Computational Testing of a Unidirectional Carbon Fiber Composite: Micromechanical Simulations and Machine Learning Approach

Yannic J. Gagnon*, Phillip H. Burnside, and Anthony S. Tai

January 18, 2024

Abstract

We present a computational method that incorporates micromechanical modeling through representative volume element theory simulations of a unidirectional carbon fiber / polymer resin composite to produce a dataset of 625 mechanical testing simulations. The focus was to first lay down the groundwork for subsequent, more sophisticated versions of the model; as such, effects such as temperature and humidity, and variable process parameters are not considered. Good agreement with room temperature, dry carbon fiber composite experimental data is found. Many constituent properties influence the final material properties of composite materials. An ML regression method that is especially useful for estimating feature importance in small datasets [1, 2] is applied to both a large dataset ($N = 625$ test points) and a smaller subset of this dataset ($N = 30$ test points) that is more representative of a dataset that could be generated through real-world experimental testing. A quantitative understanding of the influence of constituent material properties on the composite material response is achieved for both datasets. Through our empirical approach, the fiber volume fraction and fiber modulus emerge as important parameters in determining the Young's modulus of the composite along the fiber direction, as anticipated from composite theory.

1 Introduction

Composite materials combine multiple substituent materials to achieve a desired performance. Polymer composites such as carbon fiber composites are commonly used for their high strength to weight ratio. Additional applications of polymer composites includes high electromagnetic transmission (e.g., honeycomb core sandwich composites), anti-corrosive properties, and relatively low cost of manufacture [3]. Non-polymeric composite materials such as ceramic matrix composites also have benefits such as high temperature resistance, with applications to hypersonics / high-temperature radomes.

Before composite components are fielded, the composite materials that make up components are tested/qualified (e.g., mechanically or electrically) to ensure that the properties exhibited are the desired material properties. This qualification process determines the "allowable" values, corresponding to material property values that have been adjusted from ideal values to account for environmental effects (e.g., temperature and humidity), damage effects (e.g., a defect caused by routine use), or batch-to-batch / statistical variability (e.g., natural variations in the property value caused by non-ideal production processes and ma-

*To whom correspondence should be addressed: yannic.j.gagnon.civ@us.navy.mil

terial defects) [4]. Mechanical testing is typically done by producing test coupons from panels of the composite material, and conducting a series of ASTM or MIL-STD tests. The specific number of coupons needed to measure a given material property is determined by a competition of factors: while more measurements can better estimate the spread of mechanical property values, allowing a more precise material allowable value, collecting more measurements can be costly and time-consuming. Typically, dozens of samples are produced/tested, and allowable values represent overly-conservative reductions in material properties, requiring components to be more bulky than likely is necessary to accommodate the reduced property value. Difficulties in traditional mechanical testing are small-scale datasets and facility-specific measurements that do not extrapolate well to measurements conducted by other facilities. Additionally, the design of innovative composite materials suffers due to the long time and cost factors needed to test these new materials.

With these drawbacks to conventional composites testing, theory and simulation approaches have been leveraged to refine composite material property estimates [5]. Composite theory is a useful avenue for comparing experimental results with expectations, and for providing mathematical intuition regarding symmetries in the composite system. For example, considering the symmetry of a transversely isotropic unidirectional carbon fiber composite, the 36 terms of the stiffness matrix relating stresses and strains in a unidirectional composite material [6] reduce to only five independent material constants that determine the mechanical response. However, theory often employs approximations and simplifying assumptions to simplify the analysis. These simplifying approximations present a limit to the realistic situations encountered in the real-world, and emergent material properties exhibited by advanced engineering materials / metamaterials based on intricacies of the materials' internal structure may not be captured accurately by tractable theoretical approaches.

Computer simulations present an alternative to pure theory and more closely resemble real-world experiments. Ref. [7] used finite element method (FEM) simulations with micromechanical modeling and estimation of process parameters of carbon fiber composites to generate several thousand mechanical test points. These data provided a better quantification of the properties of the underlying probability distribution, and "virtual" material allowables could be extracted more simply then purely experimental values from the probability distribution, rather than the standard method outlined in the CMH-17 handbook designed for small datasets ($\sim$ dozens of test points). Ref. [5] similarly conducted simulations of unidirectional carbon fiber composites using Ansys and COMSOL, using the representative volume element (RVE) method, and demonstrated that simulations were in relatively good agreement with each other and were generally in better agreement than analytical models with experimental data. Interestingly, Ansys simulations were unable to accurately model experimental shear modulus data, while COMSOL was able to model these data. It is likely that Ansys and COMSOL may use different default assumptions relating to either the composite system or computational method. This difference between Ansys and COMSOL in modeling real-world data illustrates the point that simulations can be very useful, but caution must be exercised when setting simulation parameters and evaluating simulation results, and comparison to experimental values should be made whenever possible to validate simulations. An additional difficulty for simulations is that failure in experimental systems often begins near regions with defects such as regions with non-ideal adhesion between the filler and matrix or regions with voids. These non-idealities are not included in some simulations that assume constant material properties throughout the material, and instead model ideal interactions between constituent materials. Therefore, simulations that incorporate micromechanical modeling – models that do not assume homogeneous properties throughout the material but instead take into account locally heterogeneous properties resulting from locally heterogeneous materials, as in a composite – are especially important to develop accurate predictions of material properties and failure in experimental systems.

In addition to simulations reducing the time/cost factors of a material qualification program, a promising application to composite material simulations is the ability to produce large datasets that can be fit or classified using machine learning (ML) methods. Machine learning has revolutionized material and chemical discovery/optimization [8, 9] and will likely be a powerful tool for future composite materials' development and material property optimization. In composite research, ML may provide understanding beyond that which experiments, theory, and simulations by themselves provide. While a well-known application of ML are neural networks often used for classification or regression problems, other machine learning regression methods such as regularized regression provide a relatively simple method to determine the relative importance of constituent material properties or manufacturing process parameters (known as "features" in ML) [1]. Such regularized regression models are not "black box" methods, and results from the model are best-fit parameters that can be traced back to a fitting procedure similar to ordinary least-squares fitting. As ML methods, ML regression methods differ from conventional regression in that ML regression is not primarily focused on estimating a model's parameter set from a given dataset, but is instead focused on predicting whether a given model with fixed parameters determined by fitting a "training" dataset is able to characterize data in a separate "test" dataset. While this may appear a subtle difference, this difference in approach leads to differences in the parameter estimates, and allows regularized regression ML methods to avoid overfitting and to provide information about which parameters are most impactful in determining the response.

Here we present a computational method that incorporates micromechanical modeling through representative volume element theory simulations to produce a large (in the context of real-world mechanical tests) dataset of 625 mechanical testing simulations. Gaussian probability density functions (PDFs) are created from these data, and A- and B-basis allowables calculated directly from the PDFs. A machine learning regularized regression method - the least absolute shrinkage and selection operator (LASSO) - that is especially useful for estimating feature importance in small datasets [1, 2] is applied to both the full dataset and a smaller subset of the dataset made to represent a smaller dataset that could realistically be generated through real-world experiments. Both a qualitative and quantitative understanding of the influence of constituent material properties on the composite material response is established.

2 Computational Methods

The mechanical response of a composite is determined by local stress and strain relations. These relations are most succinctly described by the matrix form of Hooke's law for linearly elastic materials, in which stresses and strains are related by a 6x6 stiffness matrix[10]. For a single orthotropic lamina/ply, the stiffness matrix contains nine independent moduli values. For a unidirectional carbon fiber composite, the stiffness matrix reduces further, and the material response can be fully described by five independent properties. The five independent properties are typically taken to be the composite's Young's moduli parallel to and perpendicular to the fibers (E_1 and E_2, respectively), the in-plane and out-of-plane shear moduli (G_{12} and G_{23}, respectively), and the in-plane Poisson's ratio ν_{12}. The Young's and shear moduli describe the relation between stress and strain applied along different axes, while ν_{12} relates the strain along the axis parallel to the fiber with the strain along the axis perpendicular to the fiber.

The finite element method (FEM) is typically applied to characterize mechanical components that do not contain high degrees of symmetry, and for cases in which local variations in material properties are considered [11]. The FEM method transforms a continuous surface into a mesh containing a discrete number of mesh nodes and connections. Mesh geometries can be chosen to contain various mesh sizes and shapes; common geometries are square, hexagonal, and diamond. Solutions to partial differential equations at the nodes of the mesh govern the evolution of an applied stress or strain on the boundary of the material. Interpolation is used to estimate the stress and strain values between nodes in the mesh. The output of a FEM analysis is the stiffness matrix; additional properties such as the Poisson's ratio related to either the strains or stresses of the material can also be determined.

A representative volume element (RVE) model is a particularly useful model used to represent the smallest portion of composite material with mechanical properties representative of the full material. Solving the FEM for this smaller volume of the material greatly reduces the computational expense associated with simulating the full material volume. In addition, a RVE model is a type of micromechanical model that directly connects the constituent material properties with the properties of the full system. With RVE models, individual constituent parameters such as the moduli and density of the matrix and filler, in addition to properties regulating interactions between the matrix and filler such as the interfacial energy, can be investigated. Ansys Material Designer specifically contains pre-built RVE geometries, which include unidirectional carbon fiber composite and honeycomb core composite geometries. Ansys Material Designer also contains parameters that simulate experimentally realistic modifications to RVE materials, such as misalignment of fibers in a fiber composite. Once the RVE is generated and solved, the RVE material can be saved as a new material with continuous properties equal to those of the RVE, and this material can become the material used to simulate, for example, sandwich composites with a user-defined layup and ply thicknesses.

Assumptions related to the FEM and RVE model used in this work are that complete adhesion between the fiber and the matrix are assumed, fiber misalignment is considered negligible, large gradients in material properties are neglected, and no defects or other sources of local heterogeneity in the structure or properties other than the fiber and resin materials themselves are considered. These assumptions are common in the research literature, and often represent experimental fiber composite systems fairly well, while minimizing computational expense. We also use periodic boundary conditions in this work to ensure that RVE-measured properties are representative of macroscopic-scale material properties. By monitoring the material properties of the RVE as we varied the number of cells of the RVE, we determined that there was no difference in calculated material properties using only a single unit cell or multiple cells, ensuring that our results generalize to larger systems.

A unidirectional carbon fiber composite system of AS4 carbon fiber within 8552 resin amine cured epoxy coating was investigated in this work due to its numerous aerospace applications related to its cost-effectiveness and high strength-to-weight ratio, and the existence of well-characterized datasets. In addition, this unidirectional carbon fiber composite has a more simple geometry than, for example, honeycomb core composites, allowing for more intuitive results and comparison between simulation results and expectations from theory. The properties of AS4 and 8552 were collected from Refs. [12, 13, 14, 15] and are summarized in Table 1. The diameter of the AS4 fibers of 7.1 μm was used [16]. The uncertainty values represent the single standard deviation uncertainties. These values were determined in most cases by calculating the standard deviation of values provided in Refs. [12, 13, 14, 15]. In the case of the uncertainty of the volume fraction, only a range of values was provided in the literature [13]. For this case, the approximate relation between the standard deviation σ and the range R of the values $\sigma \approx R/4$ was used [17]. Material properties correspond to room temperature ($T = 70 \pm 10$ °F), dry (ambient humidity) conditions. The volume fraction ϕ of the composite expresses the ratio of the volume

Table 1: Material properties of 8552 polymer resin, AS4 carbon fiber, and unidirectional AS4/8552 lamina composite [12, 13, 14, 15].

Material	**Tensile modulus (GPa)**	**Tensile strength (GPa)**	**Poisson's ratio (in tension)**	**Density (g/cm^3)**
8552 (resin)	4.67 ± 0.38	0.121	0.37±0.01	1.30
AS4 (carbon fiber)	Fiber direction: 227.5 ± 6.9 Transverse: 22.75	Fiber direction: 4.41 ANSYS-supplied shear moduli: $G_{xy} = G_{xz} = 9$ $G_{yz} = 8.21$	$\nu_{xy} = \nu_{xz} = 0.2$ $\nu_{yz} = 0.4$ (ANSYS-supplied)	1.791± 0.043
Lamina ϕ = 0.569 ± 0.013 T = 70 ± 10 °F	Fiber direction: 137.93 ± 5.60 Transverse: 9.86 ± 0.59	Fiber direction: 2.03 ± 0.22 Transverse: 0.0599 ± 0.021	0.302	1.58

of the carbon fiber (V_{CF}) to the total volume (V_{CF} + V_{resin}): $\phi = \frac{V_{CF}}{V_{CF}+V_{resin}}$ and was calculated from Refs. [12, 13]. Block and conformal mesh settings were used with a single cell with periodic boundary conditions and a length ratio $XZ = 2$. Simulations were run on a PC with 8 CPUs and 51 GB RAM. The time for a single mechanical test simulation was approximately 90 seconds.

A dataset of 625 mechanical test simulation values was prepared by varying the fiber modulus in the direction of the fiber E_{f1}, the isotropic resin modulus E_r, the fiber volume fraction ϕ, and the resin's Poisson's ratio ν_r over five values each over the values $[\hat{x} - 2\sigma, \hat{x} - \sigma,\ \hat{x},\ \hat{x} + \sigma\ \hat{x} + 2\sigma]$. The $\hat{x}$ denotes the mean value for one of the four variables and σ denotes the standard deviation of the variable. Four variables varied over five values produces a test table of $5^4 = 625$ unique combinations of input variables. This table of test values was input into Ansys using the "Design Point Table" in conjunction with the "Parameterize" features of Ansys Material Designer. Simulations were then run in Ansys for each test point to determine the mechanical properties of the composite for each test point. Gaussian white noise with a magnitude equal to the standard deviation of the composite modulus was added to the simulation-determined composite modulus values. This addition of random noise was done to provide a noisy dataset simulating an experimental dataset that contains an upper bound of experimental testing error.

For the application of LASSO regression, the data were standardized to ensure that the magnitude of the coefficient values was not biasing the regression. LASSO regression is similar to ordinary least squares (OLS) regression, in which an objective function describing the differences between model predictions and experimental data are minimized. However, LASSO regression provides an additional constraint on the magnitude of the coefficients in the model. The best-fit values of coefficients in the model correspond to the coefficients that minimize the objective function. For LASSO regression, the minimization problem to determine the set of best-fit coefficients $\hat{c}$ of a model with variable x is[18]:

$$\hat{c} = \text{argmin}\left\{\frac{1}{2}\sum_{i=1}^{N}(y_i - c_0 - \sum_{j=1}^{p} x_{ij}c_j)^2 + \lambda\sum_{j=1}^{p}|c_j|\right\}, \tag{1}$$

where the minimization is done by varying the model's coefficients c_i, N is the number of test points (quantity of data), p is the number of coefficients (not including the c_0 term, if such an offset term is used in the model), y_i is the experimental value of the ith datapoint, and λ is a constant known as the L_1 regularization parameter. Increasing the value of λ shrinks the model's coefficients to zero. The terms that play a minor role in the response are rapidly shrunk to zero when λ is given larger values, while terms that are more important in defining the response remain non-zero for relatively larger values of λ. This shrinkage aspect of LASSO regression therefore provides a simple method of determining which terms in a polynomial model are most important.

Throughout this work, the composite mechanical response plotted as a function of constituent variables was found to be approximately linear over the ranges of values chosen. The variables of the model were the fiber modulus in the direction of the fiber E_{f1}, the isotropic resin modulus E_r, the fiber volume fraction ϕ, and the resin's Poisson's ratio ν_r. In this study the influence of these variables on the composite's Young's modulus in the direction of the fibers E_1 was investigated. The linear model used was:

$$\overline{E}_1 = c_1\overline{E}_{f1} + c_2\overline{E}_r + c_3\overline{\phi} + c_4\overline{\nu}_r, \qquad (2)$$

where c_1, c_2, c_3, and c_4 are coefficients that are determined through LASSO regression (Eq. (1)). The bars over the variables indicate that these are standardized variables.

3 Results and Discussion

An important parameter to set in FEM simulations is the mesh size (or equivalently, the mesh density). If the mesh size is too large, results may be inaccurate, but if mesh sizes are too small, simulation time increases. A proper mesh size therefore maintains the accuracy of values determined from simulations while minimizing the computational expense of the simulations. Figure 1 shows the influence of the mesh size on the composite modulus in the fiber direction E_1. Figure 1(a) shows three different square mesh RVE geometries with mesh sizes of (left-to-right) 200 nm, 1 μm, and 4 μm. From a visual inspection, the 200 nm mesh appears extremely fine, and the 4 μm mesh appears coarse. The 1 μm mesh appears to be a reasonable intermediary between these two extremes. Figure 1(b) quantifies the effect of mesh size selection by plotting the influence of mesh size on E_1. All simulations were conducted with fixed simulation/material property parameters discussed in Section 2. Blue circles are values of E_1 determined from simulations run over a range of mesh sizes from 200 nm to 4 μm. Dashed lines are to guide the eye. The solid vertical line is at a mesh size of 1 μm. Interestingly, the values of E_1 vary by less than 0.1% for the full range of mesh sizes selected, far below the range of E_1 that would be obtained in experimental measurements of nominally identical composite samples. Figure 2 similarly plots as a function of the mesh size the composite's Young's modulus perpendicular to the fiber direction E_2 (a), the in-plane Poisson's ratio ν_{12} (b) and shear modulus G_{12} (c), and the out-of-plane shear modulus G_{23} (d). In all cases, a small variation of the dependent variable of less than 1% was observed for the full range of mesh sizes selected. While even the maximum tested mesh size of 4 μm would likely have provided accurate simulations, a cautious maximum mesh size value of 1 μm was selected and held fixed for all subsequent simulations.

The volume fraction ϕ of the filler in a polymer composite is a common property to tune to alter the mechanical or electromagnetic properties of the composite for a specific application. Figure 3 a) provides a visual perspective by showing three different RVEs (RVEs not shown to scale) with volume fractions of $\phi = 0.1$, $\phi = 0.569$ (the mean value of the volume fraction used in this study), and $\phi = 0.7$. This maximum value of 0.7 is typically the limit of realistic packing efficiency in the fabrication of carbon fiber composites. Figure 3 b) shows E_1 as a function of ϕ ranging from zero (neat polymer resin value shown as a grey square) to 0.7. The colored triangles are values determined from our simulations of RVEs with square, hexagonal, or diamond mesh geometries. Experimental values from the literature [6, 12] are shown as black circles. Ref. [6] collected the majority of these experimental data (the data with no error bars), with the exception of the data located at $\phi = 0.57$. Ref. [6] collected values via ultrasonic measurements of a similar carbon fiber composite to that in this study

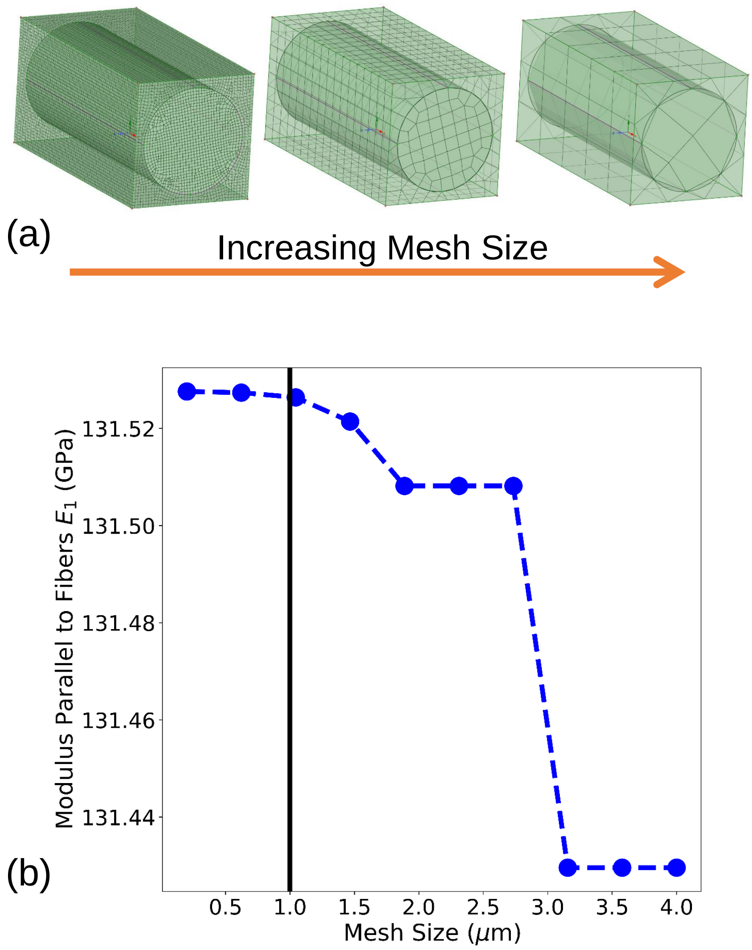

Figure 1: Impact of mesh size on the Young's modulus of the composite along the fiber direction E_1. Three different square mesh RVE geometries are shown in (a), with mesh sizes of 200 nm, 1 μm, and 4 μm. Quantification of the influence of mesh size on E_1 is shown in (b). Simulations were conducted over a range of mesh sizes, keeping other parameters constant. The fiber volume fraction was fixed at $\phi = 0.5691$. Blue circles are values of E_1 determined from simulations run over a range of mesh sizes from 200 nm to 4 μm. Dashed lines are to guide the eye. The solid vertical line highlights a mesh size of 1 μm.

consisting of a Modmor II carbon fiber / LY558 epoxy resin composite. The data at $\phi = 0.57$ corresponds to the mean value and standard deviation of four separate experimental measurements collected by Refs. [12, 13, 14, 15] of the AS4/8552 composite modeled in this work. Excellent agreement between our simulations and the experimental data are observed. An approximately linear trend in E_1 vs. ϕ is found in simulations. There also does not appear to be a strong dependence of E_1 on the mesh geometry selected.

Figure 4 shows the influence of volume fraction on E_2, ν_{12}, G_{12} and G_{23}. The colored triangles represent simulation values calculated from simulations using mesh geometries illustrated on the right hand side of Fig. 3 and the black circles are experimental values from Refs [6, 12, 13, 14, 15]. The range of composite properties with mesh geometry is small, and similar in magnitude to experimental noise. However, selection of the mesh geometry at higher volume fractions can lead to small changes in the dependent variable. Our simulation results

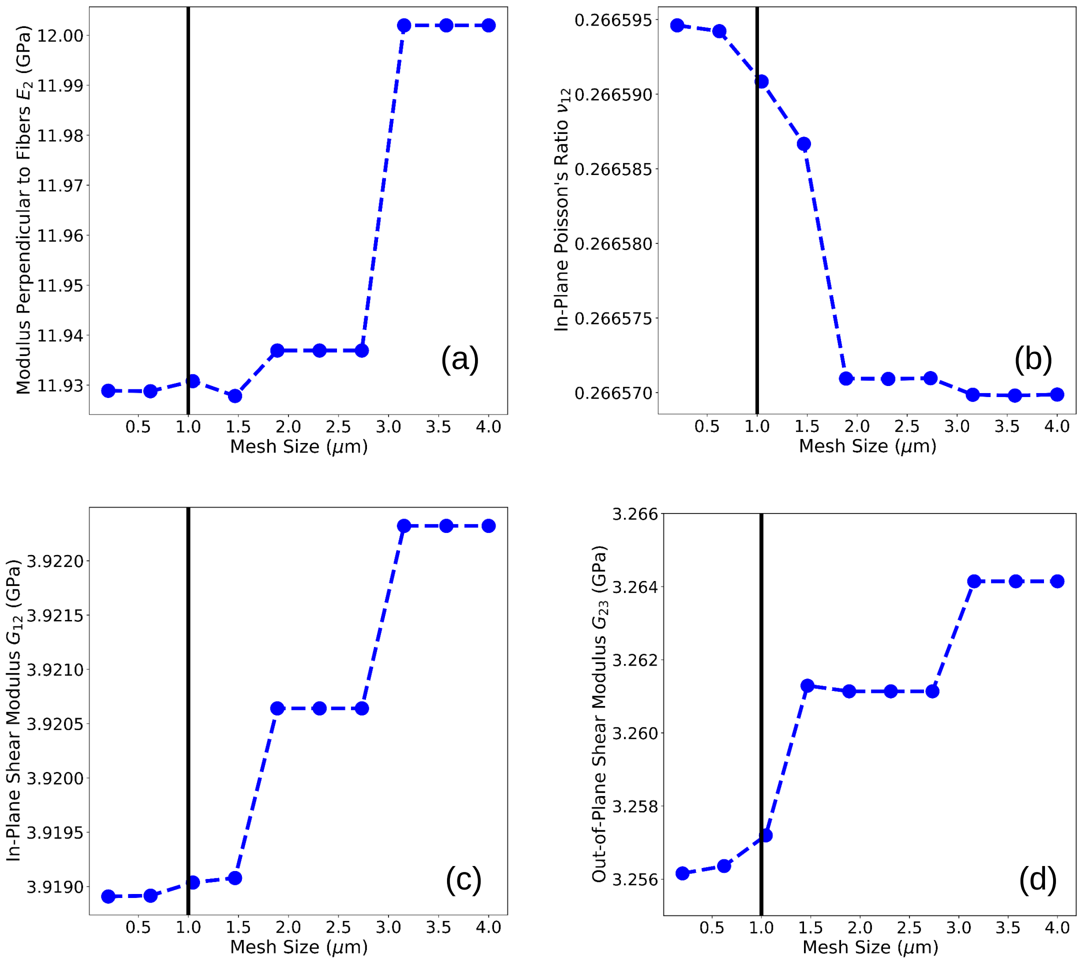

Figure 2: Quantification of the influence of mesh size on the composite's Young's modulus perpendicular to the fiber direction E_2 (a), the in-plane Poisson's ratio ν_{12} (b) and shear modulus G_{12} (c), and the out-of-plane shear modulus G_{23} (d). Simulations were conducted over a range of mesh sizes, keeping other parameters constant. The fiber volume fraction was fixed at $\phi = 0.5691$. Blue circles are the results from simulations run over a range of mesh sizes from 200 nm to 4 μm. Dashed lines are to guide the eye. The solid vertical lines highlight a mesh size of 1 μm.

are generally in good agreement with experimental values. Fig. 4(c) shows an exception to this rule, where it is observed that experimental data are consistently larger than simulation results, as also observed in Ref. [5] for Ansys simulations. There is not a clear benefit across all properties measured using a particular mesh geometry, so we perform subsequent simulations with a square mesh geometry.

Determining Allowables from Probability Distributions of Material Properties

With all simulation parameters selected and establishing that our simulation results - particularly results for the longitudinal E_1 and transverse E_2 Young's moduli of the composite - are in good agreement with experimental data from the literature, we turn to the consideration of determining material allowables. In general, the determination

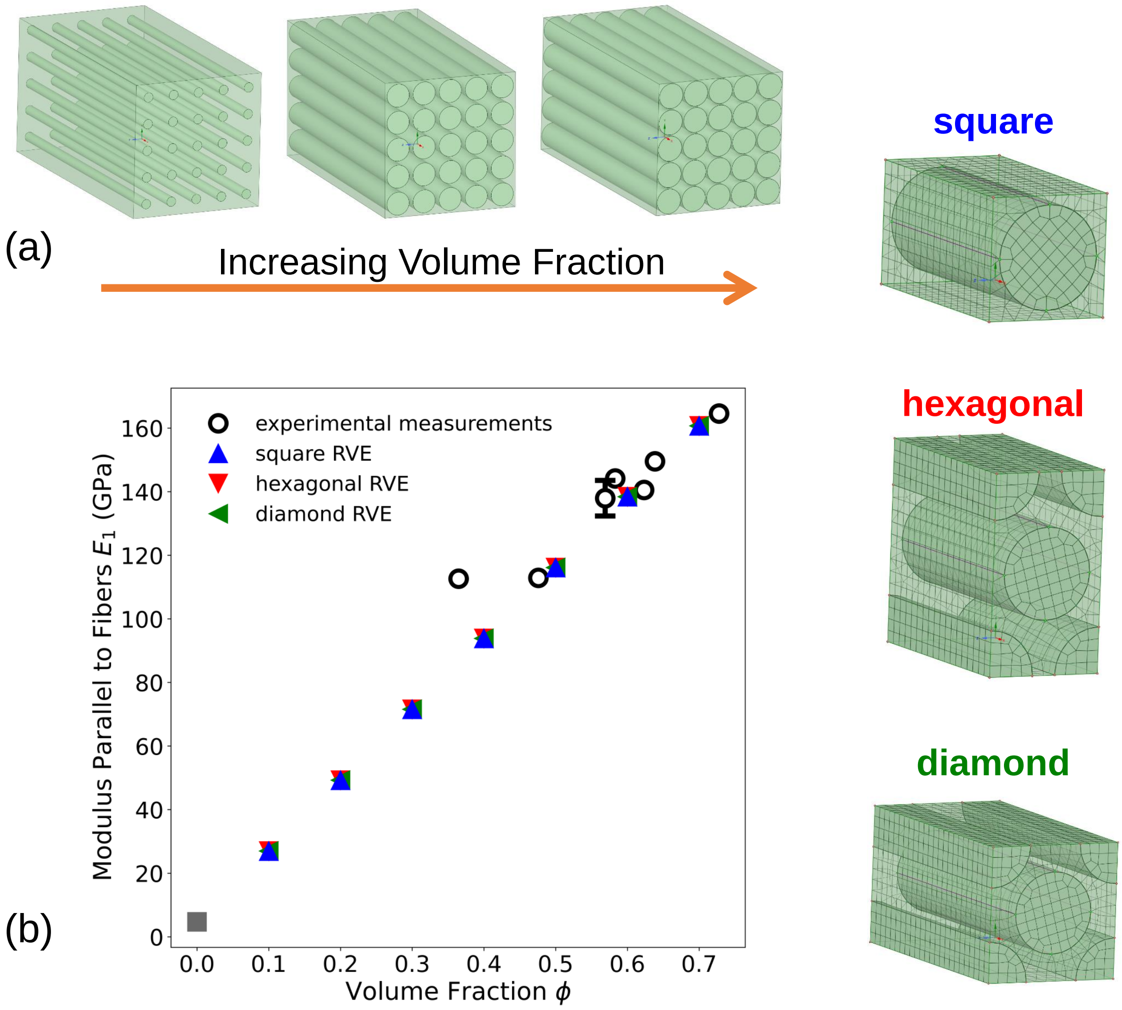

Figure 3: Young's modulus of the composite along the fiber direction E_1 as a function of fiber content (volume fraction). (a) Three different RVEs (not to scale) with volume fractions of $\phi = 0.1$, $\phi = 0.569$, and $\phi = 0.7$ with otherwise identical properties. b) E_1 as a function of volume fraction. The colored triangles are values from simulations conducted using square, hexagonal, or diamond mesh geometries for the RVE. The black circles are experimental values from Refs. [6, 12, 13, 14, 15].

of material allowables includes elevated temperature/humidity testing and open-hole testing in addition to a statistical analysis. In this study, temperature and humidity are not varied and we consider only the statistical aspect of the determination of material allowables. Figure 5 shows the probability density functions (PDFs) of the composite's material properties (E_1, E_2, G_{12}, ν_{12}) determined from 625 simulations. Inadequate data from the literature were available for the estimation of the experimental uncertainty of the in-plane shear modulus G_{12}; therefore, noise was not added to these data. All other material properties had added noise to better simulate experimental uncertainty involved in real-world measurements. The simulated data are displayed as histograms; in addition, a normal distribution was calculated and is overlayed on the data as a black curve. The cyan vertical bar denotes the "B-basis" material property values, corresponding to the 10th percentile of the probability distribution. The magenta vertical bar denotes the "A-basis" material property values, corresponding to the 1st percentile of the probability distribution. These values are deter-

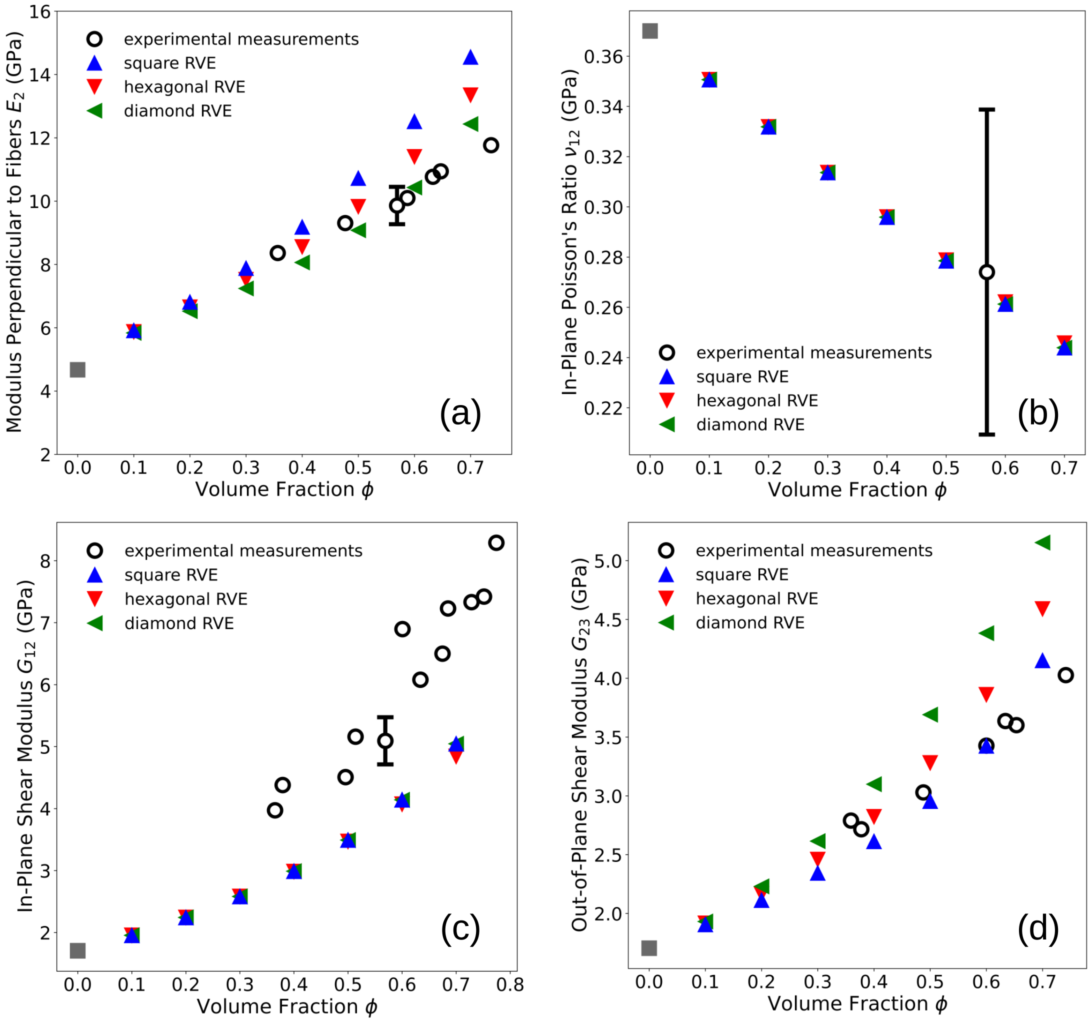

Figure 4: Quantification of the influence of volume fraction on the composite's Young's modulus perpendicular to the fiber direction E_2 (a), the in-plane Poisson's ratio ν_{12} (b) and shear modulus G_{12} (c), and the out-of-plane shear modulus G_{23} (d). Simulations were conducted over a range of volume fractions, holding other parameters constant. Colored triangles are results from simulations run using mesh geometries indicated in Fig. 3(a). Black circles are experimental measurements collected by Refs. [6, 12, 13, 14, 15].

mined by adding the standard deviation of the data multiplied by the associated z-score ($z_B = -1.282$ for the B-basis allowable and $z_A = -2.326$ for the A-basis allowable) to the mean of the data. Unfortunately, allowable measurements in the research literature for the AS4/8552 lamina do not exist for the material properties measured by simulations in this study. Instead, the longitudinal and transverse tensile strength allowables are measured in Ref. [12]. Therefore there does not exist data in the literature to the authors' knowledge for us to draw comparisons to. While problematic, this highlights an additional benefit of simulations. Provided that the simulations are benchmarked on known experimental data, such as material properties as a function of volume fraction data or tensile test data, simulations can provide estimates of allowables for other material properties that may be too difficult or time intensive to determine experimentally.

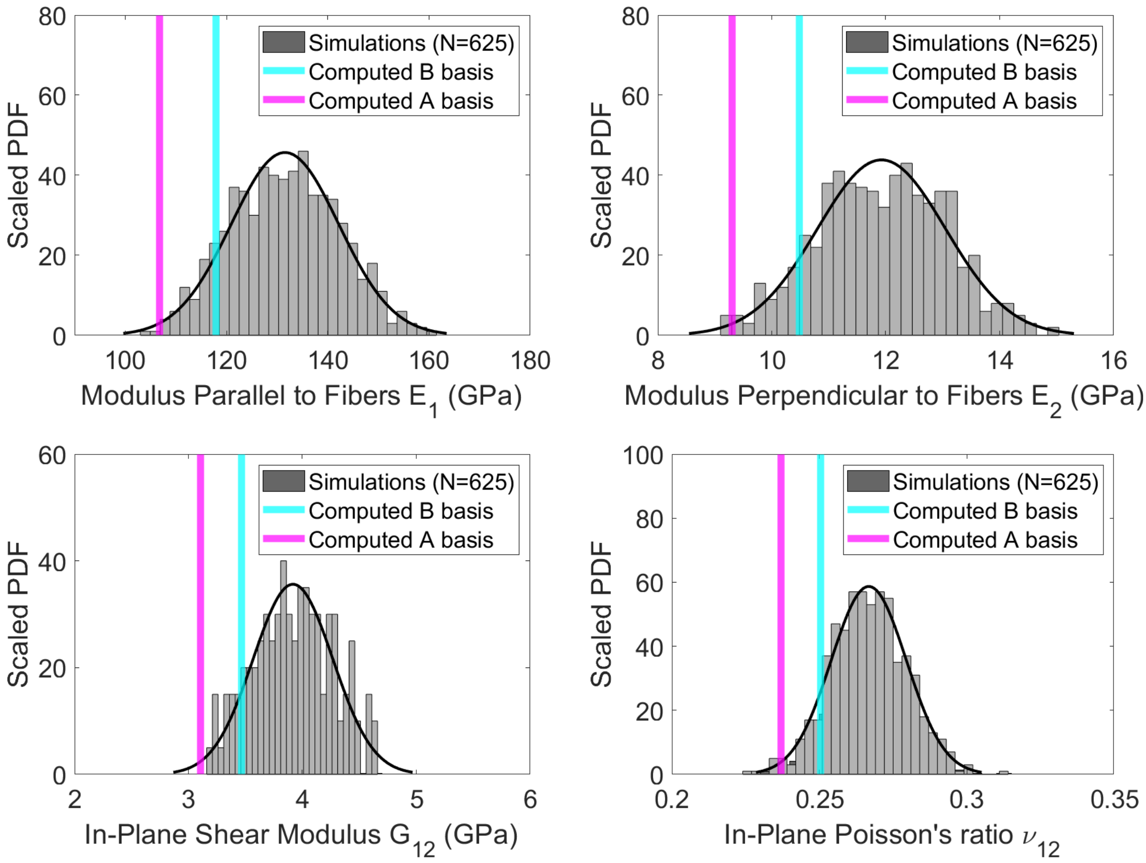

Figure 5: Determining material property allowables from the probability density functions (PDFs) of the AS4/8552 composite's material properties (E_1, E_2, G_{12}, ν_{12}) determined from 625 simulations. Histograms used MATLAB's default bin width. The vertical axis corresponds to the probability density function value, scaled by a coefficient related to the bin width. Cyan and magenta bars correspond to the B-basis (10th percentile) and A-basis (1st percentile) material property allowables, respectively.

Machine Learning Approach to Determining Relative Importance of Constituent Properties

The tendency of the modulus of a carbon fiber composite along the fiber direction to increase in response to increasing particle volume fraction (up to a point), or increasing the fiber modulus or resin modulus, is well-known. Less well-known, and often system-dependent, are the relative influences of varying constituent materials' properties on the composite modulus. For example, under which conditions does increasing the fiber volume fraction alter the resulting composite modulus to a greater extent than using a different, higher modulus fiber? Or from a manufacturing or repair perspective, what tuning of which collection of material properties or process parameters would most efficiently control the composite's modulus? With the aim to better characterize the influence of constituent material properties in determining the composite response, we applied LASSO regression (Eq. (1)) to our data. We specifically considered only linear terms in the LASSO model; in addition, preliminary modeling that considered interaction terms found these interaction terms to be relatively unimportant and results of considering these interaction terms are therefore not included in this study. Figure 6(a) shows the result of the application of this multivariable regression LASSO model fit to 80% of the full dataset, consisting of a training data size of 500 simulated measurements.

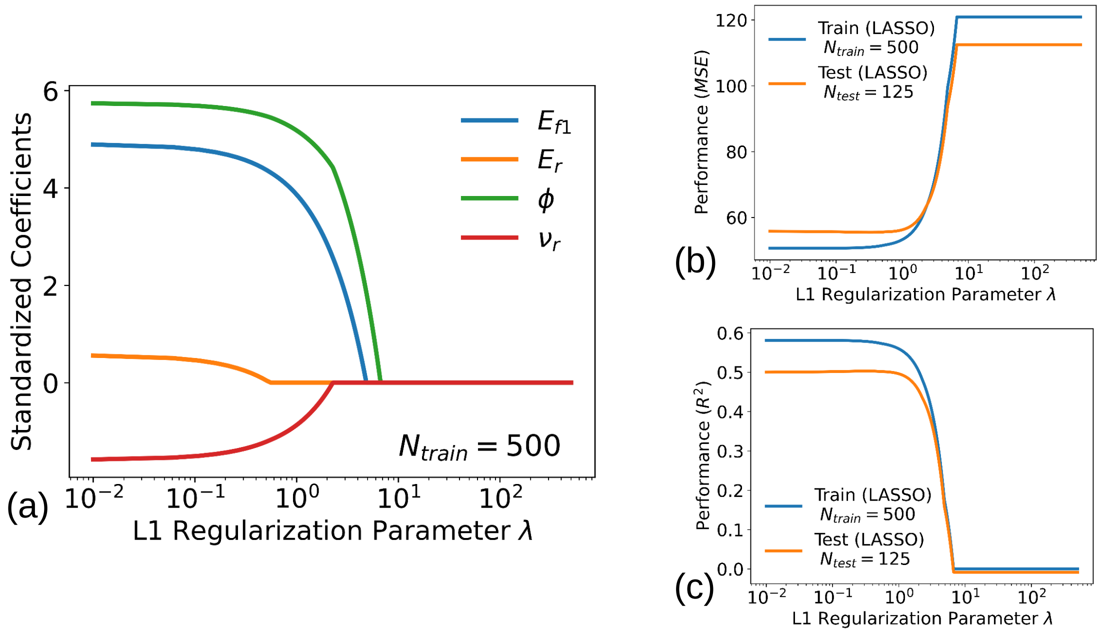

Figure 6: LASSO regression on regularized data from simulations with added noise. Coefficient values from LASSO model fit to 80% of the full dataset ($N_{train} = 500$) are shown in (a). Each curve corresponds to the coefficients determined from the fit. Increasing the L1 regularization parameter λ eventually shrinks the coefficients to zero. The mean squared error (MSE) performance metric calculated from both the training and test datasets as a function of λ is shown in (b). The minimum MSE, corresponding to the λ value that provides the best fit, occurs for both training and test datasets at $\lambda \approx 0$. The alternative R^2 performance metric calculated from both the training and test datasets as a function of λ is shown in (c). The maximum R^2, corresponding to the λ value that provides the best fit, similarly occurs for both training and test datasets at $\lambda \approx 0$.

Each curve in Figure 6(a) corresponds to a coefficient determined from the fit. It is important to note that the data are regularized, leading to, e.g., negative values for the coefficient of the regularized Poisson's ratio. This does not imply the Poisson's ratio itself is negative; instead, what is important is the magnitude of the coefficient relative to the magnitude of the other standardized coefficients. Another important aspect to the data is that Gaussian noise equal to the uncertainty in the parameters was added to the data before the data were standardized to model experimental uncertainty that is encountered in real-world measurements.

Increasing the regularization parameter eventually shrinks the coefficients in the model (c_1, c_2, c_3, and c_4 in Eq. (2)) to zero. Coefficients that remain nonzero longer than other coefficients with increasing λ are those that are particularly important in describing the mechanical response (known as "features" in ML). This method therefore provides an intuitive feature selection tool. As anticipated from composite theory, the volume fraction and fiber modulus are the most important features in determining E_1. It was not obvious, however, that the volume fraction was necessarily a more important feature in this composite than the fiber modulus until this analysis. Interestingly, the Poisson's ratio, rather than the Young's modulus, of the resin is predicted to be a more important physical property of the resin in influencing the composite's modulus along the fiber direction. There are two commonly used metrics of performance of a regression model: the mean squared error MSE (for both linear and nonlinear models) and R^2 (generally used for linear models). Smaller MSE values correspond to better fits, while larger R^2 values approaching unity correspond to better linear fits. Figure 6(b) shows the MSE that compares the linear model fit by LASSO to both training and test data. As anticipated, the MSE of the training data is lower than the test data, since the model was fit to (trained) to the training data only. The minimum MSE determined from LASSO regression of the training dataset was MSE = 50.6, occurring at the minimum value of λ tested of $\lambda = 1\times10^{-6}$. The minimum MSE determined from LASSO regression of the test dataset was MSE = 55.5, occurring at $\lambda = 0.35$. The small value of $\lambda \approx 0$ from the regression analysis of the training dataset suggests that the LASSO method produces equivalent results to the standard least squares fitting method when applied to the training dataset. This is unsurprising, since we have a relatively small number of parameters (4 parameters) fitting a large number of data (500 data in the training dataset). The trained LASSO regression model evaluated using the test data suggests a non-negligible λ value of $\lambda = 0.35$ should be used. Since we wish to maximize the performance of the model on test data, the coefficients corresponding to the value of $\lambda = 0.35$ provide the best-fit coefficients to our linear model. Figure 6(c) similarly shows performance measured by the R^2 metric. The maximum R^2 determined from LASSO regression of the training dataset was $R^2 = 0.58$, occurring at the minimum value of λ tested of $\lambda = 1 \times 10^{-6}$. The maximum R^2 determined from LASSO regression of the test dataset was $R^2 = 0.50$, occurring at $\lambda = 0.35$. The equivalent values of λ determined from performance measured by R^2 and MSE relates to both of these metrics being excellent performance metrics for linear models. The LASSO model provides the following multivariable linear relation between the standardized composite modulus parallel to the fiber direction E_1 and the standardized substituent properties varied in our simulations:

$$\overline{E}_1 = 4.53\overline{E}_{f1} + 0.205\overline{E}_r + 5.56\overline{\phi} - 1.33\overline{\nu}_r. \quad (3)$$

Standardization is denoted by bars over the standardized variables. Determinations of uncertainties of coefficient values calculated from LASSO regression are typically determined by cross-validation. While this work explored cross-validation in preliminary investigations of coefficient uncertainty calculations, we leave it to future work to determine the best method to determine uncertainties in LASSO coefficients that relate to constituent material properties of composites. In addition, only the standardized material properties are reported in this work; an additional step that will be taken in future work is to relate coefficients determined from fits to standardized data to coefficients corresponding to non-standardized material properties.

One of the benefits of the LASSO method is that it is able to minimize the number of parameters in a model by shrinking relatively unimportant parameters to zero. This can result in decreasing the amount of data needed to be collected, making LASSO a particularly attractive ML regression method for small datasets commonly encountered in the experimental lab or production environments. To model such a small dataset, a random selection of 30 total data from the full 625 mechanical test simulation dataset was taken. The data were then randomly split between a training set of 70% of the data (21 points) and a test set consisting of the remaining 30% of the data (9 points). This 70%-30% split was used to ensure enough data were available in the test dataset. Figure 7(a) shows the standardized coefficients determined from fitting the LASSO model to this smaller training data set. The trend in the coefficients are in good agreement with the trend in coefficients determined from fitting the full dataset (Fig. 6(a)). The performance as measured by MSE shown in Figure 7(b) is maximized for the training dataset at MSE = 30.1, occurring at the minimum value of λ tested of $\lambda = 1 \times 10^{-6}$. The minimum MSE determined from LASSO regression

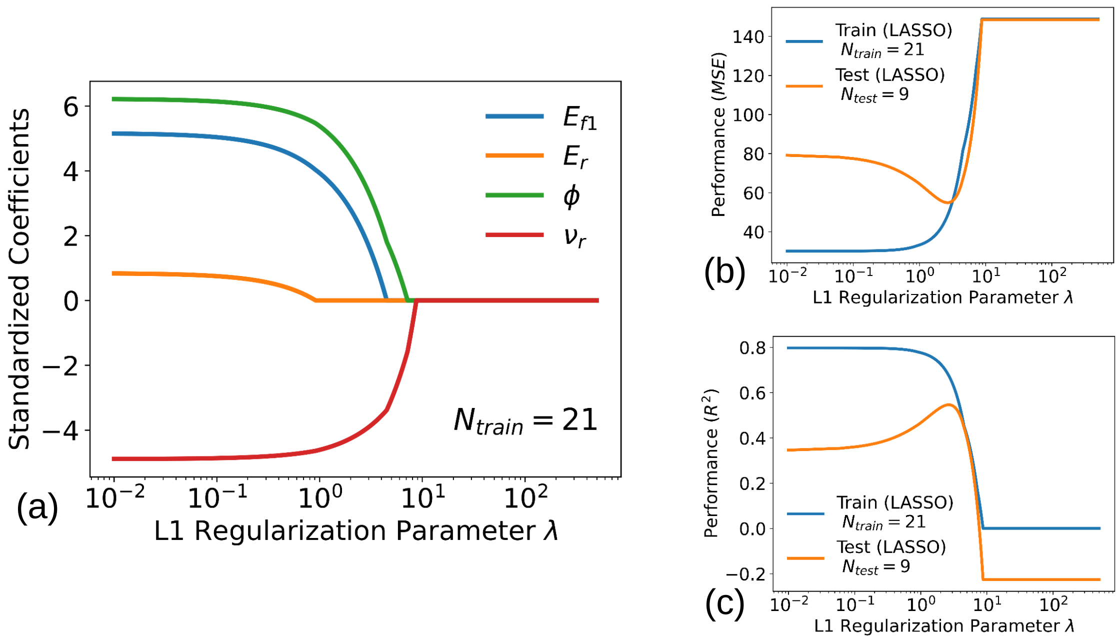

Figure 7: LASSO regression applied to a small subset of the full simulation dataset. Coefficient values from LASSO model fit to 70% of data randomly selected from the full dataset ($N_{train} = 21$) is illustrated in (a). Each curve corresponds to the coefficients determined from the fit. Increasing the L1 regularization parameter λ eventually shrinks the coefficients to zero. The mean squared error (MSE) performance metric calculated from both the training and test datasets as a function of λ is shown in (b). The minimum MSE, corresponding to the λ value that provides the best fit, occurs for both training and test datasets at $\lambda = 2.65$. The alternative R^2 performance metric calculated from both the training and test datasets as a function of λ is shown in (c). The maximum R^2, corresponding to the λ value that provides the best fit, similarly occurs for both training and test datasets at $\lambda = 2.65$.

of the test dataset was MSE $= 54.9$, occurring at $\lambda = 2.65$. The small value of $\lambda \approx 0$ from the regression analysis of the training dataset suggest that the LASSO method produces equivalent results to the standard least squares fitting method when applied to the training dataset. However, the most important metric of performance of the model is the performance of the model when applied to the test dataset, and the maximum performance will determine the optimum value of λ and therefore the material property coefficients to use to relate the substituent material properties to the composite's modulus. Figure 6(c) similarly shows performance measured by the R^2 metric. The maximum R^2 determined from LASSO regression of the training dataset was $R^2 = 0.80$, occurring at the minimum value of λ tested of $\lambda = 1 \times 10^{-6}$. The maximum R^2 determined from LASSO regression of the test dataset was $R^2 = 0.55$, occurring at $\lambda = 2.65$. The equivalent values of λ determined from performance measured by R^2 and MSE relates to both of these metrics being excellent performance metrics for linear models. The LASSO model fit to the smaller dataset of a total of $N = 30$ data provides the following multivariable linear relation between the standardized composite modulus along the fiber direction E_1 and the standardized substituent properties varied in our simulations:

$$\overline{E}_1 = 2.07\overline{E}_{f1} + 3.69\overline{\phi} - 4.03\overline{\nu}_r. \quad (4)$$

Standardization is denoted by bars over the standardized variables, and while uncertainties were not determined for the coefficients, future work will determine the best method to determine uncertainties in LASSO coefficients. In addition, only the standardized material properties are reported in this work.

Equation (4) demonstrates the ability of LASSO regression to remove unimportant features to obtain a better regression model when fitting sparse datasets. The optimum λ value ($\lambda = 2.65$) occurs in a regime of substituent material properties in which the coefficient of the modulus of the polymer resin, E_r, is zero. This suggests that the modulus of the polymer resin is relatively unimportant in determining the modulus of the composite along the fiber direction. This data-driven discovery relates to a well-known scientific understanding of load-bearing in unidirectional composites: the majority of the stresses applied to a unidirectional composite along the direction of fibers is carried by the fibers. An interesting empirical prediction of Equations (3) and (4) is the relative importance of the Poisson's ratio of the resin ν_r in influencing $\overline{E}_1$. In the continuum theory, when the composite is strained in the longitudinal (fiber) direction, contraction in the transverse axis occurs. The degree to which this contraction occurs is set by the resin's Poisson's ratio. It is well-known that a resin's Poisson's ratio is related to both the resin's shear and Young's moduli. A calculation based on Equations (3.16)-(3.17) in Ref. [10] also suggests that ν_r relates to the composite's Poisson's ratio ν_{12}. Considering that the majority of the transverse contraction of the composite that occurs upon an applied longitudinal strain (as is done to determine E_1 in a mechanical test) occurs within the resin rather than within the fibers, ν_{12} may depend primarily on the resin's shear and Young's moduli; i.e., the resin's Poisson's ratio. If this is the case, ν_r may then be strongly correlated with ν_{12} and may be an important property to consider when selecting resins or matrix materials for unidirectional composite materials.

4 Summary and Conclusions

Simulations provide an inexpensive and fast method of determining the mechanical properties of composite materials. Used in conjunction with experimental data, simulations can fill gaps in data or supply more data than can be collected experimentally. Hybrid experimental/simulation approaches may therefore reduce the quantity of experimental testing required to qualify components, reducing the time and cost of required testing. In addition to expediting material testing, the increased quantity of data possible with simulations are important for the determination of material allowables, in which case more data can provide more accurate allowable values, and therefore less conservative reductions in the material properties can be applied. These less conservative reductions in material properties typically reduce the weight of the component, which is especially important for reducing the weight of aircraft components. Combined with machine learning, simulations can reveal new physical or chemical relationships between constituent materials, paving the way for the development of materials that take advantage of these physical/chemical relationships to improve performance.

In this study we established a computational method that incorporates micromechanical modeling through representative volume element theory in Ansys and applied the method to an AS4/8552 unidirectional carbon fiber / polymer resin composite commonly used in the aerospace industry. Overall, good agreement was found between the simulation data and experimental data collected from the literature. We conducted 625 mechanical testing simulations in an automated manner to provide a relatively large dataset of the mechanical properties of the AS4/8552 composite. We determined the probability distributions of a subset of the composite's material property values and determined the A and B basis allowable values for these material properties by identifying the 1st percentile and 10th percentile values of the probability density functions of the material properties.

A standard ML regression method - the least absolute shrinkage and selection operator (LASSO) - was fit to the data to determine how the constituent properties of the fiber and resin, and the volume fraction of the fiber, influence the final mechanical properties of the composite. Through this empirical approach, the fiber volume fraction and fiber modulus emerged as the most important parameters in determining the Young's modulus of the composite along the fiber direction, as anticipated from composite theory. Interestingly, the Poisson's ratio of the polymer resin emerged as a more important parameter than the resin's modulus in the mechanical response. As a further exploration of the LASSO method, the LASSO regression analysis was repeated on a smaller subset of only 30 computational test simulation data randomly selected from the full dataset. This small subset of the data represented a quantity of data similar to what would be feasible to collect in a purely experimental mechanical test study. LASSO regression was found to provide qualitatively similar results to those determined from fitting the full dataset, illustrating the feasibility of LASSO regression to provide information about the relative importance of constituent material properties when fitting a relatively low number (N=30) of noisy data with a moderate number of parameters (4 parameters).

The present work conducted several hundred simulations of an undamaged composite at room temperature and negligible humidity, focusing on the statistical treatment of material allowables and on feature selection through machine learning regression. An important real-world aspect of material allowables generation is reducing the mechanical properties by "knock down" factors determined through open hole testing (to simulate damage to the composite) and elevated temperature and humidity testing. The sum of all knock downs (including the statistical knock down considered in this study) can reduce the allowable modulus of the composite by over 70% [10]. In the interest of aligning with real-world measurements, future work will seek to conduct computational tests of a composite lamina/laminate composite with an open hole geometry to determine damage knockdown factors and will seek to calculate temperature/humidity knockdown factors by providing simulation inputs of temperature and humidity lookup tables for the constituent material properties. In addition, the longitudinal and transverse tensile strength values will be determined to better compare results with literature data. In the current study, simulation inputs were determined from a comprehensive literature review. While this collection of material properties presents an excellent starting estimate of simulation parameters, future work will seek to calibrate the simulations to a particular experimental dataset potentially collected in-house, using, e.g., an iterative mean squared error (MSE) calculation to determine which simulation inputs provide output data that most align with experimental results. This calibration approach and a hybrid simulation/experimental approach to composites testing may help solve the long-standing problem in composites qualification / allowables determination of a lack of agreement in material property data collected from different facilities.

Acknowledgements and Funding Sources

We thank E. Rensi for helpful discussions. Funding from the Naval Surface Warfare Center, Crane Division's Naval Innovative Science & Engineering (NISE) program is gratefully acknowledged.

References

[1] P. Mehta, M. Bukov, C.-H. Wang, A. G. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, "A High-Bias, Low-Variance Introduction to Machine Learning for Physicists," *Physics Reports*, vol. 810, pp. 1–124, 2019.

[2] Y. Zhang and C. Ling, "A Strategy to Apply Machine Learning to Small Datasets in Materials Science," *npj Computational Materials*, vol. 4, no. 1, p. 25, 2018.

[3] L. Dorworth, G. Gardiner, and G. Mellema, *Essentials of Advanced Composite Fabrication and Repair*. Aviation Supplies and Academics, Incorporated, second ed., 2019.

[4] *Composite Materials Handbook*. SAE International, 2012.

[5] S. B. Rayhan and M. M. Rahman, "Modeling Elastic Properties of Unidirectional Composite Materials Using Ansys Material Designer," *Procedia Structural Integrity*, vol. 28, pp. 1892–1900, 2020.

[6] R. Kriz and W. Stinchcomb, "Elastic Moduli of Transversely Isotropic Graphite Fibers and Their Composites," *Experimental Mechanics*, vol. 19, pp. 41–49, 1979.

[7] W. Seneviratne, J. Tomblin, and T. Jayaratne, "Accelerating Composite Material Qualification Using Statistically-Significant Multi-Scale Progressive Damage Models," *The Composites and Advanced Materials Expo*, 2016.

[8] J. E. Saal, A. O. Oliynyk, and B. Meredig, "Machine Learning in Materials Discovery: Confirmed Predictions and Their Underlying Approaches," *Annual Review of Materials Research*, vol. 50, pp. 49–69, 2020.

[9] C. Trinh, D. Meimaroglou, and S. Hoppe, "Machine Learning in Chemical Product Engineering: The State of the Art and a Guide for Newcomers," *Processes*, vol. 9, no. 8, p. 1456, 2021.

[10] C. Kassapoglou, *Design and Analysis of Composite Structures: With Applications to Aerospace Structures*. John Wiley & Sons, 2013.

[11] O. C. Zienkiewicz and R. L. Taylor, "The Finite Element Method Fifth Edition Volume 1: The Basis," 2012.

[12] E. Clarkson, "Hexcel 8552 AS4 Unidirectional Prepreg Qualification Statistical Analysis Report," *FAA special project no SP4614WI-Q, Report No NCP-RP-2010-008 Rev D, Wichita State University*, 2011.

[13] "HexPly 8552 Product Data Sheet," *Hexcel Composites*, 2020.

[14] "HexPly 8552 Epoxy Matrix Product Data Sheet," *Hexcel Composites; retrieved from ORNL.gov*, 2013.

[15] J. Bonhomme, J. Viña, A. Argüelles, I. Viña, and V. Mollón, "Influence of the Matrix Toughness in Carbon-Epoxy Composites Subjected to Delamination Under Modes I, II, and Mixed I/II," *Mechanics of Advanced Materials and Structures*, vol. 20, no. 8, pp. 679–686, 2013.

[16] "HexTow AS4 Product Data Sheet," *Hexcel Composites*, 2013.

[17] S. P. Hozo, B. Djulbegovic, and I. Hozo, "Estimating the Mean and Variance From the Median, Range, and the Size of a Sample," *BMC Medical Research Methodology*, vol. 5, no. 1, pp. 1–10, 2005.

[18] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, vol. 2. Springer, 2009.

Appendix

Figure (8) provides a subset of the data after the addition of noise representing simulated experimental uncertainty was added to both the independent and dependent variables. The large spread of the data in both E_1 and the constituent values is due to the data representing an upper limit to the experimental uncertainty. This upper limit of simulated experimental uncertainty was investigated to test whether LASSO regression would still be able to extract useful information about feature importance. As discussed in the main text, LASSO regression was indeed able to provide realistic predictions of which constituent properties (features) were most important in determining the Young's modulus of the composite in the direction of the fiber alignment.

Figure (9) displays the results of LASSO regression for the standardized composite Young's modulus perpendicular to the fiber orientation E_2 and the composite's shear modulus G_{12}. These figures illustrate the well-known understanding in composite theory that loads applied perpendicular to the fiber direction in a unidirectional fiber composite are carried primarily by the polymer resin, rather than the fibers. The resin's Young's modulus is therefore more important than the fiber's Young's modulus in this situation.

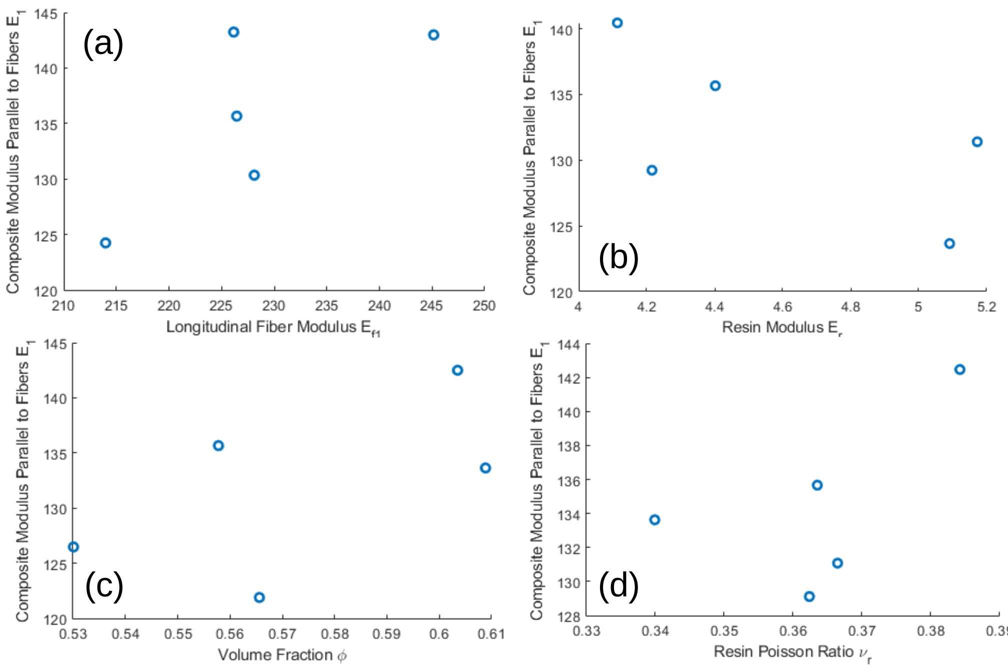

Figure 8: Subset of data after the addition of noise representing simulated experimental uncertainty in both the independent and dependent variables. Moduli are in GPa. Each figure was generated by plotting the composite modulus as a function of a constituent variable, while holding the other constituent values fixed at their mean values with added noise. The large spread of the data in both E_1 and the constituent values is due to the data representing an upper limit to the experimental uncertainty.

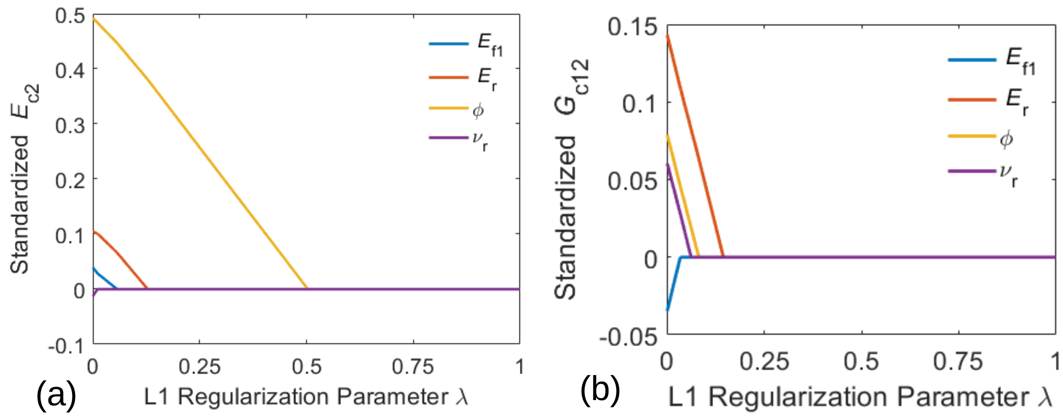

Figure 9: Coefficient values of constituent properties for the composite Young's modulus perpendicular to the fiber orientation (a) and shear modulus (b), as a function of the L_1 regularization parameter λ. The step size of λ is more coarse in these figures than in Figures (6)-(7) so the finer details of the coefficient curves are not captured; however, the trends inform the relative importance of the influence of constituent properties on the moduli of the composite.