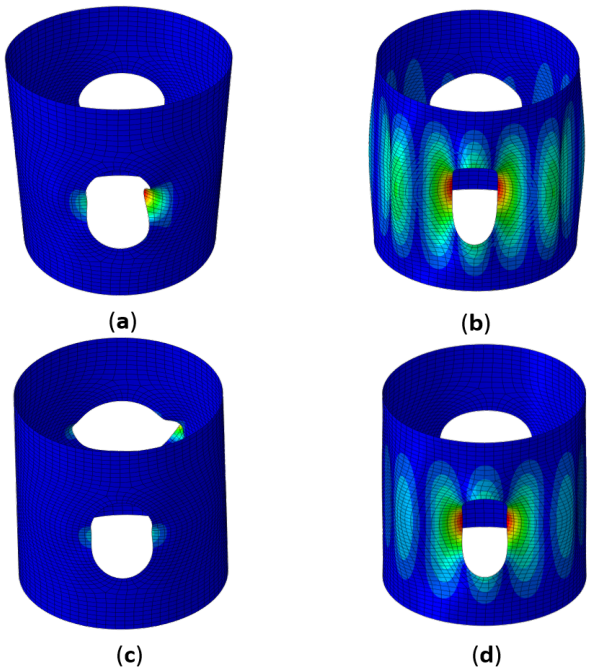


Figure 4.46: Depiction of the first modes for the objectives of (a) buckling and (b) free vibrations frequency for the optimised cut-out cylinders and likewise in (c) and (d) for the QI cut-out cylinders.

Table 4.8: The ply angles of the optimal designs for the CS case of the cut-out cylinder.

Objective	1st Ply Angle [deg]	2nd Ply Angle [deg]	3rd Ply Angle [deg]	4th Ply Angle [deg]
Optimal Buckling Angles	-71.3	53.9	82.4	-52.9
Optimal Frequency Angles	9.5	-35.7	-79.7	71.0
Optimal Axial Stiffness Angles	90.0	90.0	90.0	90.0
Optimal Lateral Stiffness Angles	54.1	-54.7	75.6	-75.5
Optimal First Ply Failure Angles	90.0	90.0	90.0	90.0

The design space of the optimisation can be seen in Figure 4.47. It is observed that plies with angles close to 0° were avoided for the most part. Furthermore, the optimal axial stiffness and FPF load designs were practically the same designs, oriented towards the cylinder's axis. Once more, the optimal lateral stiffness design was balanced, while no distinct pattern is found for the buckling and frequency objectives, which were quite different designs with respect to the pristine cylinder. For clarity, the optimal designs are given in tabular form in Table 4.8.

The frequency of the locations where FPF occurred in the cylinder is presented in Figure 4.48. The strains around the circumferentially-elongated cut-out were the highest, leading to ply failure at the ends of the major axis of the ellipse, as expected. Since the mesh was not perfectly symmetrical about the cut-out, a bias is exhibited towards its right side, although it can be regarded as insignificant.

The extremities of the design space were investigated as there is a greater chance that inaccuracies are exhibited there, between the FEM results and the ANN predictions. Additionally, the relative difference of each objective for the QI laminate and the optimised designs is displayed in Table 4.9. The relative difference between designs is comparable to the pristine cylinder case, while the ANN predictions have a slightly higher inaccuracy overall, although still within acceptable levels.

In Figure 4.49, the optimal architecture of the ANNs for each objective can be seen. The networks are very similar with respect to the pristine cylinder case, where the SELU function was preferred overall. In both cases, the ELU function was chosen only for the lateral stiffness objective.

The surrogate models RBF and Kriging were also compared with the ANN in Table 4.10. It is observed that the accuracy is excellent in all cases for all the surrogate models. The ANN models did not perform the worst for each objective, but they did not manage to perform the best either. The RBF models appear to be the most consistent and yielding the best predictions, with the Kriging models as a close second.

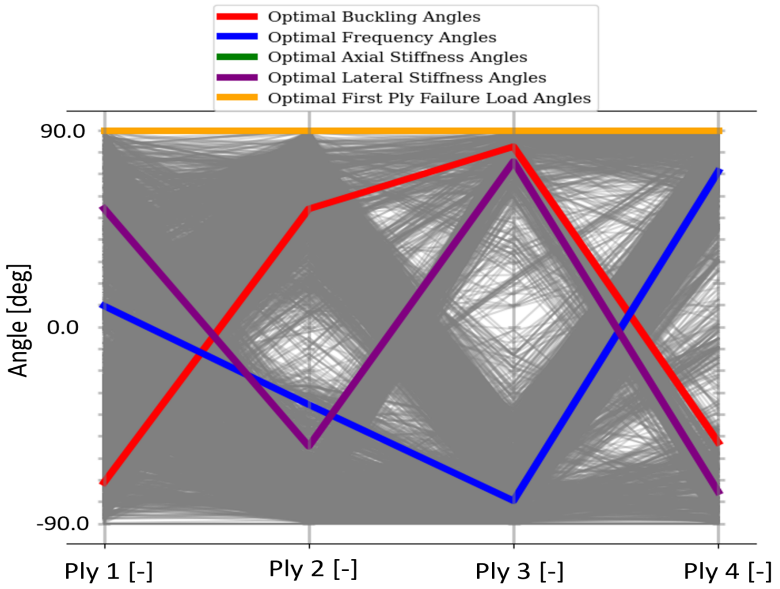


Figure 4.47: Parallel coordinate plot of the design space for the CS case of the cut-out cylinder.

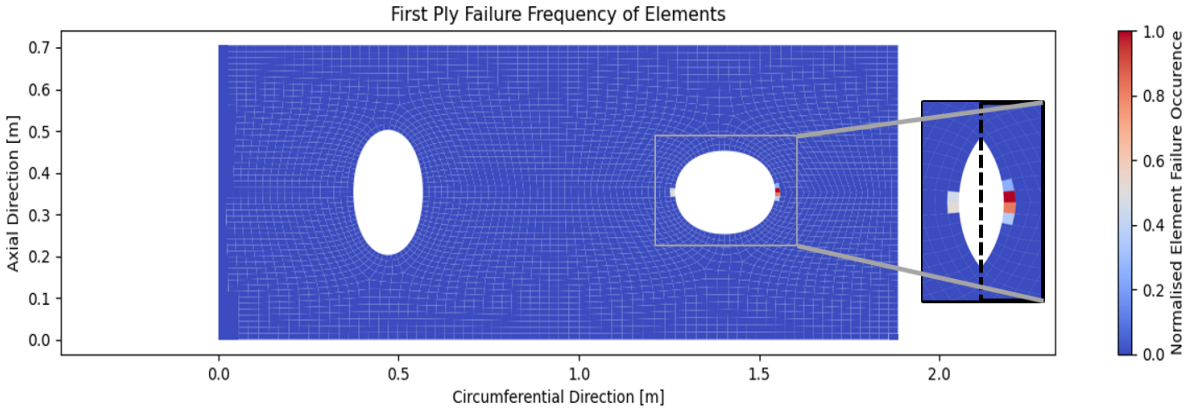


Figure 4.48: Frequency of elements where FPF took place.

Table 4.9: Comparison between predictions and analyses values for the best results of each objective in the CS case of the cut-out cylinder.

Optimal values	QI reference	Optimised FEA Designs	Relative difference between QI and optimised designs [%]	ANN prediction	Relative difference between FEM and ANN [%]
Buckling [kN]	60	77	21.3	79	0.3
Frequency [Hz]	308	319	3.3	319	0.2
Axial Stiffness [MN/m]	169	419	60.7	417	0.4
Lateral Stiffness [MN/m]	10	12	19	12	0.6
First Ply Failure [kN]	28	73	62.1	74	0.3

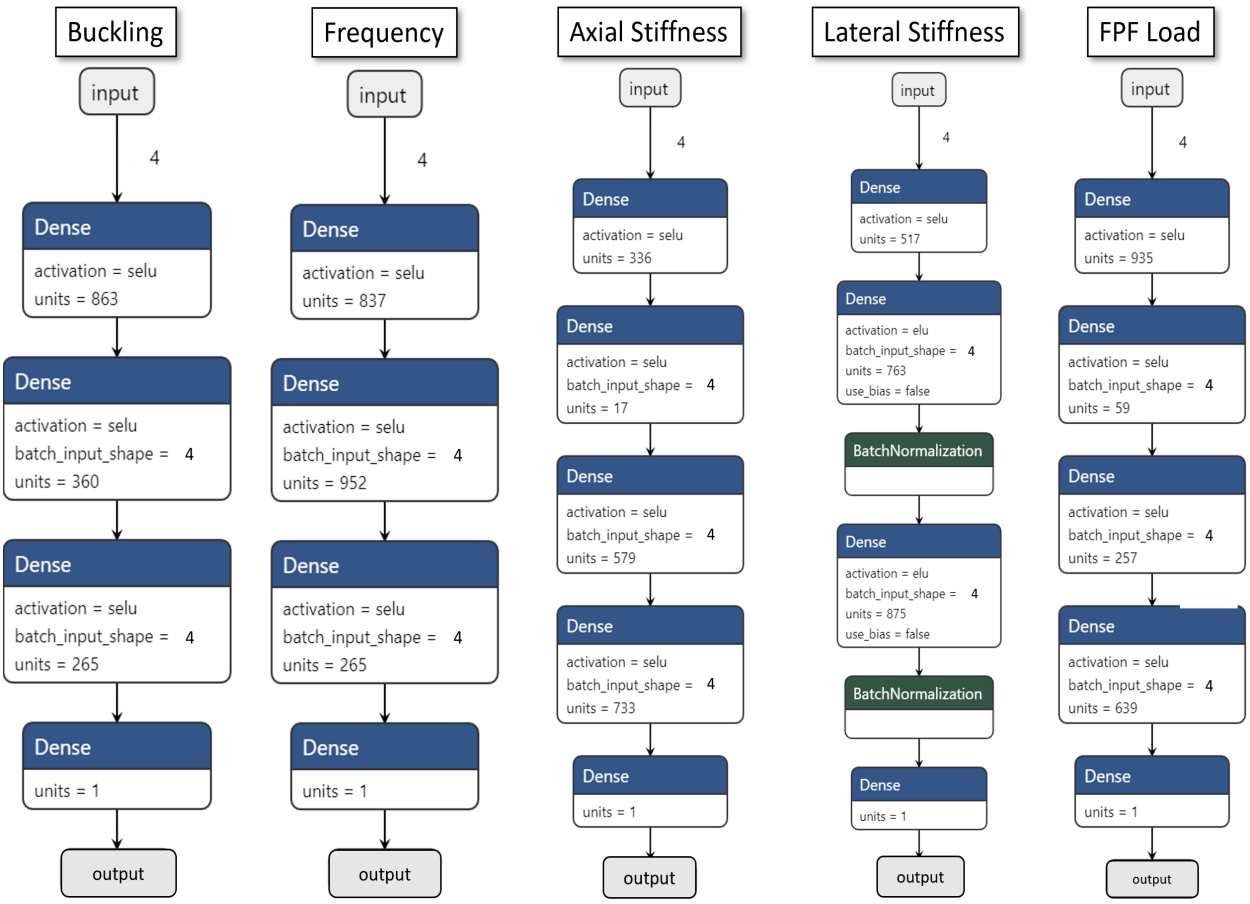


Figure 4.49: Neural network optimised architectures for each objective for the CS case of the cut-out cylinder.

Table 4.10: Comparison between the different surrogate models on the accuracy of the predictions of the objectives.

Buckling							
Metric	ANN	RBF	Kriging	Metric	ANN	RBF	Kriging
Normalised MSE	0.41	0.27	1.00	MSE	1.98E+06	1.30E+06	4.77E+06
Normalised MAE	1.00	0.66	0.75	MAE	983.00	650.00	741.00
Normalised R^2	0.99	1.00	0.95	R^2	0.97	0.98	0.93
Frequency							
Metric	ANN	RBF	Kriging	Metric	ANN	RBF	Kriging
Normalised MSE	1.00	0.92	0.56	MSE	17.4	15.94	9.76
Normalised MAE	1.00	0.83	0.46	MAE	2.61	2.16	1.20
Normalised R^2	0.99	0.99	1.00	R^2	0.97	0.98	0.99
Axial Stiffness							
Metric	ANN	RBF	Kriging	Metric	ANN	RBF	Kriging
Normalised MSE	1.00	0.28	0.52	MSE	1.92E+14	5.48E+13	1.01E+14
Normalised MAE	1.00	0.46	0.36	MAE	9.72E+06	4.49E+06	3.52E+06
Normalised R^2	0.97	1.00	0.99	R^2	0.96	0.99	0.98
Lateral Stiffness							
Metric	ANN	RBF	Kriging	Metric	ANN	RBF	Kriging
Normalised MSE	1.00	0.69	0.16	MSE	2.85E+11	1.96E+11	4.66E+10
Normalised MAE	1.00	0.72	0.2	MAE	3.61E+05	2.62E+05	7.38E+04
Normalised R^2	0.93	0.95	1.00	R^2	0.91	0.94	0.98
FPF load							
Metric	ANN	RBF	Kriging	Metric	ANN	RBF	Kriging
Normalised MSE	0.4	0.3	1.00	MSE	5.36E+06	4.07E+06	1.36E+07
Normalised MAE	0.9	0.73	1.00	MAE	1769.00	1449.00	1969.00
Normalised R^2	0.99	1.00	0.88	R^2	0.94	0.95	0.84

4.2.2. Variable stiffness cylinder

The results for the VS cylinder optimisations with cut-outs are presented in this section. The format of the section is similar to the one followed for the pristine cylinder, with the exception that the designs with 4 CPs are shown first, since they contain the FEM analyses used to train the surrogate models.

Axial Variation - 4 Control Points

The convergence of the optimisation for 4 CPs for the AV case is presented first. Relevant to the convergence is the normalised hypervolume and the individual objective values for each generation, which are shown in Figure 4.50. The hypervolume remains constant over many generations, indicating that the algorithm has sufficiently converged, as is also evident by the objective values that exhibit similar behaviour.

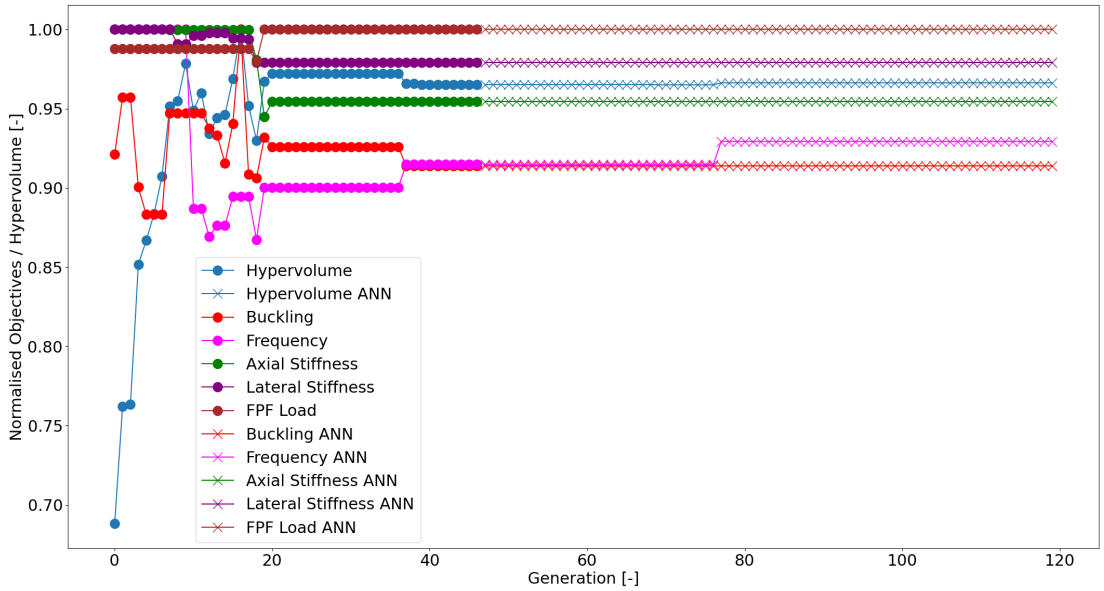


Figure 4.50: Generational objective change for the AV case with 4 CPs for the cut-out cylinder.

The relation among the objectives and the differences between the FEM and ANN predictions are given in Figure 4.51. It is observed that the ANN predictions yielded some higher mass designs than the FEM data but nevertheless their density peak matched. It also seems that the ANN predictions avoided the extremities and were confounded within the respective cloud of points, especially for the axial stiffness objective.

The feasibility of the generated designs can be seen in Figure 4.52. The results are very similar to what was observed in the pristine cylinder counterpart, where the feasible designs yielded superior results over the infeasible ones and also that the constraint violation was mainly driven by the excess mass. The only exception where the feasible and infeasible designs overlap, in terms of how optimal they are, is for the frequency objective. This possibly indicates that the additional mass does not reduce the effectiveness of this objective to a significant degree.

The Spearman correlation method is utilised to quantify the presented relationships in Figure 4.53. Unlike the pristine cylinder results, the correlations are more significant this time around, although the same trends are displayed in both instances. The greatest difference is found for the correlation between the buckling and lateral stiffness objectives, which now have a strong positive correlation, while in the pristine case there was no correlation to speak of. A negative correlation between the mass and the objectives is observed, except for the frequency objective as it was hinted at in Figure 4.52.

In Figure 4.54, a star coordinate plot depicts the objective space of the optimisation. The plot suggests that most of the designs possess good strength values in a relative sense, particularly in the expense of axial stiffness. Additionally, the goodness of the buckling results is significantly more visible than the pristine case.

A focus on the optimal designs is given in Figure 4.55, where the relative superiority of the objective values is shown in a petal plot. The objectives to focus on are the buckling and FPF load, as they are the ones that differ from the optimal designs found in the CS cylinder. The respective objectives were improved by these VS designs and in addition they exhibit great values for the rest of the objectives. The buckling design obtained a 14% buckling increase in the expense of a 7.5% increase in mass after being mass standardised. The increase

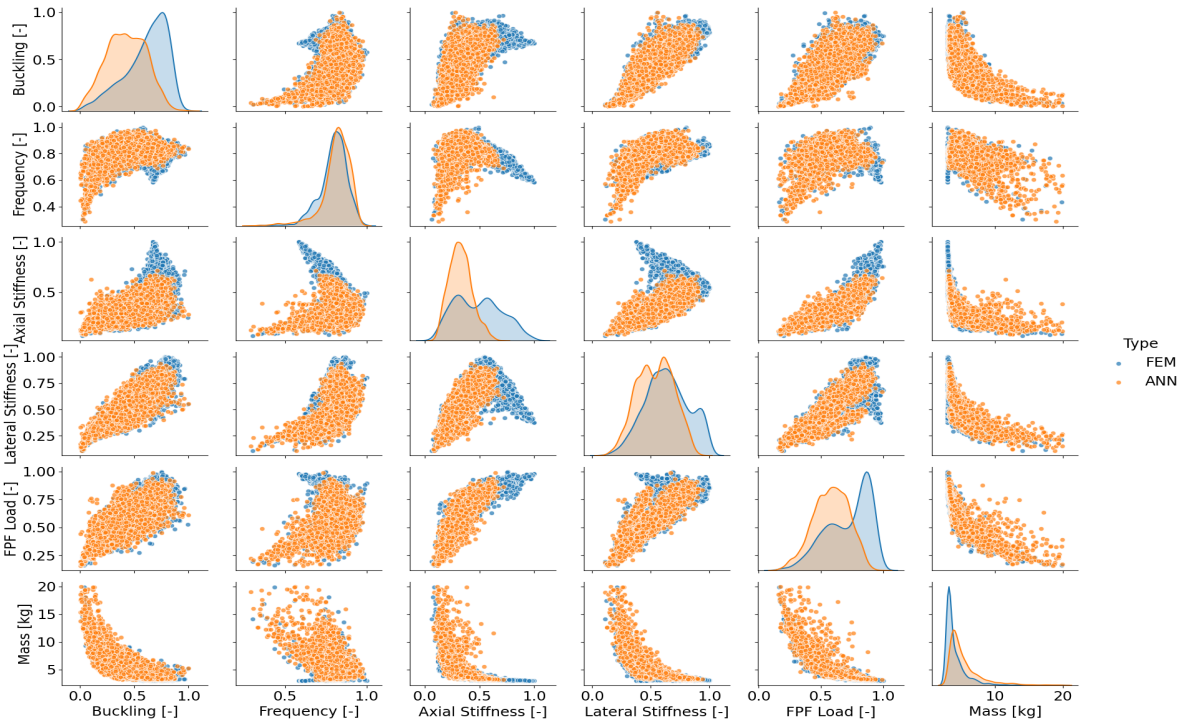


Figure 4.51: Objective relationship for both FEM and ANN results in the AV case with 4 CPs for the cut-out cylinder.

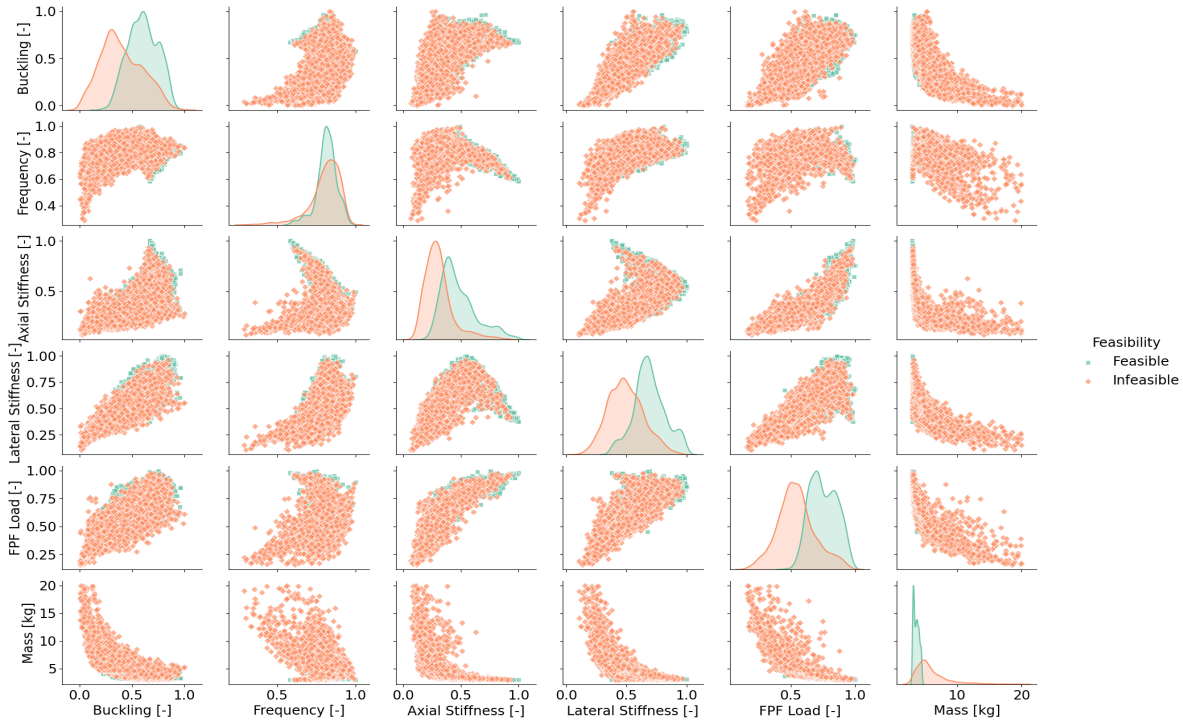


Figure 4.52: Design feasibility with respect to the imposed constraints for the AV case with 4 CPs for the cut-out cylinder.

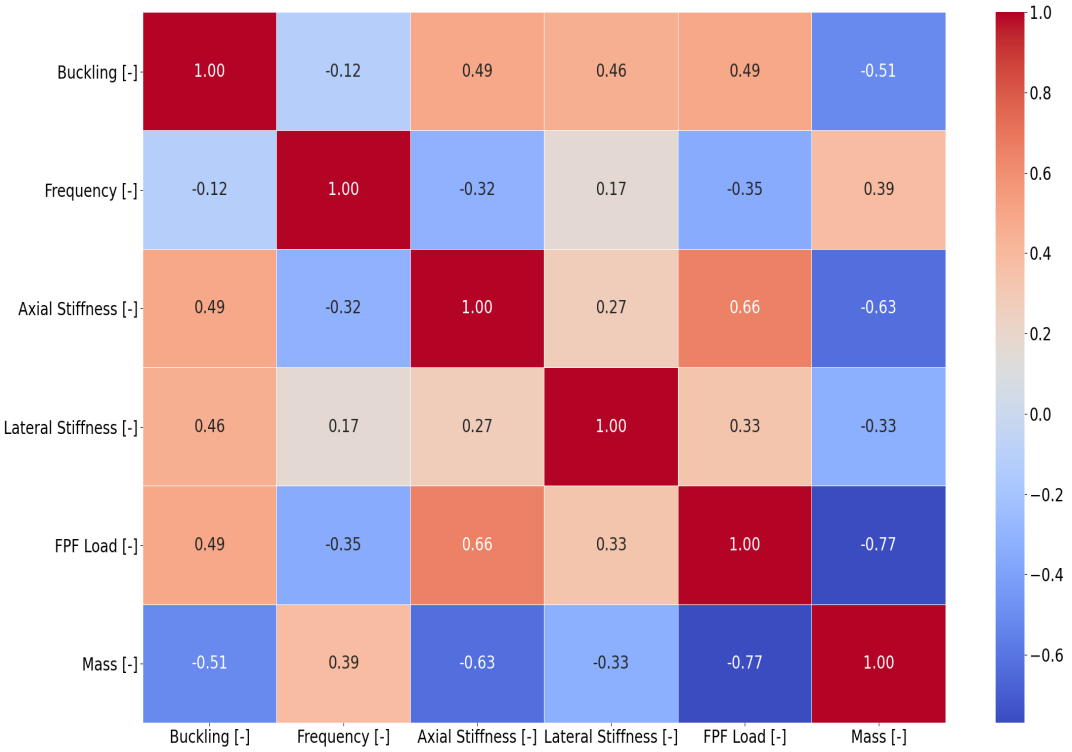


Figure 4.53: Correlation heatmap between objectives for the AV case with 4 CPs for the cut-out cylinder.

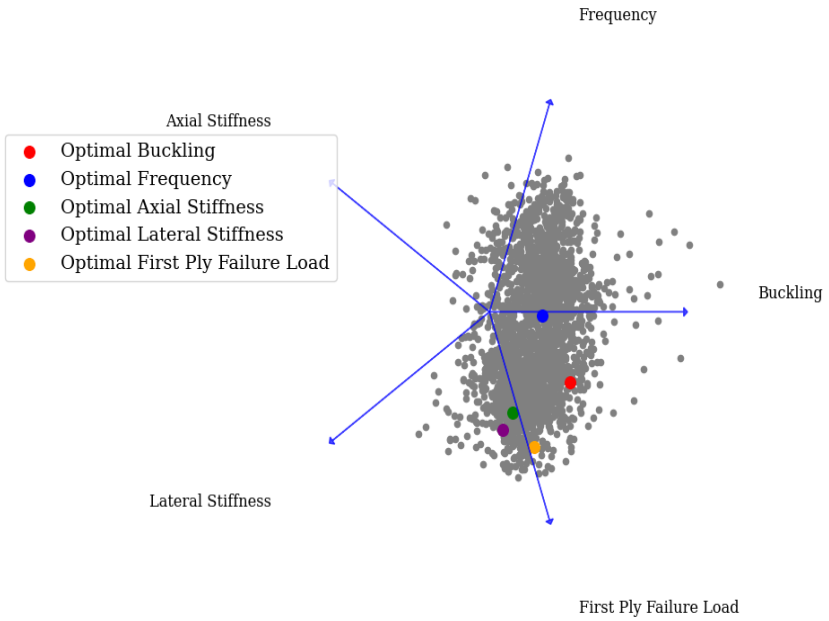


Figure 4.54: 2-D representation of the objective space for the AV case with 4 CPs for the cut-out cylinder.

for the FPF load is not as significant but it comes at a lesser mass increase, where the gain is 2.7% against a 1% mass increase.

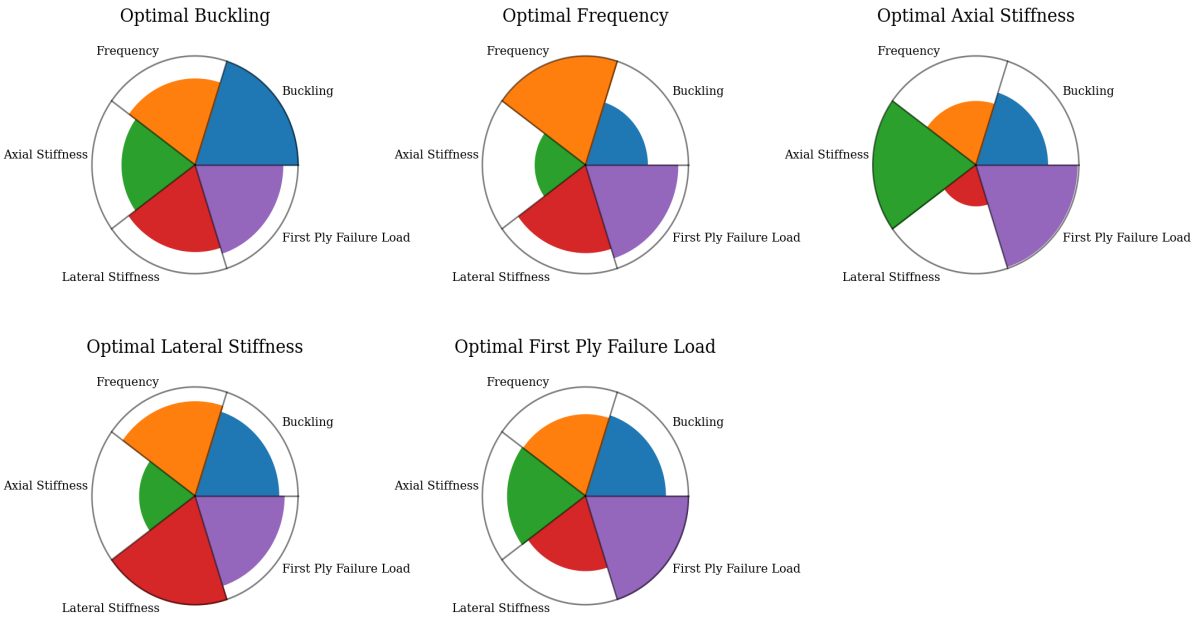


Figure 4.55: Petal plot of the optimal solution for the lateral stiffness objective in the AV case with 4 CPs for the cut-out cylinder.

The optimal designs, as well as the rest of the design space is depicted in Figure 4.56. It can be seen that designs with plies around the 0° region were avoided in general. The optimal buckling design shows the highest angle variation with the angles not straying too far away from the $90°/-90°$ domain. Therefore, the angles are more aligned with the axis of the cylinder and diverge starting from the bottom up. The optimal FPF load design exhibited the same trend, although the angle variation was minimal.

Circumferential Variation - 4 Control Points

The case of the circumferential angle variation for 4 CPs is presented in this section. Firstly, the optimisation's hypervolume is investigated in Figure 4.57. Unlike the axial angle variation case, the algorithm has not converged as is evident by the sudden drop of the hypervolume towards the last generations. The FPF load was stagnant over most of the generations and was the culprit for the hypervolume drop, in combination with a reduction in axial stiffness.

A pair plot between each objective is shown in Figure 4.58. Once more, the ANN predictions mainly occupy the interior of each cloud of points and the distribution curves align to a greater degree than in the AV case. Additionally, designs with a high mass are generated by both FEM and ANN predictions, while in the AV case it was mainly the ANN that produced high mass results.

Figure 4.59 depicts whether the designs considered in the optimisation were viable or not. It is found that the extra mass resulted in most of the infeasibility. The feasible designs were superior in terms of objective values over the infeasible ones. A notable difference with the AV case is that the frequency objective does improve for feasible designs, although this is not exhibited for the FPF load objective this time around.

The correlation between objectives and the mass is presented in Figure 4.60. In this instance, the mass only had a positive correlation with the axial stiffness and FPF load objectives, even though it was not significant for either of them. This highlights the contrast between axial and circumferential angle variations in how they affect objectives differently. Furthermore, the positive correlation between objectives is reduced overall with the most considerable decline for the buckling objective, which now displays negative correlations compared to the positive ones in the AV case.

The objective space is more spread out with respect to the AV case, as shown in Figure 4.61. A greater draw towards the axial stiffness axis is observed, while it seems that the frequency axis is not pulling as many objectives along its direction.

A petal plot of the optimal designs is given in Figure 4.62. The only designs that surpassed the ones achieved in the CS optimisation were for the frequency and FPF load objectives. An increase of 0.7% for

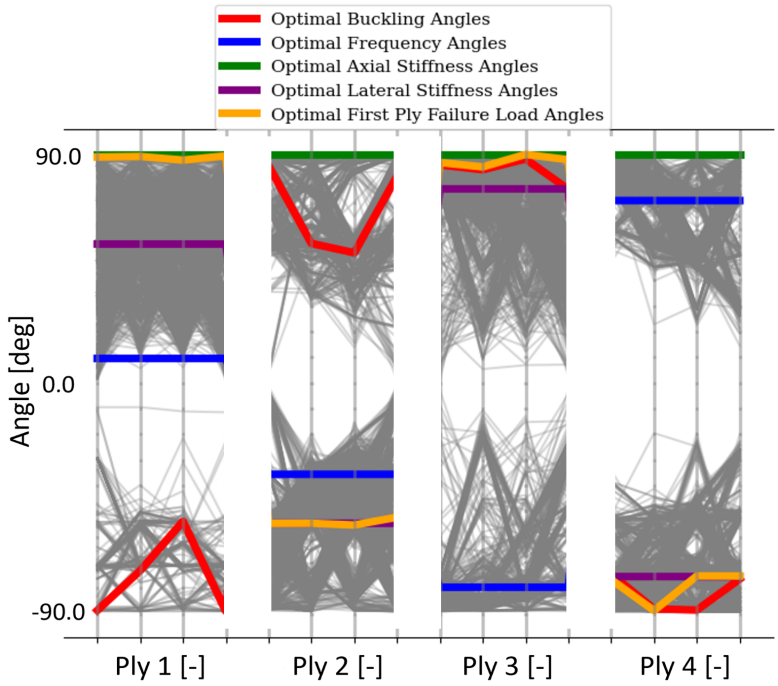


Figure 4.56: Parallel coordinate plot of the design space for the AV case with 4 CPs for the cut-out cylinder.

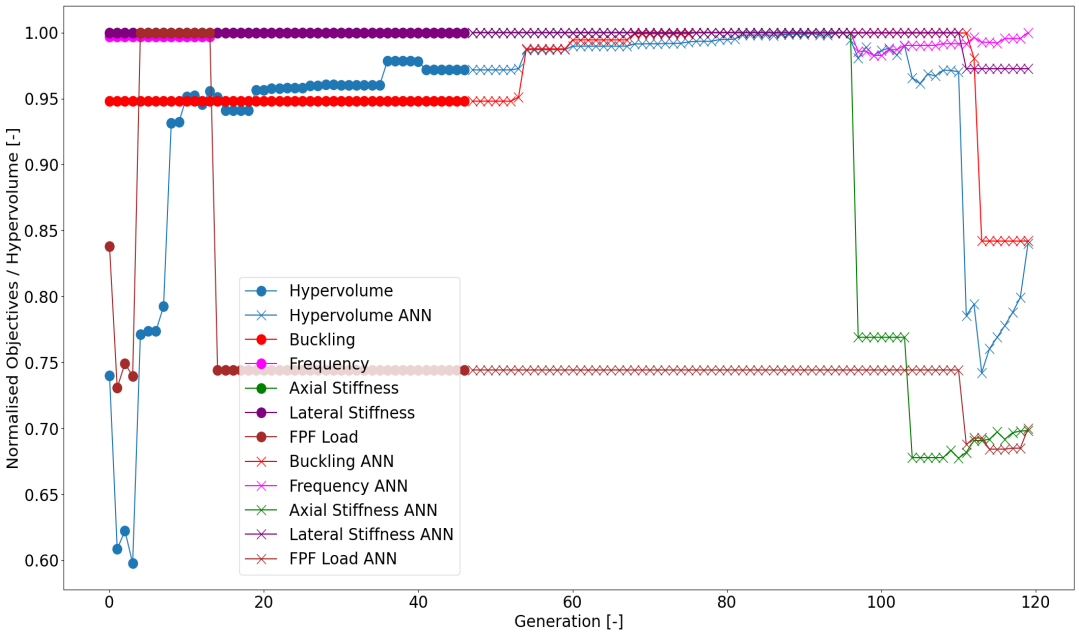


Figure 4.57: Generational objective change for the CV case with 4 CPs for the cut-out cylinder.

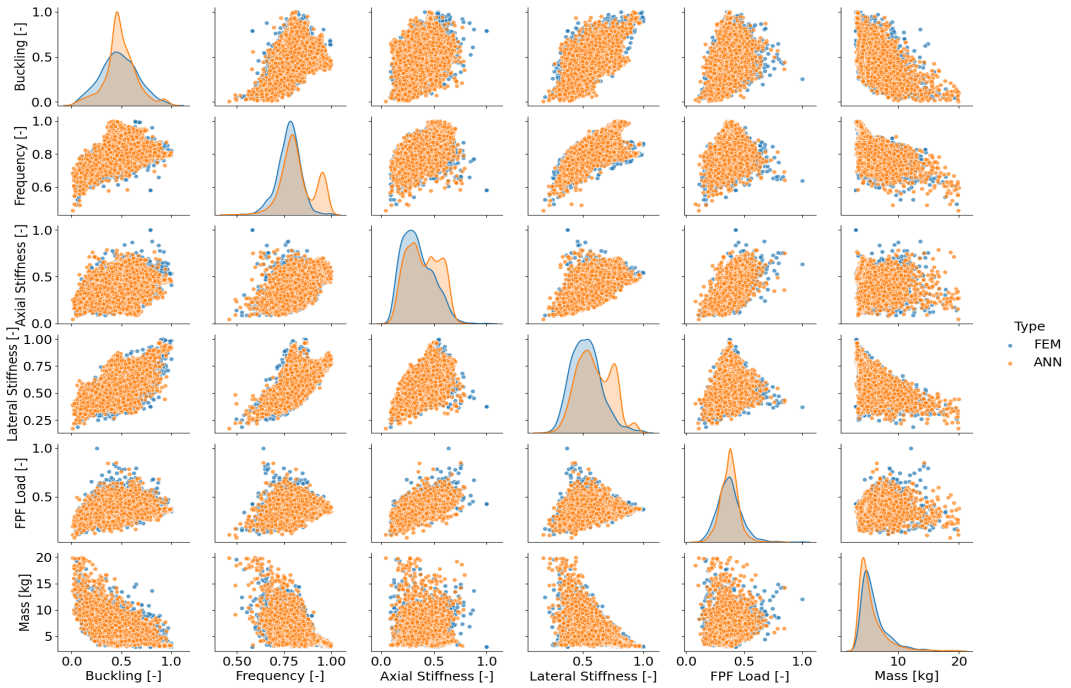


Figure 4.58: Objective relationship for both FEM and ANN results in the CV case with 4 CPs for the cut-out cylinder.

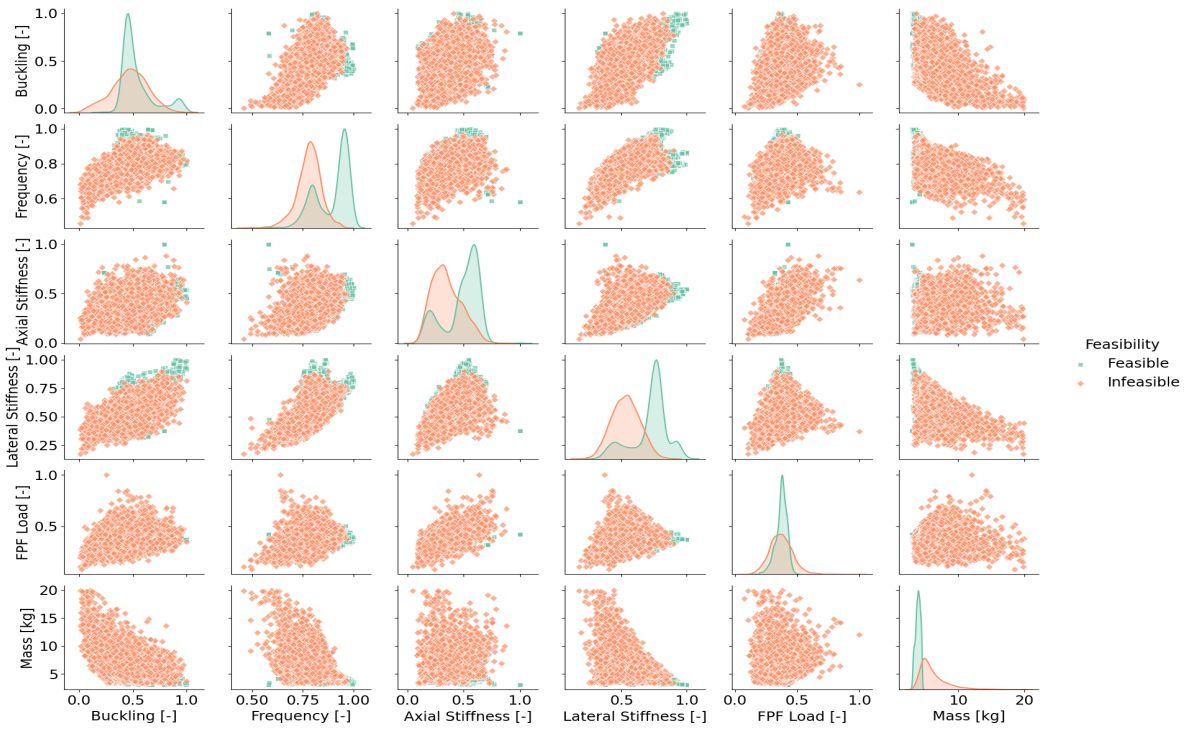


Figure 4.59: Design feasibility with respect to the imposed constraints for the CV case with 4 CPs for the cut-out cylinder.

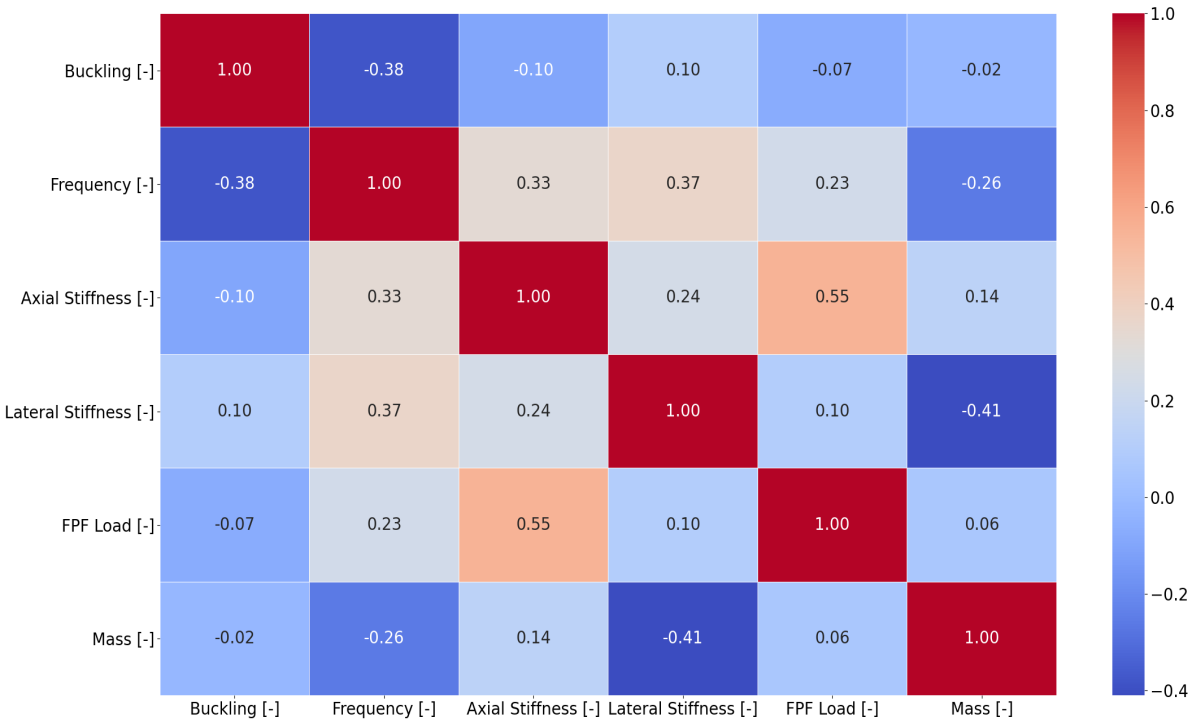


Figure 4.60: Correlation heatmap between objectives for the CV case with 4 CPs for the cut-out cylinder.

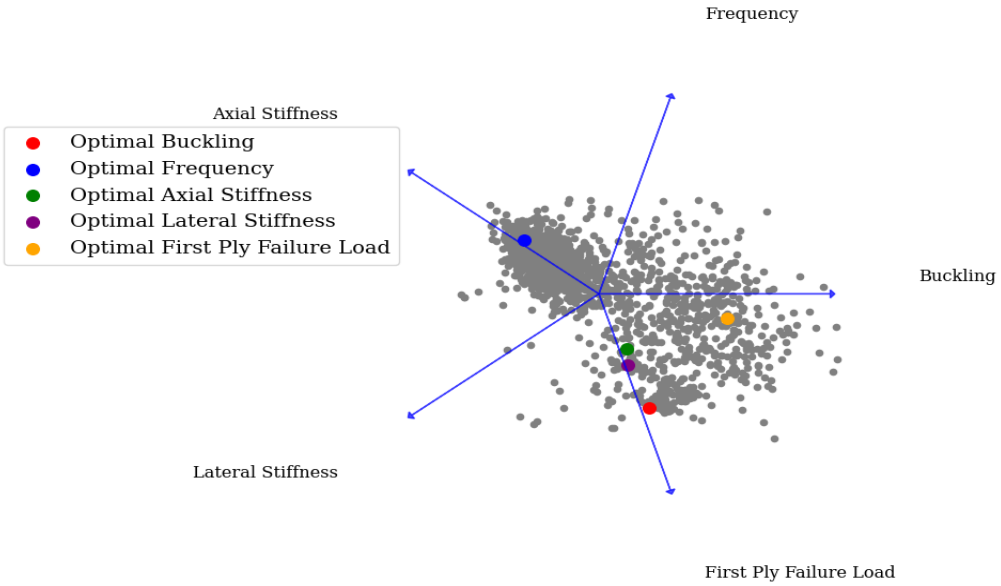


Figure 4.61: 2-D representation of the objective space for the CV case with 4 CPs for the cut-out cylinder.

the frequency objective was obtained against a 16.6% rise in mass. The significant increase in mass does not make this design particularly appealing if an improvement in frequency is desired, although the rest of the objectives are also improved in relation to the respective CS optimal design. On the other hand, a significant increase of 32% for the FPF load against a 47% gain in mass is observed. This increase is the most important yet, albeit it comes at the cost of the highest mass for an optimal design. However, it is noted that designs with lesser masses were also found to perform better than the respective optimal CS design, but the increase was not as pronounced.

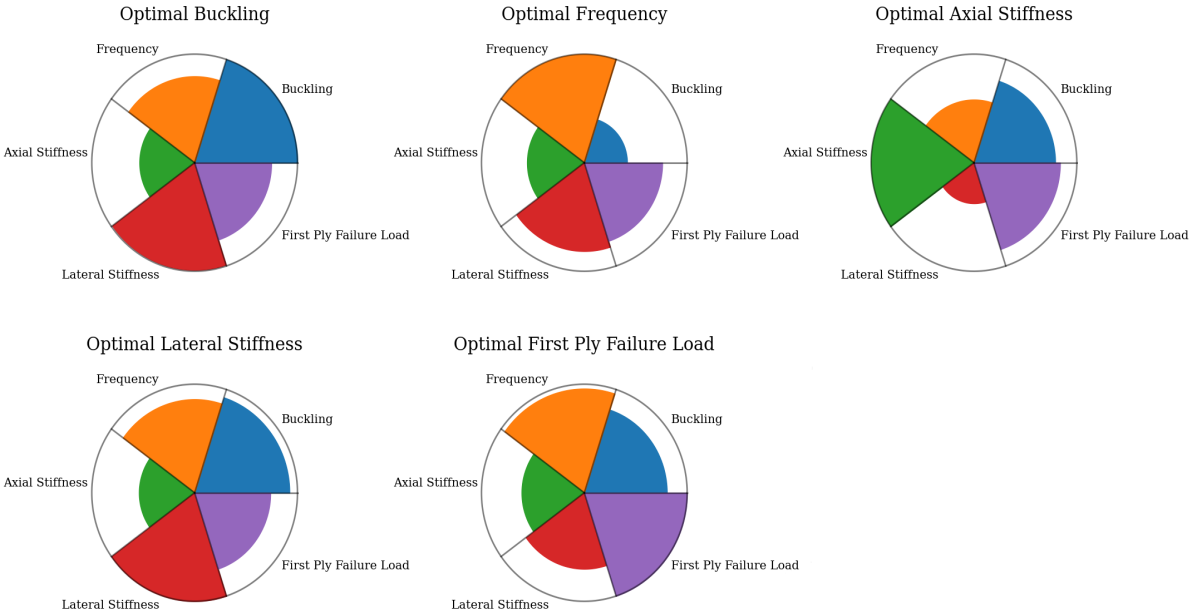


Figure 4.62: Petal plot of the optimal solution for the lateral stiffness objective in the CV case with 4 CPs for the cut-out cylinder.

Figure 4.63 shows the design space for the currently examined case. A stark difference between the AV and CV optimisations is displayed, where the latter preferred to explore and exploit angles closer to the region around 0° and avoiding the area near 90°, reversely to the former.

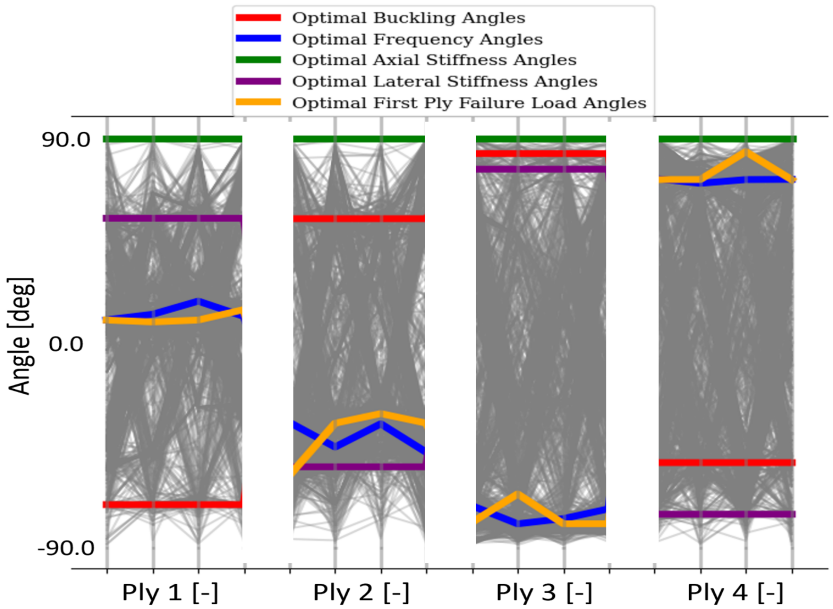


Figure 4.63: Parallel coordinate plot of the design space for the CV case with 4 CPs for the cut-out cylinder.

Axial Variation - 2 and 3 Control Points

The combined optimisations for 2 and 3 CPs for the axial variation are presented in this section, where it is reminded that they were attained by ANN predictions.

In Figure 4.64 (a) and (b), the hypervolume convergence plots of the 2 and 4 CPs cases can be found respectively. Neither of them have converged, although the CV case was still rising, which indicates that given a few more generations it might have found better designs. Similarly to the AV case with 4 CPs, the optimisation with 2 CPs showcases a struggling behaviour for the frequency objective for most designs. This is not observed for the 3 CPs case where frequency values are consistently high in contrast to the buckling objective, making it hard to recognise a trend.

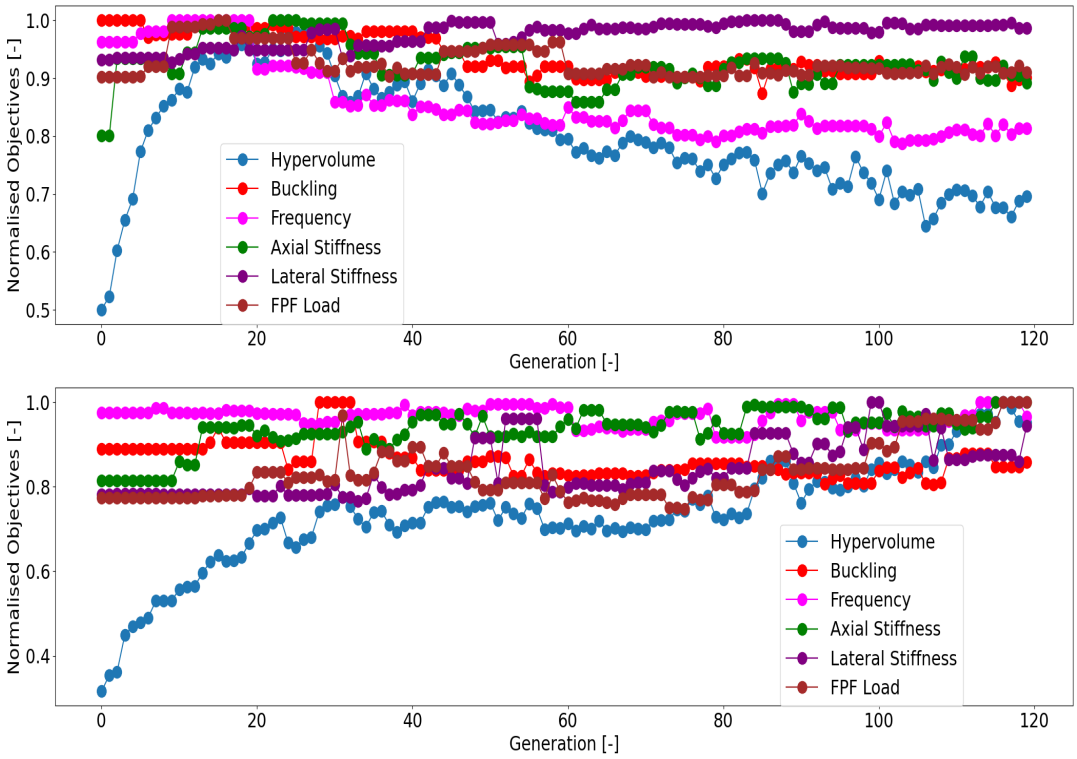


Figure 4.64: Generational objective change for the AV case with 2 CPs (top) and 3 CPs (bottom) for the cut-out cylinder.

The feasibility of the designs is shown in Figure 4.65 (a) and (b) for 2 and 3 CPs, respectively. The abundance of feasible designs over the case of 4 CPs is clearly seen. An additional difference can be noticed for the frequency objective, showcasing higher values for infeasible designs as evidenced by its density plots. Seeing as this is more pronounced for 2 CPs, and that in the case of 4 CPs the density plot curves overlapped, it can be stipulated that this effect vanishes with higher degrees of freedom for the Lagrange polynomial that represents the reference curves.

The relationship between the objectives is quantified in Figure 4.66. The aforementioned remark about the 3 and 4 CPs exhibiting more similarities is highlighted once more, since the objective correlations are nearly identical between the two cases. For the designs with 2 CPs, the positive correlations are significantly reduced, except for the relationship between frequency and lateral stiffness.

The tug-of-war between the objectives can be seen in Figure 4.67 for (a) 2 and (b) 3 CPs. Both of them show strong FPF load designs, as was displayed in designs with 4 CPs. In many respects all the cases are similar except for the significant attraction towards the lateral stiffness axis for 2 CPs, which draws points away from the frequency axis.

The optimal designs are shown in petal plots in Figure 4.68 for (a) 2 and (b) 3 CPs. Although certain objectives may seem to have different designs due to the changes in the petals between the two cases, this is attributed to the overall best designs in each optimisation being different and thus filling the petals accordingly. The design that was superior to the ones seen in the CS optimisation, was only the buckling objective for 3 CPs. In that instance, the buckling load and mass were raised by 9% and 2% respectively, which is lower than what was observed in the optimal design for 4 CPs.

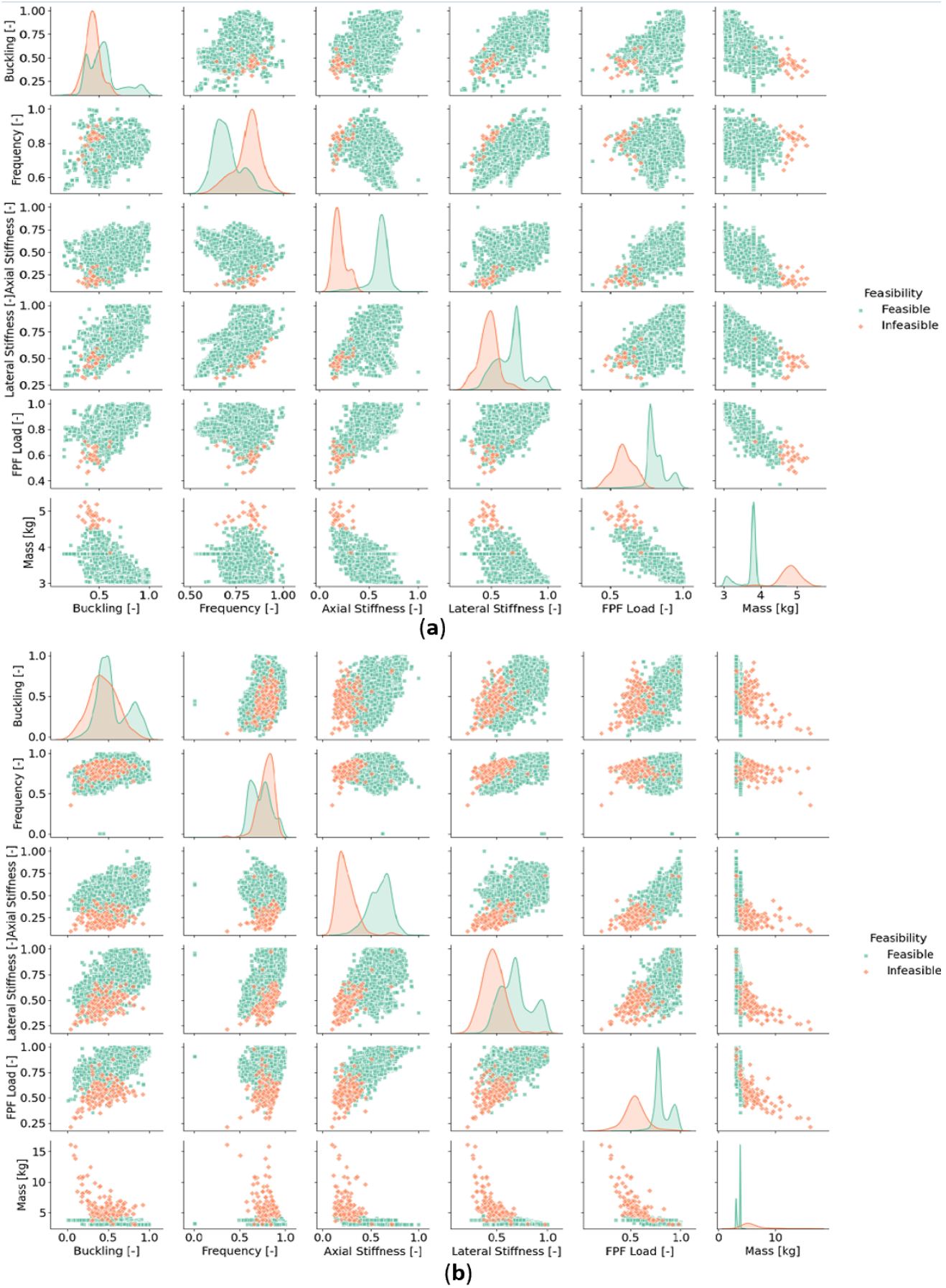


Figure 4.65: Design feasibility with respect to the imposed constraints for the AV case with a) 2 and b) 3 CPs for the cut-out cylinder.

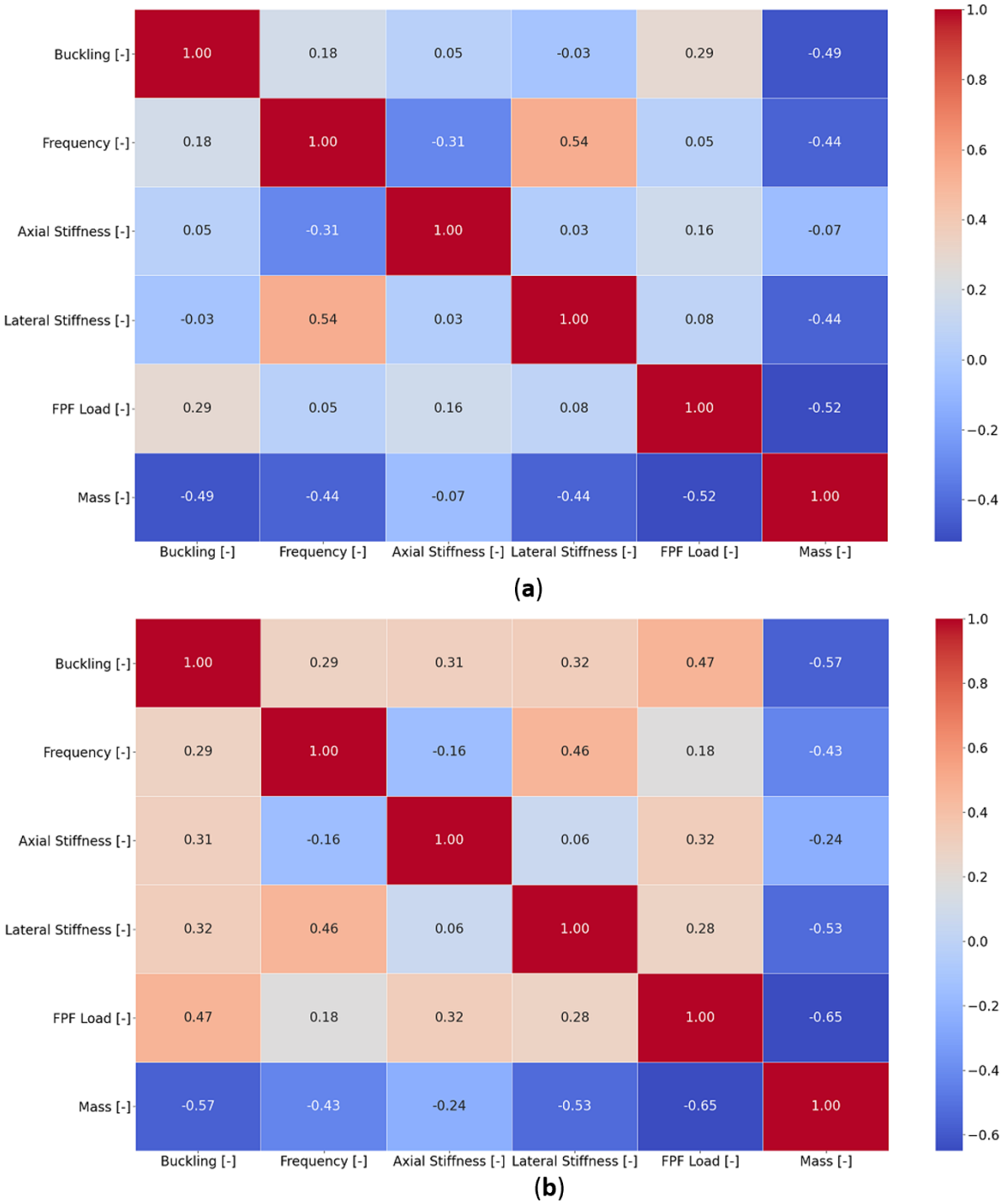


Figure 4.66: Correlation heatmap between objectives for the AV case with (a) 2 and (b) 3 CPs for the cut-out cylinder.

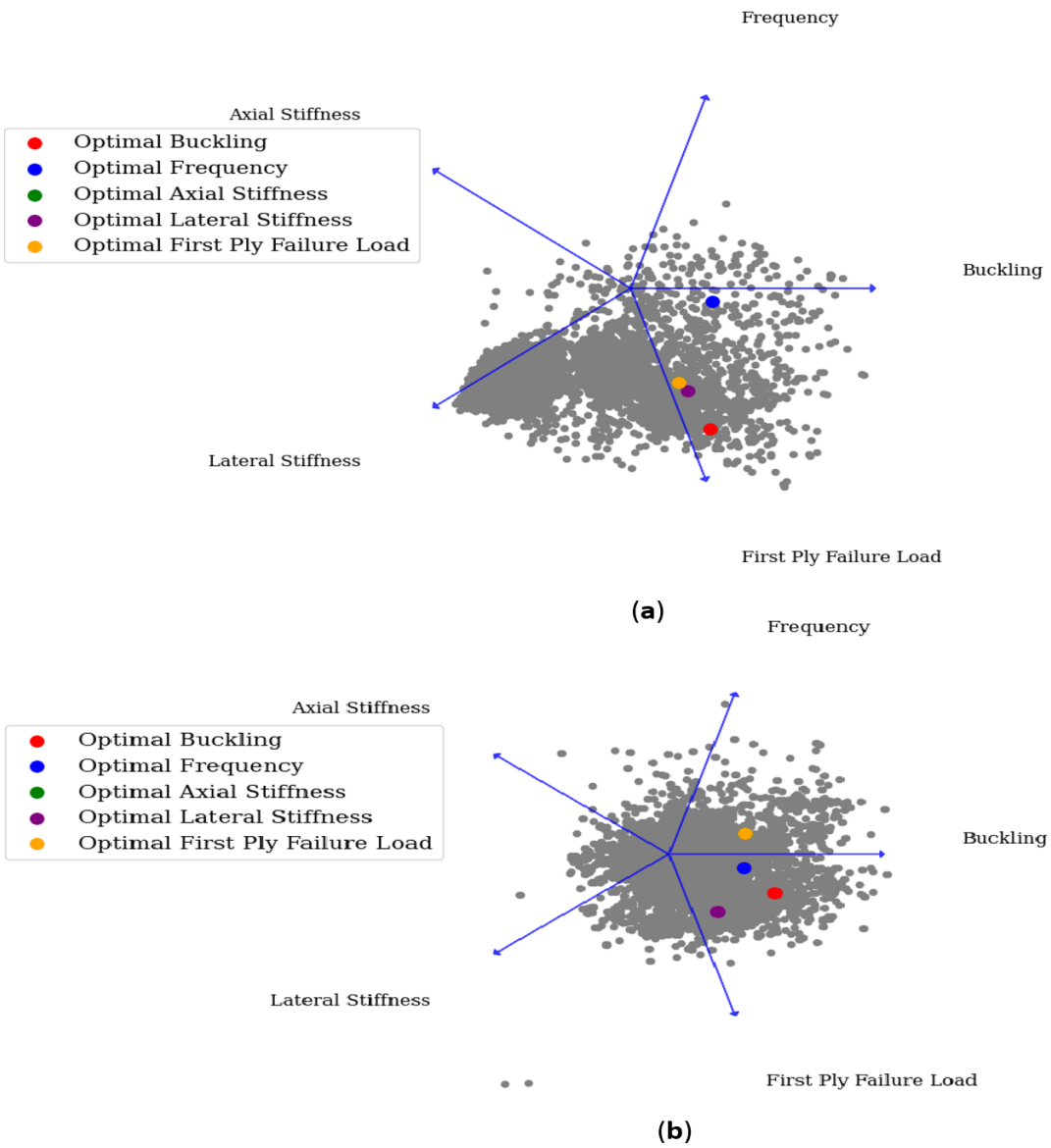


Figure 4.67: 2-D representation of the objective space for the AV case with (a) 2 and (b) 3 CPs for the cut-out cylinder.

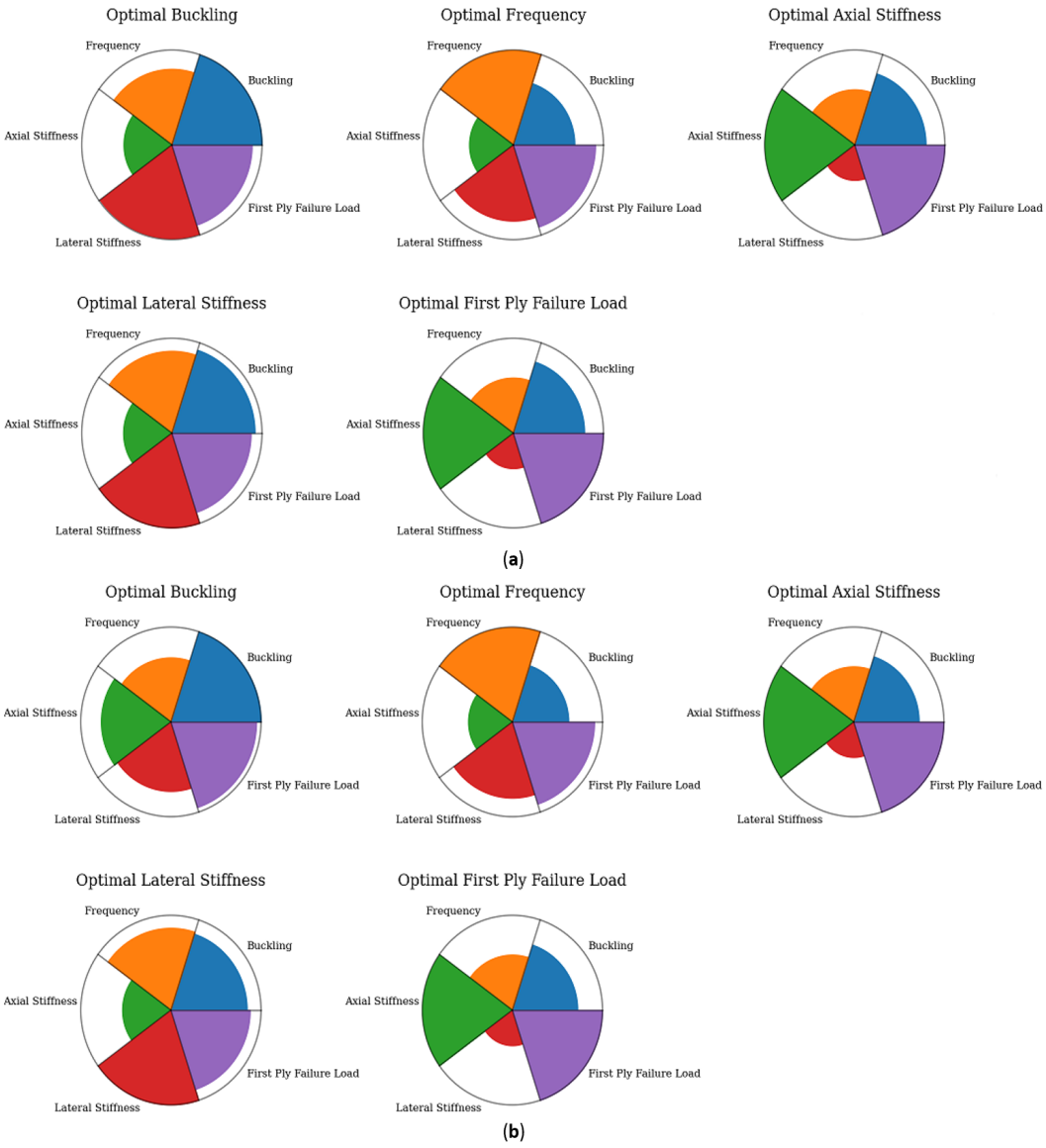


Figure 4.68: Petal plot of the optimal solution for the lateral stiffness objective for the AV case with (a) 2 and (b) 3 CPs for the cut-out cylinder.

In Figure 4.69 (a) and (b), the design space for the cases of 2 and 3 CPs is presented, respectively. The similarity between the two instances is clearly noticed, where the designs were constructed with angles closer to 90° and avoided the 0° region, as was also seen in the 4 CPs design space.

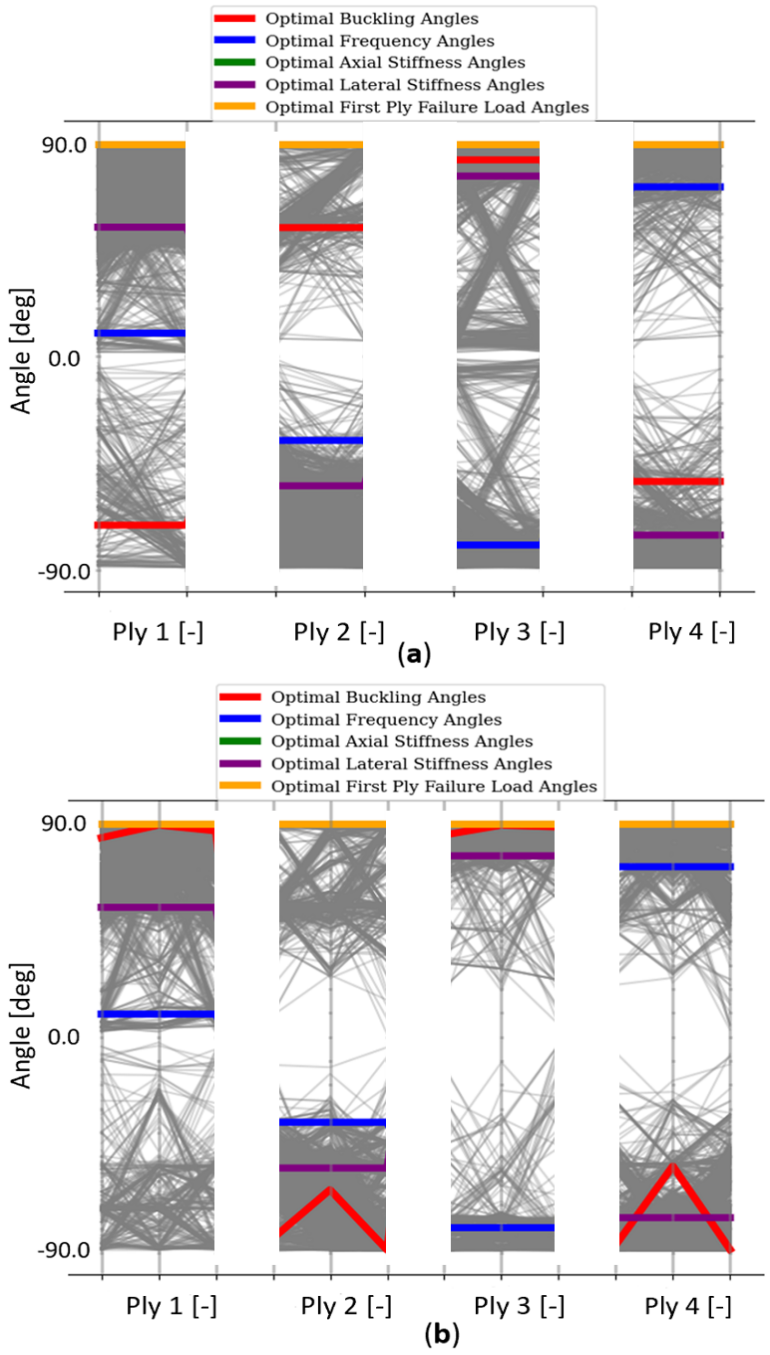


Figure 4.69: Parallel coordinate plot of the design space for the AV case with (a) 2 and (b) 3 CPs for the cut-out cylinder.

Circumferential Variation - 2 and 3 Control Points

The convergence plots of the optimisation for 2 and 3 CPs for the circumferential angle variation are shown in Figure 4.70 (a) and (b), respectively. The hypervolume and the objectives for the case of 3 CPs behave in a similar way as the 4 CPs optimisation. However, in the instance of 2 CPs a peculiar behaviour is witnessed, where most of the objectives get significantly better in the span of one generation. The reason behind this is unclear as a more gradual increase is typically expected. Nevertheless, the hypervolume appears to have converged for that specific case.

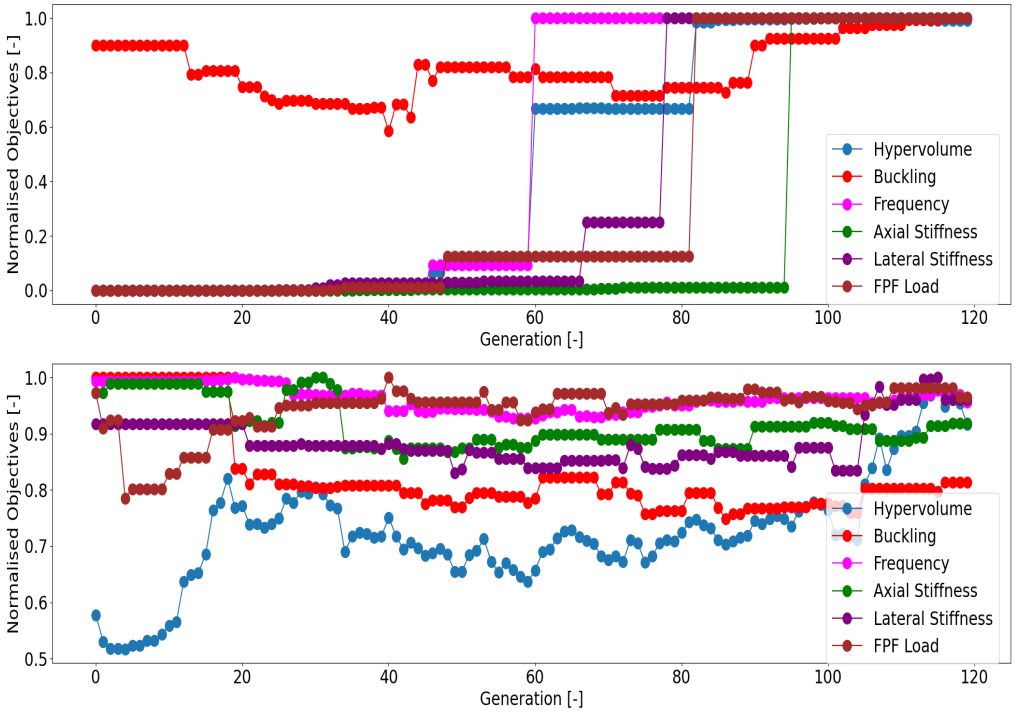


Figure 4.70: Generational objective change for the CV case with 2 CPs (top) and 3 CPs (bottom) for the cut-out cylinder.

The feasible designs for 2 and 3 CPs are inspected in Figure 4.71 (a) and (b), respectively. When compared to the corresponding AV plot, it can be said that CV produces more infeasible designs overall. For 2 and 3 CPs, that is mainly due to the extra mass, but in the latter case the curvature constraint plays a bigger role. The feasible designs yield the optimal results for the most part, albeit this is not entirely the case for the axial stiffness and FPF load objectives which have some of their best designs being infeasible. It is noted that the same similarity was observed for the case of 4 CPs.

The objectives' relationship is given in Figure 4.72 (a) and (b) for 2 and 3 CPs, respectively. A more positive correlation between the objectives is noticed for the designs with 2 CPs and a clear negative correlation between objectives and mass. The same positive correlation between objectives is found in the instance of 3 CPs but it is not as significant. It is noted that there is an exception between the axial stiffness and FPF load objectives, where their correlation increased with respect to the reported one for 2 CPs. Additionally, a strong correlation between the same objectives and the structural mass is observed.

In Figure 4.73 (a) and (b) the objective space can be seen for 2 and 3 CPs, respectively. It seems that the objectives spaces are distinctly different between the two cases. When it comes to the results with 2 CPs, the axes that exhibit the best results in a relative manner are axial stiffness, frequency and buckling. This changes for the case of 3 CPs, where the point cloud is more aligned along the FPF load and axial stiffness axes. This trend is visible for 4 CPs as well, where this focus along these two axes is even more pronounced. This highlights the distinct capabilities of each case concerning the objective results that can be obtained.

The optimal designs for each objective category are presented in a relational manner in the petal plot of Figure 4.74. It is reminded that a petal plot can be different even though the optimal design is the same in both cases, due to the relative best objectives in the respective optimisation. The designs that are found to perform better than the CS designs is the FPF load objective for 2 and 3 CPs, and the frequency objective for 3 CPs. The frequency increase is a marginal value of 0.5% for a mass gain of 18.5%, which is worse than the optimal design observed for 4 CPs for both objective value and mass, respectively. On the other hand, the improvement of the FPF load is 32% and 33% for an additional mass of 48% and 49%, for the cases of 2 and 3 CPs, respectively. Thus, the latter case achieved the best FPF load boost, albeit at the expense of the highest mass value, which barely satisfies the mass optimisation constraint.

The designs for 2 and 3 CPs can be seen in Figure 4.75 (a) and (b), respectively. The explored design space is similar for all CPs, as angles around 90° were avoided. The optimal frequency design for 3 CPs closely follows the CS optimal design, which is also the same for 2 CPs, with the only difference being a slight variation

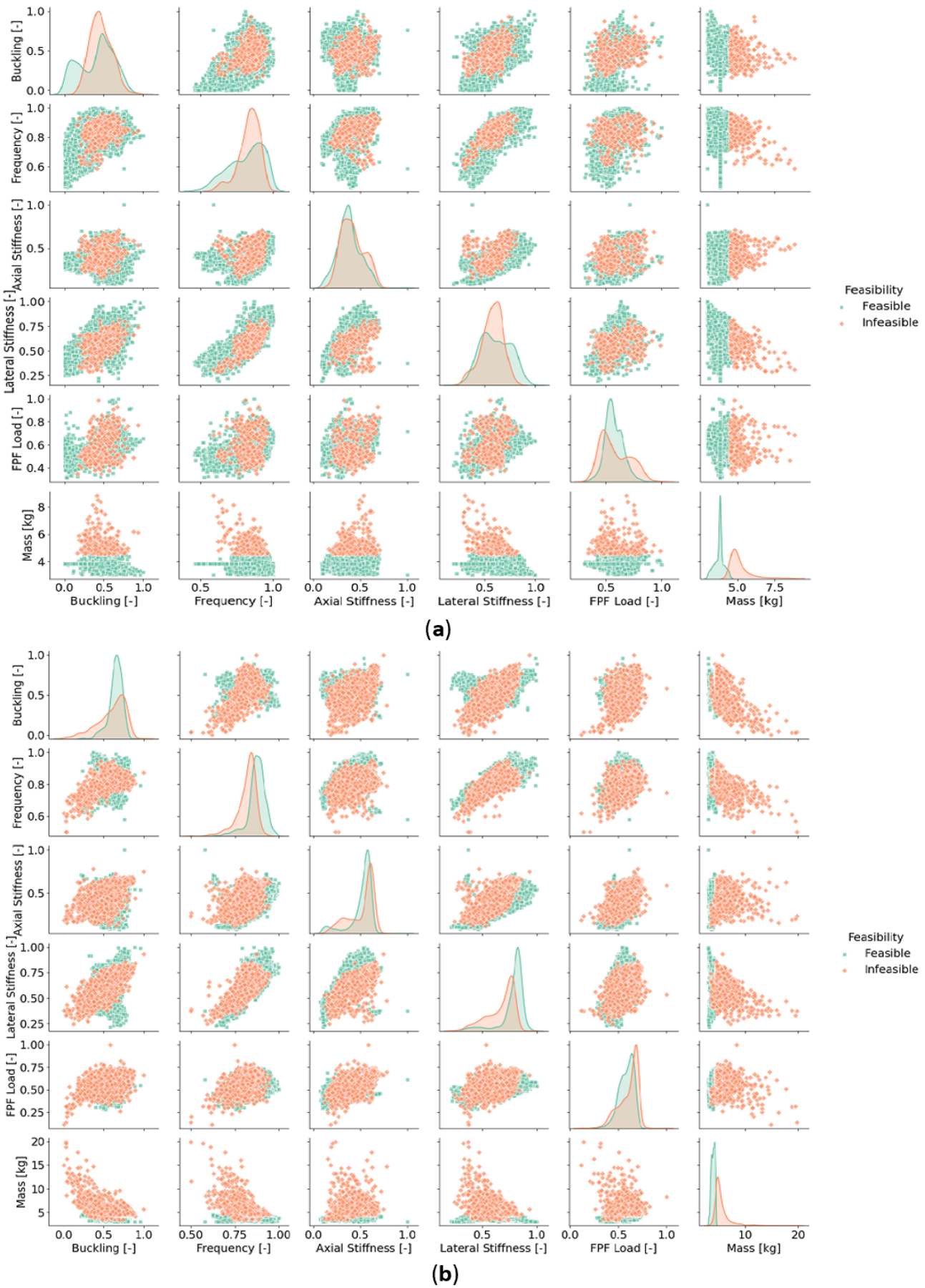


Figure 4.71: Design feasibility with respect to the imposed constraints for the CV case with a) 2 and b) 3 CPs for the cut-out cylinder.

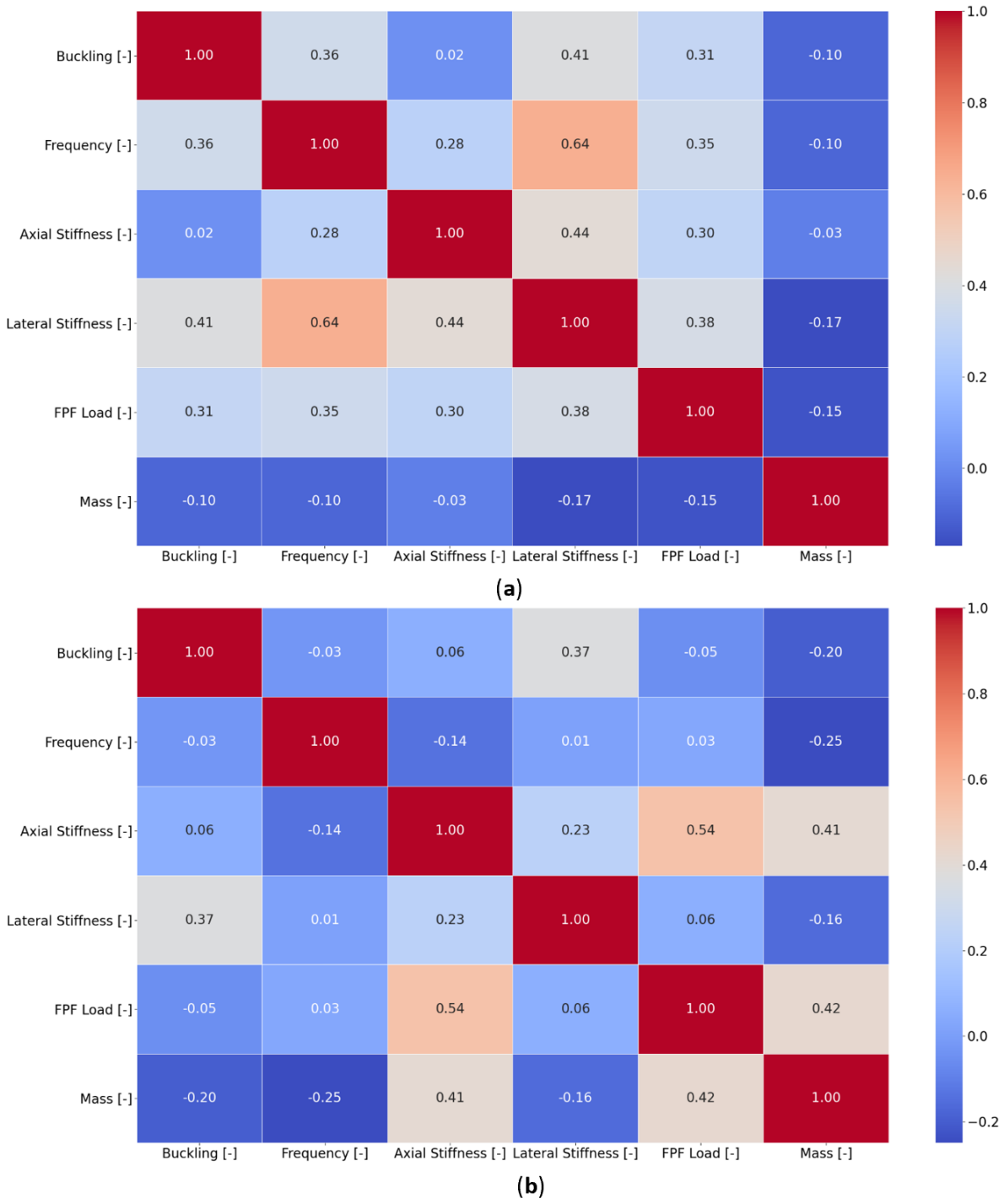
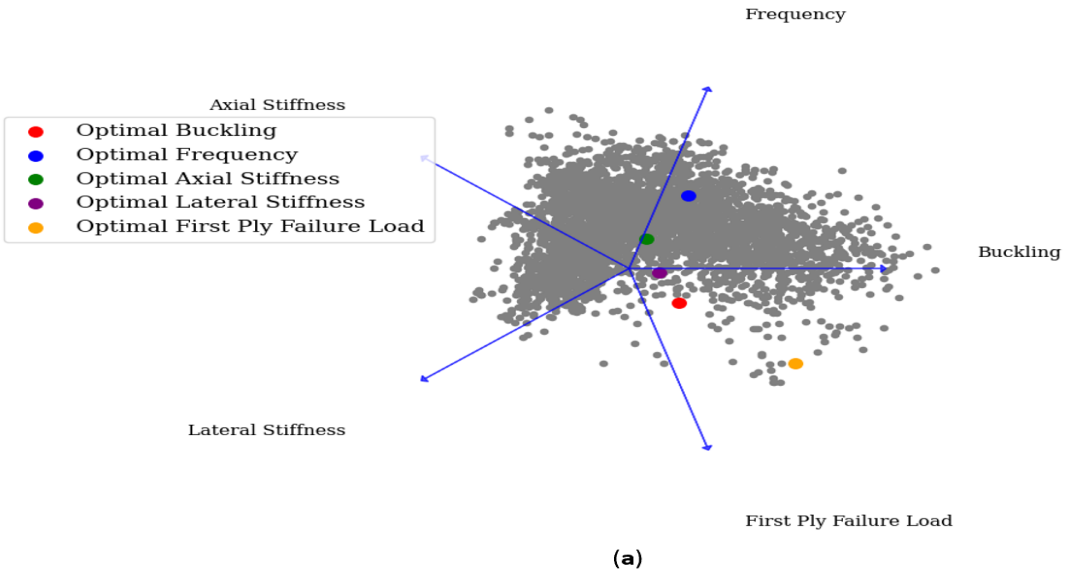
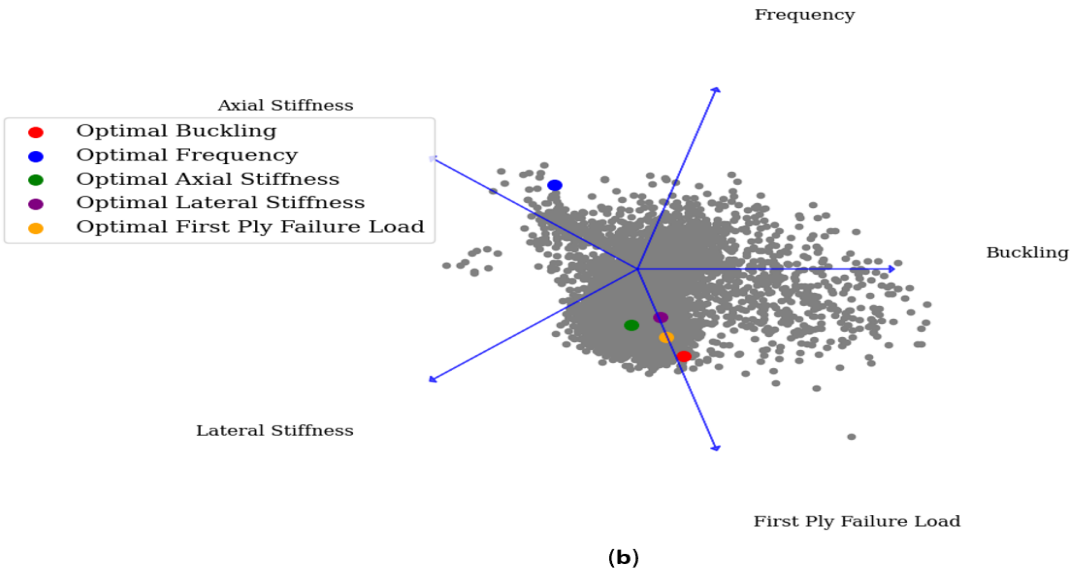


Figure 4.72: Correlation heatmap between objectives for the CV case with (a) 2 and (b) 3 CPs for the cut-out cylinder.



(a)



(b)

Figure 4.73: 2-D representation of the objective space for the CV case with (a) 2 and (b) 3 CPs for the cut-out cylinder.

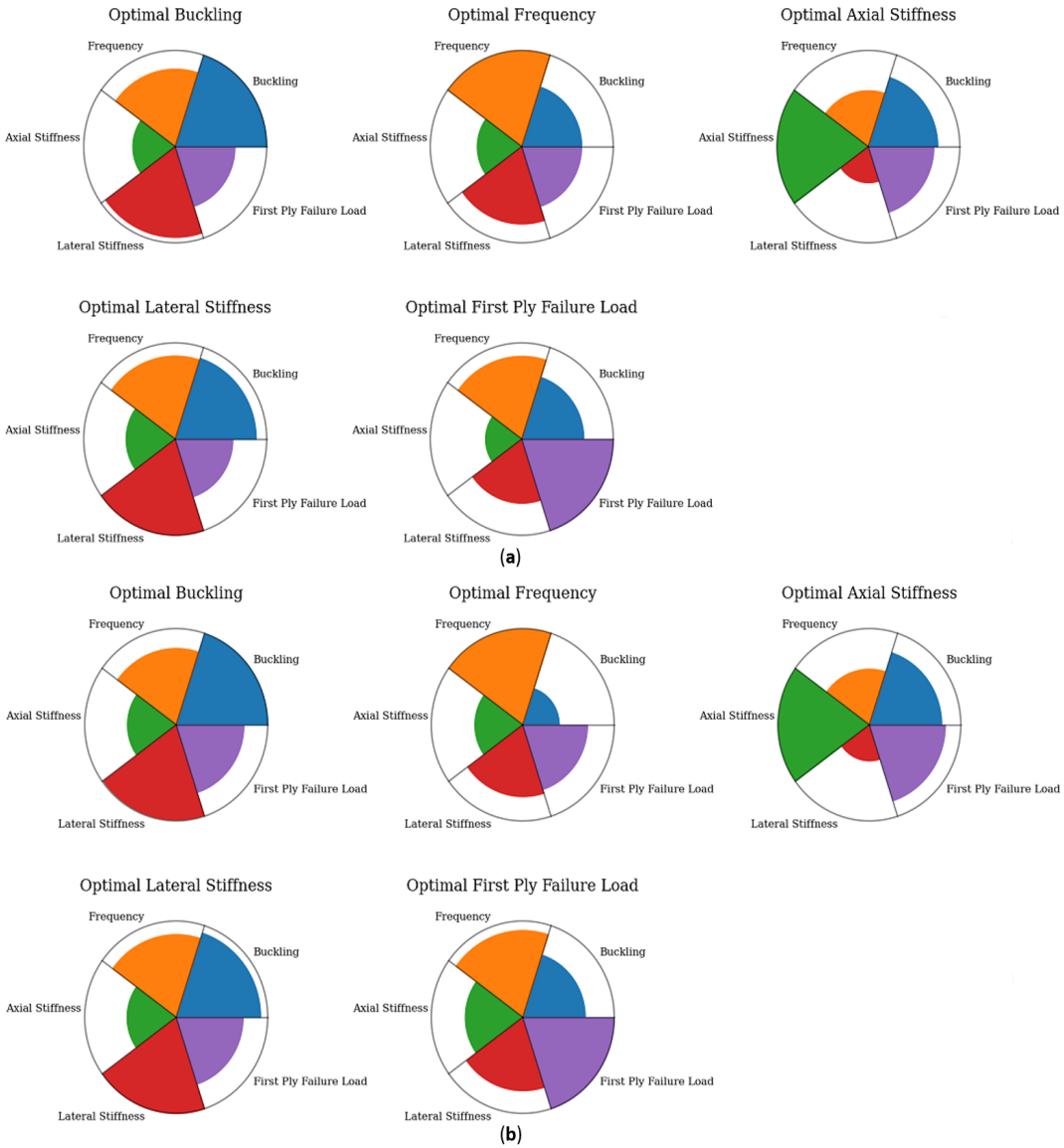


Figure 4.74: Petal plot of the optimal solution for the lateral stiffness objective for the CV case with (a) 2 and (b) 3 CPs for the cut-out cylinder.

in the 1^{st} and 4^{th} plies. On the contrary, the FPF load optimal designs exhibit a much greater variation within the plies, and although the optimal designs for this objective are distinct, a similar trend is identified for all of them.

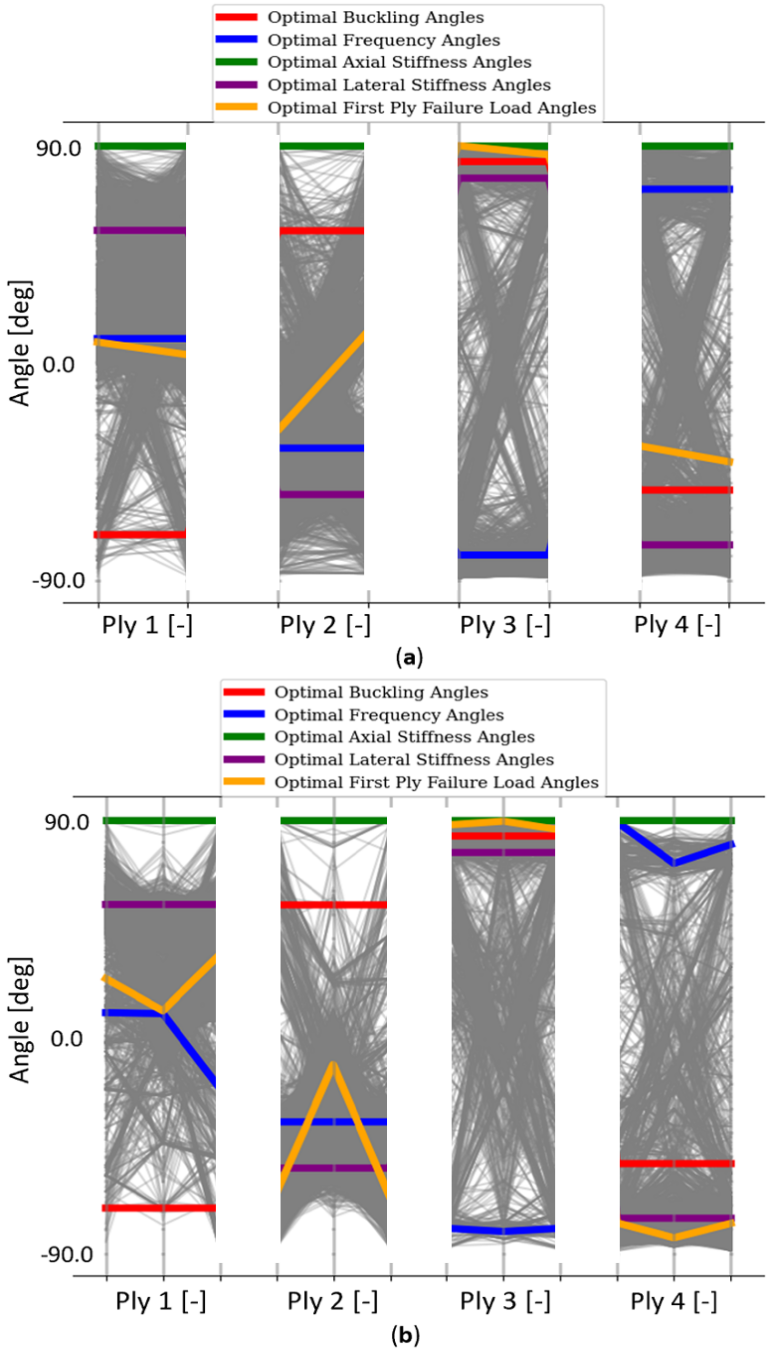


Figure 4.75: Parallel coordinate plot of the design space for the CV case with (a) 2 and (b) 3 CPs for the cut-out cylinder.

Surrogate model performance

In this section, the accuracy of the surrogate models is assessed, as well as the architectures of the neural networks. First off, the optimised ANN architectures that made the element properties predictions are reported in 4.76. As with the pristine cylinder, the network that was responsible for the angle predictions was shallow, while the thickness predictions required a deeper network. In both cases the SELU architecture was preferred over ELU.

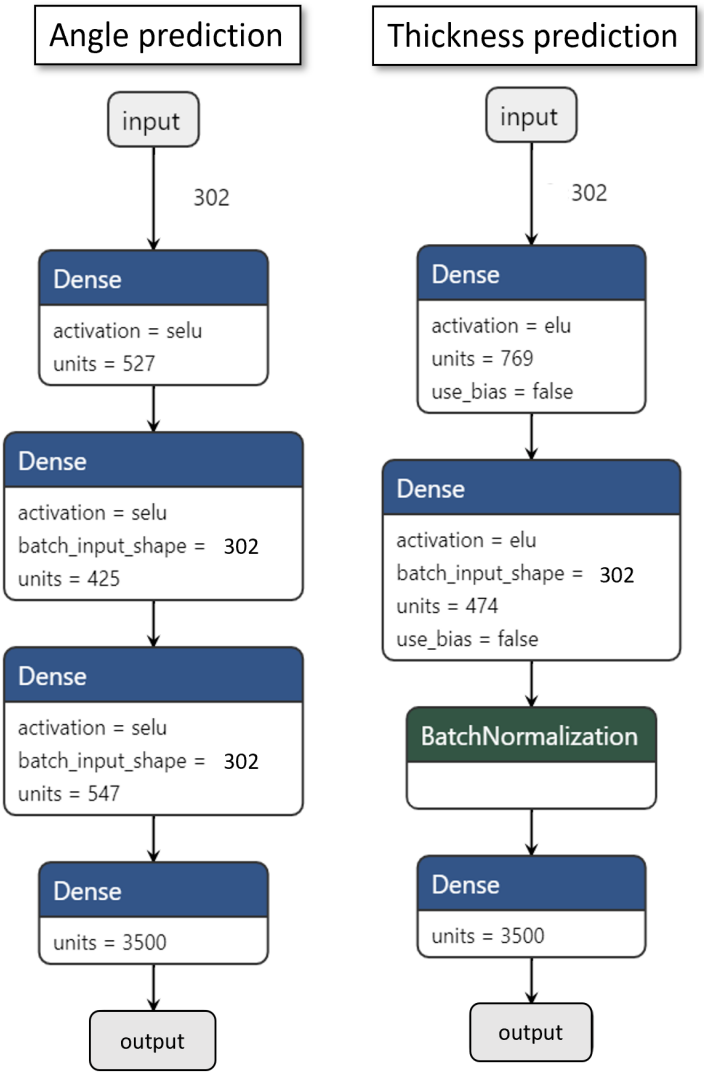


Figure 4.76: Neural network optimised architectures for the angle and thickness predictions of each element for the cut-out cylinder.

The prediction accuracy of the element properties can be seen in Table 4.11. The same trends that were identified in the case of the pristine cylinder are also noticed in this instance. The accuracy of the angle prediction is high for all the surrogate models, while there is a significant reduction for the prediction of the thickness field. It is noted that the ANN was the best overall, although its prediction for the thickness properties is worse for the cut-out cylinder than it was for the pristine counter-part.

The optimal architectures for the ANNs that performed predictions on the optimisation objectives are showcased in Figure 4.77. It is interesting to note that the buckling architecture was the shallowest among them, as was the case with the pristine cylinder. Similarly the other architectures are consistent with their pristine cylinder counterparts, with the FPF load one exhibiting the deepest architecture overall. Additionally, only the SELU architecture variants are observed in this case, unlike the pristine cylinder where there were 3 instances of ELU architectures.

The accuracy of the prediction of the optimisation objectives is shown in Table 4.12. The results closely follow what was established for the pristine cylinder, with the objectives exhibiting a coefficient of determi-

Table 4.11: Comparison between different surrogate models on the accuracy of the predictions of the element properties for the cut-out cylinder.

Angle							
Metric	ANN	RBF	Kriging	Metric	ANN	RBF	Kriging
Normalised MSE	0.43	1.00	0.92	MSE	0.89	2.06	1.89
Normalised MAE	0.39	1.00	0.84	MAE	0.57	1.44	1.21
Normalised R^2	1.00	0.89	0.93	R^2	1.00	0.89	0.93
Thickness							
Metric	ANN	RBF	Kriging	Metric	ANN	RBF	Kriging
Normalised MSE	0.10	1.00	0.43	MSE	6.05E-08	6.13E-07	2.64E-07
Normalised MAE	0.36	1.00	0.86	MAE	6.96E-05	1.94E-04	1.67E-04
Normalised R^2	1.00	0.01	0.28	R^2	0.83	0.01	0.23

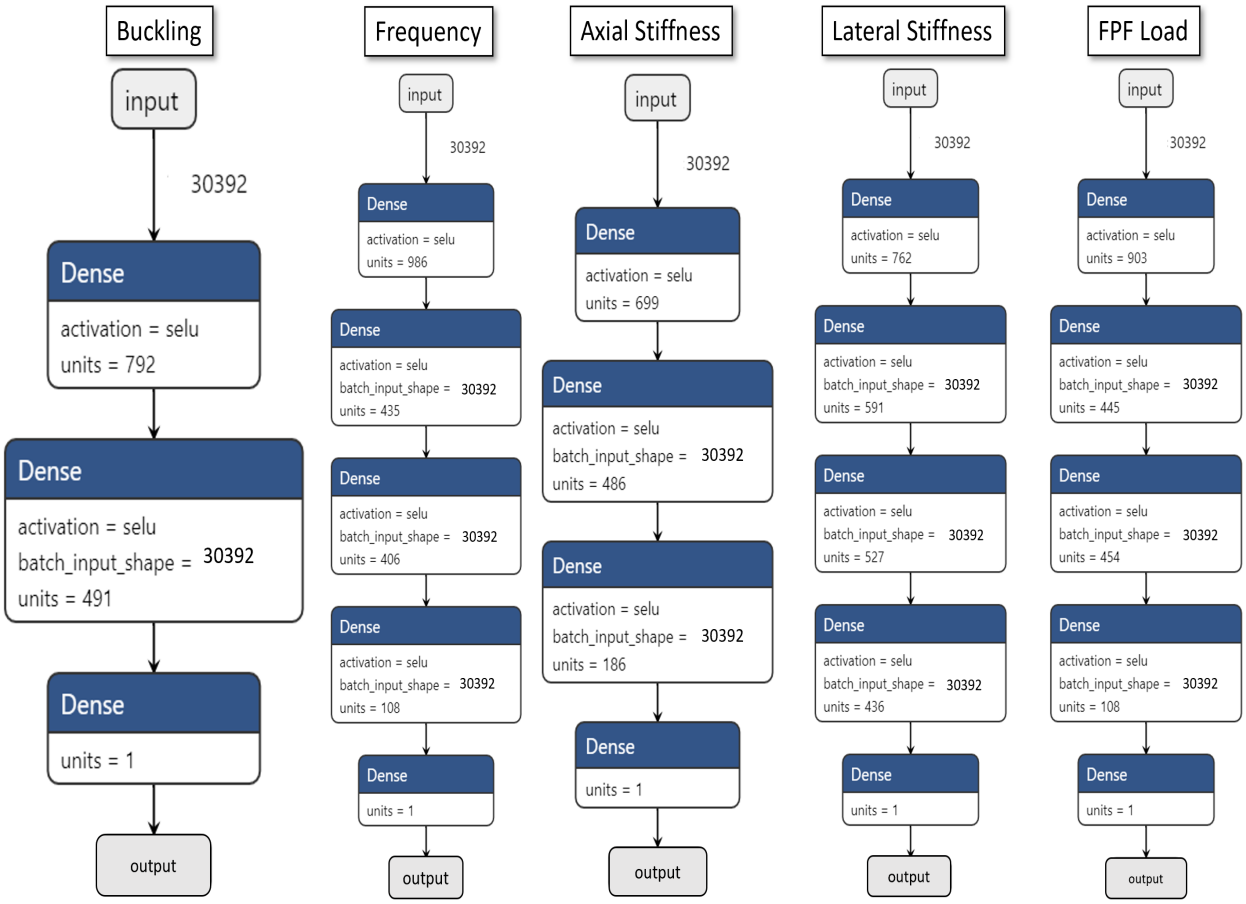


Figure 4.77: Neural network optimised architectures for each objective.

Table 4.12: Comparison between different surrogate models on the accuracy of the predictions of the objectives for the VS cut-out cylinder case.

Buckling							
Metric	ANN	RBF	Kriging	Metric	ANN	RBF	Kriging
Normalised MSE	0.20	1.00	0.57	MSE	2.84E+09	1.45E+10	8.31E+09
Normalised MAE	0.35	1.00	0.60	MAE	26923.74	75935.44	45769.74
Normalised R^2	1.00	0.00	0.53	R^2	0.80	0.00	0.43
Frequency							
Metric	ANN	RBF	Kriging	Metric	ANN	RBF	Kriging
Normalised MSE	0.10	1.00	0.49	MSE	164.51	1617.51	798.84
Normalised MAE	0.28	1.00	0.63	MAE	9.14	32.16	20.32
Normalised R^2	1.00	0.01	0.57	R^2	0.90	0.01	0.51
Axial Stiffness							
Metric	ANN	RBF	Kriging	Metric	ANN	RBF	Kriging
Normalised MSE	0.11	1.00	0.47	MSE	2.78E+15	2.49E+16	1.16E+16
Normalised MAE	0.30	1.00	0.52	MAE	3.19E+07	1.07E+08	5.52E+07
Normalised R^2	1.00	0.00	0.55	R^2	0.89	0.00	0.49
Lateral Stiffness							
Metric	ANN	RBF	Kriging	Metric	ANN	RBF	Kriging
Normalised MSE	0.09	1.00	0.51	MSE	3.36E+12	3.74E+13	1.91E+13
Normalised MAE	0.26	1.00	0.67	MAE	1.07E+06	4.09E+06	2.74E+06
Normalised R^2	1.00	0.00	0.54	R^2	0.91	0.00	0.49
FPF load							
Metric	ANN	RBF	Kriging	Metric	ANN	RBF	Kriging
Normalised MSE	0.14	1.00	0.49	MSE	1.04E+10	7.59E+10	3.68E+10
Normalised MAE	0.31	1.00	0.61	MAE	58382.76	187826.74	115177.50
Normalised R^2	1.00	0.00	0.60	R^2	0.86	0.00	0.51

nation ranging between 80% and 90%, with the buckling and FPF load objectives being the most difficult to predict accurately.

The computational times for the various surrogate models are omitted, since the order of the problem with the cut-out cylinder is of the same magnitude as the pristine counterpart and therefore no significant differences are expected.

Angle variation model comparison

Similarly to the pristine case, the strengths and weaknesses of each angle approach were determined by performing a comparison of their decomposed values. For the sake of clarity, it is reiterated that the decomposed values of the top 100 designs following a weighted-sum scalar decomposition with equal weights for each objective were investigated, as presented at the top left of Figure 4.78. Furthermore, every objective was examined independently to acquire further insight into the objective space.

The decomposed values suggest that the AV cases with 2, 3 and 4 CPs contain the best all-around designs, followed by the CV cases. Subsequently, the CS designs can be found at the bottom of the graph. These findings indicate that designs which are great overall can be achieved by allowing for an angle variation within the laminate, although the circumferential angle variation performs more poorly, as was also showcased for the pristine cylinder. However, unlike the pristine cylinder, the greatest amount of CPs that was considered, does not yield the best designs overall this time. This can be interpreted by the high CPs designs catering to

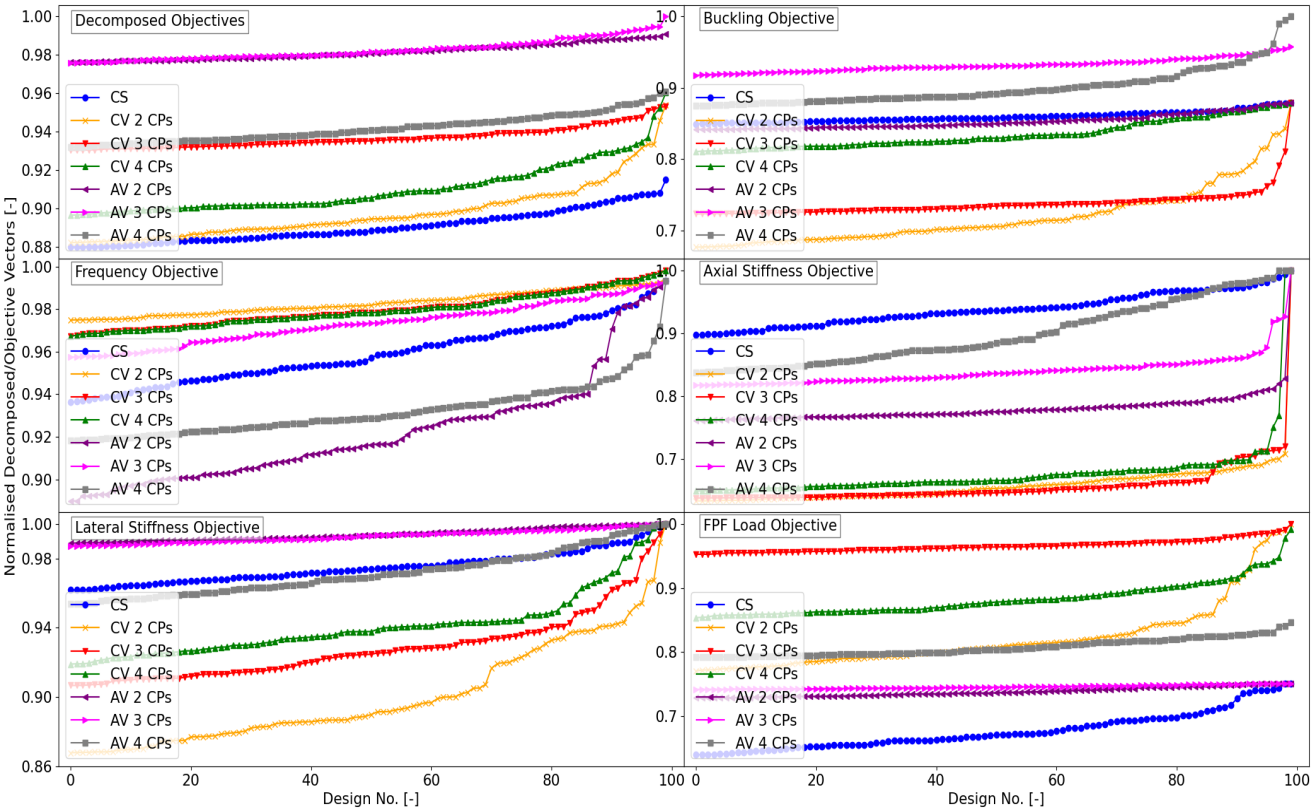


Figure 4.78: Comparison between angle variation models.

the improvement of specific objectives more easily in the presence of cut-outs, and therefore to the detriment of the rest of the objectives, with the increased freedom they possess.

The individual objectives suggest which angle variation technique is best suited to improve their performance. The buckling objective is significantly enhanced by the AV cases with a higher amount of CPs, although the instance with 2 CPs is effectively tied with the CS designs. On the other hand, the CV designs are vastly superior with respect to the FPF load objective, however the AV case with 4 CPs also managed to surpass the optimal CS design. The CV designs also perform very well for the frequency objective, albeit in a less pronounced manner. The only instance where the CS case yielded superior values for the considered designs was the axial stiffness objective. This suggests that the angle variation cases focused on improving the rest of the objectives in the detriment of the axial stiffness objective, which cannot be improved any further by any angle variation.

Angle variation for the optimal designs

It is intuitive to show how the angle orientation varies throughout the cylinder for the optimal designs that exhibit high angle variation. Therefore, the optimal buckling and FPF load designs are investigated from the axial and circumferential angle variations, respectively. Two images are taken in each case, which focus on the area around the two cut-outs, and the variations in each symmetric ply are also displayed. The optimal buckling design can be seen in Figure 4.79 for the longitudinally elongated elliptical cut-out and in Figure 4.80 for the circumferentially elongated one.

The plies are unique in their variation, although they all showcase the same behaviour regarding the ends of the cylinder, where the fibre orientation is aligned with longitudinal axis to better transfer the load. In Figure 4.79, plies (a) and (b) have fibre orientations with opossite directions around the cut-out, which follow the cut-out tangentially from different sides each. The same is true for plies (c) and (d), which highlights that each ply tried to tangentially follow specific sections of the cut-out. The aforementioned remarks hold true for the second cut-out of the same cylinder in Figure 4.80. The fibre orientations tangentially follow a larger section of the circumferentially elongated cut-out, which means that this cut-out was the driving force for the exhibited angle variation. This was to be expected however, since cut-outs that occupy more space circumferentially have a greater effect on the buckling characteristics of a cylinder.

The optimal FPF load design for both cut-outs is shown in Figure 4.81 and Figure 4.82. The fibre orientations are once more opposite of one another in each symmetric ply. The circumferentially elongated cut-out determined the fibre orientations, as the highest strains that lead to FPF can be found there. It is interesting to see that the angle changed circumferentially from one end of the cut-out to the other (e.g. from left to right) to tangentially follow the cut-out at both of its sides. Plies (a) and (b) focused on the lower and upper portions of the cut-out, while plies (c) and (d) opted to follow the middle sections.

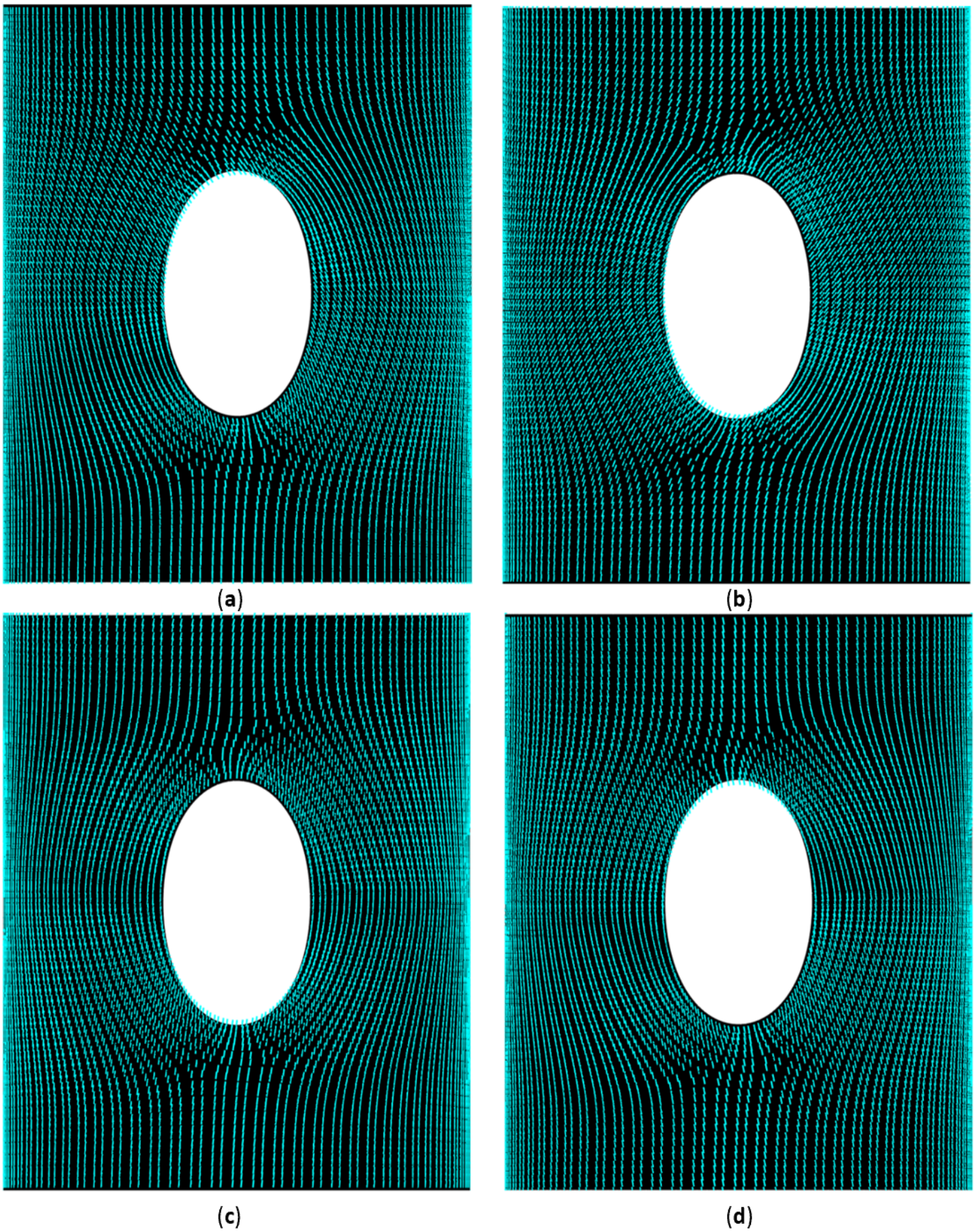


Figure 4.79: Angle variation for each symmetric ply (a), (b), (c), and (d) for the optimal buckling design which shows the longitudinally elongated cut-out.

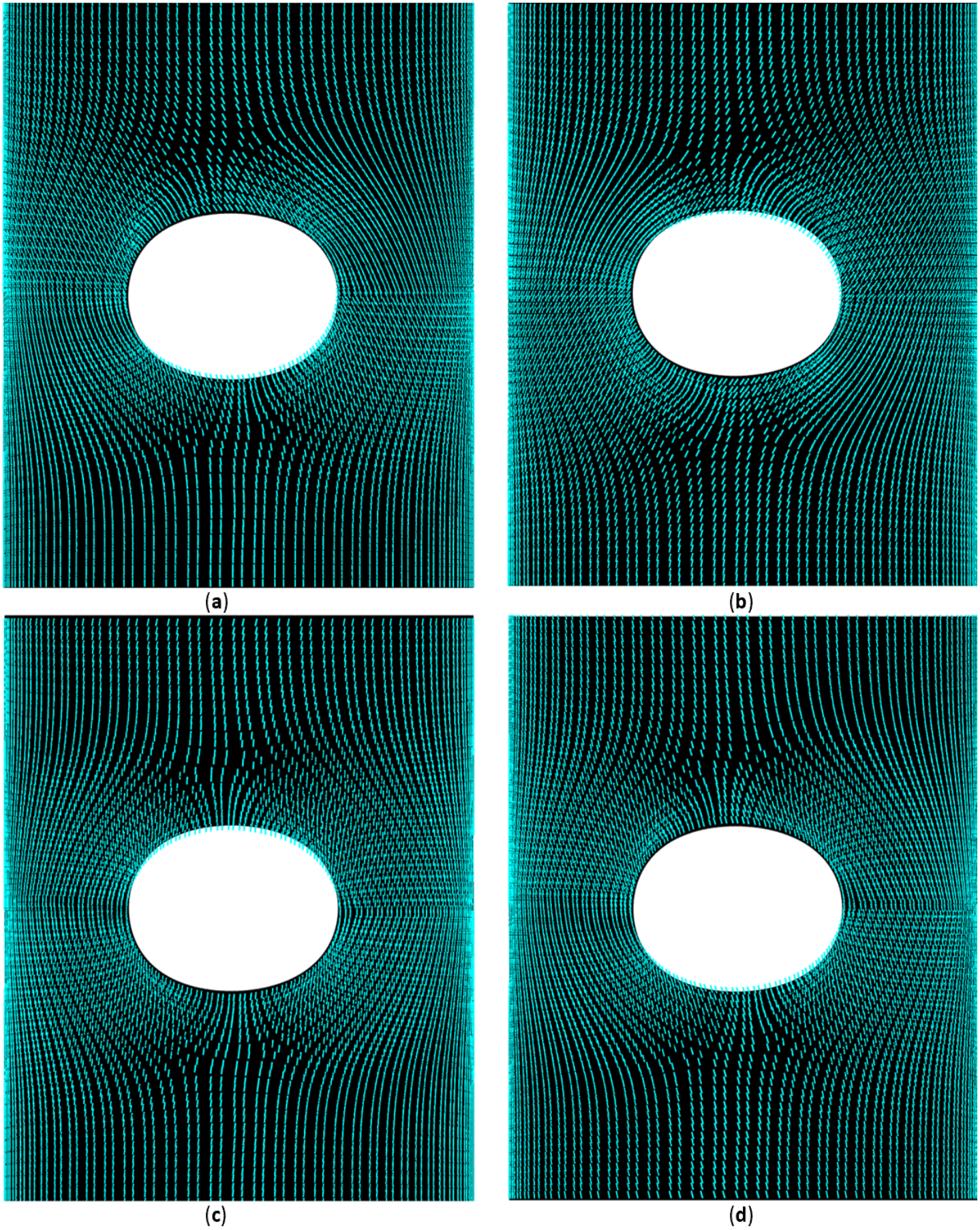


Figure 4.80: Angle variation for each symmetric ply (a), (b), (c), and (d) for the optimal buckling design which shows the circumferentially elongated cut-out.

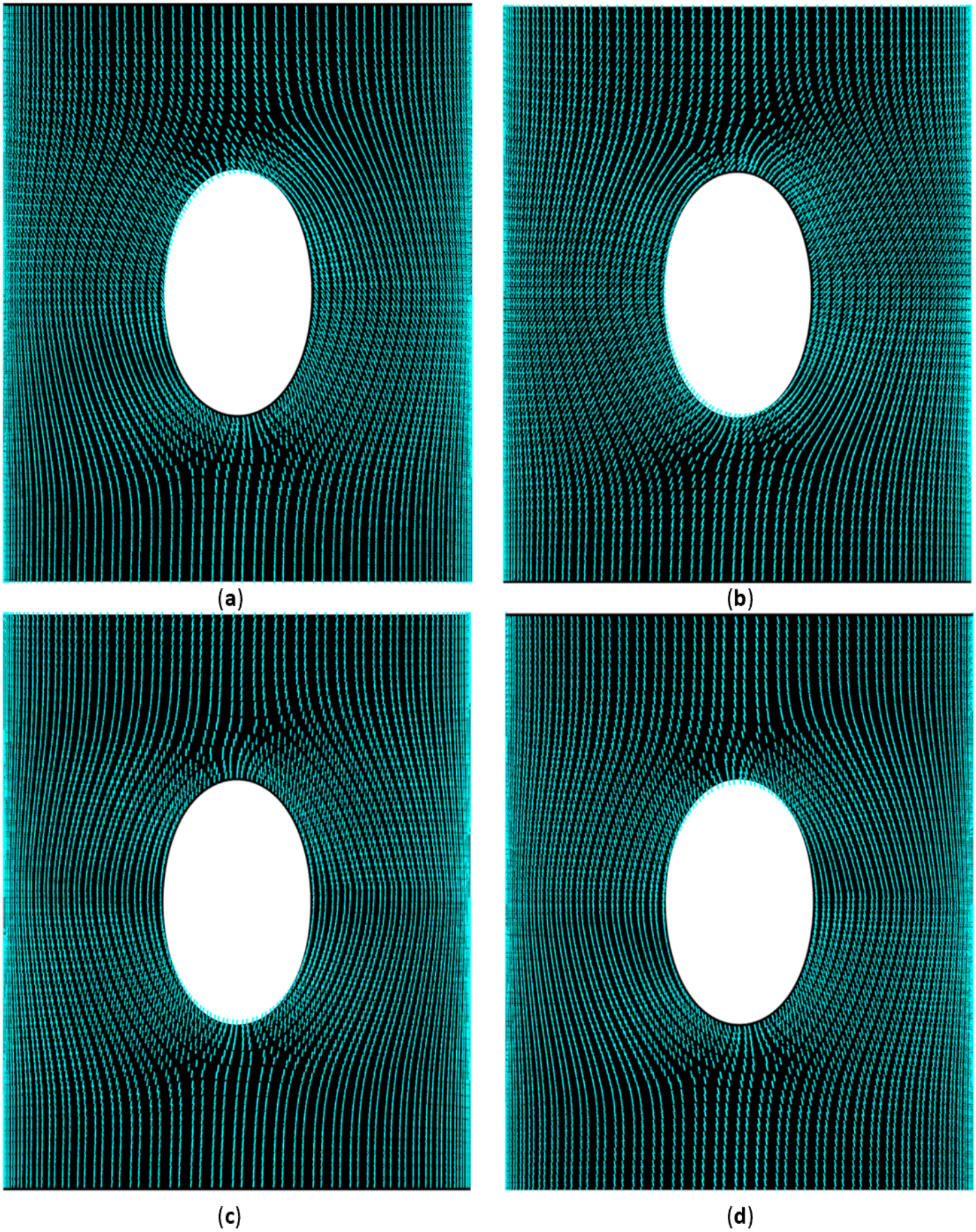


Figure 4.81: Angle variation for each symmetric ply (a), (b), (c), and (d) for the optimal FPF load design which shows the longitudinally elongated cut-out.

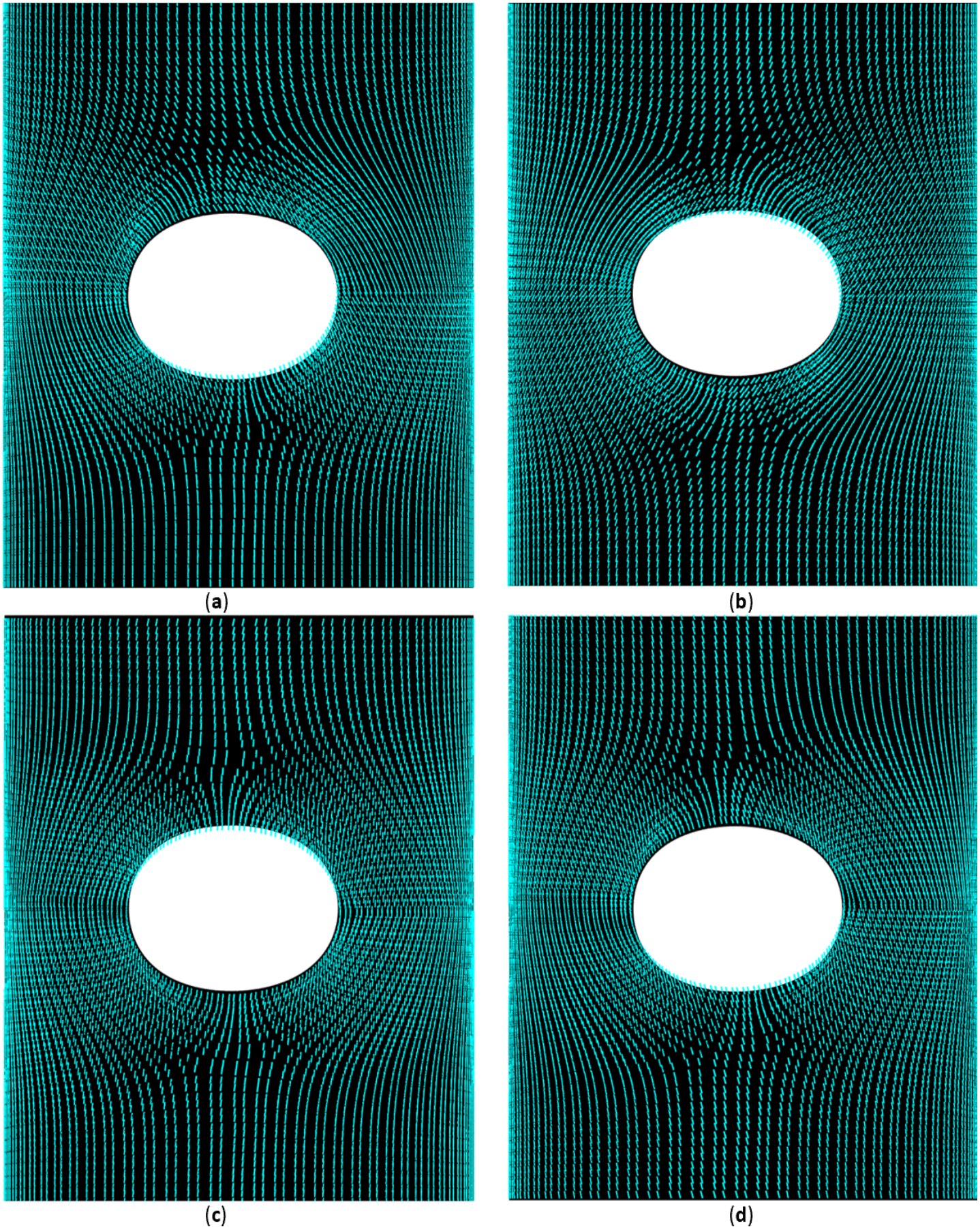


Figure 4.82: Angle variation for each symmetric ply (a), (b), (c), and (d) for the optimal FPF load design which shows the circumferentially elongated cut-out.

5

Conclusions and recommendations for future work

In this chapter, the conclusions of the thesis are presented and recommendations for possible future work are laid out subsequently. A brief summary of the work that was performed is given first, followed by the answers to the research questions that were posed at the beginning of the project.

5.1. Conclusions

The aim of this thesis was to investigate the prospects of a variable stiffness cylinder for multiple objectives and the extend to which machine learning can be utilised to expedite the optimisation process. Advanced applications of machine learning, such as artificial neural networks, were employed in order to make accurate predictions on various steps along the optimisation. Some of these predictions were pertinent to the analysis of variable stiffness structures, along with their manufacturing implications. An investigation of a pristine cylinder and one containing cut-outs was performed, so that their differences with respect to their variation in stiffness can be studied. A constant stiffness cylinder served as a baseline reference in each case. The axial and circumferential shifting manufacturing techniques were also implemented so that their influence on the structural behaviour can be compared. Design spaces of various sizes and varying complexity were also incorporated, so that a comparison between manufacturing feasibility, Pareto optimality and machine learning accuracy, could be performed. Apart from artificial neural networks, surrogate models that find popular use in literature were identified and their performance was compared on multiple tasks, in order to establish on which parts of the process they perform best.

A number of research questions were contrived to direct the focus of the project and bring it into successful completion. The primary search question was:

Within the time-span of this thesis, the objective is to explore the extent of the applicability of artificial intelligence techniques in the multi-objective optimisation of variable stiffness cylinders, carried out in an efficient manner to yield enhanced structural performance, over a basis constant stiffness cylinder.

This main question was then supported by auxiliary questions, which will be repeated and answered in the following paragraphs of this section.

Question 1

1. What is an effective modelling technique to create a variable stiffness cylinder?
   (a) What kind of fibre path formulation is best suited for the optimisation problem in terms of performance and efficiency?
   (b) To what extent can a surrogate model be exploited to include the manufacturing defects in the model in a more efficient manner?

Three different classifications of modelling the VS structure were identified and explained in the literature study. All of them have their own advantages and disadvantages, but the functional fibre representation is the

one that finds greatest usage owing to its continuity and greater control over the trajectory and curvature of the tow. Under this category, many different equations can be chosen to formulate the tow path, where they pose a trade-off between limited variables and an expanded design space. Therefore, a Lagrange polynomial that is able to represent the design space with varying control points was selected. Thus, the influence of the control points will pin-point the optimum balance for the trade-off.

The design space that is best suited for the structure is case-dependent. For the two distinct cylinders that were considered, the tow formulations that yielded the optimal results were not the same. Additionally, the shifting technique during manufacturing influences which formulation is optimal. In order to quantify the optimality of the structure, the objectives can either be considered separately or scalarised to provide a value for comparison. For the pristine cylinder following the axial angle variation technique, a greater number of control points resulted in better values in both instances. On the other hand, the circumferential angle variation performed the worst with more control points. This can be interpreted such that the design space was enlarged so much that the greatest portion of it did not contain optimal or feasible designs. When it comes to the cut-out cylinder, an increase in the number of control points was beneficial for certain individual objectives but that was not the case if overall good designs were sought-after. This can be attributed to the greater directed efficiency that the higher angle variation can achieve, resulting in designs that are intended to enhance a single objective to the detriment of the rest.

The acquisition of the thickness and angle fields for each element were an essential part to perform the analysis of the variable stiffness cylinder. The variation of the thickness field was due to the overlaps between courses, as exhibited during the manufacturing process. These two fields were calculated with a high degree of accuracy and for that reason they were a computationally expensive part of the optimisation process. Therefore, surrogate models were employed to make predictions on these properties. It was found that the angle field can be predicted with an excellent accuracy across various metrics. On the other hand, the thickness field could only be adequately predicted for the case where the number of control points was small, as the expansion of the design space requires more data to improve the surrogate models. Due to that fact, a less exact calculation of the thickness field via analytical means might prove a better alternative for a high number of control points.

Question 2

2. What is the optimal approach to the training of the artificial neural network architecture?

    (a) What is the impact of a deep neural network architecture on the multi-objective optimisation problem?

    (b) What is the most effective network architecture on average?

    (c) How much data is needed from the high-fidelity simulations in order to acquire an accurate surrogate model?

Out of all the surrogate models considered, the one with the most hyperparameters was the artificial neural network. Thus, the possible combinations regarding its architecture result in a significantly extensive design space, whose exploration is often-times intractable. For that reason, it is pertinent to consider which architectures tend to give the best results and limit the design space around such values. Although an extensive optimisation on the network architecture was not permitted by the allocated time, the non-converged results highlight the accuracy of deeper networks.

On average, architectures with 4 self-normalising hidden layers with the SELU activation function were preferred. Additionally, lower learning rates were needed for these architectures as divergence of the derivative-based optimisation of the chosen metric was observed otherwise.

The number of data that is needed for an accurate neural network varied greatly between the variables that were to be predicted, and whether the stiffness was constant or variable. It is reported that the constant stiffness cylinder ended up with a very similar number of samples with respect to the variable stiffness kind, but this is mainly attributed to the discrete training intervals that were set to be identical in both cases. In other words, the neural network achieved greater accuracy with the same amount of samples in the constant stiffness case, although it could be considered accurate for less samples than that. The networks for the variable stiffness cylinder required 3000 samples to be accurate for the objectives of frequency, axial and lateral stiffness. The predictions for buckling and first ply failure were not particularly accurate, owing to the objectives' greater complexity, and thus more samples are needed.

Question 3

3. What is the difference in the objective space between a variable stiffness cylinder and a reference constant stiffness cylinder?

The Pareto fronts and the relationship between objectives were explored for the constant and variable stiffness cylinders in order to evaluate their differences. The results were standardised with respect to the additional mass of the variable stiffness cylinders due to the overlaps. It was found that the results were very similar between the constant stiffness case and the variable stiffness instances with 2 control points. A greater discrepancy is observed for a higher number of control points and especially for the case of circumferential angle variation, which obtained a much more spread-out Pareto front than the other cases. A greater number of infeasible designs were found for the same case, with respect to the constraints that were set.

In all instances, the design space was not exploited thoroughly for fibre directions along the circumferential direction (0°), since the optimisation algorithm preferred higher values instead. The best designs for the axial stiffness objective were always identified accurately to be aligned with the cylinder's axis. The best lateral stiffness designs were balanced in every case, while no kind of pattern could be discerned for the rest of the objectives.

The constant stiffness cylinder was able to obtain the best results in almost every category for the optimisation objectives for the pristine cylinder, although its designs were generally not great when it came to overall objective performance. The variable stiffness designs with a high number of control points were the best in that regard, and they were also found to be consistent with a wider range of designs. In particular, the axial angle variation outperformed the circumferential angle variation both in overall performance and best designs. Nevertheless, both of them surpassed the best constant stiffness design for the lateral stiffness objective. An improvement of the mass-standardised lateral stiffness objective was observed, with a 2% increase of the objective value.

A significant difference is observed for the cut-out cylinder, where the variable stiffness designs were able to improve the performance of multiple objectives. The designs with a high number of control points yielded the most optimal objective values but they were not the most robust when it came to overall good designs. The constant stiffness objective space proved to be the worst, although it can still be considered a viable choice if the increased efficiency of the variable stiffness scheme does not make up for the additional mass that is gained for a given design. Drastic improvements were witnessed for the buckling and strength objectives, since enhancements of 27% and 33% were found respectively, although for different designs. Furthermore, the frequency of the free vibrations was increased by a lesser value of 0.6%.

Question 4

4. Which artificial intelligence techniques should be included in the comparison and which one yields the best results?

In order to determine which surrogate model performs best for a given task, it was decided that different surrogate models should be compared. In aerospace in general, and in variable stiffness structures specifically, two surrogate models are commonly employed, namely the gaussian radial basis functions model and the Kriging model. Thus, the artificial neural network was pitted against them, so that the strengths and weaknesses of each surrogate model could be identified within the framework of the project.

When it comes to predicting the structural objectives of a constant stiffness cylinder, they all performed exceptionally well with satisfying accuracies. This is attributed to the low number of dimensionality of the problem. A time comparison was also performed, where apart from the aforementioned models, two different neural networks were included, where they had a shallow and a deep architecture respectively. Both neural networks proved to be slower at training and making predictions, while having a similar accuracy to the rest. This highlights a potential drawback for the neural network when employed for tasks with low-dimensionality.

The variable stiffness cylinders contained a significantly higher amount of inputs, leading to a completely different problem. The first task for the surrogate models was to predict the properties of each element of the discretised geometry, such as the angle and thickness values. Once again, the angle field was predicted with superb accuracy by all the considered surrogate models, although the neural network performed the best. On the other hand, the thickness field could only be predicted accurately by the neural network, since the other models showcased very poor performance and could not be used for that task. The structural objectives were

also predicted for the variable stiffness cylinders, albeit with a much greater dimensionality. This time, the radial basis functions exhibited the worst performance, while the Kriging model performed significantly better. Nevertheless, the artificial neural network yielded the best performance for every objective. A limitation of the Kriging model was observed, where it would take orders of magnitude more computational time to train than the rest of the models. Both neural networks scored the best time in the same category and specifically the shallower variant. In regards to the prediction time, the radial basis functions performed the poorest, while the Kriging model was the fastest. The neural networks had very similar prediction times and were a close second to the Kriging model. Overall, all of the above highlights that the neural networks are able to outperform conventional surrogate models in accuracy, training and prediction time for high-dimensionality tasks, but not necessarily for lower ones.

5.2. Recommendations for future work

Owing to the vast possibilities of the investigated fields of variable stiffness and machine learning, the current work focused on a number of aspects of interest. Therefore, the work can be expanded and different considerations can be addressed by acting on the recommendations that are presented in this section.

A decision was made on the equation that characterises the functional representation of the curve of the angle variation based on a trade-off between time and the degrees of freedom of the curves. An alternative to the current methodology could incorporate non-uniform rational B-splines curve instead of a Lagrange polynomial, so that the design space can be enlarged significantly. In this case, the increased control points can serve as an input to an artificial neural network, which handles high-dimensional problems well, in order to make a prediction on a structural objective. However, it is recommended that only a single-objective optimisation is performed in this instance, since the time taken for the optimisation algorithm to converge could prove to be prohibitively high due to the enlarged design space.

In this work, the manufacturing method of choice was the automated fibre placement, albeit other manufacturing methods possess certain advantages over it. It seems that the continuous tow shearing approach is very promising, as the increasing research on the topic suggests. This method allows for a higher curvature limit and an analytical way to compute the thickness variation across the structure. The latter point can save a significant amount of time, since the numerical calculation of the thickness field is a computationally expensive process. Notwithstanding additional manufacturing considerations, the analyst can potentially yield superior results with the continuous tow shearing method.

The optimisation was performed for a certain number of objectives on a specific structure, highlighting the improvements that can be achieved by a variable stiffness application. Nevertheless, the extent of improvement for other objectives, loading conditions, and for cylinders of different geometries should also be explored. Therefore, combinations of internal pressure, bending, torsional, axial, and thermal loads can be applied simulteneously. In regards to geometry, a parametric analysis of different aspect ratios and different cylindrical elliptical shapes could also be performed. Additionally, the inclusion of curvilinear stiffeners against traditional stiffeners is another interesting topic that is likely to further enhance the structural performance.

Finally, the explosion of the popularity of neural networks primarily comes from the success of convolutional neural networks in image classification. A possible way to improve the accuracy of the surrogate model is by utilising such a network. In order to do that, the problem could be cast into a different form by mapping the spatial coordinates of the curves that define the position of the curve, or the values of the element properties, as images. In that way, the network may be able to better understand existing symmetries and features at the micro and macro levels.

Bibliography

[1] Z. Gurdal and R. Olmedo. In-Plane Response of Laminates with Spatially Varying Fiber Orientations - Variable Stiffness Concept. *AIAA Journal*, 31:751–758, 1993.

[2] E. Labans and C. Bisagni. Buckling and Free Vibration Study of Variable and Constant-Stiffness Cylindrical Shells. *Composite Structures*, 210:446–457, 2019.

[3] M. W. Hyer and R. F. Charette. Use of Curvilinear Fiber Format in Composite Structure Design. *AIAA Journal*, 29:1011–1015, 1991.

[4] Y. Xu, J. Zhu, Z. Wu, Y. Cao, Y. Zhao, and W. Zhang. A Review on the Design of Laminated Composite Structures: Constant and Variable Stiffness Design and Topology Optimization. *Advanced Composites and Hybrid Materials*, 1(3):460–477, 2018.

[5] P. Zhu, F. Pan, W. Chen, and F. A. C. Viana. Lightweight Design of Vehicle Parameters Under Crashworthiness Using Conservative Surrogates. *Computers in Industry*, 64:280–289, 2013.

[6] X. Fu, S. Ricci, and C. Bisagni. Minimum-Weight Design for Three Dimensional Woven Composite Stiffened Panels Using Neural Networks and Genetic Algorithms. *Composite Structures*, 134:708–715, 2015.

[7] K. Kalnins, R. Rikards, J. Auzins, C. Bisagni, H. Abramovich, and R. Degenhardt. Metamodeling Methodology for Postbuckling Simulation of Damaged Composite Stiffened Structures with Physical Validation. *International Journal of Structural Stability and Dynamics*, 10:705–716, 2010.

[8] M. J. Asher, B. F. W. Croke, A. J. Jakeman, and D. J. M. Peeters. A Review of Surrogate Models and their Application to Groundwater Modeling. *Water Resources Research*, 51:5957–5973, 2015.

[9] M. A. Bouhlel, J. T. Hwang, N. Bartoli, R. Lafage, J. Morlier, and J. R. R. A. Martins. A Python Surrogate Modeling Framework with Derivatives. *Advances in Engineering Software*, 135:1–13, 2019.

[10] K. Hinkelmann. *Design and Analysis of Experiments, Volume 3: Special Designs and Applications*. John Wiley & Sons, 2011.

[11] N. V. Queipo, R. T. Haftka, W. Shyy, T. Goel, R. Vaidyanathan, and P. K. Tucker. Surrogate-Based Analysis and Optimization. *Progress in Aerospace Sciences*, 41:1–28, 2005.

[12] J. E. Gentle. *Random Number Generation and Monte Carlo Methods*. Springer Science & Business Media, 2006.

[13] A. Géron. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, 2019.

[14] M. J. D. Powell. The Theory of Radial Basis Function Approximation in 1990. In *Advances in Numerical Analysis*, pages 105–210. Clarendon Press, 1992.

[15] L. Peng, L. Liu, and T. Long. Study of Sequential Radial Basis Function for Computation-Intensive Design Problem. In *12th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference and 14th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, pages 1–12, 2017.

[16] R. Yondo, E. Andres, and E. Valero. A Review on Design of Experiments and Surrogate Models in Aircraft Real-Time and Many-Query Aerodynamic Analyses. *Progress in Aerospace Sciences*, 96:23–61, 2018.

[17] O. Barak. Recurrent Neural Networks as Versatile Tools of Neuroscience Research. *Current Opinion in Neurobiology*, 46:1–6, 2017.

[18] D. Fisch, A. Hofmann, and B. Sick. On the Versatility of Radial Basis Function Neural Networks: A Case Study in the Field of Intrusion Detection. *Information Sciences*, 180:2421–2439, 2010.

[19] L. K. Song, G. C. Bai, and C. W. Fei. Multi-Failure Probabilistic Design for Turbine Bladed Disks Using Neural Network Regression with Distributed Collaborative Strategy. *Aerospace Science and Technology*, 92:464–477, 2019.

[20] F. Rosenblatt. The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review*, 65:386–408, 1958.

[21] M. Minsky and S. A. Papert. *Perceptrons: An Introduction to Computational Geometry*. MIT Press, 2017.

[22] H. Hassan, N. Abdelazim, Z. Mohamed, and S. Oliver. Assessment of Artificial Neural Network for Bathymetry Estimation Using High Resolution Satellite Imagery in Shallow Lakes: Case Study El Burullus Lake. *International Water Technology*, 7:248–259, 2015.

[23] J. Heaton. *Introduction to Neural Networks with Java*. Heaton Research, Inc., 2008.

[24] K. He, X. Zhang, S. Ren, and J. Sun. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[25] F. Chollet. *Deep Learning with Python*. MITP-Verlags GmbH & Co. KG, 2018.

[26] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning Internal Representations by Error Propagation. Technical report, California University San Diego La Jolla Institute for Cognitive Science, 1985.

[27] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016.

[28] M. A. Nielsen. *Neural Networks and Deep Learning*. Determination Press, 2015.

[29] B. Xu, N. Wang, T. Chen, and M. Li. Empirical Evaluation of Rectified Activations in Convolutional Network. *Cornell University Library*, 2015.

[30] D. A. Clevert, T. Unterthiner, and S. Hochreiter. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). *Cornell University Library*, 2015.

[31] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter. Self-Normalizing Neural Networks. In *Advances in Neural Information Processing Systems*, pages 971–980, 2017.

[32] P. Y. Papalambros and D. J. Wilde. *Principles of Optimal Design: Modeling and Computation*. Cambridge University Press, 2000.

[33] C. Bisagni, L. Lanzi, and S. Ricci. Optimization of Helicopter Subfloor Components Under Crashworthiness Requirements Using Neural Networks. *Journal of Aircraft*, 39:296–304, 2002.

[34] V. Chandwani, V. Agrawal, and R. Nagar. Modeling Slump of Ready Mix Concrete Using Genetic Algorithms Assisted Training of Artificial Neural Networks. *Expert Systems with Applications*, 42:885–893, 2015.

[35] T. T. Truong, S. Lee, and J. Lee. An Artificial Neural Network-Differential Evolution Approach for Optimization of Bidirectional Functionally Graded Beams. *Composite Structures*, 233:1–11, 2019.

[36] S. F. Pitton, S. Ricci, and C. Bisagni. Buckling Optimization of Variable Stiffness Cylindrical Shells Through Artificial Intelligence Techniques. *Composite Structures*, 230:1–9, 2019.

[37] A. I. Galushkin. *Neural Networks Theory*. Springer Science & Business Media, 2007.

[38] X. Glorot and Y. Bengio. Understanding the Difficulty of Training Deep Feedforward Neural Networks. In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics*, pages 249–256, 2010.

[39] S. Ioffe and C. Szegedy. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *Cornell University Library*, 2015.

[40] L. Prechelt. Early Stopping – But When? In *Neural Networks: Tricks of the Trade*, pages 55–69. Springer, 1998.

[41] Four Years Remaining » Blog Archive » The Mystery of Early Stopping. URL `http://fouryears.eu/2017/12/06/the-mystery-of-early-stopping/Accessed[2019-11-17]`.

[42] B. T. Polyak. Some Methods of Speeding up the Convergence of Iteration Methods. *USSR Computational Mathematics and Mathematical Physics*, 4:1–17, 1964.

[43] Y. E. Nesterov. A Method for Solving the Convex Programming Problem with Convergence Rate O (1/kˆ 2). In *Proceedings of the USSR Academy of Sciences*, pages 543–547, 1983.

[44] J. Duchi, E. Hazan, and Y. Singer. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. *Journal of Machine Learning Research*, 12:2121–2159, 2011.

[45] G. Hinton, N. Srivastava, and K. Swersky. Neural Networks for Machine Learning – Overview of Mini-Batch Gradient Descent. *University of Toronto Computer Science Lecture Notes*, 2012.

[46] D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. *Cornell University Library*, 2014.

[47] T. Dozat. Incorporating Nesterov Momentum Into Adam. In *International Conference on Learning Representations*, 2016.

[48] D. H. Wolpert and W. G. Macready. No Free Lunch Theorems for Optimization. *IEEE Transactions on Evolutionary Computation*, 1:67–82, 1997.

[49] R. T. Marler and J. S. Arora. Survey of Multi-Objective Optimization Methods for Engineering. *Structural and Multidisciplinary Optimization*, 26:369–395, 2004.

[50] K. Mason, J. Duggan, and E. Howley. Evolving Multi-Objective Neural Networks Using Differential Evolution for Dynamic Economic Emission Dispatch. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, pages 1287–1294, 2017.

[51] M. J. Kochenderfer and T. A. Wheeler. *Algorithms for Optimization*. MIT Press, 2019.

[52] S. Luke. *Essentials of Metaheuristics*. Springer, 2011.

[53] I. Rechenberg. *Evolutionsstrategie; Optimierung Technischer Systeme nach Prinzipien der Biologischen Evolution*. Frommann–Holzboog, 1973.

[54] K. Sörensen. Metaheuristics – The Metaphor Exposed. *International Transactions in Operational Research*, 22:3–18, 2015.

[55] J. McCall. Genetic Algorithms for Modelling and Optimisation. *Journal of Computational and Applied Mathematics*, 184:205–222, 2005.

[56] J. H. Holland. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. University of Michigan Press, 1975.

[57] D. Simon. *Evolutionary Optimization Algorithms*. John Wiley & Sons, 2013.

[58] D. E. Goldberg and K. Deb. A Comparative Analysis of Selection Schemes Used in Genetic Algorithms. In *Foundations of Genetic Algorithms*, pages 69–93. Elsevier, 1991.

[59] J. F. Crow and M. Kimura. Efficiency of Truncation Selection. *Proceedings of the National Academy of Sciences*, 76:396–399, 1979.

[60] A. Lipowski and D. Lipowska. Roulette-Wheel Selection via Stochastic Acceptance. *Physica A: Statistical Mechanics and its Applications*, 391:2193–2196, 2012.

[61] B. L. Miller and D. E. Goldberg. Genetic Algorithms, Tournament Selection, and the Effects of Noise. *Complex systems*, 9:193–212, 1995.

[62] A. Abraham, N. Nedjah, and L. de Macedo Mourelle. Evolutionary Computation: From Genetic Algorithms to Genetic Programming. In *Genetic Systems Programming*, pages 1–20. Springer, 2006.

[63] G. Syswerda. Simulated Crossover in Genetic Algorithms. In *Foundations of Genetic Algorithms*, pages 239–255. Elsevier, 1993.

[64] J. Magalhaes-Mendes. A Comparative Study of Crossover Operators for Genetic Algorithms to Solve the Job Shop Scheduling Problem. *WSEAS Transactions on Computers*, 12:164–173, 2013.

[65] A. J. Umbarkar and P. D. Sheth. Crossover Operators in Genetic Algorithms: A Review. *ICTACT Journal on Soft Computing*, 6:1083–1092, 2015.

[66] K. Deb and R. B. Agrawal. Simulated Binary Crossover for Continuous Search Space. *Complex Systems*, 9:115–148, 1995.

[67] K. Deb, K. Sindhya, and T. Okabe. Self-Adaptive Simulated Binary Crossover for Real-Parameter Optimization. In *Proceedings of the 9th Annual Conference on Genetic and Evolutionary Computation*, pages 1187–1194, 2007.

[68] O. Kramer and H. P. Schwefel. On Three New Approaches to Handle Constraints Within Evolution Strategies. *Natural Computing*, 5:363–385, 2006.

[69] R. Storn and K. Price. Differential Evolution – A Simple and Efficient Heuristic for Global Optimization Over Continuous Spaces. *Journal of Global Optimization*, 11:341–359, 1997.

[70] K. V. Price, R. M. Storn, and J. A. Lampinen. *Differential Evolution: A Practical Approach to Global Optimization*. Springer, 2005.

[71] S. Das and P. N. Suganthan. Differential Evolution: A Survey of the State-of-the-Art. *IEEE Transactions on Evolutionary Computation*, 15:4–31, 2010.

[72] S. M. Elsayed, R. A. Sarker, and D. L. Essam. Differential Evolution with Multiple Strategies for Solving CEC2011 Real-World Numerical Optimization Problems. In *2011 IEEE Congress of Evolutionary Computation (CEC)*, pages 1041–1048. IEEE, 2011.

[73] N. Srinivas and K. Deb. Multi-Objective Optimization Using Nondominated Sorting in Genetic Algorithms. *Evolutionary Computation*, 2:221–248, 1994.

[74] K. Deb, A. Pratap, S. Agarwal, and T. A. M. T. Meyarivan. A Fast and Elitist Multi-Objective Genetic Algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6:182–197, 2002.

[75] Q. Zhang and H. Li. MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition. *IEEE Transactions on Evolutionary Computation*, 11:712–731, 2007.

[76] A. Haghighi and A. Z. Asl. Uncertainty Analysis of Water Supply Networks Using the Fuzzy Set Theory and NSGA-II. *Engineering Applications of Artificial Intelligence*, 32:270–282, 2014.

[77] T. Vo-Duy, D. Duong-Gia, V. Ho-Huu, H. C. Vu-Do, and T. Nguyen-Thoi. Multi-Objective Optimization of Laminated Composite Beam Structures Using NSGA-II Algorithm. *Composite Structures*, 168:498–509, 2017.

[78] P. Jiang, C. Wang, Q. Zhou, X. Shao, L. Shu, and X. Li. Optimization of Laser Welding Process Parameters of Stainless Steel 316L Using FEM, Kriging and NSGA-II. *Advances in Engineering Software*, 99:147–160, 2016.

[79] K. Abhishek, S. Chatterjee, S. Datta, and S. S. Mahapatra. Application of NSGA II for Optimization of Multi-Performance Characteristics During Machining of GFRP (Epoxy) Composites. *Materials Today: Proceedings*, 2:2353–2358, 2015.

[80] A. Kamaloo, M. Jabbari, M. Y. Tooski, and M. Javadi. Optimization of Thickness and Delamination Growth in Composite Laminates Under Multi-Axial Fatigue Loading Using NSGA-II. *Composites Part B: Engineering*, 174:1–9, 2019.

[81] K. Deb and H. Jain. An Evolutionary Many-Objective Optimization Algorithm Using Reference-Point-Based Nondominated Sorting Approach, Part I: Solving Problems with Box Constraints. *IEEE Transactions on Evolutionary Computation*, 18:577–601, 2013.

[82] H. Jain and K. Deb. An Evolutionary Many-Objective Optimization Algorithm Using Reference-Point-Based Nondominated Sorting Approach, Part II: Handling Constraints and Extending to an Adaptive Approach. *IEEE Transactions on Evolutionary Computation*, 18:602–622, 2013.

[83] I. Das and J. E. Dennis. Normal-Boundary Intersection: A New Method for Generating the Pareto Surface in Nonlinear Multicriteria Optimization Problems. *SIAM Journal on Optimization*, 8:631–657, 1998.

[84] H. Seada and K. Deb. U-NSGA-III: A Unified Evolutionary Algorithm for Single, Multiple, and Many-Objective Optimization. Technical report, COIN Laboratory of Michigan State University, 2014.

[85] K. Deb and H. Jain. An Evolutionary Many-Objective Optimization Algorithm Using Reference-Point-Based Nondominated Sorting Approach, Part I: Solving Problems with Box Constraints. *IEEE Transactions on Evolutionary Computation*, 18:577–601, 2013.

[86] E. Mezura-Montes and C. A. C. Coello. Constraint-Handling in Nature-Inspired Numerical Optimization: Past, Present and Future. *Swarm and Evolutionary Computation*, 1:173–194, 2011.

[87] P. D. Surry and N. J. Radcliffe. The COMOGA Method: Constrained Optimisation by Multi-Objective Genetic Algorithms. *Control and Cybernetics*, 26:391–412, 1997.

[88] C. A. C. Coello. Treating Constraints as Objectives for Single-Objective Evolutionary Optimization. *Engineering Optimization+ A35*, 32:275–308, 2000.

[89] J. A. Joines and C. R. Houck. On the Use of Non-Stationary Penalty Functions to Solve Nonlinear Constrained Optimization Problems with GA's. In *Proceedings of the 1st IEEE Conference on Evolutionary Computation*, pages 579–584, 1994.

[90] O. Kramer. *Genetic Algorithm Essentials*. Springer, 2017.

[91] R. Farmani and J. A. Wright. Self-Adaptive Fitness Formulation for Constrained Optimization. *IEEE Transactions on Evolutionary Computation*, 7:445–455, 2003.

[92] C. M. Fonseca, L. Paquete, and M. López-Ibánez. An Improved Dimension-Sweep Algorithm for the Hypervolume Indicator. In *IEEE International Conference on Evolutionary Computation*, pages 1157–1163, 2006.

[93] G. G. Lozano, A. Tiwari, C. Turner, and S. Astwood. A Review on Design for Manufacture of Variable Stiffness Composite Laminates. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 230:981–992, 2016.

[94] L. Bittrich, A. Spickenheuer, J. H. S. Almeida, S. Müller, L. Kroll, and G. Heinrich. Optimizing Variable-Axial Fiber-Reinforced Composite Laminates: The Direct Fiber Path Optimization Concept. *Mathematical Problems in Engineering*, 2019:1–11, 2019.

[95] D. O. Evans, M. M. Vaniglia, and P. C. Hopkins. Fiber Placement Process Study. *Tomorrow's Materials: Today*, 34:1822–1833, 1989.

[96] H. J. A. L. Dirk, C. Ward, and K. D. Potter. The Engineering Aspects of Automated Prepreg Layup: History, Present and Future. *Composites Part B: Engineering*, 43:997–1009, 2012.

[97] A. W. Blom, C. S. Lopes, P. J. Kromwijk, Z. Gurdal, and P. P. Camanho. A Theoretical Model to Study the Influence of Tow-Drop Areas on the Stiffness and Strength of Variable-Stiffness Laminates. *Journal of Composite Materials*, 43:403–425, 2009.

[98] B. F. Tatting and Z. Gürdal. *Design and Manufacture of Elastically Tailored Tow Placed Plates*. National Aeronautics and Space Administration, Langley Research Center, 2002.

[99] G. Marsh. Automating Aerospace Composites Production with Fibre Placement. *Reinforced Plastics*, 55:32–37, 2011.

[100] C. Waldhart. *Analysis of Tow-Placed, Variable-Stiffness Laminates*. PhD thesis, Virginia Tech, 1996.

[101] F. Heinecke and C. Willberg. Manufacturing-Induced Imperfections in Composite Parts Manufactured via Automated Fiber Placement. *Journal of Composites Science*, 3:56, 2019.

[102] S. G. P. Castro, M. V. Donadon, and T. A. M. Guimarães. ES-PIM Applied to Buckling of Variable Angle Tow Laminates. *Composite Structures*, 209:67–78, 2019.

[103] B. C. Kim, K. Potter, and P. M. Weaver. Continuous Tow Shearing for Manufacturing Variable Angle Tow Composites. *Composites Part A: Applied Science and Manufacturing*, 43:1347–1356, 2012.

[104] B. F. Tatting and Z. Gürdal. Automated Finite Element Analysis of Elastically-Tailored Plates. Technical report, National Aeronautics and Space Administration, 2003.

[105] A. W. Blom, P. B. Stickler, and Z. Gürdal. Optimization of a Composite Cylinder Under Bending by Tailoring Stiffness Properties in Circumferential Direction. *Composites Part B: Engineering*, 41:157–165, 2010.

[106] C. S. Lopes. *Damage and Failure of Non-Conventional Composite Laminates*. PhD thesis, Delft University of Technology, 2009.

[107] A. Sabido, L. Bahamonde, R. Harik, and M. J. L. van Tooren. Maturity Assessment of the Laminate Variable Stiffness Design Process. *Composite Structures*, 160:804–812, 2017.

[108] B. C. Kim, P. M. Weaver, and K. Potter. Computer Aided Modelling of Variable Angle Tow Composites Manufactured by Continuous Tow Shearing. *Composite Structures*, 129:256–267, 2015.

[109] B. C. Kim, P. M. Weaver, and K. Potter. Manufacturing Characteristics of the Continuous Tow Shearing Method for Manufacturing of Variable Angle Tow Composites. *Composites Part A: Applied Science and Manufacturing*, 61:141–151, 2014.

[110] S. T. IJsselmuiden. *Optimal Design of Variable Stiffness Composite Structures Using Lamination Parameters*. PhD thesis, Delft University of Technology, 2011.

[111] Z. Gürdal, B. F. Tatting, and C. K. Wu. Variable Stiffness Composite Panels: Effects of Stiffness Variation on the In-Plane and Buckling Response. *Composites Part A: Applied Science and Manufacturing*, 39: 911–922, 2008.

[112] S. Nagendra, S. Kodiyalam, J. Davis, and V. Parthasarathy. Optimization of Tow Fiber Paths for Composite Design. In *36th Structures, Structural Dynamics and Materials Conference*, pages 1031–1040, 1995.

[113] L. Demasi, G. Biagini, F. Vannucci, E. Santarpia, and R. Cavallaro. Equivalent Single Layer, Zig-Zag, and Layer Wise Theories for Variable Angle Tow Composites Based on the Generalized Unified Formulation. *Composite Structures*, 177:54–79, 2017.

[114] S. Honda and Y. Narita. Natural Frequencies and Vibration Modes of Laminated Composite Plates Reinforced with Arbitrary Curvilinear Fiber Shape Paths. *Journal of Sound and Vibration*, 331:180–191, 2012.

[115] T. A. Guimaraes, S. G. Castro, D. A. Rade, and C. E. Cesnik. Panel Flutter Analysis and Optimization of Composite Tow Steered Plates. In *58th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, pages 1–14, 2017.

[116] Z. Wu, P. M. Weaver, G. Raju, and B. C. Kim. Buckling Analysis and Optimisation of Variable Angle Tow Composite Plates. *Thin-Walled Structures*, 60:163–172, 2012.

[117] C. S. Lopes, Z. Gürdal, and P. P. Camanho. Tailoring for Strength of Composite Steered-Fibre Panels with Cutouts. *Composites Part A: Applied Science and Manufacturing*, 41:1760–1767, 2010.

[118] B. F. Tatting. *Analysis and Design of Variable Stiffness Composite Cylinders*. PhD thesis, Virginia Tech, 1998.

[119] S. Guldu and A. Kayran. Maximizing Buckling Load Factors of Fiber-Placed Composite Cylindrical Shells by Particle Swarm Optimization. In *56th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, pages 1–25, 2015.

[120] M. Rouhi, H. Ghayoor, J. Fortin-Simpson, T. T. Zacchia, S. V. Hoa, and M. Hojjati. Design, Manufacturing, and Testing of a Variable Stiffness Composite Cylinder. *Composite Structures*, 184:146–152, 2018.

[121] D. M. J. Peeters, S. Hesse, and M. M. Abdalla. Stacking Sequence Optimisation of Variable Stiffness Laminates with Manufacturing Constraints. *Composite Structures*, 125:596–604, 2015.

[122] J. L. Grenestedt. Layup Optimization Against Buckling of Shear Panels. *Structural Optimization*, 3: 115–120, 1991.

[123] C. Kassapoglou. *Design and Analysis of Composite Structures: With Applications to Aerospace Structures*. John Wiley & Sons, 2013.

[124] B. Esp. *Practical Analysis of Aircraft Composites*. Grand Oak Publishing, 2017.

[125] H. T. Hahn and S. W. Tsai. *Introduction to Composite Materials*. CRC Press, 1980.

[126] V. B. Hammer, M. P. Bendsøe, R. Lipton, and P. Pedersen. Parametrization in Laminate Design for Optimal Compliance. *International Journal of Solids and Structures*, 34:415–434, 1997.

[127] D. M. J. Peeters, F. X. Irisarri, C. Groenendijk, and R. Ružek. Optimal Design, Manufacturing and Testing of Non-Conventional Laminates. *Composite Structures*, 210:29–40, 2019.

[128] J. M. J. F. Van Campen, C. Kassapoglou, and Z. Gürdal. Generating Realistic Laminate Fiber Angle Distributions for Optimal Variable Stiffness Laminates. *Composites Part B: Engineering*, 43:354–360, 2012.

[129] X. Y. Zhou and P. D. Gosling. Towards an Understanding of Variations in the Buckling of Tailored Variable Angle Tow Composite Plates. *Composite Structures*, 203:797–809, 2018.

[130] P. T. Langley. *Finite Element Modeling of Tow-Placed Variable-Stiffness Composite Laminates*. PhD thesis, Virginia Tech, 1999.

[131] K. Fayazbakhsh, M. A. Nik, D. Pasini, and L. Lessard. Defect Layer Method to Capture Effect of Gaps and Overlaps in Variable Stiffness Laminates Made by Automated Fiber Placement. *Composite Structures*, 97:245–251, 2013.

[132] A. Noevere and C. Collier. Mapping Manufacturing Data for Stress Analysis of Automated Fiber Placement Structures. In *2018 AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, pages 1–21, 2018.

[133] A. Marouene, P. Legay, and R. Boukhili. Experimental and Numerical Investigation on the Open-Hole Compressive Strength of AFP Composites Containing Gaps and Overlaps. *Journal of Composite Materials*, 51:3631–3646, 2017.

[134] A. Marouene, R. Boukhili, J. Chen, and A. Yousefpour. Effects of Gaps and Overlaps on the Buckling Behavior of Optimally Designed Variable-Stiffness Composite Laminates – A Numerical and Experimental Study. *Composite Structures*, 140:556–566, 2016.

[135] M. A. Nik, K. Fayazbakhsh, D. Pasini, and L. Lessard. Optimization of Variable Stiffness Composites with Embedded Defects Induced by Automated Fiber Placement. *Composite Structures*, 107:160–166, 2014.

[136] V. Mishra, D. M. J. Peeters, and M. M. Abdalla. Stiffness and Buckling Analysis of Variable Stiffness Laminates Including the Effect of Automated Fibre Placement Defects. *Composite Structures*, 226:1–20, 2019.

[137] H. Ghiasi, K. Fayazbakhsh, D. Pasini, and L. Lessard. Optimum Stacking Sequence Design of Composite Materials Part II : Variable Stiffness Design. *Composite Structures*, 93:1–13, 2010.

[138] X. Legrand, D. Kelly, A. Crosky, and D. Crépin. Optimisation of Fibre Steering in Composite Laminates Using a Genetic Algorithm. *Composite Structures*, 75:524–531, 2006.

[139] J. Huang and R. T. Haftka. Optimization of Fiber Orientations Near a Hole for Increased Load-Carrying Capacity of Composite Laminates. *Structural and Multidisciplinary Optimization*, 30:335–341, 2005.

[140] D. M. J. Peeters, F. X. Irisarri, C. Groenendijk, and R. Růžek. Optimal Design, Manufacturing and Testing of Non-Conventional Laminates. *Composite Structures*, 210:29–40, 2019.

[141] M. Rouhi, H. Ghayoor, S. V. Hoa, and M. Hojjati. Computational Efficiency and Accuracy of Multi-Step Design Optimization Method for Variable Stiffness Composite Structures. *Thin-Walled Structures*, 113: 136–143, 2017.

[142] M. A. Nik, K. Fayazbakhsh, D. Pasini, and L. Lessard. A Comparative Study of Metamodeling Methods for the Design Optimization of Variable Stiffness Composites. *Composite Structures*, 107:494–501, 2014.

[143] T. A. M. Guimarães, S. G. P. Castro, C. E. S. Cesnik, and D. A. Rade. Supersonic Flutter and Buckling Optimization of Tow-Steered Composite Plates. *AIAA Journal*, 57:397–407, 2019.

[144] M. Rouhi, H. Ghayoor, S. V. Hoa, and M. Hojjati. Multi-Objective Design Optimization of Variable Stiffness Composite Cylinders. *Composites Part B: Engineering*, 69:249–255, 2015.

[145] G. S. Babu, P. Zhao, and X. L. Li. Deep Convolutional Neural Network Based Regression Approach for Estimation of Remaining Useful Life. In *International Conference on Database Systems for Advanced Applications*, pages 214–228. Springer, 2016.

[146] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, and M. Isard. Tensorflow: A System for Large-Scale Machine Learning. In *12th Symposium on Operating Systems Design and Implementation*, pages 265–283, 2016.

[147] J. Blank and K. Deb. Pymoo: Multi-Objective Optimization in Python. *IEEE Access*, 8:89497–89509, 2020.

[148] E. Clarkson. Hexcel 8552 AS4 Unidirectional Prepreg Qualification Statistical Analysis Report. Technical report, National Institute for Aviation Research, Wichita State University, 2011.

[149] R. Hooke and T. A. Jeeves. "Direct Search" Solution of Numerical and Statistical Problems. *Journal of the Association for Computing Machinery*, 8:212–229, 1961.

[150] M. J. D. Powell. A Direct Search Optimization Method that Models the Objective and Constraint Functions by Linear Interpolation. In *Advances in Optimization and Numerical Analysis*, pages 51–67. Springer, 1994.

[151] Dassault Systèmes Simulia Corp. *ABAQUS 2016 User's Guide Volume II: Analysis*. Providence, RI, USA, 2016.

[152] R. Burden and J. D. Faires. *Numerical Analysis*. Cengage Learning, 2004.

[153] Dassault Systèmes Simulia Corp. *ABAQUS 2016*. Providence, RI, USA, 2016.

[154] J. L. Loeppky, J. Sacks, and W. J. Welch. Choosing the Sample Size of a Computer Experiment: A Practical Guide. *Technometrics*, 51:366–376, 2009.