Reduced-order models for microstructure-sensitive effective thermal conductivity of woven ceramic matrix composites with residual porosity

Adam P. Generale, Surya R. Kalidindi *

George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

ARTICLE INFO

ABSTRACT

This paper presents a data-driven framework for the development of reduced-order models to predict microstructure-sensitive effective thermal conductivity of woven ceramic matrix composites (CMCs) with residual porosity. The main components of the proposed framework include (i) digital generation of representative volume elements (RVEs), (ii) estimation of the effective thermal conductivities of the RVEs using finite element (FE) models, (iii) low dimensional representation of the microstructure in the RVEs using 2-point spatial correlations and principal component analysis (PCA), and (iv) an active learning strategy based on Gaussian process regression (GPR) that minimizes the size of the training dataset through the selection of microstructures with the highest potential for information gain. The reduced-order models are demonstrated to provide high fidelity predictions on new RVEs.

1. Introduction

Ceramic-matrix composites (CMCs) using textile architectures are of tremendous interest for a variety of high-temperature structural applications [1–4]. The use of textile architectures offers an extremely large design space for the optimization of the component performance through varied stitching patterns (e.g., weaving, braiding, knitting) [5,6] in addition to the selection of the fiber and the matrix materials. These varied parameters control the salient details of the internal structure of CMCs, and consequently their effective properties[7]. Our primary focus in this paper will be on establishing reliable and robust low-computational cost reduced-order models between the three-dimensional microstructure of CMCs and their effective thermal conductivity.

The effective thermal conductivity of composite materials can be predicted by solving the governing heat conduction equation while accounting for the conductivities of the constituent phases [8,9]. However, these rigorous solutions have been largely restricted to highly idealized microstructures, such as small concentrations of dispersed non-interacting spherical or cylindrical filler material in a continuous matrix. These classical solutions are not directly applicable to CMCs with complex architectures. Several extensions have been proposed to bridge this gap to account for interfacial resistance [10–12], varied filler material shapes and packing [13–15], regularly arranged pores and radiative heat transfer [16], and experimentally calibrated trans- verse conductivity [17,18]. However, most of these models continue to utilize mainly the volume fractions of the major constituents. There is little reason to believe that such simple models would be able to provide reliable predictions over the wide range of morphologies and pore arrangements observed in woven CMCs [19,20]. Indeed, prior studies have reported very different calibrated values of the parameters involved in such models [15,21,22].

Establishing microstructure-sensitive reduced-order models for the effective properties of porous woven CMCs is quite challenging. These composites often exhibit complex microstructures with very high contrast in the conductivities of the constituents (i.e., reinforcement, matrix, and porosity). It is well established that the elementary (but rigorous) bounds for high contrast composites are often widely separated [19,20]. Similarly, the spatial distribution of pores (e.g., aligned pores) are known to induce high levels of anisotropy to the effective conductivity. Most of the simplified models described earlier only predict the isotropic effective conductivity of the composite material.

Numerical approaches based on finite element method (FEM) offer a different avenue for the prediction of the effective properties. These approaches allow one to capture the complex spatial arrangements in the material microstructure and the assignment of complex physics for describing the behavior of the individual constituents. The FEM approach has provided reliable predictions of the effective thermal conductivity for a variety of composite microstructures [23–26]. The main difficulty with the use of the FEM approach comes from its high

computational cost, especially for large three-dimensional (3-D) RVEs. This precludes its effective use for the optimization of the material microstructure for a specified application. Hybrid approaches, where the unknown parameters in the simplified analytical/empirical models are calibrated using FEM simulations, have also been explored [27,28]. These approaches are still limited, as their parametric model forms have been largely aimed at very simple microstructures.

Recently, a data-driven framework called Materials Knowledge System (MKS) [29–31] has been proposed to address the gaps identified above. This framework offers novel avenues for building low-computational cost reduced-order models between the complex 3-D composite microstructure and its effective properties. The MKS models are calibrated to simulation datasets generated by incorporating the requisite physics (e.g., FEM). The MKS approach featurizes the microstructure using n-point spatial correlations (also called n-point statistics) [32,33] and principal component analyses (PCA) [34]. The n-point statistics provide a hierarchy of increasingly complex descriptions of the material microstructure. For example, 1-point statistics capture the probabilities of finding a specific local material state (i.e., microscale constituent) at any randomly selected voxel in a 3-D RVE. Therefore, they capture information such as the phase volume fractions. At the next higher level, 2-point statistics capture the probability of finding two specified local states at the head and tail of a randomly placed vector in the 3-D RVE. Generally, n-point statistics provide an extremely high-dimensional representation of the microstructure, where each dimension captures one statistic related to one specified relative spatial arrangement of local states. PCA provides a low-dimensional representation of the n-point statistics in a form which maximizes the explained variance in any given ensemble of 3-D RVEs in the minimum number of features (called PC scores). The MKS approach has successfully produced robust structure–property linkages for a broad range of materials and their effective properties, including the elastic and inelastic properties of composites [35,36] and polycrystalline materials [37,38], fatigue of polycrystalline materials [39,40], and diffusivities in porous electrolytes [41]. These previous efforts have primarily used standard regression techniques (e.g., polynomial regression). Very recently, these approaches have been extended to include modern machine learning approaches based on Gaussian process regression (GPR) [42]. Beyond removing the need for known parametric model forms, GPR offers opportunities for the rigorous treatment of prediction uncertainty for new inputs. This allows for the optimal generation of training data by selecting specific sets of inputs that exhibit the most potential for improving the fidelity of the GPR model; such strategies minimize the computational cost of generating training data using sophisticated physics-based simulations.

In this paper, the MKS approach is extended to the prediction of microstructure-sensitive effective thermal conductivity of woven CMCs with residual porosity. Specifically, we will demonstrate and validate the approach for the anisotropic effective thermal conductivity of an 8-ply stack of 5-harness satin (5HS) weave with varying constituent morphology and porosity. It is emphasized that the methodology can be generalized to include widely varied textile architectures. Specifically, we will explore the viability and benefits of GPR for optimally generating a reduced-order model trained on FEM simulations.

2. Background methods

2.1. Statistical representation of microstructure

Quantification of the microstructure is essential to formulate reliable structure–property linkages. As discussed earlier, the MKS framework [29–31] will be employed for this purpose. This framework utilizes a voxelized representation of the RVE as an array m_s^h [29],

reflecting the volume fraction of material local state h in voxel s. For describing 3-D RVEs, it is convenient to use $s = \{s_1, s_2, s_3\}$. Furthermore, since each voxel is assigned to a single local state, the value of m_s^h is either zero or one [29]. The strategies and protocols for computing 2-point statistics of a microstructure have been described in detail in prior literature [29,31]. The 2-point statistics of an individual microstructure can be expressed as

$$f_r^{hh'} = \frac{1}{|S|} \sum_s^S m_s^h m_{s+r}^{h'} \tag{1}$$

where S denotes the set of voxels in the 3-D RVE of the microstructure, and $|S|$ is the total number of voxels. $f_r^{hh'}$ can be computed efficiently using the Fast Fourier Transform (FFT) algorithm [36,43], and serve as the initial set of features representing the microstructure in formulating the desired structure–property linkages.

2.2. Dimensionality reduction

Principal component analysis (PCA) offers a robust dimensionality reduction method for the full set of n-point statistics [44]. This method transforms the original data into a new frame that is arranged to maximize the capture of variance within the dataset in the minimum number of terms. This transformation can be mathematically described as

$$f_r^{(j)} = \sum_{k=1}^{min((J-1),R)} \alpha_k^{(j)} \varphi_{kr} + \bar{f}_r \tag{2}$$

where $f_r^{(j)}$ is a vectorized representation of the spatial correlations of the jth microstructure, J is the total number of microstructures, R is the number of dimensions in the original dataset, $\alpha_k^{(j)}$ are the principal component weights (i.e., PC scores) of the jth microstructure, φ_{kr} is the principal components (i.e., basis vectors of the transformed space), and $\bar{f}_r$ represents the ensemble averaged values of $f_r^{(j)}$. PCA is performed through a singular value decomposition (SVD) on the dataset X, where each row represents one mean-subtracted data point. The SVD results in $X = U\Sigma V^T$, where the basis vectors φ_{kr} are V^T, and the scores α_k are $U\Sigma$. The principal components are sorted by eigenvalues (i.e., the singular values on the diagonal of Σ). The number of principal components to be retained (i.e., R) can be objectively determined through the evaluation of the explained variance by each basis vector as a portion of the total variance in the full dataset. Prior work [45] found that more than 90% of the variance in most microstructure ensembles can be explained within the first five principal components. This reduction in dimensionality without appreciable loss in representation is essential for the development of high fidelity reduced-order structure–property models.

2.3. Gaussian process regression (GPR)

GPR is employed in this work for building the reduced-order models of interest. GPR is a non-parametric Bayesian approach that is particularly advantageous when there is little prior knowledge of the model form and the computational cost of generating training data is large. The application of Bayesian inference in model-building allows for tracking model uncertainty and utilizing the estimated uncertainty to select the optimal locations in the input domain where new training data is to be generated [46].

A Gaussian process (GP) can be formally defined as a collection of random variables of which any finite subset follow a joint multivariate normal distribution (denoted by $\mathcal{N}(\bar{x}, X)$, where $\bar{x}$ denotes the mean and X is the covariance) [47]. A flexible and interpretable GP can be expressed as [47,48]

$$f(x) = g(x) + h(x)^T \beta \tag{3}$$

with $g(x) \sim \mathcal{N}(0, K(X, X'))$, $h(x) = \{h_1(x), h_2(x), \cdots, h_M(x)\}$ representing the set of fixed basis functions, and $\beta = \{\beta_1, \beta_2, \cdots, \beta_M\}$ the associated coefficients. $X = \{x_1, x_2, \cdots, x_N\}$ defines the set of multi-dimensional training inputs, and $y = \{y_1, y_2, \cdots, y_N\}$ their corresponding outputs. $K(X, X')$ represents the covariance matrix of the GP, whose elements will be defined using a kernel function $k(x, x')$. In this work, the automatic relevance determination squared exponential (ARD-SE) [48] defined below is employed:

$$k(x, x') = \sigma_f^2 exp\left(-\frac{1}{2}\sum_{m=1}^{R}\frac{(x-x')^T(x-x')}{\lambda_m^2}\right) \tag{4}$$

The ARD-SE allows separate interpolation lengthscale parameters for each of the input variables. In the present study, the inputs variables will include the ordered PC scores representing the material microstructure, which exist over varying ranges, and therefore are likely to require distinct interpolation hyperparameters. The ARD-SE kernel function is modulated by the hyperparameter set $\theta = \{\sigma_f, \lambda_1, \lambda_2, \cdots, \lambda_R\}$. The amplitude, σ_f, controls the amplitude of the output function and reflects its ability to reach outliers in the training dataset, at the expense of increased noise in predictions. The length-scale λ controls the interpolation range, such that larger values result in smoother interpolations between points and smaller values result in more variable functions.

Assuming that the priors on β are also normally distributed (i.e., $\beta \sim \mathcal{N}(b, B(\beta, \beta'))$), the target can be modeled as another GP with a non-zero mean and variance as

$$f(x) \sim \mathcal{N}(h(x)^T b, k(x, x') + h(x)^T B h(x')) \tag{5}$$

Our goal is to form a conditional distribution on new input values (i.e., new PC scores representing unseen microstructures). Let X^* denote the set of test points, y^* the corresponding outputs. Let $K_*(X_*, X)$ denote the covariance matrix between test and training set, and $K_{**}(X_*, X_*)$ the covariance in the test points, all of which are evaluated using the kernel function (Eq. (4)). Let σ_n denote the homoscedastic noise in the training set. Employing a non-informative prior on the weights β (i.e., assignment of a large variance of B such that $B^{-1} \approx 0$), the predictive distribution of test points can be established with the mean $\bar{f}$ and covariance Σ_* expressed as

$$\bar{f}(X_*) = K_* K_y^{-1} y + R^T \bar{\beta} \tag{6a}$$

$$\Sigma_* = K_{**} - K_* K_y^{-1} K_* + R^T \left(H K_y^{-1} H^T\right)^{-1} R \tag{6b}$$

where $K_y = \left[K(X, X') + \sigma_n^2 I\right]$, H and H_* are matrices collecting the vectors of the evaluated basis functions for the training and test sets, and $R = H_* - H K_y^{-1} K_*$.

It should be noted that the underlying model form in the GPR is controlled by a set of parameters and hyperparameters, $\theta = \{\sigma_f, \lambda_1, \lambda_2, \cdots, \lambda_R\}$, observational noise in the training data, σ_n, and the model parameter coefficients β. This set of hyperparameters and mean function model parameters must then be optimized to best match the observed training data. This is accomplished by maximizing the log marginal likelihood defined as

$$\log p(y|X, \theta) = -\frac{1}{2}y^T K_y^{-1} y + \frac{1}{2}y^T C y - \frac{1}{2}\log|K_y| - \frac{1}{2}\log|A|$$
$$-\frac{N-M}{2}\log 2\pi \tag{7}$$

where $A = H K_y H^T$ and $C = K_y^{-1} H^T A^{-1} H K_y^{-1}$. In this work, this optimization was performed using the quasi-Newton algorithm [49,50].

The performance of the GPR model can be evaluated through the Normalized Mean Absolute Error (NMAE) defined as

$$NMAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - f(x_i)|/\bar{y} \tag{8}$$

where $f(x_i)$ is the predicted expected value for the input vector x_i, and $\bar{y}$ denotes the mean of the test outputs.

2.4. Active learning and experimental design

The high computational cost of performing FE simulations on microstructural RVEs of interest requires strategies to maximize the informational gain of each training microstructure. Clearly, a random selection methodology would be suboptimal. Bayesian approaches facilitate an efficient selection through the use of the posterior distribution over all available candidate training data. Multiple active learning strategies exist for maximizing model accuracy with the smallest subset of training points. Examples of which include, uncertainty sampling [51] and entropy-based selection [46], where new training points are identified based on the largest posterior variance, along with selection by maximum difference from current estimates [52], or selection by maximum expected improvement in the model fit [53], among many others [53,54]. Current work selects each subsequent training point through the evaluation of their uncertainty in the posterior prediction. The process involves seeding the input domain of interest with a large number of potential training points - a small subset of which is selected as the initial training data (i.e., FE simulations are performed for these) for the GPR model. Predictions are then made for all of the remaining candidate training points using the most recently updated GPR model. The candidate training point exhibiting the maximum predicted variance is then selected as the new training point (to perform FE analysis). The cycle is repeated as many times as needed – usually until the model hyperparameters and the error show the desired convergence characteristics.

2.5. Benchmarks for prediction of effective thermal conductivity

The GPR model developed in this work for the prediction of effective thermal conductivity is benchmarked against the analytical hierarchical two-layer (H2L) model [55], due to its prior use on similar woven composite architectures with residual porosity. The model can be viewed as an extension of the Hassleman-Johnson (H-J) model [10] with correction factors for both the filler material and porosity. In the H2L model, the microstructure is idealized into alternating layers of dense plies and porous matrix material. The effective conductivity, k_{eff}, in this model is expressed as

$$\frac{1}{k_{eff}} = \frac{V_t}{K_t} + \frac{V_M}{K_M} \tag{9}$$

where V_t and V_M are the tow and matrix volume fractions, respectively, and K_t and K_M are the conductivities of the dense ply and the porous matrix layers. A Maxwell-Eucken (ME) correction factor is used to adjust the pure dense matrix conductivity, k_m, into one exhibiting macro-porosity, through the expression, $K_M = k_m[(1-P)/(1+\beta P)]$, where P is the porosity content in the absence of tows, and β is a porosity shape factor. The effective dense ply layer conductivity is then estimated as

$$K_t = k_m \left[\frac{(1-P')}{(1+\beta' P')}\right]\left[\frac{1-(B/A)V_t}{1+(B/A)V_t}\right] \tag{10}$$

where $A = 1 + k_t/ah + k_t/K_M$, $B = 1 + k_t/ah - k_t/K_M$, ah is the product of an effective tow radius and interfacial conductance, and P' and β' are additional ply-level correction factors, corresponding to microporosity content within the effective dense ply and the microporosity shape factor, respectively.

3. FEM for effective conductivity and dataset generation

3.1. Microstructure generation

Our interest in this study will be focused on an 8-ply stack of a 5HS weave CMC microstructure with the repeating unit cell (RUC) shown in Fig. 1 [56], with the goal of producing training data covering a broad range of variations within this microstructural class. It is emphasized here that the training dataset does not need to replicate exactly the microstructures of the samples produced in the laboratory. Instead, our goal here is to cover a broader space of microstructure statistics so that the effective properties of the potential test microstructures (representing the laboratory samples) can be interpolated from the model calibrated on the training set.

The overall microstructure generation workflow presented in Fig. 1 used in this work is comprised of two main steps. The first is performed utilizing the open source software TexGen [57], and starts with the specification of a RUC as a single ply of the 5HS woven textile using the geometric parameters defined in Fig. 2a and b. The typical ranges of the values for these parameters are provided in Table 1, and were extracted from prior work on similar material systems [58,59]. Several simplifying assumptions were made in the generation of this stochastic dense RVE. Variation in tow positioning and cross section was neglected, and assumed to be completely periodic with uniform elliptical shape. The composite lay-up in the RVE (shown in the middle of the TexGen procedure in Fig. 1) was formed through duplication of the 5HS ply in the Z direction as defined by the local coordinate system in Fig. 1. The distance from the centerline of each duplicated ply to its neighbor was set by the defined ply spacing, p_c (see Fig. 2b). Each of the duplicated 5HS plies were also stochastically translated in the XY plane, where the translation components were randomly selected. The strategy described above produced an ensemble of RVEs. Each resulting RVE was discretized into $100 \times 100 \times 100$ uniform voxels of size 55 μm × 55 μm × 19.5 μm.

The voxelated microstructure generated by TexGen consisted of only two local states, the woven tows and matrix material. Pores were then inserted into the dense RVE through reassignment of the matrix voxels beyond a specified distance from the tows into pore voxels, where the specified distance is defined by the matrix thickness parameter m_t in Table 1. This approach is consistent with observed locations of residual macro-pores in manufactured CVI SiC/SiC CMCs [1,55,60] left over from processing, where a CVI-SiC layer is deposited on the preform. The process of CVI for densification of composites is exceedingly complex where the matrix thickness and distribution are functions of the reactor processing parameters, microstructure, along with the geometry of the preform [61]. It should be highlighted that the generation methodology described in this work produces an approximation of the deposited matrix, as the thickness is uniform throughout the RVE considered. This is a reasonable assumption for relatively thin panels, although fails to be representative of thicker specimens [61]. A spatial map of distances of all matrix voxels to the nearest matrix/tow interfaces is required for the implementation of this pore placement criteria. This map can be computed efficiently using an Euclidean Distance Transform (EDT) [62]. For a given binary voxelated image, EDT outputs a Euclidean distance map quantifying the closest distance of the matrix/tow interface to each voxel in the TexGen-produced voxelated RVE. The end result of the protocol described in Fig. 1 is a 3-D voxelated RVE, where each voxel is labeled as either a tow, matrix, or pore (i.e., three distinct local states). The generated 3-D RVE depicts a tow architecture based on the specified RUC, with deposited matrix surrounding the tows, layup stochasticity in the form of variable ply to ply positioning, and the resulting stochastic placement of pores due to the placement criteria.

The microstructure generation process described above was used to create an ensemble of 3125 digital microstructures. The generated microstructure ensemble spanned a five-dimensional input space that included suitable ranges for the tow major and minor axes, tow spacing, ply spacing, and matrix thickness, as shown in Table 1. A Latin hypercube design (LHD) [63] was used to ensure a relatively uniform
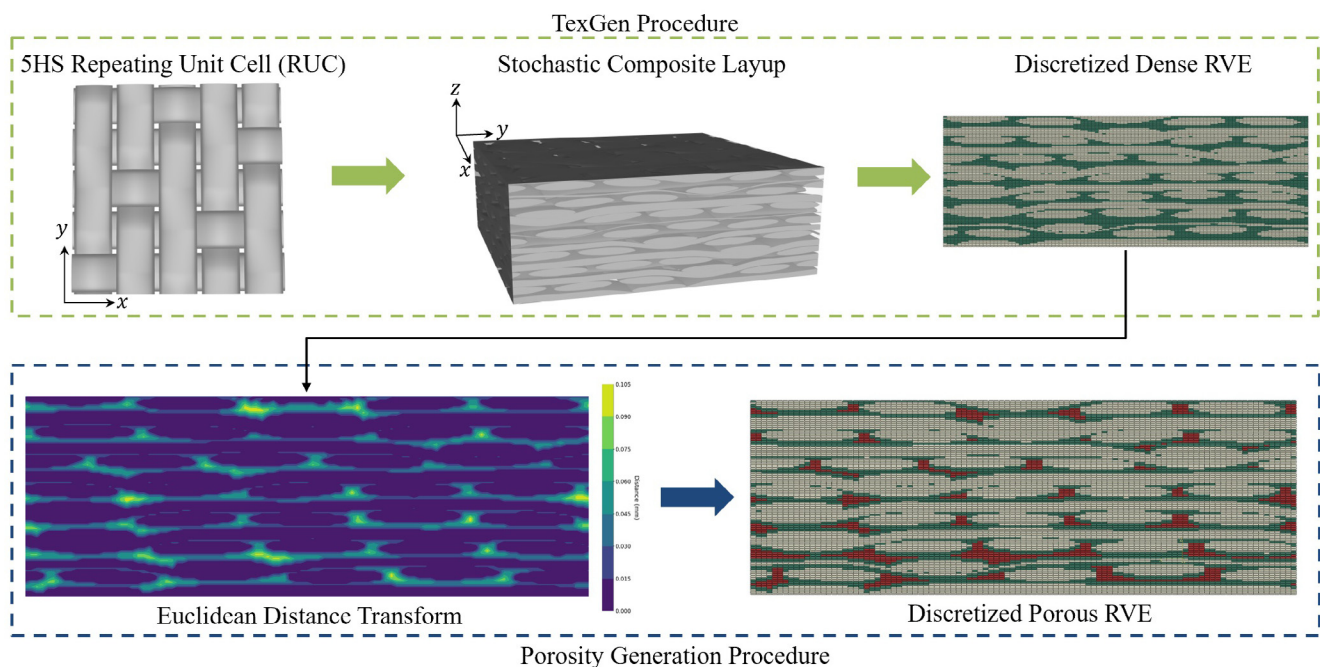

TexGen Procedure

5HS Repeating Unit Cell (RUC)   Stochastic Composite Layup   Discretized Dense RVE

Euclidean Distance Transform   Discretized Porous RVE

Porosity Generation Procedure

Fig. 1. Illustration of the overall workflow used in this study for microstructure generation. It is comprised of two steps. The first step is performed in TexGen and results in the generation of a dense RVE. The second step adds pores using a criterion based on Euclidean Distance Transforms. Constituents of the final RVE are colored green for the CVI-SiC matrix, beige for the SiC Sylramic tows, and red for residual macro-porosity.
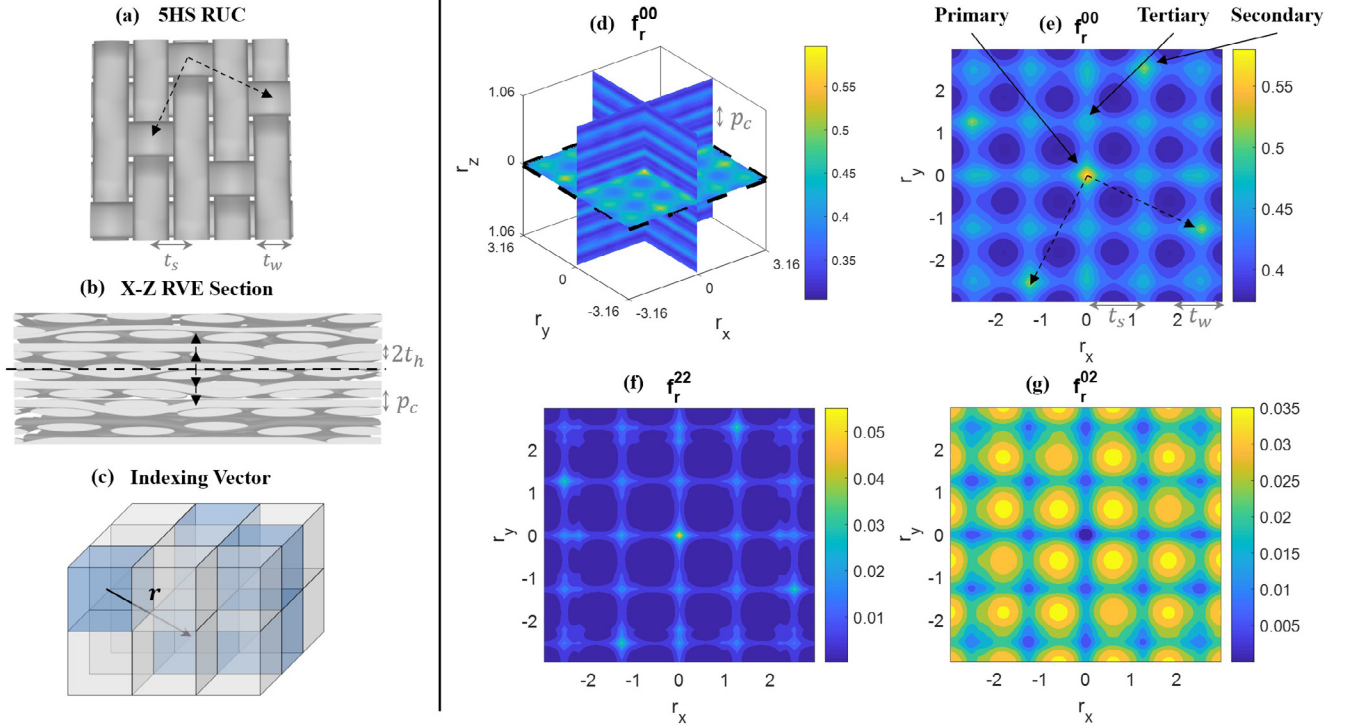
Fig. 2. Statistical quantification of an exemplary 5 harness-satin woven composite RVE with residual macro-porosity. (a) 5HS RUC with vectors demonstrating location of weft tows and tow width and spacing dimensions. (b) X-Z section of exemplar RVE with vectors highlighting central ply positions and dimensions of major tow axis and ply spacing. (c) Illustration of an indexing vector used to define a spatial correlation between voxels of the microstructure. (d) Auto-correlations of tow voxels (local state labelled as 0) shown only on the central planes of the full 3-D set. (e) Tow auto-correlations on the X-Y plane with features highlighted. (f) Pore auto-correlations, and (g) tow-pore cross-correlations on the X-Y plane. All vector dimensions are in mm.

Table 1
Ranges of values used in the present study for the different geometric parameters involved in the specification of RUC and the building of the woven RVE (see also Figs. 1, 2a, and b).

Microstructural Dimension	Minimum (µm)	Nominal (µm)	Maximum (µm)
Tow Major Axis -t_w	788	985	1182
Tow Minor Axis -t_h	92	115	138
Tow Spacing -t_s	880	1100	1320
Ply Spacing -p_c	257	275	292
Matrix Thickness -m_t	47	78	109

Table 2
Thermal conductivities of constituent materials at 20 °C.

Material	$k(W/mK)$
Matrix Conductivity -k_m	65.0
Tow Longitudinal -$k_{\parallel t}$	32.6
Tow Transverse -$k_{\perp t}$	5.1
Air -k_a	0.001

distribution of the generated structures over the selected input domain. For each microstructure, a FE mesh was created by converting each voxel to a linear hexahedral diffusive heat transfer element of type DC3D8 in ABAQUS [64]. Although voxel-based FE meshes result in an approximation of geometry in woven composites (e.g., the elliptical tows), they have been shown to provide good estimates of the effective properties [65,66]. The use of voxel-based meshing enables implementation of automated workflows for data generation and reduced-order model development.

3.2. Material properties

Material properties assigned to the tow, matrix, and pore constituents are summarized in Table 2 and assumed to be temperature independent. The matrix conductivity used is that of CVI-SiC from Murthy et al. (1999) [58]. The tows were defined as transversely isotropic. Their conductivity was determined based on the values reported by Pugilia et al. [25,26] for a CVI-SiC matrix and C tow, with a suitable scaling based on the values reported previously for Sylramic SiC fiber [58].

3.3. Model boundary conditions

A temperature gradient of 100 °C was imposed across the RVE by specifying the temperatures at nodes located on the opposite faces of the RVE. The resulting temperature field in the RVE was solved using the FE solver ABAQUS [64]. Effective conductivity was computed from the FE predicted heat flux (integrated over the relevant surface). This process was repeated in each of the three principal directions to establish the effective orthotropic thermal conductivity ($k = \{k_{11}, k_{22}, k_{33}\}$). Surfaces not oriented perpendicular to the direction of heat flow in each simulation were assumed to be adiabatic. Such insulated perpendicular faces to the heat flow direction as boundary conditions have been extensively used in the calculation of composite RVE effective thermal conductivity [23,26,67]. Periodic boundary conditions have also been extensively applied in the calculation of effective properties of RVEs [68–70]. In this work, the effect of the different boundary conditions on the predicted effective thermal conductivity was found to be very minimal (less than 1%).

4. Reduced-Order model

As the first step towards building the desired structure–property linkage, the set of 2-point statistics are computed for each of the

3125 generated microstructures. Each voxel in these RVEs is wholly assigned to one of the three local states, i.e., $h \in H = \{0, 1, 2\}$, with the value zero associated with the SiC Sylramic tows, the value one associated with the CVI-SiC matrix, and the value two associated with the residual macro-porosity. With these three local states, a total of nine sets of $f_r^{hh'}$ can be defined. Of these, three are referred as auto-correlations (when $h = h'$), and six are referred as cross-correlations (when $h \neq h'$). It has been shown previously that there exist a large number of interdependencies within this complete set of correlations [34,43]. It is necessary to include at least two of these sets [43] in representing the microstructure, although including three sets has resulted in a better representation of the microstructural features [45]. Building on this prior experience, the auto-correlations of the Sylramic tows (i.e., f_r^{00}), the auto-correlations of the pores (i.e., f_r^{22}), and the cross-correlation between the tows and pores (i.e., f_r^{02}) were utilized to represent the microstructures in this work. These specific correlations were selected as the tows constitute the highest volume fraction local state in the RVE across the range of microstructures studied, and the pore local state introduces the largest contrast in the constituent thermal conductivities.

Fig. 2 illustrates the 2-point statistics computed in this work. For any selected pair of local states h and h', $f_r^{hh'}$ can be visualized as a mapping between a specified vector r and a statistic that informs the probability of finding the specified local states separated by this vector in the microstructure (see Fig. 2c). Therefore, in the r-space, $f_r^{hh'}$ can be visualized as a 3-D contour. Fig. 2d shows three of the orthogonal planes for the auto-correlations of the tow voxels, f_r^{00}. The intersection of the three orthogonal planes identifies the location of $r = 0$. For auto-correlations value of the auto-correlation at $r = 0$ is the volume fraction of the corresponding local state (i.e., 1-point statistic). While f_r^{00} contains information for all possible values of r, only three planes are shown in Fig. 2d for visual clarity. The patterns observed in Fig. 2d arise naturally from the periodic nature of the woven composite. Looking closely at the contours in the X-Y plane of f_r^{00} in Fig. 2e, one can see that features correlate to specific details of the textile architecture of the 5HS weave pattern in Fig. 2a. For example, the pattern in Fig. 2e suggests that tow voxels are correlated to each other in X and Y directions, with the peak width and the inter-peak distances in both directions matching very well with the tow width, t_w, and tow spacing, t_s, respectively. The peak strengths decrease systematically as one moves away from the center peak corresponding to the vector $r = 0$. This reduction of the peak strengths is due to the stochasticity introduced during the generation of the RVEs. A set of secondary peaks are also seen in the auto-correlations on the X-Y plane that are weaker than the primary central peak, but significantly stronger than their neighboring peaks (labelled as tertiary peaks in Fig. 2e). These secondary peaks reflect the warp tow pattern in the 5HS RUC (see the dashed-line vectors on the 5HS RUC in Fig. 2a); for these specific vectors there is an increased correlation between tow voxels because of the increased density of the tow voxels at such intersections. The auto-correlations on the Y-Z and X-Z planes in Fig. 2d indicate that the tow voxels are significantly more correlated in X and Y directions, when compared to the Z direction. This is easily confirmed in the X-Z section of the RVE shown in Fig. 2b. Furthermore, the thickness of the bands and their spacing in the X-Z and Y-Z planes corresponded well with the ply spacing, p_c, and tow height, t_h, respectively.

While the discussion above was restricted to f_r^{00}, features seen in f_r^{22} and f_r^{02} can also be interpreted in a similar manner, and are shown in the X-Y plane in Fig. 2f and g, respectively. The auto-correlations of pore voxels exhibit similar features as those seen in the tow auto-correlations. Note that secondary peaks emerge in these plots due to the 5HS weave architecture, as described earlier. The high similarity between the auto-correlations of the tow voxels and the auto-

correlations of the pore voxels is because the locations of the pores are highly correlated to the locations of the tows, as seen in the pore-tow cross-correlations in Fig. 2g. In fact, the peaks in f_r^{02} identify the vectors that most commonly separate the pore voxels and the tow voxels. These vectors reflect that the pore centers are roughly at $\pm 45°$ to the tow cross-over points, and their distances to the tows correspond roughly to half the tow spacing in both X and Y directions. It can also be seen that the effects of the warp tow pattern are apparent in both f_r^{22} and f_r^{02}.

The discretization scheme used in this study resulted in a total of 100^3 voxels for each microstructure, and with the set of correlations considered in this work, the complete set of the descriptors for each microstructure results in a vector of size 3×100^3; this vector represents a flattened set of correlations comprising $\{f_r^{00}, f_r^{22}, f_r^{02}\}$. Utilizing this high-dimensional representation in downstream model building efforts would be computationally prohibitive, highlighting the need for dimensionality reduction. This was accomplished here using PCA. Scaling factors were applied to sets of correlations such that any one set of correlations did not dominate the transformation. In other words, the vectorized representation for this work was established as $\{f_r^{00}, \alpha_{22} f_r^{22}, \alpha_{02} f_r^{02}\}$ with $\alpha_{22} = 21.58$ and $\alpha_{02} = 8.44$; these values ensured that the variance of each of the three sets of scaled spatial correlations for the full dataset of 3125 microstructures were equivalent to one another.

The efficiency of PCA in providing low-dimensional representations can be visualized through the scree-plot shown in Fig. 3a, which depicts the variance explained by each of the first five PCs. It is seen that the first three components account for more than 98% of the variance in the dataset, signifying that the 2-point statistics of the full ensemble of microstructures can be effectively differentiated using three PC scores. Selected 2-D sections of the corresponding PC basis are shown in Fig. 3b. Inspecting the basis vector maps illustrates that PC1 is strongly correlated with the pore volume fraction, evidenced by the strong central peak in the basis vector map for f_r^{22}. Since each PC basis represents a weighted set of three million spatial statistics, it is very difficult to fully interpret all the features captured in each PC basis. In addition to the pore volume fractions, PC1 basis also captures many of the secondary and tertiary peaks in the tow auto-correlations in specified weighted combinations. The PC score simply reflects the strength of the corresponding basis in any given microstructure (i.e., the PC score of a microstructure is simply the inner product of the PC basis with its spatial correlations). PC2 then captures a weighted set of spatial correlations orthogonal to PC1 basis. In the present study, PC2 basis appears to capture the main details of the pore placement relative to the tows. In general, the higher order basis capture increasingly complex spatial patterns in the microstructure.

The distribution of the microstructure ensemble in PC space for the first three PCs can be seen in Fig. 4. Cross sections of four selected microstructures through the RVE ply stack have been highlighted to illustrate the descriptive ability of the three PC scores. Characteristic dimensions used during the generation of these four microstructures are also reported alongside the resulting pore volume fractions. These microstructures were selected because they lie on the boundaries in the PC subspace shown in Fig. 4. Specifically, comparing microstructures #2660 and #124, it can be confirmed that increasing values of PC1 correlate strongly with increasing levels of porosity. A comparison of microstructures #126 and #2600 indicates that increasing values of PC2 correlate strongly with an increase in tow spacing combined with a decrease in the matrix deposition thickness.

The structure–property models were built in this work using the first three PC scores of each microstructure as inputs. A sequential design of the numerical (FE) experiments was performed to generate the needed training data. The protocols used were described in Sections 2.3 and 2.4. A subset of 1000 microstructures were selected from the complete set of 3125 generated microstructures, and were desig-
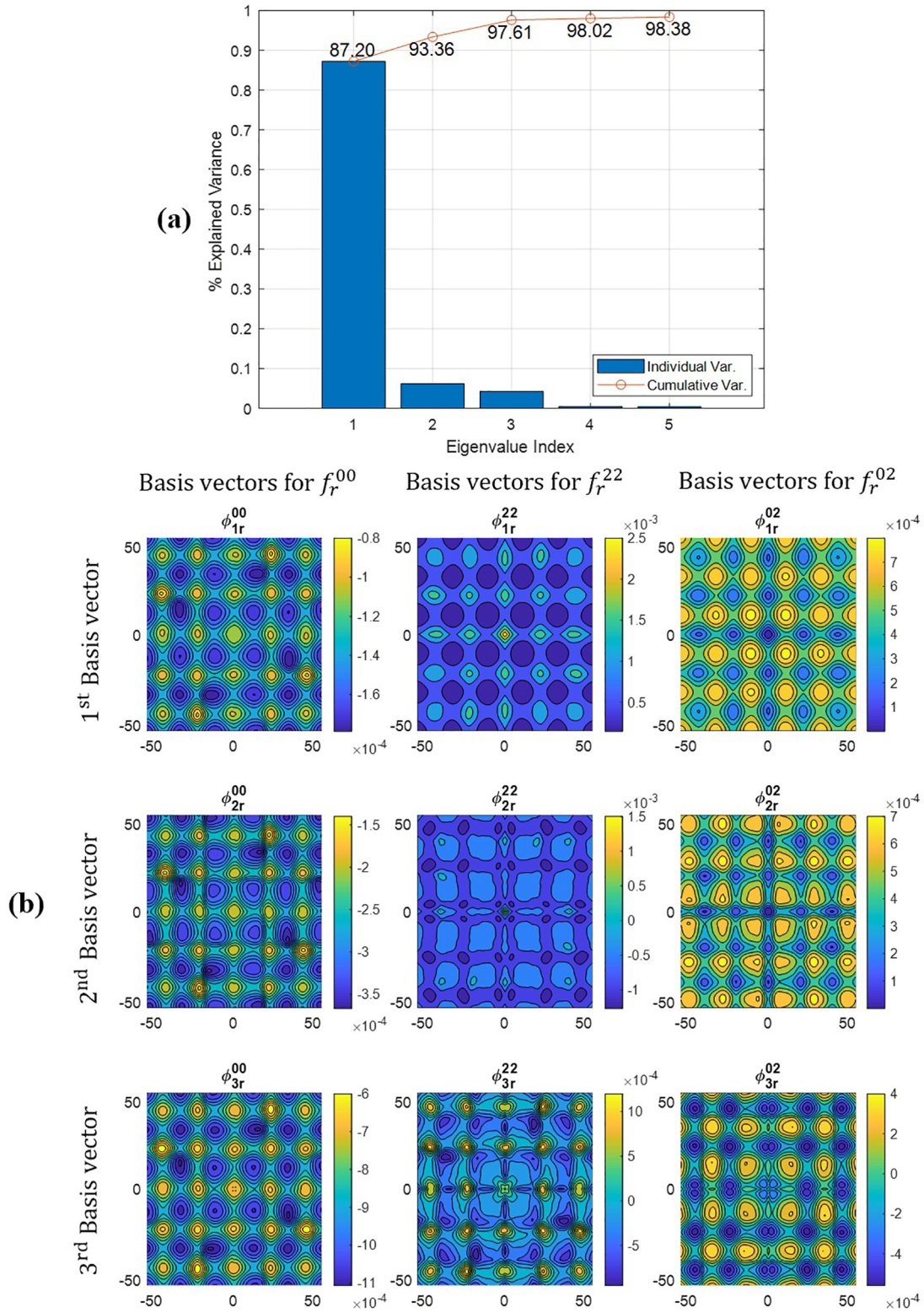
Fig. 3. (a) Scree-plot showing variance explained by the first 5 principal components. Bars represent the variance captured with each principal component, and the line tracks the cumulative variance with the increasing number of principal components. (b) Selected 2-D sections of the first three PC basis.

nated as the test set. This test set was consistently utilized to quantify the error in the GPR model being built, but the data for the test set was never exposed in the training of the GPR model. The test set was selected to ensure an uniform distribution in the three-dimensional PC-subspace through a nearest neighbor analysis [71]. The active learning methodology optimally selected new training points that best

informed the calibration of the GPR model using the concept of differential entropy score [46,47]. A schematic of the workflow utilized in this study is presented in Fig. 5, and can be summarized as follows:

1. Generate an ensemble of candidate training microstructures and the statistical quantification of each microstructure.
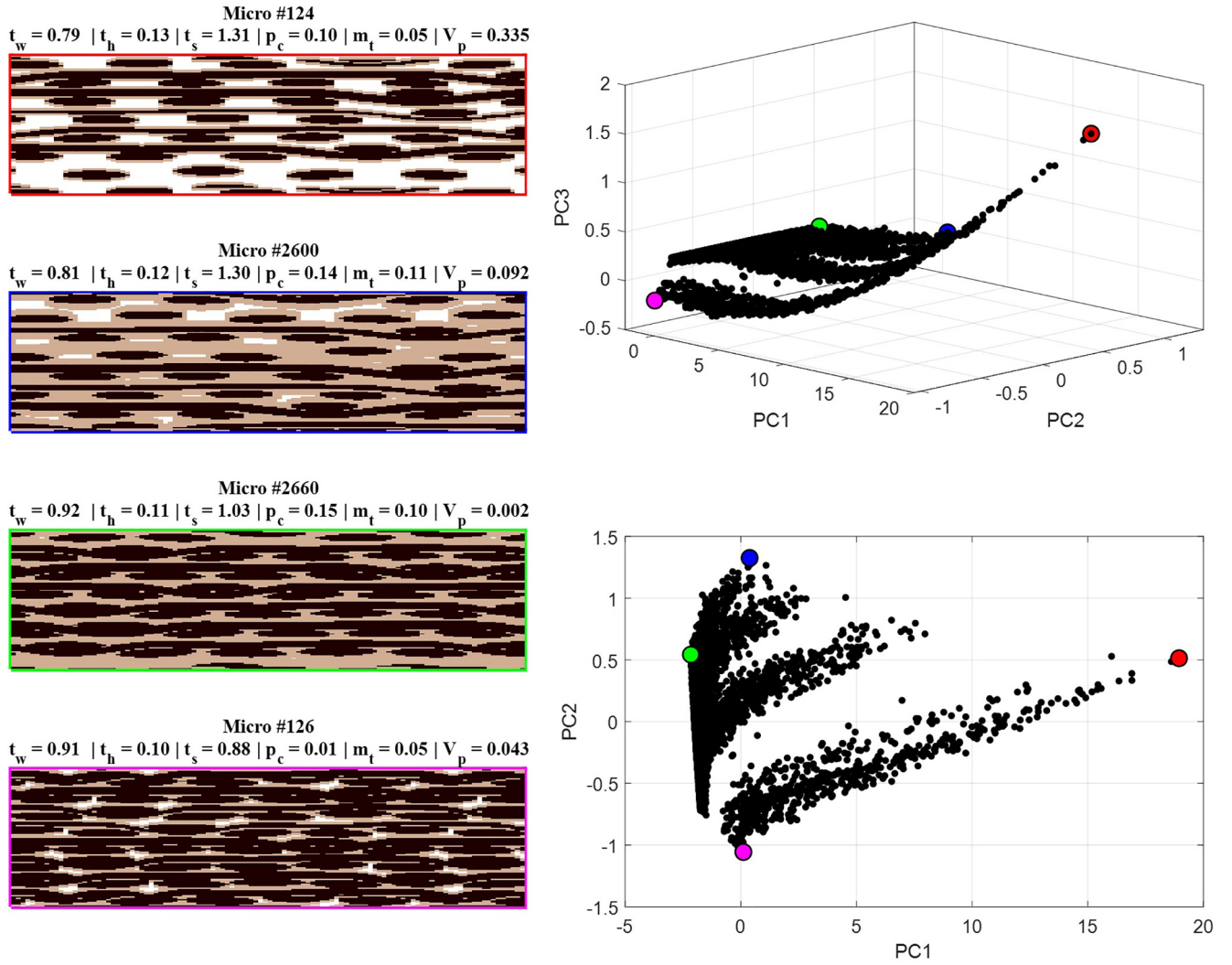
Micro #124
$t_w = 0.79 \mid t_h = 0.13 \mid t_s = 1.31 \mid p_c = 0.10 \mid m_t = 0.05 \mid V_p = 0.335$

Micro #2600
$t_w = 0.81 \mid t_h = 0.12 \mid t_s = 1.30 \mid p_c = 0.14 \mid m_t = 0.11 \mid V_p = 0.092$

Micro #2660
$t_w = 0.92 \mid t_h = 0.11 \mid t_s = 1.03 \mid p_c = 0.15 \mid m_t = 0.10 \mid V_p = 0.002$

Micro #126
$t_w = 0.91 \mid t_h = 0.10 \mid t_s = 0.88 \mid p_c = 0.01 \mid m_t = 0.05 \mid V_p = 0.043$

Fig. 4. Distribution of the complete set of 3125 microstructures shown with the first three PCs (top right) and in the first two PCs (bottom right). Select microstructures displayed on the left correspond to their color-coded locations in the PC space, indicating correlation of microstructural features to PC scores. Generation parameters (with dimensions in mm) and pore volume fraction are also listed for each microstructure.

2. Select a small initial subset of training microstructures. In this work, this group consisted of 5 microstructures chosen through the application of a sparse selection methodology [71] on the PC representation of the candidate training microstructures.
3. Perform FE analysis for each microstructure in the initialization subset, and use the three computed orthogonal effective thermal conductivities as targets for the three corresponding GPR models.
4. Estimate the expected value and its variance for all unevaluated (i.e., FE analyses has not been performed) candidate training microstructures using the GPR models (see Eq. (6)). The unevaluated candidate microstructure with the highest mean uncertainty across all three GPR models is identified as the next training point.
5. Evaluate the identified microstructure using the FE solver, i.e., its effective thermal conductivity is computed and added to the training set. The GPR models are then updated and Step 4 is repeated. At each update, the GPR model hyperparameters and the NMAE evaluated on the test set are recorded.
6. Repeat the GPR update cycles until the NMAE and GPR model hyperparameters exhibit convergence.

The predictive accuracy of the trained reduced-order models for k_{11} and k_{33} thermal conductivities are visualized in Fig. 6a and b, where the expected values are plotted for both training and test sets, with error bars showing one standard deviation of the predicted distribution. The parity plot for k_{22} is omitted as the results were similar to those for k_{11}, as expected. The converged values of NMAE for the test set were 0.81%, 0.76%, and 2.94% for k_{11}, k_{22}, and k_{33}, respectively. While the morphologies of the microstructures considered are quite varied spanning k_{33} values from 5.3 to 17.2 W/mK, experimentally observed results on the same material system in both 5HS and 8HS architectures fall within the microstructures considered. Values from Youngblood et al. (2004) show a 5HS ply layup with a value of 12.8 W/mK [55] and an 8HS and 5HS ply layup provided by Katoh et al. (2014) resulting in 8.5 W/mK and 10.6 W/mK, respectively [1]. Therefore, it is seen that the FE predicted values in this study are quite consistent with the related experimental values from prior literature.

The lower performance for the k_{33} model is attributed to the higher levels of contrast seen in the local values of thermal conductivity in the 3-direction compared to those seen in the 1–2 plane. For in-plane conductivity, the tow longitudinal conductivity is close to the matrix conductivity, and there exist several continuous pathways for thermal transport in both the matrix and the tow. Consequently, the effective in-plane thermal conductivity is fairly close to the longitudinal tow conductivity. For out-of-plane conductivity, such continuous pathways do not exist. The disc-shape of the pores with large projected areas in
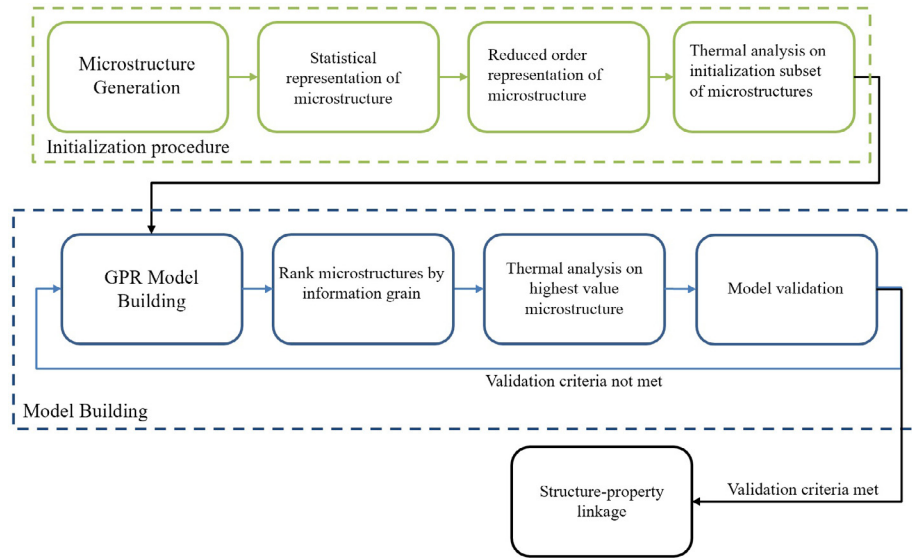
Fig. 5. Sequential design of experiments workflow for the optimized generation of the training data needed for model building. The green components represent the initialization procedure to select the first subset of microstructures for model building, while the blue components demonstrate the active learning methodology to select subsequent microstructure training points.
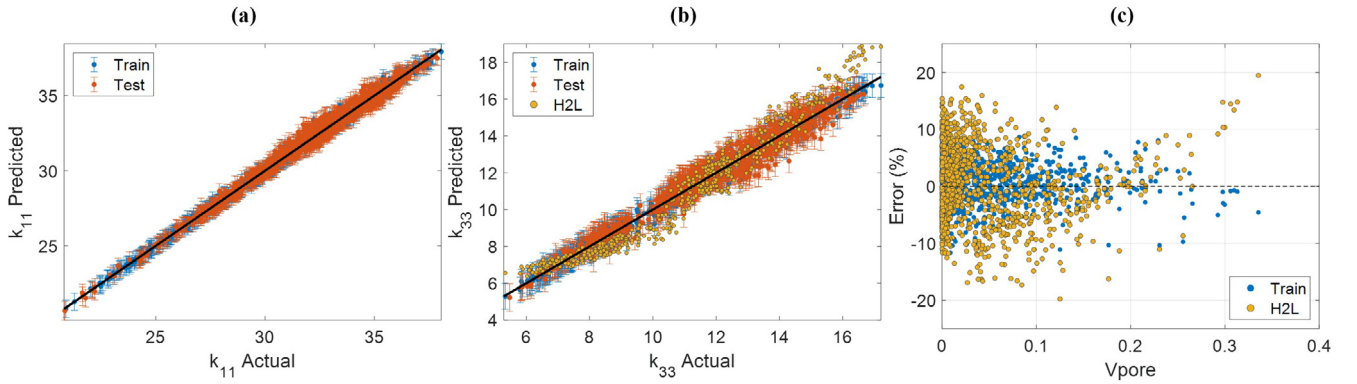

Fig. 6. Summary of the reduced-order models predictions of the effective thermal conductivity in the principal directions of the composite. All conductivity values are reported in W/mK. (a) Comparison of the GPR predictions with FE thermal analysis results for conductivity in the 1-direction. (b) Comparison of GPR and the calibrated H2L model predictions with FE thermal analysis results for conductivity in the 3-direction, along with experimental results for similar composite systems. (c) Percentage errors in GPR and the calibrated H2L model predictions for conductivity in the 3-direction as a function of pore volume fraction for the test set.

the 1–2 plane and the significantly lower transverse conductivity of the tow (compared to its longitudinal conductivity) significantly limit effective thermal transport in the 3-direction. Although a significant number of these microstructural features are captured in the 2-point statistics used in this work, it is also clear that some of these features require higher-order statistics (i.e., 3-point and higher). Recent work on microstructure-dependent effective transport properties such as permeability [72] has highlighted the importance of capturing interface voxels distinctly in computing the microstructure descriptors for improved model performance. The identification of interface voxels as distinct local states using local neighborhood information and then computing 2-point statistics is tantamount to computing certain subsets of higher-order spatial correlations. Therefore, for further improvement in k_{33} model performance, it might be necessary to systematically include higher-order microstructure statistics.

For the present study, improvements in model accuracy were sought through increasing the captured variance through the number of PC descriptors. It was observed that the inclusion of additional PC scores beyond the first three did not produce any significant improve-

ments in model accuracies. Table 3 summarizes the NMAE for models built with increasing numbers of PCs.

Despite the higher NMAE for the k_{33} model, the MKS framework performed significantly better than the currently used analytical H2L model in a benchmarking exercise. This comparison is shown in Fig. 6b, where the correction factors in the analytical H2L model ($P' = 0.33$, $\beta = 0.62$) were calibrated through a grid search on the same dataset utilized to train the GPR models. The poor performance of the H2L model is attributed to its reliance on simple constituent vol-

Table 3
Test set NMAEs for GPR models built with an increasing number of PC scores.

	k_{11} NMAE (%)	k_{22} NMAE (%)	k_{33} NMAE (%)
1 PC	5.82	5.80	13.02
2 PCs	0.83	0.77	3.12
3 PCs	0.81	0.76	2.94
4 PCs	0.79	0.74	2.95

ume fractions as microstructure descriptors. Fig. 6c plots the percent errors for both models on the test dataset, further demonstrating the improved accuracy of the GPR model over the H2L model, especially for woven architectures with both minimal and greater than 20% porosity. The test set NMAE for the H2L model was 5.73%, compared to 2.94% for the GPR model. It is remarkable that the non-parametric GPR model with only three PC scores is able to capture accurately the inherent link between the microstructure and its effective conductivity.

The efficiency of the sequential design strategy can be observed in the rapid decay and asymptotic behavior of the test set NMAE shown in Fig. 7a, where it becomes apparent that the addition of new microstructures past a training set of approximately 150 microstructures provides little additional benefit. This model building strategy is contrasted against random selection (repeated 20 times) for learning the GPR model for k_{33} in Fig. 7b. In random selection, each new training microstructure is selected randomly from the unevaluated pool of candidate microstructures. It is seen that the sequential design strategy significantly outperforms all of the random selection trials. The sequential design strategy reaches within 0.3% of the asymptotic NMAE value with about 100 training microstructures, while of the 20 random selection trials, only 2 attained comparable accuracy after about 500 training microstructures. Even then, the sequential design strategy results in more consistent and stable error improvement with each model update. Occasionally, random selection trials can be observed to outperform the sequential design methodology initially (at less than 50 training microstructures). However, this advantage is ephemeral, with the sequential design significantly outperforming the random trials by about a training size of 100 microstructures. Overall, our study indicates over 5X savings in the cost of producing training data with the use of sequential design strategies.

The convergence of the hyperparameters and parameters in the GPR model updates was tracked by evaluating their respective L_2 norm, and the angle change in their respective directions (utilizing a vector representation of the hyperparameters and parameters). The stabilization of the vector representation of the hyperparameters can be seen in Fig. 7c and d, and those of the parameters in Fig. 7e and f. It is seen from these figures that the mean function parameters converge much faster than the GPR interpolation hyperparameters, attesting to the efficiency of the proposed protocols. The converged values of GPR model hyperparameters and parameters corresponding to each model are summarized in Table 4. The mean function used in this work was formulated as a linear function with the three PCs, producing four parameters (β_1 to β_4) in Table 4 describing the mean linear form of the reduced-order model. As expected, β_1 (roughly corresponds to the average for the training set) for k_{11} and k_{22} are similar to each other, and higher than the corresponding value for k_{33}. β_2 and β_3 indicate a strong linear correlation with PC1 and PC2, respectively, while β_4 values indicate a weak linear correlation with PC3. The lower absolute values of the interpolation hyperparameters λ_i indicate a higher sensitivity of the input on the output. They also serve to control fluctuations from the mean function. For example, for all models, it is seen that the output exhibits relatively lower sensitivity to PC3. It is also observed that k_{33} exhibits a much higher sensitivity to microstructure descriptors when compared with k_{11} and k_{22}.

The data-driven approach presented in this work is highly scalable and extendable to different classes of microstructure morphologies, well beyond woven composites. The central advantage is that the models obtained using these protocols exhibit broader applicability compared to currently used empirical/analytical approaches. This work has demonstrated the applicability of this data-driven framework to a new material class, governing physics not previously considered,
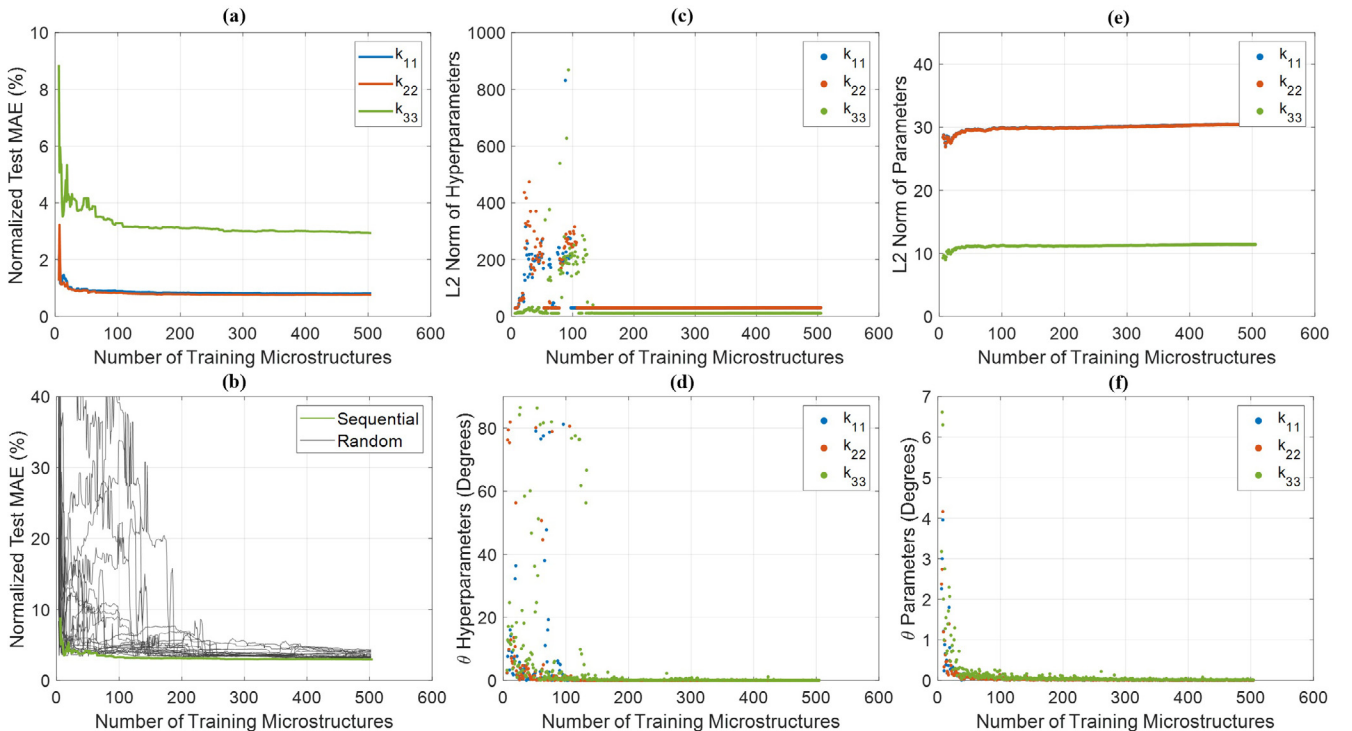

Fig. 7. Characteristics of model updates with sequential design of experiments strategy for identifying new training points. (a) The variation of test set NMAE for thermal conductivity in orthogonal directions with each model update. (b) Comparison with random selection for the model predicting conductivity in the 3-direction. (c) The variation of the L_2 norm of the GPR hyperparameter vector with each model update. (d) The variation of the angle change of the GPR hyperparameter vector with each model update. (e) The variation of the L_2 norm of the GPR mean function parameter vector with each model update. (f) The variation of the angle change of the GPR mean function parameter vector with each model update.

Table 4
Hyperparameter and parameter values for GPR models built in this study.

	λ_1	λ_2	λ_3	σ_f	β_1	β_2	β_3	β_4	σ_n
k_{11}	0.39	0.37	0.46	0.26	30.18	-3.16	2.36	0.03	0.24
k_{22}	0.38	0.46	0.72	0.25	30.15	-3.17	2.33	0.03	0.23
k_{33}	0.24	0.21	0.28	0.26	11.05	-2.02	1.95	0.01	0.30

along with demonstrating an ability for effective property prediction with an extremely large constituent contrast ratio.

5. Conclusions

A data-driven framework has been presented and demonstrated for the robust formulation of high-fidelity reduced-order models for the prediction of microstructure-sensitive effective anisotropic thermal conductivity of woven composite architectures. The proposed framework featurizes the material microstructure using PC representations of 2-point spatial correlations, and produces better models compared to current approaches for woven CMCs with residual porosity. Furthermore, the use of GPR offers a non-parametric approach to formulating the complex microstructure-property relationships in woven composites, while providing objective guidance in the optimal selection of the training points. The latter is of particular value when the cost of generating the training data is high. A sequential design strategy based on maximum variance was found to work well with the GPR modeling approaches in the present work, producing about 5X savings in the number of training data points.

6. Data availability

The raw data required to reproduce these findings are available to download from https://doi.org/10.17632/6x7jrn5fhm.1. The reduced-order model was built with the use of the built-in function in MATLAB [73] *fitrgp*. Details of which can be found at https://www.mathworks.com/help/stats/fitrgp.html.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

References

[1] Katoh Y, Ozawa K, Shih C, Nozawa T, Shinavski RJ, Hasegawa A, et al. Continuous SiC fiber, CVI SiC matrix composites for nuclear applications: Properties and irradiation effects. J Nucl Mater 2014;448:448–76. https://doi.org/10.1016/j.jnucmat.2013.06.040.

[2] DiCarlo JA. Advances in SiC/SiC Composites for Aero-Propulsion, 2014. https://doi.org/10.1002/9781118832998.ch7.

[3] GE Aviation. In aviation's material world, GE's CMCs unlock opportunity. The GE Aviation Blog | Aerospace & Flight News 2018. https://blog.geaviation.com/product/in-aviations-material-world-ges-cmcs-unlock/ (accessed May 18, 2020).

[4] Snead LL, Nozawa T, Katoh Y, Byun T-S, Kondo S, Petti DA. Handbook of SiC properties for fuel performance modeling. J Nucl Mater 2007;371:329–77. https://doi.org/10.1016/j.jnucmat.2007.05.016.

[5] Long AC. Design and manufacture of textile composites. Woodhead Publishing Limited; 2005.

[6] Crookston JJ, Long AC, Jones IA. A summary review of mechanical properties prediction methods for textile reinforced polymer composites. Proceedings of the IMechE 2005;219:91–109. https://doi.org/10.1243/146442005X10319.

[7] Bansal NP, Lamon J, editors. Ceramic Matrix Composites: Materials, Modeling and Technology. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2014. https://doi.org/10.1002/9781118832998.

[8] Maxwell CJA. Treatise on Electricity and Magnitism, vol. 1. Oxford Clareden Press; 1873.

[9] Rayleigh JS. On the influence of obstacles arranged in rectangular order upon the properties of a medium. Phi Mag 1892;481.

[10] Hasselman DPH, Johnson LF. Effective thermal conductivity of composites with interfacial thermal barrier resistance. J Compos Mater 1987. https://doi.org/10.1177/002199838702100602.

[11] Markworth AJ. The transverse thermal conductivity of a unidirectional fibre composite with fibre-matrix debonding: a calculation based on effective-medium theory. J Mater Sci Lett 1993;12:1487–9. https://doi.org/10.1007/BF00277073.

[12] Landauer R. The electrical resistance of binary metallic mixtures. J Appl Phys 1952;23:779–84. https://doi.org/10.1063/1.1702301.

[13] Lewis TB, Nielsen LE. Dynamic mechanical properties of particulate-filled composites. J Appl Polym Sci 1970;14:1449–71. https://doi.org/10.1002/app.1970.070140604.

[14] Fricke H. A Mathematical treatment of the electric conductivity and capacity of disperse systems I. The Electric Conductivity of a Suspension of Homogeneous Spheroids. Phys Rev 1924;24:575–87. https://doi.org/10.1103/PhysRev.24.575.

[15] Böhm HJ, Nogales S. Mori-Tanaka models for the thermal conductivity of composites with interfacial resistance and particle size distributions. Compos Sci Technol 2008;68:1181–7. https://doi.org/10.1016/j.compscitech.2007.06.009.

[16] Loeb AL. Thermal Conductivity: VIII, A theory of thermal conductivity of porous materials. J Am Ceram Soc 1954;37:96–9. https://doi.org/10.1111/j.1551-2916.1954.tb20107.x.

[17] Clayton W. Constituent and composite thermal conductivities of phenolic-carbon and phenolic- graphite ablators. 12th Structures, Structural Dynamics and Materials Conference. Am Inst Aeronaut Astronaut 1971. https://doi.org/10.2514/6.1971-380.

[18] Pilling MW, Yates B, Black MA, Tattersall P. The thermal conductivity of carbon fibre-reinforced composites. J Mater Sci 1979;14:1326–38. https://doi.org/10.1007/BF00549304.

[19] Yu H, Heider D, Advani S. Prediction of effective through-thickness thermal conductivity of woven fabric reinforced composites with embedded particles. Compos Struct 2015;127:132–40. https://doi.org/10.1016/j.compstruct.2015.03.015.

[20] Wang J, Carson JK, North MF, Cleland DJ. A new approach to modelling the effective thermal conductivity of heterogeneous materials. Int J Heat Mass Transf 2006;49:3075–83. https://doi.org/10.1016/j.ijheatmasstransfer.2006.02.007.

[21] Progelhof RC, Throne JL, Ruetsch RR. Methods for predicting the thermal conductivity of composite systems: A review. Polym Eng Sci 1976;16:615–25. https://doi.org/10.1002/pen.760160905.

[22] Marino GP. The porosity correction factor for the thermal conductivity of ceramic fuels. J Nucl Mater 1971;38:178–90. https://doi.org/10.1016/0022-3115(71)90041-9.

[23] Alghamdi A, Mummery P, Sheikh MA. Multi-scale 3D image-based modelling of a carbon/carbon composite. Modell Simul Mater Sci Eng 2013;21. https://doi.org/10.1088/0965-0393/21/8/085014.

[24] Ali J, Farooqi JK, Buckthorpe D, Cheyne A, Mummery P. Comparative study of predictive FE methods for mechanical properties of nuclear composites. J Nucl Mater 2009;383:247–53. https://doi.org/10.1016/j.jnucmat.2008.09.020.

[25] Del Puglia P, Sheikh MA, Hayhurst DR. Classification and quantification of initial porosity in a CMC laminate. Compos A Appl Sci Manuf 2004;35:223–30. https://doi.org/10.1016/j.compositesa.2003.09.026.

[26] Puglia PD, Sheikh MA, Hayhurst DR. Modelling the degradation of thermal transport in a CMC material due to three different classes of porosity. Modelling Simul Mater Sci Eng 2004;12:357–72. https://doi.org/10.1088/0965-0393/12/2/014.

[27] Guan K, Wu J, Cheng L. Modeling of Thermal Conductivity of CVI-Densified composites at fiber and bundle level. Materials 2016;9:1011. https://doi.org/10.3390/ma9121011.

[28] Zhou L, Sun X, Chen M, Zhu Y, Wu H. Multiscale modeling and theoretical prediction for the thermal conductivity of porous plain-woven carbonized silica/phenolic composites. Compos Struct 2019;215:278–88. https://doi.org/10.1016/j.compstruct.2019.02.053.

[29] Kalidindi SR. Hierarchical Materials Informatics: Novel Analytics for Materials Data. 1 edition. Amsterdam: Butterworth-Heinemann; 2015.

[30] Fast T, Niezgoda SR, Kalidindi SR. A new framework for computationally efficient structure–structure evolution linkages to facilitate high-fidelity scale bridging in multi-scale materials models. Acta Mater 2011;59:699–707. https://doi.org/10.1016/j.actamat.2010.10.008.

[31] Kalidindi S, Niezgoda S, Landi G, Vachhani S, Fast T. A Novel Framework for Building Materials Knowledge Systems. Computers Mater Continua 2010;17:103–25.

[32] Gray AG, Moore AW. "N-body" problems in statistical learning. Proceedings of the 13th International Conference on Neural Information Processing Systems, Denver, CO: MIT Press; 2000, p. 500–6.

[33] Torquato S. Random Heterogeneous Materials: Microstructure and Macroscopic Properties. New York: Springer-Verlag; 2002. https://doi.org/10.1007/978-1-4757-6355-3.

[34] Fullwood DT, Niezgoda SR, Adams BL, Kalidindi SR. Microstructure sensitive design for performance optimization. Prog Mater Sci 2010;55:477–562. https://doi.org/10.1016/j.pmatsci.2009.08.002.

[35] Cecen A, Dai H, Yabansu YC, Kalidindi SR, Song L. Material structure-property linkages using three-dimensional convolutional neural networks. Acta Mater 2018;146:76–84. https://doi.org/10.1016/j.actamat.2017.11.053.

[36] Gupta A, Cecen A, Goyal S, Singh AK, Kalidindi SR. Structure–property linkages using a data science approach: Application to a non-metallic inclusion/steel composite system. Acta Mater 2015;91:239–54. https://doi.org/10.1016/j.actamat.2015.02.045.

[37] Paulson NH, Priddy MW, McDowell DL, Kalidindi SR. Reduced-order structure-property linkages for polycrystalline microstructures based on 2-point statistics. Acta Mater 2017;129:428–38. https://doi.org/10.1016/j.actamat.2017.03.009.

[38] Latypov MI, Toth LS, Kalidindi SR. Materials knowledge system for nonlinear composites. Computer Methods in Applied Mechanics and Engineering 2019;346:180–96. https://doi.org/10.1016/j.cma.2018.11.034.

[39] Paulson NH, Priddy MW, McDowell DL, Kalidindi SR. Data-driven reduced-order models for rank-ordering the high cycle fatigue performance of polycrystalline microstructures. Mater Des 2018;154:170–83. https://doi.org/10.1016/j.matdes.2018.05.009.

[40] Paulson NH, Priddy MW, McDowell DL, Kalidindi SR. Reduced-order microstructure-sensitive protocols to rank-order the transition fatigue resistance of polycrystalline microstructures. Int J Fatigue 2019;119:1–10. https://doi.org/10.1016/j.ijfatigue.2018.09.011.

[41] Çeçen A, Fast T, Kumbur EC, Kalidindi SR. A data-driven approach to establishing microstructure–property relationships in porous transport layers of polymer electrolyte fuel cells. J Power Sources 2014;245:144–53. https://doi.org/10.1016/j.jpowsour.2013.06.100.

[42] Yabansu YC, Rehn V, Hötzer J, Nestler B, Kalidindi SR. Application of Gaussian process autoregressive models for capturing the time evolution of microstructure statistics from phase-field simulations for sintering of polycrystalline ceramics. Modell Simul Mater Sci Eng 2019;27. https://doi.org/10.1088/1361-651X/ab413e.

[43] Niezgoda SR, Fullwood DT, Kalidindi SR. Delineation of the space of 2-point correlations in a composite material system. Acta Mater 2008;56:5285–92. https://doi.org/10.1016/j.actamat.2008.07.005.

[44] Niezgoda SR, Kanjarla AK, Kalidindi SR. Novel microstructure quantification framework for databasing, visualization, and analysis of microstructure data. Integrat Mater 2013;2:54–80. https://doi.org/10.1186/2193-9772-2-3.

[45] Yabansu YC, Steinmetz P, Hötzer J, Kalidindi SR, Nestler B. Extraction of reduced-order process-structure linkages from phase-field simulations. Acta Mater 2017;124:182–94. https://doi.org/10.1016/j.actamat.2016.10.071.

[46] Herbrich R, Lawrence ND, Seeger M. Fast Sparse Gaussian Process Methods: The Informative Vector Machine. In: Becker S, Thrun S, Obermayer K, editors. Advances in Neural Information Processing Systems 15, MIT Press; 2003, p. 625–32.

[47] Rasmussen CE, Williams CKI. Gaussian processes for machine learning. Cambridge, Mass: MIT Press; 2006.

[48] Bishop CM. Pattern recognition and machine learning. New York: Springer; 2006.

[49] Gill PE, Murray W. Quasi-Newton methods for unconstrained optimization. IMA J Appl Math 1972;9:91–108. https://doi.org/10.1093/imamat/9.1.91.

[50] Broyden CG, Dennis JE, Moré JJ. On the local and superlinear convergence of quasi-newton methods. IMA J Appl Math 1973;12:223–45. https://doi.org/10.1093/imamat/12.3.223.

[51] Lewis DD, Catlett J. Heterogeneous Uncertainty Sampling for Supervised Learning. In: Cohen WW, Hirsh H, editors. Machine Learning Proceedings 1994, San Francisco (CA): Morgan Kaufmann; 1994, p. 148–56. https://doi.org/10.1016/B978-1-55860-335-6.50026-X.

[52] Wang H, Li J. Adaptive gaussian process approximation for bayesian inference with expensive likelihood functions. Neural Comput 2018;30:3072–94. https://doi.org/10.1162/neco_a_01127.

[53] Yue X, Wen Y, Hunt JH, Shi J. Active learning for gaussian process considering uncertainties with application to shape control of composite fuselage. IEEE Trans Automat Sci Eng 2020:1–11. https://doi.org/10.1109/TASE.2020.2990401.

[54] Ginsbourger D. Sequential Design of Computer Experiments. In: Balakrishnan N, Colton T, Everitt B, Piegorsch W, Ruggeri F, Teugels JL, editors. Wiley StatsRef: Statistics Reference Online, Chichester, UK: John Wiley & Sons, Ltd; 2018, p. 1–9. https://doi.org/10.1002/9781118445112.stat08124.

[55] Youngblood GE, Senor DJ, Jones RH. Modeling the Transverse Thermal Conductivity of 2-D SiC/SiC Composites Made with Woven Fabric. Fusion Science and Technology 2004;45:583–91. https://doi.org/10.13182/FST04-A533.

[56] Naik RA. Failure analysis of woven and braided fabric reinforced composites. J Compos Mater 1995;29:2334–63. https://doi.org/10.1177/002199839502901706.

[57] Lin H, Brown LP, Long AC. Modelling and simulating textile structures using TexGen. Adv Mater Res 2011;331:44–7. https://doi.org/10.4028/www.scientific.net/AMR.331.44.

[58] Murthy P, Mital S, DiCarlo A. Characterizing the Properties of a Woven SiC/SiC Composite Using W-CEMCAN. Computer Code. 1999.

[59] Lang J, DiCarlo J. Design Curve Generation for 3D SiC Fiber Architecture 2014.

[60] Morscher GN. Advanced Woven SiC/SiC Composites for High Temperature Applications 2013.

[61] Tai N-H, Chou T-W. Analytical modeling of chemical vapor infiltration in fabrication of ceramic composites. J Am Ceram Soc 1989;72:414–20. https://doi.org/10.1111/j.1151-2916.1989.tb06145.x.

[62] Rosenfeld A, Pfaltz JL. Sequential operations in digital picture processing. J ACM 1966;13:471–94. https://doi.org/10.1145/321356.321357.

[63] McKay MD, Beckman RJ, Conover WJ. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 1979;21:239–45. https://doi.org/10.2307/1268522.

[64] Abaqus/CAE User's Manual 2019:1174.

[65] Ma D, Manes A, Giglio M. The effect of mesh morphologies on the mesoscale Finite Element modelling of woven composites. Procedia Struct Integrity 2019;24:80–90. https://doi.org/10.1016/j.prostr.2020.02.007.

[66] Kim HJ, Swan CC. Voxel-based meshing and unit-cell analysis of textile composites. Int J Numer Meth Eng 2003;56:977–1006. https://doi.org/10.1002/nme.594.

[67] Sheikh MA, Taylor SC, Hayhurst DR, Taylor R. Microstructural finite-element modelling of a ceramic matrix composite to predict experimental measurements of its macro thermal properties. Modell Simul Mater Sci Eng 2000;9:7–23. https://doi.org/10.1088/0965-0393/9/1/302.

[68] Tsukrov I, Piat R, Novak J, Schnack E. Micromechanical Modeling of Porous Carbon/Carbon Composites. Mech Adv Mater Struct 2005;12:43–54. https://doi.org/10.1080/15376490490492034.

[69] Tian W, Qi L, Chao X, Liang J, Fu M. Periodic boundary condition and its numerical implementation algorithm for the evaluation of effective mechanical properties of the composites with complicated micro-structures. Compos B Eng 2019;162:1–10. https://doi.org/10.1016/j.compositesb.2018.10.053.

[70] Chao X, Qi L, Cheng J, Tian W, Zhang S, Li H. Numerical evaluation of the effect of pores on effective elastic properties of carbon/carbon composites. Compos Struct 2018;196:108–16. https://doi.org/10.1016/j.compstruct.2018.05.014.

[71] Altman NS. An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. Am Statist 1992;46:175–85. https://doi.org/10.1080/00031305.1992.10475879.

[72] Yabansu YC, Altschuh P, Hötzer J, Selzer M, Nestler B, Kalidindi SR. A digital workflow for learning the reduced-order structure-property linkages for permeability of porous membranes. Acta Mater 2020;195:668–80. https://doi.org/10.1016/j.actamat.2020.06.003.

[73] MATLAB (2019a). The MathWorks Inc.; 2019.