DOI: https://doi.org/10.4038/sljas.v25i2.8127

Sri Lankan Journal of Applied Statistics, Vol (25-2)

Comparative Study on the Kernel Performance of Gaussian Processes for Mechanical Property Predictions of Woven Carbon Fibre Composites

N. Ariyasinghe , J. Yasantha, and S. Herath*

Department of Civil Engineering, University of Moratuwa, Sri Lanka

*Corresponding author: sumuduh@uom.lk

Received: 16th May 2024/ Revised: 19th September 2024/ Published: 9th October 2024

ABSTRACT

Carbon fibre woven composites owe their exceptional mechanical properties to the characteristics of their constituent materials: fibres and resin. However, uncertainties in these constituent properties introduce variations in the macroscale mechanical performance of composites. Especially with multiple plies, and the orthotropic nature of the material, prediction and quantification of the uncertain mechanical properties at the macroscale pose a challenge and remain computationally expensive during analysis. In this study, Gaussian Process Regression (GPR) is used to predict and quantify uncertainties propagated to the macroscale mechanical properties expressed in terms of the ABD stiffness matrix due to the uncertainties in the constituent material properties at the microscale. The GPR model incorporates 16 input uncertainties for each tow in a two-ply composite and outputs the relevant ABD stiffness matrix. We investigate the applicability of different kernels in capturing the underlying relationship between microscale material properties and macroscale stiffnesses, which is crucial as the kernel choice governs the ability of the model to capture these underlying relationships. Our findings show that the predictions for Direct axial stiffness (A_{11} and A_{22}) showed excellent agreement for all kernels with NRMSE values below 8×10^{-4} and near-unity R2 scores in interpolation and good agreement in extrapolation. The deviation of NRMSE values of the ESS kernel for D_{11} (direct bending stiffness) and D_{12} (bending coupling) is comparatively high. D_{22} (direct bending stiffness) represents a good agreement for all kernels with near-unity R^2 scores. The GPR model's ability to provide outputs over a distribution of functions allowed for 95% confidence intervals for each prediction, with most stiffness predictions exhibiting variations as low as ±0.2%. Despite a 5% microscale

uncertainty, the actual macroscale uncertainty was lower, with minimal variations in A_{12}, A_{66}, D_{12}, and D_{66} even for 20% material uncertainty. These results demonstrate the effectiveness of Gaussian process regression in predicting and quantifying uncertainties in the mechanical properties of carbon fibre composites.

Keywords: Woven Composites, Gaussian Process Regression, Statistical Learning, Uncertainty Quantification, Extrapolatory Predictions

1 Introduction

Woven carbon fibre composites are gaining popularity due to their unique advantages, including a high strength-to-weight ratio, flexibility in shaping, environmental resistance, and compact storage (Deleo et al., 2020; Gangineni et al., 2022). The combination of precise weave patterns and strategically designed plies allow carbon fibre composites to excel in delivering lightweight yet robust solutions in numerous industries (Manocha and Bahl, 1988). The structure of woven composites involves a material hierarchy, with larger (macro) scales depending on the arrangement and behaviour of smaller (meso and micro) scales (Ullah et al., 2019). At the microscale, fibres and resin form tows, while at the mesoscale, these tows are woven to create the macroscale composite (Campbell, 2004). Due to their woven structure, these composites exhibit near-anisotropic properties, with complex load transfer mechanisms. Moreover, the constituent properties and their variations could also lead to complex material constitutive models with complicated mechanical responses of the composite on its macroscale. Therefore, it is vital to understand the micromechanics of the woven and composite actions at both the design and manufacturing stages.

This highlights the importance of prediction and Uncertainty Quantification (UQ). Numerical models are generally used in mechanical analysis, though, the assumption of ideal conditions makes it difficult to predict the various discrepancies that are observed between them and the built environment. The main reason for these observations is the uncertainties that are introduced at different scales/stages; starting from sourcing materials to manufacturing processes and other physical parameters that change during the operational period (Naskar et al., 2018; Alazwari and Rao, 2019). Therefore, quantification of these uncertainties is key to connecting the idealised models of composite material behaviour with how they actually perform in real-world conditions. Sullivan (2015) describes UQ as the merger of probability theory with statistical methods to address real-world challenges. Consequently, by combining probability theory with data-driven approaches, we can achieve both predictive capabilities and UQ, providing a deeper understanding of material responses and allowing for better-informed design decisions.

To address UQ, uncertainties must first be integrated into the model using appropriate techniques. Woven composites contain various sources of uncertainty, spanning across morphological, geometrical, and material properties

(Bostanabad et al., 2018; Clément et al., 2013) resulting in uncertainties in the overall behaviour of the material. UQ carried out on woven composites by Bostanabad et al. (2018) uses Gaussian random fields to introduce uncertainty due to spatial fluctuations in yarn angles. In another study, Huang et al. (2021) employed an image-based approach to generate data samples incorporating uncertainty in fibre waviness. The introduced uncertainties are then propagated along the lengthscales using various techniques. Having a Representative Volume Element (RVE) of a multiscale simulation consisting of integration points (IP) where the properties of an IP depend on the RVE at the finer scale is the uncertainty propagation (UP) technique used by Bostanabad et al. (2018). Tao et al. (2020) have used statistical volume elements to extract the probability distributions of microscale and mesoscale properties which are used as inputs for the mesoscale and macroscale as their UP technique. The uncertainties that are propagated are later quantified using data-driven approaches. Balokas, Kriegesmann and Rolfes (2021) contributed to this field by using Polynomial Chaos Expansion (PCE) based surrogate models to predict mechanical properties and conduct UQ analyses in carbon fibre-reinforced composites. In a similar study by them (Balokas, Kriegesmann, Czichon and Rolfes, 2021), they utilised a mix of Artificial Neural Network (ANN) and Gaussian Process Regression (GPR) in a surrogate model to capture the nonlinear behaviour of braided composites. Furthermore, studies relying on GPR-based surrogate models show promise, with Tao et al. (2020) achieving low prediction errors (under 4%), and Li et al. (2022) using GPR for optimal woven composite design.

Despite various studies focusing on uncertainty quantification in mechanical properties of woven composites, such as Young's modulus and tensile strength (Tao et al., 2020), damage initiation and strength failure criteria (Zhu et al., 2019), and spatial stress field variations (Bostanabad et al., 2018), to the author's best knowledge, no research has been carried out to predict and quantify the uncertainties in the generalised ABD stiffness matrix which considers the constituent material (i.e. fibre and matrix) properties of the woven composite as the sources of uncertainties. Since the ABD stiffness matrix effectively captures the mechanical properties of laminated composites, which serve as an idealized representation of woven fabrics. This matrix provides a comprehensive understanding of the material's mechanics, enabling a detailed analysis of its structural behavior under various loading conditions. Therefore, a study to quantify the uncertainties of the ABD stiffness matrix due to the material properties is an important aspect that needs to be considered. In this paper, we present a GPR-based data-driven model for predicting the ABD stiffness matrix entries of woven composites at the macroscale, along with the associated uncertainties in the stiffness values. Given the critical influence of kernel selection on the performance of the GPR model, a comparative study is conducted to identify the most effective kernel for this application. A multiscale modelling technique coupled with a sampling method is used for the database generation which will then be used to train the predictive

model. Once trained, this model demonstrates the capability to predict stiffness values at the macroscale, accommodating arbitrary material uncertainties at the microscale, and subsequently quantifying potential uncertainties at the macroscale, as depicted in Figure 1.

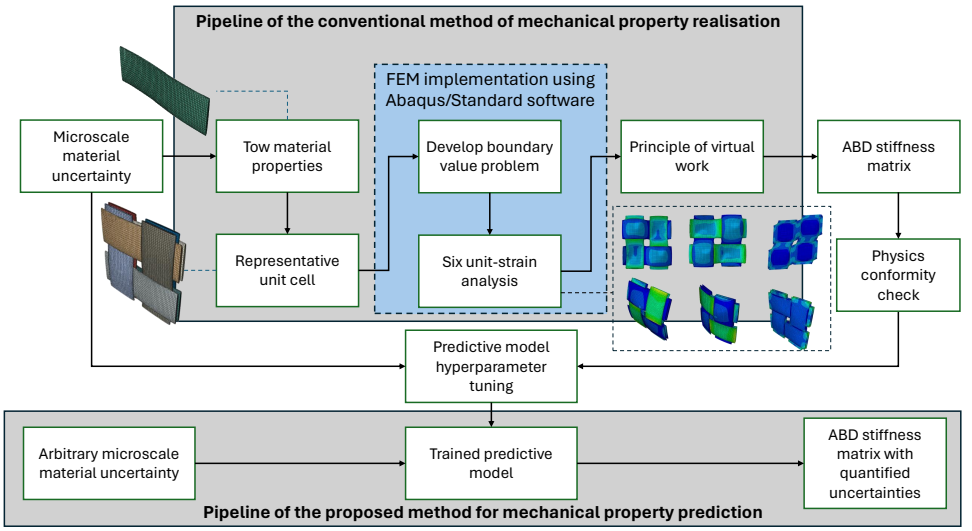

Fig. 1: Comparison of the mechanical property prediction using GPR with the conventional approach

2 Gaussian Process Regression

GPR is a type of Bayesian regression analysis that falls within the spectral probabilistic methodology and is included in computational mechanics' UQ and UP (Sullivan, 2015). The special feature of GPR is the ability to select the input variables as correlated. In contrast, the output variables are assumed to be not a single function but a distribution over functions (Rasmussen and Williams, 2006; Wang, 2023). The covariance matrix represents the correlation structure and propagates that uncertainty in the GPR. Also, a set of basis functions is generated, where the covariance matrix undergoes the spectral decomposition, and that function set is used to model the output distribution (Rasmussen and Williams, 2006). These functions are obtained using mean and variance that indicate how accurate the predictions are, these functions are based on the probability distribution of all suitable functions as expressed in Equation (1).

$$f \sim \mathrm{GP}(m, k) \tag{1}$$

where m is the mean function $m(x) = \mathbb{E}[f(x)]$ and k is the covariance (kernel) function $k(x, x') = \mathbb{E}[(f(x) - m(x))(f(x') - m(x'))]$. Here $\mathbb{E}[*]$ is the expected value of $*$, and x and x' are two arbitrary data points from an input space (X).

Kernel functions are broadly classified into two categories: stationary and non-stationary kernel functions (Rasmussen and Williams, 2006; Genton, 2001). Given the nature of this study (i.e. the material property uncertainty at the microscale for a given data point, as described in Section 3.1 does not exhibit spatial correlation with the uncertainty of another data point that has a different material property uncertainty), stationary kernel functions have been selected. The performance of the Gaussian Process Regression (GPR) model heavily relies on the choice of the appropriate kernel function. In this research, five kernel functions were employed to assess their effectiveness in predicting the mechanical properties of woven carbon fibre fabrics. These kernels include the Radial Basis Function (RBF), the Matérn function (MAT) with $\nu = 3/2$, the Rational Quadratic function (RQ), Exponential Sine Squared function (ESS) and the Dot Product function (DP) (Chen and Wang, 2018).

The selection of these kernel types aims to evaluate which one best captures the underlying variation in the mechanical responses discussed in Section 3.2, considering the uncertainties introduced. Given the high-dimensional nature of the input space, it is difficult to discern the initial patterns or variations within the mechanical properties. Therefore, these five kernels were chosen for their ability to capture distinct patterns in the data. The RBF kernel is well-suited for modelling highly nonlinear relationships due to its inherent smoothness; the MAT kernel offers flexibility between smooth and rough predictions, making it useful for capturing various degrees of variation in the mechanical responses; the RQ kernel, with its capacity to combine different length scales, accounts for varying sensitivity levels in the input data; the ESS kernel is designed to model repeating or periodic mechanical responses; lastly, the DP kernel captures linear relationships, making it valuable when the mechanical properties are dependent on simple input features (Rasmussen and Williams, 2006). By employing this diverse set of kernels, each with its own unique strengths, this study effectively conducts a comprehensive evaluation to determine which kernel performs best in predicting the mechanical behaviour of woven carbon fibre fabrics.

$$k_{RBF}(\boldsymbol{x}, \boldsymbol{x'}) \;=\; \sigma_f^2 \exp\left(-\frac{(\boldsymbol{x}-\boldsymbol{x'})^2}{2l^2}\right) \tag{2}$$

$$k_{Mat}(\boldsymbol{r}) \;=\; \sigma_f^2 \,\frac{2^{1-\nu}}{\Gamma(\nu)}\left(\sqrt{2\nu}\,\frac{\boldsymbol{r}}{l}\right)^{\nu} C_\nu\left(\sqrt{2\nu}\,\frac{\boldsymbol{r}}{l}\right) \tag{3}$$

$$k_{RQ}(\boldsymbol{x}, \boldsymbol{x'}) \;=\; \sigma_f^2\left(1 \;+\; \frac{(\boldsymbol{x}-\boldsymbol{x'})^2}{2\alpha l^2}\right)^{-\alpha} \tag{4}$$

$$k_{ESS}(\boldsymbol{r}) = {\sigma_f}^2 \exp\left(-2\sin^2\left({}^{\pi\boldsymbol{r}}/_{p}\right)/l^2\right) \tag{5}$$

$$k_{DP}(\boldsymbol{x}, \boldsymbol{x'}) \;=\; {\sigma_o}^2 + \boldsymbol{x}.\boldsymbol{x'} \tag{6}$$

In the above equations, the length scale hyperparameter (l), and scaling hyperparameter (σ_f^2) are the two hyperparameters used in Equation (2). The mean square derivatives of every order of Gaussian processes that use the RBF function as the covariance function are extremely smooth since it is infinitely differentiable. Here, the Matérn kernel function (equation (3)) has three parameters which are l (length scale hyperparameter), v, and σ_f^2 (scaling hyperparameter). r is the Euclidean distance between two points, Γ is the Gamma function and C_ν is the modified Bessel function of the second kind Pandit and Infield (2019). When v goes to infinity the Matérn function converges to the Squared exponential covariance function (Pandit and Infield, 2019). The RQ kernel function has two hyperparameters; l (length scale hyperparameter) and σ_f^2 (scaling hyperparameter). $\boldsymbol{x}$ and $\boldsymbol{x'}$ are two input points. The function is defined in Equation (4). Here, the scale-mixture parameter (α) governs the weightage for the coarser and finer variations. The α value is positive (Pandit and Infield, 2019). The exponential sine squared kernel function has three hyperparameters which are the length scale hyperparameter (l), periodicity hyperparameter (p), and scaling hyperparameter (σ_f^2). Equation (5) defines the kernel function. As before, positive α is the scale-mixture parameter that determines the weightage for the coarser and finer variations. ${\sigma_o}^2$ is the only hyperparameter in the DP kernel function defined in Equation (6).

The joint distribution of $\boldsymbol{y}$ (outputs of training data points) and $\boldsymbol{y_*}$ (expected output of testing data points) is defined for n number of points in training data, n_* number of points in testing data, and d number of features in the inputs where $\boldsymbol{X}$ is the $d \times n$ matrix with all training input data points and $\boldsymbol{X_*}$ is the $d \times n_*$ matrix with all testing input data points. This joint distribution is expressed as,

$$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{y_*} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{m}(\boldsymbol{X}) \\ \boldsymbol{m}(\boldsymbol{X_*}) \end{bmatrix}, \begin{bmatrix} \boldsymbol{K} & \boldsymbol{K_*} \\ \boldsymbol{K_*}^{\mathrm{T}} & \boldsymbol{K_{**}} \end{bmatrix} \right), \tag{7}$$

where [$\boldsymbol{K}$ or $\boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X})$] is the $n \times n$ covariance (or Gram) matrix among training input data, [$\boldsymbol{K_*}$ or $\boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X_*})$] is the $n \times n_*$ covariance matrix between training and testing input data and [$\boldsymbol{K_{**}}$ or $\boldsymbol{K}(\boldsymbol{X_*}, \boldsymbol{X_*})$] is the $n_* \times n_*$ covariance matrix among testing input data. The mean function is assumed zero for the prior given that the data is often normalised to a zero mean which is $(m(X), m(X_*)) = 0$, therefore the joint distribution of training and test outputs can be written as,

$$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{y_*} \end{bmatrix} \sim \mathcal{N}\left(\boldsymbol{0}, \begin{bmatrix} \boldsymbol{K} & \boldsymbol{K_*} \\ \boldsymbol{K_*}^{\mathrm{T}} & \boldsymbol{K_{**}} \end{bmatrix} \right). \tag{8}$$

Hyperparameters are initially assumed in the prior of a GPR model. An effective fit to the data (i.e., the posterior) requires precise adjustment of these hyperparameters. In hyperparameter estimation, the log marginal likelihood

of the observed data is usually maximised. The marginal likelihood can be expressed as,

$$P\left(\boldsymbol{y}|\ \boldsymbol{X}\right)\ =\ \int P\left(\boldsymbol{y}|\ \boldsymbol{X},\boldsymbol{\Theta}\right)P(\boldsymbol{\Theta})\mathrm{d}\boldsymbol{\Theta}, \tag{9}$$

where $\boldsymbol{\Theta}$ is the vectorised hyperparameters. The log marginal likelihood is expressed in Equation (10).

$$\log P\left(\boldsymbol{y}|\ \boldsymbol{X}\right)\ =\ -\frac{1}{2}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{K}^{-1}\boldsymbol{y}\ -\ \frac{1}{2}\log\ |\boldsymbol{K}|\ -\frac{n}{2}\log 2\pi \tag{10}$$

The approach of hyperparameter tuning involves maximising the log marginal likelihood using a gradient-based method, whereby the derivative is determined concerning the hyperparameters as follows,

$$\frac{\partial}{\partial\Theta_{j}}\log P\left(\boldsymbol{y}|\ \boldsymbol{X}\right)\ =\ \frac{1}{2}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{K}^{-1}\frac{\partial\boldsymbol{K}}{\partial\Theta_{\ j}}\ \boldsymbol{K}^{-1}\boldsymbol{y}\ -\ \frac{1}{2}\mathrm{Tr}\left(\boldsymbol{K}^{-1}\frac{\partial\boldsymbol{K}}{\partial\Theta_{\ j}}\right) \tag{11}$$

The ideal hyperparameters can be determined by modifying Equations in (11) to zero. The trained GPR model can be derived by substituting these hyperparameters in the Gram matrix in (8). The following are then analytically derived from the Multivariate Normal (MVN) functions of (8)and their marginal and conditional distributions (Rasmussen and Williams, 2006; Wang, 2023).

$$\bar{\boldsymbol{y}}_{*}|\boldsymbol{X}\ ,\ \boldsymbol{y},\ \boldsymbol{X}_{*}\ \sim\mathcal{N}\left(\boldsymbol{K}_{*}^{\mathrm{T}}\boldsymbol{K}^{-1}\ \boldsymbol{y}\ ,\ \boldsymbol{K}_{**}-\ \boldsymbol{K}_{*}^{\mathrm{T}}\boldsymbol{K}^{-1}\boldsymbol{K}_{*}\right). \tag{12}$$

When the conditional distribution is derived, the predictive equations for GPR are given by, $\bar{\boldsymbol{y}}_{*}\mid\boldsymbol{X},\boldsymbol{y},\boldsymbol{X}_{*}\sim\mathcal{N}(\bar{\boldsymbol{y}}_{*},cov(\boldsymbol{y}_{*}))$. Hence, the best estimate for $\boldsymbol{y}_{*}$ is the mean of the distribution (denoted as $\bar{\boldsymbol{y}}_{*}$) where,

$$\bar{\boldsymbol{y}}_{*}=\boldsymbol{K}_{*}^{\mathrm{T}}[\boldsymbol{K}]^{-1}\boldsymbol{y} \tag{13}$$

and the uncertainty of the estimate is given by the covariance which is

$$cov(\boldsymbol{y}_{*})=\boldsymbol{K}_{**}-\boldsymbol{K}_{*}^{\mathrm{T}}(\boldsymbol{K}^{-1}\boldsymbol{K}_{*}). \tag{14}$$

3 Initiation and propagation of uncertainty

3.1 Initiation of material property uncertainty

As mentioned in Section 1, different sources of uncertainty can affect the overall mechanical response of the composite. In this study, we consider the uncertainties that are introduced through the material property variations. The woven composite considered in this study is composed of T300-1k fibres and Hexply 913 epoxy as the constituent materials and their constitutive material

properties are provided in Table 1. The variations in these material properties are due to inconsistencies in the manufacturing process (Fu and Yao, 2022), though the material property probability distributions for this specific material are not found in the literature. Therefore, this study considers the material properties of both the fibre and epoxy to be normally distributed around the mean values provided in Table 1 with a 5% material uncertainty. This assumption has been previously used in other research as well (Omairey et al., 2018; Herath, 2022).

Table 1: Reference material properties of microscale constituents

Material properties	T300 fibres	HexPly 913 resin
Longitudinal stiffness, E_1 (N/mm^2)	233,000	3390
Transverse stiffness, $E_2 = E_3$ (N/mm^2)	23,100	3390
Shear stiffness, $G_{12} = G_{13}$ (N/mm^2)	8963	1210
Poisson's ratio, $\nu_{12} = \nu_{13}$	0.2	0.41

3.2 Propagation of uncertainty

The focus of this research is on a two-ply plain woven carbon fibre composite. This composite is made from carbon fibres and epoxy, which form the tows that are woven together to create the final material. This study utilises the multiscale modelling approach introduced by Mallikarachchi (2019), as illustrated in Figure 1, to analyse the true mechanical responses of the composite. This method acknowledges the hierarchical structure of the composite by defining its material properties and geometries at smaller scales. It then uses this detailed information to derive the mechanical behaviour at the macroscale, represented by the ABD stiffness matrix. Since the tows consist of fibre and epoxy with a fibre volume fraction of 0.62, the material properties of the tows are determined using the rules of mixtures (Tam et al., 2012) and the Halpin-Tsai equations (Budarapu et al., 2019). These methods enable the propagation of material uncertainties from the fibre and resin to the tow properties. Given that the material is geometrically homogeneous, a representative unit cell (RUC) with a 2×2 tow arrangement and periodic boundary conditions is employed. The cross-section of the tow is modelled as a power-ellipse, and since the tows are interwoven, their longitudinal profile is idealised as a cubic Bézier spline curve.

The rendered representative unit cell (RUC) is then modelled and analysed numerically using the Finite Element Method (FEM), implemented with Abaqus/Standard software. The success of this analysis depends significantly on defining the correct boundary value problem, starting with the identification of the composite's true neutral plane. Once this neutral plane is established, control points (serving as reference points) and dummy nodes are positioned accordingly (see Figure 2). Surface-to-surface tie constraints are used to join the tows, preventing them from intersecting or penetrating each other.

Multiple point constraints connect the control points to the tow edge cross-sections, ensuring they move cohesively. Equation constraints are employed to enforce the periodic boundary conditions, contributing to the model's homogenisation.

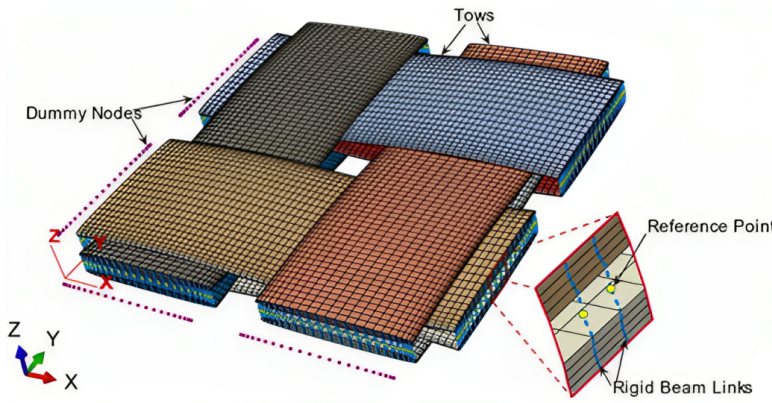

Fig. 2: Representative unit cell of the composite

With the boundary value problem defined, the representative unit cell (RUC) is subjected to six unit strains: extensional strains in the x and y directions, shear strain, rotational strains in the x and y directions, and twisting strain. Following this analysis, boundary forces and displacements are collected for post-processing. Using the principle of virtual work, the ABD stiffness matrix is derived, as represented by (15). The ABD stiffness matrix is able to express the constitutive relationship between stress and strain of a laminate, where $\boldsymbol{N}$ - force per unit length, $\boldsymbol{M}$ - moment per unit length, $\boldsymbol{\varepsilon}$ - mid-plane strains, $\boldsymbol{\kappa}$ - mid-plane curvatures, and $[\boldsymbol{A}]$ - extensional stiffness matrix, $[\boldsymbol{B}]$ - extension-bending coupling matrix, and $[\boldsymbol{D}]$ - bending stiffness matrix.

$$\left\{ \begin{array}{c} \boldsymbol{N} \\ -- \\ \boldsymbol{M} \end{array} \right\} = \left(\begin{array}{c|c} \boldsymbol{A} & \boldsymbol{B}^{\mathrm{T}} \\ -- & -- \\ \boldsymbol{B} & \boldsymbol{D} \end{array} \right) \left\{ \begin{array}{c} \boldsymbol{\varepsilon} \\ -- \\ \boldsymbol{\kappa} \end{array} \right\} \tag{15}$$

With the nature of the material taken into consideration (i.e. effects of extension-bending and extension-shear coupling for homogeneous tows woven in two orthogonal directions are negligible, hence $[\boldsymbol{B} = \boldsymbol{0}]$), an idealised simplified form of the expanded ABD stiffness matrix can be expressed as in Equation (15).

$$\left\{ \begin{array}{c} N_x \\ N_y \\ N_{xy} \\ M_x \\ M_y \\ M_{xy} \end{array} \right\} = \left[\begin{array}{cccccc} \mathrm{A}_{11} & \mathrm{A}_{12} & 0 & 0 & 0 & 0 \\ \mathrm{A}_{21} & \mathrm{A}_{22} & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathrm{A}_{66} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathrm{D}_{11} & \mathrm{D}_{12} & 0 \\ 0 & 0 & 0 & \mathrm{D}_{21} & \mathrm{D}_{22} & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathrm{D}_{66} \end{array} \right] \left\{ \begin{array}{c} \varepsilon_x \\ \varepsilon_y \\ \varepsilon_{xy} \\ \kappa_x \\ \kappa_y \\ \kappa_{xy} \end{array} \right\} \tag{16}$$

This approach ensures that any uncertainties originating from the microscale material properties are carried forward to the macroscale mechanical properties of the composite. As these tows inherit uncertainties from the lower scale, the analysis with Abaqus and subsequent post-processing will reflect these variations in the resulting ABD stiffness matrix. This way, the propagated uncertainties are consistently accounted for throughout the modelling process. The reference ABD stiffness matrix (without any uncertain constituent parameters) to be compared with the uncertainty propagated ABD stiffness entries is provided in equation (17). It is highlighted that this ABD stiffness matrix is already validated with experimental results (Mallikarachchi, 2019; Jayasekara et al., 2021; Herath et al., 2020; Mallikarachchi and Pellegrino, 2013), hence directly used in this study.

$$\mathrm{ABD}_{\mathrm{ref.}} = \begin{bmatrix} 8535.88 & 2117.74 & 0 & 0 & 0 & 0 \\ 2117.74 & 8535.88 & 0 & 0 & 0 & 0 \\ 0 & 0 & 313.56 & 0 & 0 & 0 \\ 0 & 0 & 0 & 36.38 & 4.41 & 0 \\ 0 & 0 & 0 & 4.41 & 36.38 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1.61 \end{bmatrix} \tag{17}$$

3.3 Data Generation

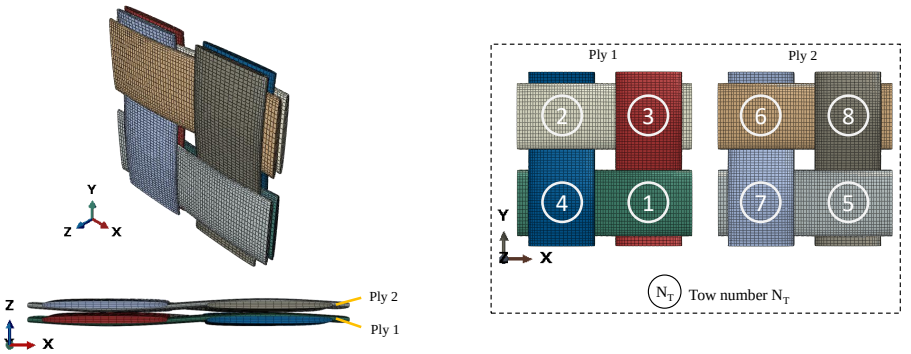

Fig. 3: Tow numbering system of the top and bottom plies of the RUC

To create a reliable data-driven model, a comprehensive database is essential. The data for this purpose was generated using the method described in Sections 3.1 and 3.2. To incorporate uncertainties and ensure a balanced representation of values across the range of [-5%, +5%], the Latin hypercube sampling technique was employed. This technique provides the advantage of randomised stratified sampling (Wang, 2022), which greatly enhances the efficiency of the data set by ensuring that data points are evenly distributed

across the input domain. With the uncertainty introduced via coefficients (i.e.: $c_{f,i}$ and $c_{r,i}$ where subscripts f and r refer to fibre and resin, respectively, and $i = 1, 2, ...8$ refers to the tow under consideration which is shown in Figure 3). The corresponding ground truth ABD stiffness matrix was then generated. A sample data point demonstrating how microscale uncertainty is introduced and its subsequent propagation to macroscale mechanical behaviour is presented in Table **??** for clearer understanding. In this study, several material property relationships must be satisfied for a material to be considered orthotropic. As outlined by Jones (Jones, 1998), the engineering constants must meet the conditions specified in equations (18) and (19). Each generated data point is examined to ensure compliance with these inequalities. This validation process is conducted to confirm that the results derived from the numerical model are consistent with the physical behaviour of real-world materials, addressing potential concerns about the model's physical fidelity.

$$|\nu_{21}| < \sqrt{\frac{E_2}{E_1}}, \quad |v_{32}| < \sqrt{\frac{E_3}{E_2}}, \quad |v_{13}| < \sqrt{\frac{E_1}{E_3}}, \tag{18a}$$

$$|v_{12}| < \sqrt{\frac{E_1}{E_2}}, \quad |v_{23}| < \sqrt{\frac{E_2}{E_3}}, \quad |v_{31}| < \sqrt{\frac{E_3}{E_1}} \tag{18b}$$

$$v_{21} > -\left[v_{32}v_{13}\frac{E_2}{E_1} + \sqrt{1 - v_{32}^2\frac{E_2}{E_3}}\sqrt{1 - v_{13}^2\frac{E_3}{E_1}}\sqrt{\frac{E_2}{E_1}}\right] \tag{19a}$$

$$v_{21} < -\left[v_{32}v_{13}\frac{E_2}{E_1} - \sqrt{1 - v_{32}^2\frac{E_2}{E_3}}\sqrt{1 - v_{13}^2\frac{E_3}{E_1}}\sqrt{\frac{E_2}{E_1}}\right]. \tag{19b}$$

4 Predictive model development and assessment

This study explores the application of GPR-based models to predict and quantify the uncertainties, recognising the inherent material uncertainties at the microscale. The theoretical foundations of GPR are outlined in Section 2. Our goal is to use the mean function for predicting mechanical properties and the covariance function to quantify the associated uncertainties, as described in Equation 1. The five kernel types employed in this study—detailed in Equations 2, 3, 4, 5 and 6 —were used to develop GPR models for predicting the stiffness components of the ABD stiffness matrix (A_{11}, A_{12}, A_{22}, A_{66}, D_{11}, D_{12}, D_{22}, and D_{66}). This resulted in a total of 40 model combinations, considering all stiffness components and kernel types. The models were developed using a Python environment, leveraging libraries from (Pedregosa et al., 2011). A comprehensive discussion of the kernel hyperparameters can be found in Section 2, which covers the initial assumptions (Equation 9), and the subsequent hyperparameter tuning process, as described in Equations 7 to 11.

Before training a predictive model, data preprocessing is a crucial step in statistical learning that significantly affects model performance (Wang et al., 2021). The main goals of data preprocessing are to minimise overfitting by reducing data bias and to avoid numerical instabilities during algebraic operations such as matrix decomposition and inversion. The standard scaler method was used in this study to preprocess the data, which ensures that all features are on a consistent scale. This step is essential to prevent features with larger magnitudes from disproportionately influencing the learning process (Herath et al., 2022; Rasmussen and Williams, 2006; Bishop, 2006).

In the model training phase, after preprocessing the dataset, hyperparameters for the Gaussian Process Regression (GPR) model were optimised starting from initial values set to unity. The optimization was carried out using the Maximum Likelihood Marginalization of Likelihood (MLML) method described in Section 2. Given the risk of converging to local minima or maxima, we adopted a stochastic optimization approach with multiple random starting points to improve the chances of identifying the global optimum (Rasmussen and Williams, 2006). Specifically, we utilised the L-BFGS-B (Limited-memory Broyden-Fletcher-Goldfarb-Shanno with Box constraints) algorithm, setting the number of restarts to 10. The L-BFGS-B algorithm was selected for its efficiency in handling large-scale datasets, outperforming traditional Bayesian optimization both in computational cost and accuracy (Isabona et al., 2023). It is particularly advantageous for problems where obtaining Hessian information is challenging or for large, dense problems (Zhu et al., 1997). The standard practice typically involves using 10 restarts for optimization. While increasing the number of restarts can potentially improve convergence in some cases, as highlighted by Dick et al. (2014), it also significantly increases computational time. We experimented with higher restart values and observed a convergence in error values with additional restarts. Therefore, to minimize computational time, the number of restarts was selected as 10. To adjust the smoothness of the posterior mean, a virtual noise factor called alpha was used (Pedregosa et al., 2011). By carefully selecting the appropriate value for alpha through a trial-and-error approach, the model avoided overfitting and underfitting issues (Rasmussen and Williams, 2006; Pedregosa et al., 2011). For instance, with the Radial Basis Function (RBF) kernel, an alpha value of 0.1 for A_{11} and 1 for A_{12} produced optimal results, whereas higher values led to overfitting. This approach ensured that the final GPR model achieved the desired balance of accuracy and robustness (Rasmussen and Williams, 2006; Pedregosa et al., 2011).

4.1 Error metrics for model evaluation

Every data-driven model performance is evaluated based on evaluation metrics. For a perfect prediction model, the actual value should be equal to the predicted value and the relationship between the actual and predicted values should ideally be a linear relationship with the gradient equal to 1 and intercept equal to zero. In this study, Normalised Root Mean Squared Error

(NRMSE) and R-squared value are chosen as the error metrics to evaluate how the GPR model had predicted the mechanical properties for arbitrary input uncertainties when compared to the actual mechanical properties obtained using the FEM based method as depicted in Figure 1.

NRMSE, which is the normalised variant of the Root Mean Squared Error (RMSE) (Khan and Noor, 2019), serves as a valuable metric to assess the concentration of data around the model of best fit. For a sample comprising of N_{obs} observations, RMSE effectively quantifies the standard deviation of residuals (here, the residual is the difference between the observed (or actual) values, $Y_{obs,i}$, and their corresponding estimated values derived from the GPR model, $Y_{pred,i}$, (i = 1, 2, ..., N_{obs})), as described in (20). A lower RMSE signifies a more favourable model fit, but direct comparisons across diverse datasets remain ambiguous due to the sensitivity of the metric to the scale of data. Consequently, to facilitate meaningful comparisons among models operating on different data scales, this study utilises NRMSE derived from (20) where normalisation is carried out with respect to the mean ($Y_{obs,mean}$) as given in Equation (21).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N_{obs}} \left(Y_{obs,i} - Y_{pred,i}\right)^2}{N_{obs}}} \tag{20}$$

$$NRMSE = RMSE / Y_{obs,mean} \tag{21}$$

In addition, the R-squared (R^2) value (also known as the coefficient of determination) is calculated for better interpretability of the fit of the model. The R^2 value ([min, max] = $[-\infty, 1]$) is calculated using Equation (22). Here SS_{RES} is the residual sum of squared errors of the regression model and SS_{TOT} is the total sum of squared errors. An R^2 value of 1 indicates a perfect fit, with values closer to unity suggesting a strong model fit. In this context, an R^2 value close to 1 implies that the predicted values $Y_{pred,i}$ from the GPR model for arbitrary inputs closely match the actual values $Y_{obs,i}$ for those inputs.

$$R^2 = 1 - SS_{RES}/SS_{TOT} \tag{22}$$

4.2 k-fold cross validation

Cross-validation is a widely used across-fold self-verification technique for independently distributed observations (Schnaubelt, 2019). Any effect of data bias, model overfitting, or the existence of outlier data is revealed during the process. The k-fold cross-validation is chosen among other methods for this purpose (Maleki et al., 2020). The data points are randomly divided among k number of distinct folds. One of the k folds is selected as the testing data and the remaining folds are used as training data. This process is iterated k times until each fold becomes the testing data fold. The variances observed in the error metrics and hyperparameters for each iteration of the process flag any issues encountered during training (Rasmussen and Williams, 2006; Bishop, 2006). Thus, lower variances across folds indicate an accurate

model. According to the comprehensive study by Maleki et al. (2020), the recommended number of folds lies between 5 and 10. Therefore, in this study, the number of folds (k) selected for cross-validation is 5.

5 Results and Discussion

5.1 Kernel performance in training and testing

Given that Gaussian Process Regression (GPR) requires a substantial number of data points to achieve optimal performance, we conducted an error convergence study to determine the necessary number of data points for effective model training and testing. Using k-fold cross-validation, the dataset was divided into multiple folds. It is important to note that the number of data points indicated on the x-axis of Figures 4a to 4c represents the total dataset size before splitting. Our study revealed a clear trend of error convergence and reduction as the dataset size increased, consistent with the observations in (Barr and Rabitz, 2023). In that study, the mean square errors of the GPR model, utilizing an RBF kernel, similarly decreased as the sample size expanded. While Figures 4a to 4c are presented as reference samples, similar trends were observed across other stiffness models and kernel functions utilised in our analysis. Based on our analysis, we concluded that a sample size of 3,000 data points is optimal for the remainder of the study. While a sample size of 3,500 produced error values comparable to those of 3,000, the training time increased significantly, making the smaller sample size a more efficient choice.

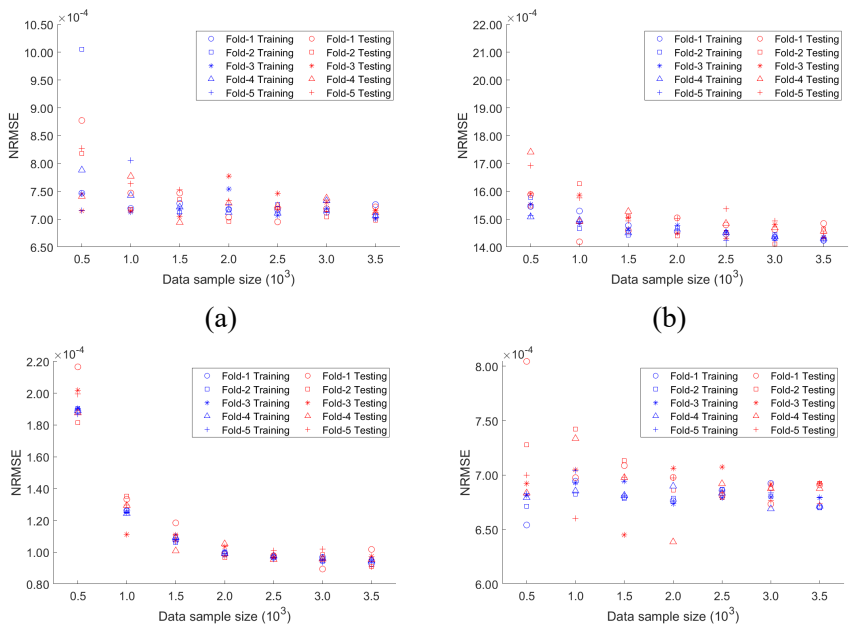

Fig. 4: Error convergence plot for (a) A_{11} with RQ kernel, (b) A_{66} with MAT kernel, (c) D_{11} with RBF kernel, and (d) D_{66} with DP kernel

Table 3: Training times for different kernels

Stiffness Output	Training time (s)				
	RBF	MAT	ESS	RQ	DP
A_{11}	172.07	520.36	246.79	643.85	10.31
A_{12}	265.31	98.90	298.33	373.37	10.34
A_{22}	174.60	164.97	233.76	558.92	10.84
A_{66}	180.78	239.45	249.07	354.19	10.14
D_{11}	195.15	329.80	225.78	520.02	10.44
D_{12}	170.12	385.31	221.43	461.50	10.26
D_{22}	224.41	335.32	189.22	892.13	10.56
D_{66}	163.52	357.91	174.37	1704.04	10.86

With a data sample size set to 3000 points and the use of k-fold cross-validation, each fold in the training process utilised 2400 data points for training and 600 for testing. As explained in Section 4.2, minimizing variation in error metrics among the folds is crucial. Figure 5 visually represents the variations in Normalised Root Mean Squared Error (NRMSE) values through box plots for both training and testing folds across different kernels used in the data-driven models. In Figure 5, the NRMSE values for A_{11} and A_{22} during training and testing exhibit low variability across all kernels. However, the models using the Matérn kernel show relatively higher NRMSE values for A_{22}. The NRMSE values for A_{12} and A_{66} demonstrate chaotic behaviour, evidenced by higher degrees of variation, a greater number of outliers, and significantly higher NRMSE values. Examining the NRMSE variations in the bending submatrix shown in Figure 5, we observe that the direct axial stiffnesses (D_{11} and D_{22}) exhibit lower error variations for models utilizing the RBF, Matérn, and DP kernels. The bending-bending coupling D_{12} shows very low NRMSE values across most models, except for those using the ESS kernel. However, models employing the Matérn and ESS kernels display instability in the torsional stiffness D_{66}, indicated by the presence of outliers with high error variations. Training a Gaussian Process Regression (GPR) model is computationally intensive, involving complex matrix operations on the Gram matrix to compute the mean and covariance functions (see Equations (13) and (14)). Table 3 provides a comparison of the training times for different models. Notably, models using the DP kernel required the least time to train, while those using the RQ kernel took the longest.

5.2 Model prediction

Model training was conducted on a dataset with 5% microscale material uncertainty. The NRMSE and R^2 values for the testing phase are presented in Tables 4 and 5. These metrics were calculated by comparing the model predictions with the ground truth from the testing fold. Figure 6 provides a visual

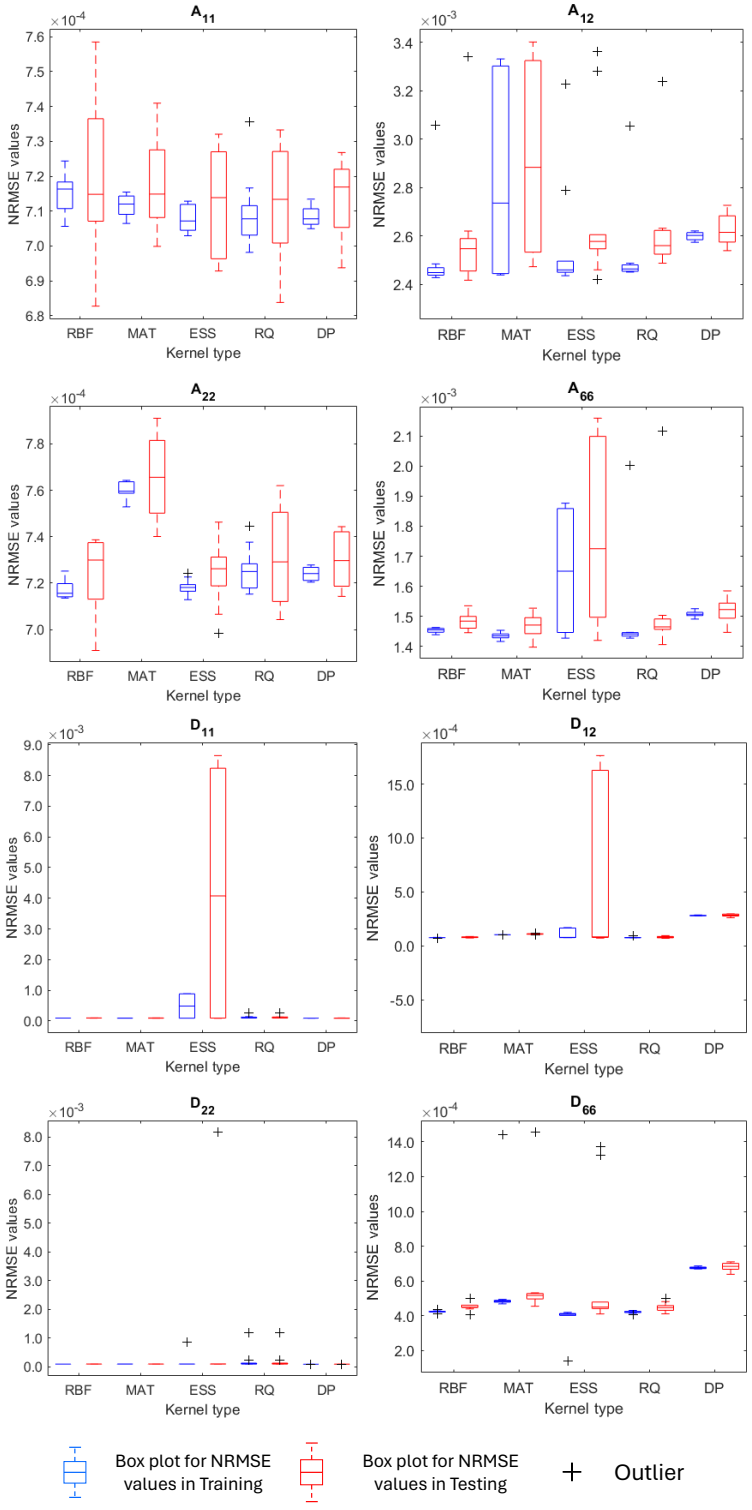

Fig. 5: NRMSE distribution in 5-fold cross-validation in training and testing

representation of the actual versus predicted stiffness distribution, specifically for the model trained with 5% material uncertainty using the RBF kernel.

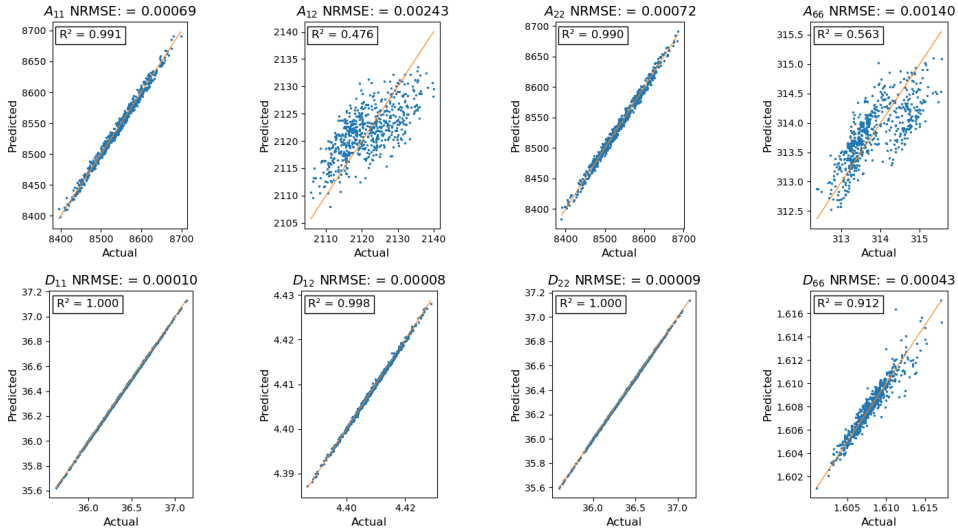

Fig. 6: Actual vs predicted plots for testing for GPR model using RBF kernel for 5% microscale material uncertainty

The figure shows that the predictability of the bending stiffnesses (D_{ij}) is high, as indicated by their low NRMSE values compared to the extensional stiffnesses (A_{ij}). This high predictability is attributed to the strong correlation between the bending stiffness values and the microscale material properties, which is evident from the high R^2 values for the D_{ij} stiffnesses in Table 5. The lowest error values and highest correlations are observed in D_{11}, D_{12}, and D_{22}. Conversely, the highest error values occur in the A_{12} and A_{66} stiffnesses, which is also reflected in Figure 6 and the low R^2 values in Table 5. This suggests that the GPR model struggled to capture the chaotic relationship between microscale material uncertainties and the Poisson's effect (A_{12}) and shear (A_{66}) stiffnesses, resulting in lower correlation values. Compared to the NRMSE values reported in Vipulananthan et al. (2023), the GPR model has outperformed both Artificial Neural Network (ANN) and Multiple Linear Regression (MLR) models. For example, the NRMSE for A_{11} in the ANN model is 1.0×10^{-3}, and 7.2×10^{-4} for MLR. As shown in Table 4, under 5% uncertainty, all the GPR models using different kernels have either matched or surpassed these error values. Although A_{12} exhibits more erratic behaviour, the GPR models still achieved lower NRMSE values than the 2.6×10^{-3} recorded in both the ANN and MLR models.

In addition to evaluating the interpolation capabilities of the models, we also assessed their extrapolation capabilities. Models trained on data with 5% input material uncertainties were used to predict stiffnesses for material property uncertainties beyond 5% (specifically, 10%, 15%, and 20%). Error

Table 4: NRMSE values for stiffness prediction

Material uncertainty	Kernel	Stiffness							
		A_{11}	A_{12}	A_{22}	A_{66}	D_{11}	D_{12}	D_{22}	D_{66}
5	RBF	0.00069	0.00243	0.00072	0.00140	0.00010	0.00008	0.00009	0.00043
	MAT	0.00071	0.00245	0.00075	0.00147	0.00009	0.00010	0.00010	0.00042
	ESS	0.00071	0.00246	0.00072	0.00144	0.00010	0.00008	0.00010	0.00043
	RQ	0.00071	0.00246	0.00072	0.00145	0.00010	0.00008	0.00009	0.00043
	DP	0.00072	0.00259	0.00073	0.00151	0.00009	0.00028	0.00009	0.00067
10	RBF	0.00116	0.00490	0.00093	0.00241	0.00081	0.00135	0.00080	0.00199
	MAT	0.00115	0.00387	0.00110	0.00202	0.00081	0.00119	0.00080	0.00123
	ESS	0.00115	0.00448	0.00094	0.00245	0.00080	0.00132	0.00080	0.00187
	RQ	0.00116	0.00473	0.00089	0.00220	0.00081	0.00098	0.00080	0.00191
	DP	0.00116	0.00490	0.00093	0.00241	0.00081	0.00135	0.00080	0.00199
15	RBF	0.00217	0.00736	0.00228	0.00554	0.00217	0.00377	0.00229	0.00690
	MAT	0.00220	0.00549	0.00255	0.00544	0.00217	0.00389	0.00229	0.00405
	ESS	0.00216	0.00722	0.00227	0.00563	0.00217	0.00376	0.00229	0.00673
	RQ	0.00216	0.00714	0.00225	0.00585	0.00217	0.00281	0.00228	0.00667
	DP	0.00217	0.00736	0.00228	0.00554	0.00217	0.00377	0.00229	0.00690
20	RBF	0.00474	0.01480	0.00380	0.01167	0.00393	0.00705	0.00381	0.00996
	MAT	0.00472	0.00999	0.00386	0.00925	0.00392	0.00730	0.00380	0.00522
	ESS	0.00471	0.01475	0.00378	0.01229	0.00393	0.00689	0.00381	0.00974
	RQ	0.00478	0.01562	0.00364	0.01130	0.00392	0.00528	0.00381	0.00949
	DP	0.00474	0.01480	0.00380	0.01167	0.00393	0.00705	0.00381	0.00996

and correlation metrics were calculated by comparing these predictions with ground truth values obtained using the conventional method pipeline illustrated in Figure 1. For each level of uncertainty (10%, 15%, and 20%), 40 data points were generated to conduct the extrapolation prediction analysis. The calculated NRMSE and R^2 values are presented in Tables 4 and 5, respectively. It is evident that errors increase significantly when making predictions in the unseen space. For instance, the NRMSE value for the model using the RBF kernel increased by 68.5% when material uncertainty rose by 5%, and further increased by 87.8% and 117.8% for the subsequent increments.

Despite this, the correlation between material uncertainty and macroscale direct axial stiffnesses (A_{11} and A_{22}) and direct bending stiffnesses (D_{11} and D_{22}) remained relatively high during extrapolation. However, while the bending-bending coupling (D_{12}) and torsional (D_{66}) stiffnesses initially showed low NRMSE values and high correlations, their predictability significantly deteriorated during extrapolation. An exception was observed with the model using the RQ kernel for D_{12}, which maintained a moderate correlation at 20% material uncertainty. The Poisson effect-related stiffness (A_{12}) and shear stiffness (A_{66}), which already exhibited low correlation and high NRMSE values at 5% microscale material uncertainty, showed even higher NRMSE values during extrapolation. This can be attributed to the negative correlation values observed in Table 5 during the extrapolation phase.

Table 5: R^2 values for stiffness prediction

Material uncertainty	Kernel	Stiffness							
		A_{11}	A_{12}	A_{22}	A_{66}	D_{11}	D_{12}	D_{22}	D_{66}
5	RBF	0.9907	0.4755	0.9897	0.5634	0.9999	0.9981	0.9999	0.9116
	MAT	0.9902	0.4503	0.9884	0.5594	0.9999	0.9967	0.9999	0.9211
	ESS	0.9904	0.4479	0.9895	0.5512	0.9999	0.9979	0.9999	0.9147
	RQ	0.9901	0.4409	0.9888	0.5757	0.9999	0.9981	0.9999	0.9133
	DP	0.9900	0.4107	0.9898	0.5577	0.9999	0.9749	0.9999	0.7909
10	RBF	0.9900	-0.4274	0.9940	0.4982	0.9969	0.8193	0.9971	0.7160
	MAT	0.9902	-0.0902	0.9913	0.5800	0.9969	0.8506	0.9971	0.8462
	ESS	0.9902	-0.2907	0.9938	0.4887	0.9969	0.8259	0.9971	0.7406
	RQ	0.9899	-0.4184	0.9944	0.5327	0.9969	0.9038	0.9971	0.7382
	DP	0.9900	-0.4274	0.9940	0.4982	0.9969	0.8193	0.9971	0.7160
15	RBF	0.9855	-0.0908	0.9880	-0.2471	0.9908	0.5283	0.9918	-0.0167
	MAT	0.9852	0.1741	0.9845	-0.2523	0.9908	0.4655	0.9919	0.1741
	ESS	0.9857	-0.1796	0.9880	-0.3802	0.9907	0.5330	0.9918	0.0415
	RQ	0.9857	-0.1256	0.9883	-0.3441	0.9908	0.7367	0.9919	0.0381
	DP	0.9855	-0.0908	0.9880	-0.2471	0.9908	0.5283	0.9918	-0.0167
20	RBF	0.9710	-1.0869	0.9819	-0.7105	0.9874	0.0953	0.9883	-0.1383
	MAT	0.9712	-1.5214	0.9809	-0.8181	0.9875	-0.5309	0.9883	0.0722
	ESS	0.9713	-1.0071	0.9821	-0.6641	0.9874	0.1325	0.9882	-0.1304
	RQ	0.9703	-1.1384	0.9832	-0.7270	0.9875	0.4896	0.9883	-0.0556
	DP	0.9710	-1.0869	0.9819	-0.7105	0.9874	0.0953	0.9883	-0.1383

5.3 Uncertainty quantification

The use of GPR models for predicting stiffnesses offers the advantage of providing an error margin with each prediction. As detailed in Section 2, GPR outputs a distribution of possible functions, allowing us to determine the 95% confidence interval for each prediction. For example, an A_{11} prediction of 8510.15 N/mm might with the consideration of 95% confidence interval results in an upper bound of 8513.31 N/mm and a lower bound of 8506.99 N/mm, indicating a variation of ±0.04%. These confidence intervals are available for each prediction and Table 6 lists the highest uncertainties observed in the predictions. As expected, uncertainties increase with extrapolation. The lowest uncertainties for predictions with 5% material uncertainty are found in D_{12} and D_{66}, while the highest uncertainties are observed in A_{12} and A_{66}. For predictions made with 20% material uncertainty, the uncertainty values for A_{12} and A_{66} increase significantly, reaching up to 2.01% for the model using the RBF kernel.

Table 6: Uncertainty values for stiffness prediction

Material uncertainty	Kernel	Extreme macroscale uncertainties (±%)							
		A_{11}	A_{12}	A_{22}	A_{66}	D_{11}	D_{12}	D_{22}	D_{66}
5	RBF	0.05	0.14	0.05	0.08	0.06	0.03	0.06	0.03
	MAT	0.05	0.15	0.15	0.08	0.06	0.03	0.06	0.05
	ESS	0.05	0.14	0.05	0.08	0.06	0.03	0.06	0.03
	RQ	0.05	0.14	0.04	0.08	0.03	0.03	0.05	0.03
	DP	0.05	0.14	0.05	0.08	0.06	0.03	0.06	0.03
10	RBF	0.10	0.52	0.11	0.26	0.12	0.11	0.12	0.13
	MAT	0.11	0.55	0.31	0.25	0.12	0.13	0.13	0.25
	ESS	0.10	0.50	0.10	0.27	0.12	0.11	0.12	0.14
	RQ	0.12	0.52	0.16	0.25	0.11	0.13	0.13	0.13
	DP	0.10	0.52	0.11	0.26	0.12	0.11	0.12	0.13
15	RBF	0.18	1.14	0.18	0.59	0.19	0.26	0.20	0.33
	MAT	0.18	1.16	0.51	0.51	0.20	0.29	0.20	0.57
	ESS	0.18	1.13	0.18	0.61	0.19	0.25	0.20	0.33
	RQ	0.24	1.18	0.36	0.56	0.18	0.31	0.23	0.33
	DP	0.18	1.14	0.18	0.59	0.19	0.26	0.20	0.33
20	RBF	0.28	2.01	0.29	1.03	0.30	0.45	0.30	0.69
	MAT	0.28	1.91	0.76	0.82	0.31	0.49	0.30	1.00
	ESS	0.28	1.95	0.29	1.06	0.30	0.45	0.30	0.67
	RQ	0.41	2.03	0.62	0.97	0.30	0.55	0.37	0.69
	DP	0.28	2.01	0.29	1.03	0.30	0.45	0.30	0.69

In addition to expressing prediction uncertainties through confidence intervals, it is crucial to understand the resulting macroscale uncertainty caused by the propagation of microscale uncertainties. The uncertainties introduced to the reference material properties listed in Table 1 will propagate, affecting the reference macroscale stiffnesses provided in Equation (17). For different combinations of material uncertainties, the model will predict stiffness values using the GPR mean function described in Equation (13). By examining these predictions, we can identify upper and lower bounds. Comparing these bounds to the reference macroscale stiffnesses in Equation (17) reveals the extent of macroscale uncertainty resulting from microscale uncertainty. Table 7 illustrates these upper and lower bounds as percentage variations from the reference stiffnesses, providing a clear representation of the induced macroscale uncertainties. It is important to note that the direct axial stiffnesses, both extensional and bending, exhibit the highest uncertainties propagated from the microscale material uncertainty. Additionally, introducing a 5% material uncertainty at the microscale does not translate to a 5% uncertainty at the macroscale; the actual macroscale uncertainty is lower. In the study of Herath et al. (2020), with minimum uncertainty values (-5% fibre and resin material property uncertainty), the extensional stiffness decreased by 4.6%. Conversely, with maximum uncertainty values (+5% fibre and resin material property uncertainty), an increase of 4.9%, which is an overestimation of the mechanical property variation attributed to microscale uncertainty.

Table 7: Minimum and maximum deviation of the prediction from the reference stiffness

Material uncertainty	Stiffness Output	Kernel type									
		RBF		MAT		ESS		RQ		DP	
		Max	Min	Max	Min	Max	Min	Max	Min	Max	Min
5%	A_{11}	1.90	-1.77	1.90	-1.77	1.91	-1.77	1.90	-1.77	1.90	-1.77
	A_{12}	0.85	-0.51	0.81	-0.47	0.86	-0.48	0.90	-0.50	0.84	-0.45
	A_{22}	1.69	-1.91	1.66	-1.87	1.69	-1.91	1.68	-1.91	1.69	-1.91
	A_{66}	0.56	-0.43	0.54	-0.45	0.54	-0.43	0.52	-0.45	0.54	-0.47
	D_{11}	2.17	-2.23	2.17	-2.23	0.50	-0.65	2.17	-2.23	2.18	-2.24
	D_{12}	0.50	-0.65	0.50	-0.65	1.90	-2.00	0.50	-0.65	0.49	-0.65
	D_{22}	2.20	-2.24	2.20	-2.24	2.20	-2.24	2.20	-2.24	2.21	-2.25
	D_{66}	0.59	-0.40	0.54	-0.40	0.58	-0.40	0.60	-0.41	0.46	-0.41
10%	A_{11}	2.32	-2.39	2.31	-2.38	2.32	-2.38	2.32	-2.35	2.32	-2.38
	A_{12}	1.25	-0.88	1.16	-0.86	1.16	-0.98	1.34	-1.04	1.27	-0.78
	A_{22}	2.43	-2.59	2.32	-2.56	2.43	-2.59	2.42	-2.59	2.44	-2.59
	A_{66}	0.53	-0.93	0.47	-0.95	0.59	-0.91	0.51	-0.90	0.47	-0.92
	D_{11}	2.83	-2.97	2.83	-2.97	2.83	-2.96	2.83	-2.97	2.83	-2.97
	D_{12}	0.55	-0.70	0.55	-0.71	0.56	-0.70	0.52	-0.74	0.61	-0.68
	D_{22}	3.16	-3.08	3.16	-3.08	3.16	-3.08	3.16	-3.08	3.17	-3.08
	D_{66}	1.12	-0.69	0.78	-0.63	1.05	-0.70	1.10	-0.69	0.61	-0.64
15%	A_{11}	4.27	-3.05	4.25	-3.04	4.28	-3.04	4.20	-3.04	4.26	-3.04
	A_{12}	1.76	-1.13	1.27	-1.27	1.23	-1.19	1.38	-1.17	1.50	-1.28
	A_{22}	3.35	-5.02	3.28	-4.95	3.35	-5.03	3.35	-5.03	3.35	-5.03
	A_{66}	1.13	-0.73	0.89	-0.77	1.11	-0.83	1.10	-0.68	0.97	-0.84
	D_{11}	5.24	-3.83	5.24	-3.83	5.24	-3.83	5.24	-3.83	5.23	-3.83
	D_{12}	0.81	-1.34	0.79	-1.33	0.79	-1.35	0.68	-1.43	0.97	-1.20
	D_{22}	4.19	-6.03	4.19	-6.03	4.19	-6.03	4.20	-6.03	4.20	-6.04
	D_{66}	2.54	-1.00	1.13	-0.64	2.51	-0.94	2.42	-0.95	0.87	-0.65
20%	A_{11}	3.77	-5.48	3.77	-5.47	3.77	-5.50	3.77	-5.45	3.78	-5.49
	A_{12}	3.41	-1.76	2.20	-0.83	2.93	-1.90	2.77	-1.21	2.11	-1.14
	A_{22}	5.32	-4.42	5.22	-4.36	5.33	-4.43	5.30	-4.46	5.32	-4.43
	A_{66}	1.88	-1.42	1.41	-1.56	1.95	-1.56	1.90	-1.38	0.54	-0.42
	D_{11}	4.46	-7.15	4.46	-7.15	4.46	-7.14	4.46	-7.15	4.47	-7.15
	D_{12}	1.80	-1.54	1.32	-1.39	1.81	-1.53	1.62	-1.72	1.28	-1.18
	D_{22}	7.01	-5.15	7.02	-5.15	7.02	-5.15	7.02	-5.15	7.02	-5.16
	D_{66}	2.83	-1.47	1.05	-1.08	2.76	-1.47	2.85	-1.70	1.04	-1.05

The microscale material uncertainty results in minimal variation from the reference stiffnesses for A_{12}, A_{66}, D_{12}, and D_{66}, with variations remaining below 5% even for a 20% material uncertainty.

6 Conclusion

This study employs Gaussian Process Regression (GPR) to predict and quantify the uncertainty in the mechanical response of two-ply plain woven carbon fibre composites at the macroscale, considering uncertainties in the microscale constituent material properties. The model incorporates 16 input parameters to account for the uncertainties in the material properties of the fibre and resin of each tow. A robust database was generated using a physics-based conventional pipeline that integrates finite element analysis with various combinations of material property uncertainties. Our study revealed several key findings. We observed clear error convergence for a sample size of 3000 data points. Using 5-fold cross-validation, we showed that computational efficiency varied, with the DP kernel being the fastest to train and the RQ kernel the slowest. The study showed that predictions for A_{11}, A_{22}, and D_{22} stiffness entries demonstrated excellent agreement across all kernels. However, ESS kernel values show considerable deviations for D_{11} and D_{12}. Stiffness entries for A_{11}, A_{22}, D_{11}, and D_{22} emphasise near-unity R^2 scores in interpolative predictions. However, the predictability of D_{12} and D_{66} stiffnesses deteriorated during extrapolation, except for moderate performance in D_{12} with the RQ kernel. Due to the inherent chaotic behaviour of A_{12} and A_{66}, all kernels only achieved a medium correlation between microscale and macroscale mechanical properties, with predictability drastically decreasing in extrapolation.

The ability of the Gaussian processes to provide outputs over a distribution of functions allowed the calculation of the 95% confidence interval for each prediction. Most stiffness predictions exhibited variations as low as ±0.2%. Except for A_{12} and A_{66}, all other stiffnesses exhibited less than ±1% variation even in extrapolation. When considering the overall uncertainty caused by microscale uncertainties, the study showed that direct axial stiffnesses, both extensional and bending (i.e., A_{11}, A_{22}, D_{11} and D_{22}), exhibited the highest propagated uncertainties. Despite introducing a 5% uncertainty at the microscale, the actual macroscale uncertainty was lower, with minimal variations from reference stiffnesses. Especially in A_{12}, A_{66}, D_{12}, and D_{66}, the macroscale uncertainties were less than 5% even for 20% material uncertainty. Overall, these results demonstrate the capability of GPR to effectively predict and quantify uncertainties in the mechanical properties of carbon fibre composites, providing valuable insights for the design and analysis of composite materials under varying material property uncertainties.

References

Alazwari, M. A. and Rao, S. S. (2019), 'Modeling and analysis of composite laminates in the presence of uncertainties', *Composites Part B: Engineering* **161**, 107–120.

Balokas, G., Kriegesmann, B., Czichon, S. and Rolfes, R. (2021), 'A variable-fidelity hybrid surrogate approach for quantifying uncertainties in the non-

linear response of braided composites', *Computer Methods in Applied Mechanics and Engineering* **381**, 113851.

Balokas, G., Kriegesmann, B. and Rolfes, R. (2021), 'Data-driven inverse uncertainty quantification in the transverse tensile response of carbon fiber reinforced composites', *Composites Science and Technology* **211**, 108845.

Barr, J. and Rabitz, H. (2023), 'Kernel-based global sensitivity analysis obtained from a single data set', *Reliability Engineering & System Safety* **235**, 109173.

Bishop, C. M. (2006), *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag, Berlin, Heidelberg.

Bostanabad, R., Liang, B., Gao, J., Liu, W. K., Cao, J., Zeng, D., Su, X., Xu, H. and Chen, W. (2018), 'Uncertainty quantification in multiscale simulation of woven fiber composites', *Computer Methods in Applied Mechanics and Eng.* **338**, 506–532.

Budarapu, P. R., Zhuang, X., Rabczuk, T. and Bordas, S. P. (2019), Chapter one - multiscale modeling of material failure: Theory and computational methods, *in* M. I. Hussein, ed., 'Advances in Crystals and Elastic Metamaterials, Part 2', Vol. 52 of *Advances in Applied Mechanics*, Elsevier, pp. 1–103.

Campbell, F. (2004), Chapter 2 - fibers and reinforcements: The string that provides the strength, *in* F. Campbell, ed., 'Manufacturing Processes for Advanced Composites', Elsevier Science, Amsterdam, pp. 39–62.

Chen, Z. and Wang, B. (2018), 'How priors of initial hyperparameters affect gaussian process regression models', *Neurocomputing* **275**, 1702–1710.

Clément, A., Soize, C. and Yvonnet, J. (2013), 'Uncertainty quantification in stochastic multiscale analysis of nonlinear elastic materials', *Computer Methods in Applied Mechanics and Eng.* **254**, 61–82.

Deleo, A. A., O'Neil, J., Yasuda, H., Salviato, M. and Yang, J. (2020), 'Origami-based deployable structures made of carbon fiber reinforced polymer composites', *Composites Science and Technology* **191**, 108060.

Dick, T., Wong, E. and Dann, C. (2014), How many random restarts are enough.
URL: *https://api.semanticscholar.org/CorpusID:9473630*

Fu, Y. and Yao, X. (2022), 'A review on manufacturing defects and their detection of fiber reinforced resin matrix composites', *Composites Part C: Open Access* **8**.

Gangineni, P. K., Gupta K., B. N. V. S. G., Patnaik, S., Prusty, R. K. and Ray, B. C. (2022), 'Recent advancements in interface engineering of carbon fiber reinforced polymer composites and their durability studies at different service temperatures', *Polymer Composites* **43**(7), 4126–4164.

Genton, M. G. (2001), 'Classes of kernels for machine learning: A statistics perspective', *Journal of Machine Learning Research* **2**, 299–312.

Herath, S. (2022), 'Material orientation optimisation of finite deformable orthotropic thin-shells', *Mechanics Research Communications* **119**, 103811.

Herath, S., Jayasekara, M. and Mallikarachchi, C. (2020), Parametric study on the homogenized response of woven carbon fibre composites, IEEE Xplore, pp. 36–41.

Herath, S., Xiao, X. and Cirak, F. (2022), 'Computational modeling and data-driven homogenization of knitted membranes', *International Journal for Numerical Methods in Engineering* **123**(3), 683–704.

Huang, T., Gao, J., Sun, Q., Zeng, D., Su, X., Liu, W. K. and Chen, W. (2021), 'Stochastic nonlinear analysis of unidirectional fiber composites using image-based microstructural uncertainty quantification', *Composite Structures* **260**.

Isabona, J., Imoize, A. L., Ojo, S., Do, D.-T. and Lee, C.-C. (2023), 'Machine learning-based gpr with lbfgs kernel parameters selection for optimal throughput mining in 5g wireless networks', *Sustainability* **15**(2).

Jayasekara, M., Herath, S., Gowrikanthan, N. and Mallikarachchi, C. (2021), Size effect and fibre arrangement on meso-mechanical modelling of woven fibre composites, Institute of Electrical and Electronics Engineers Inc., pp. 124–129.

Jones, R. M. (1998), *Mechanics of Composite Materials*, sec. edn.

Khan, M. and Noor, S. (2019), 'Performance analysis of regression-machine learning algorithms for predication of runoff time', *Agrotechnology* **08**.

Li, H., Bacarreza, O., Khodaei, Z. S. and Ferri Aliabadi, M. (2022), 'Probabilistic multi-scale design of 2d plain woven composites considering meso-scale uncertainties', *Composite Structures* **300**, 116099.

Maleki, F., Muthukrishnan, N., Ovens, K., Reinhold, C. and Forghani, R. (2020), 'Machine learning algorithm validation: From essentials to advanced applications and implications for regulatory certification and deployment', *Neuroimaging Clinics of North America* **30**, 433–445.

Mallikarachchi, H. (2019), 'Predicting mechanical properties of thin woven carbon fiber reinforced laminates', *Thin-Walled Structures* **135**, 297–305.

Mallikarachchi, H. and Pellegrino, S. (2013), 'Failure criterion for two-ply plain-weave cfrp laminates', *Journal of Composite Materials* **47**(11), 1357–1375.

Manocha, L. and Bahl, O. (1988), 'Influence of carbon fiber type and weave pattern on the development of 2d carbon-carbon composites', *Carbon* **26**(1), 13–21.

Naskar, S., Mukhopadhyay, T. and Sriramula, S. (2018), 'Probabilistic micromechanical spatial variability quantification in laminated composites', *Composites Part B: Engineering* **151**, 291–325.

Omairey, S. L., Dunning, P. D. and Sriramula, S. (2018), 'Influence of microscale uncertainties on the reliability of fibre-matrix composites', *Composite Structures* **203**, 204–216.

Pandit, R. K. and Infield, D. (2019), 'Comparative analysis of gaussian process power curve models based on different stationary covariance functions for the purpose of improving model accuracy', *Renewable Energy* **140**, 190–202.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E. (2011), 'Scikit-learn: Machine learning in Python', *Journal of Machine Learning Research* **12**, 2825–2830.

Rasmussen, C. E. and Williams, C. K. I. (2006), *Gaussian processes for machine learning*, MIT Press.

Schnaubelt, M. (2019), *A comparison of machine learning model validation schemes for non-stationary time series data*, Institute for Economics.

Sullivan, T. (2015), *Introduction to Uncertainty Quantification*, Vol. Volume 63, Springer.

Tam, D., Ruan, S., Gao, P. and Yu, T. (2012), 10 - high-performance ballistic protection using polymer nanocomposites, *in* E. Sparks, ed., 'Advances in Military Textiles and Personal Equipment', Woodhead Publishing Series in Textiles, Woodhead Publishing, pp. 213–237.

Tao, W., Zhu, P., Xu, C. and Liu, Z. (2020), 'Uncertainty quantification of mechanical properties for three-dimensional orthogonal woven composites. part ii: Multiscale simulation', *Composite Structures* **235**.

Ullah, Z., Zhou, X. Y., Kaczmarczyk, L., Archer, E., McIlhagger, A. and Harkin-Jones, E. (2019), 'A unified framework for the multi-scale computational homogenisation of 3d-textile composites', *Composites Part B: Engineering* **167**, 582–598.

Vipulananthan, V., Weerasinghe, U., Ariyasinghe, N., Mallikarachchi, C. and Herath, S. (2023), 'Homogenized material property prediction of carbon fiber composites using data-driven methods', *Journal of Sustainable Civil and Environmental Engineering Practices* **1**(1), 16–24.

Wang, J. (2023), 'An intuitive tutorial to gaussian process regression', *Computing in Science & Engineering* **25**(04), 4–11.

Wang, S., Celebi, M. E., Zhang, Y.-D., Yu, X., Lu, S., Yao, X., Zhou, Q., Miguel, M.-G., Tian, Y., Gorriz, J. M. and Tyukin, I. (2021), 'Advances in data preprocessing for biomedical data fusion: An overview of the methods, challenges, and prospects', *Information Fusion* **76**, 376–421.

Wang, Z. (2022), 'Comparative study of latin hypercube sampling and monte carlo method in structural reliability analysis', *Highlights in Science, Engineering and Technology AGECT* **2022**.

Zhu, C., Byrd, R. H., Lu, P. and Nocedal, J. (1997), 'Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization', *ACM Trans. Math. Softw.* **23**, 550–560.

Zhu, C., Zhu, P. and Liu, Z. (2019), 'Uncertainty analysis of mechanical properties of plain woven carbon fiber reinforced composite via stochastic constitutive modeling', *Composite Structures* **207**, 684–700.