

**ORIGINAL ARTICLE**

# Statistical analysis of the influence of microstructure on damage in fibrous ceramic matrix composites

Jared Tracy<sup>1</sup> | Samantha Daly<sup>2</sup>

<sup>1</sup>Department of Materials Science and Engineering, The University of Michigan, Ann Arbor, MI, USA

<sup>2</sup>Department of Mechanical Engineering, The University of California at Santa Barbara, Santa Barbara, CA, USA

**Correspondence**

Samantha Daly  
Email: [samdaly@engineering.ucsb.edu](mailto:samdaly@engineering.ucsb.edu)

**Abstract**

The effect of microstructure on cracking was analyzed in a CMC using statistical methods. It was determined that the amounts of coating surrounding fibers and their dispersion within the matrix influenced where cracks evolved in transverse plies. Linear models predicted that maximum principal strains in transverse fiber coatings increased as (i) the fiber coating area increased and (ii) the length of matrix ligament between fibers decreased. Logistic models indicated that the likelihood of transverse fibers residing on a matrix crack increased as the (i) ratio of coating to filament decreased, (ii) distance between fibers decreased, or (iii) coating area increased.

**KEYWORDS**

ceramic matrix composites, cracks/cracking, damage, silicon carbide

## 1 | INTRODUCTION

Fiber-reinforced ceramic matrix composites (CMCs) exhibit an unusually good combination of stiffness, strength, and toughness, enabling their use in high-temperature structural applications. CMCs are tougher than their monolithic counterparts as a result of mechanisms introduced by the incorporation of fibers. Modern CMCs often contain stiff, strong fiber filaments encased in a weak coating, both of which are embedded in a stiff matrix. The combination of strong filament with weak coating permits crack deflection, widespread matrix cracking, and the delayed onset of catastrophic failure. Currently, limitations in material processing make the fabrication of perfectly uniform coating thicknesses and fiber distributions impractical. If these properties are found to strongly influence damage initiation and accumulation, an educated decision can be made about time and resource investment into processing refinements for improved microstructures.

Experimental investigations of matrix cracking phenomena in CMCs<sup>1-6</sup> have demonstrated that fiber inclusions greatly influence damage progression, and they have consequently prompted development of analytical and numerical models that incorporate fiber statistics into predictions of CMC fracture behavior. For example, Curtin<sup>7</sup> used fiber

strength (estimated using Weibull statistics) and fiber properties (radius, volume fraction, and sliding friction) to analytically derive the ultimate tensile strengths of continuous fiber-reinforced CMCs. The model did not account for randomness in the composite architecture, and its application to experimental data ranged from near perfect agreement to errors in predicted strengths as large as 30%. Calard and Lamon<sup>8</sup> used probabilistic functions to predict matrix cracking as a function of flaw distribution. They followed this with an assessment of scatter in failure strengths (strength distributions were measured using Weibull statistics) of SiC/SiC CMCs loaded in tension and flexure.<sup>9</sup> Their analysis suggested that heterogeneities in fiber arrangement and matrix crack distributions were responsible for deviations in the local stress state, which ultimately impacted failure strength variability. Calard and Lamon<sup>10</sup> used Bayesian statistics to determine failure probabilities by implementing finite element models to predict failure in representative volume elements of a CMC microstructure. The probability functions applied to each element were dependent on parameters derived from macroscopic tensile and flexural stress distributions, as opposed to local microstructural parameters.

Sorenson and Talreja<sup>11</sup> reported that residual stresses in the vicinity of fibers are influenced by the distance between

fibers. Using finite element models to simulate residual stress fields around fibers, they found that the radial compressive stress showed the most sensitivity to fiber distribution, noting that variability in stress increased in the vicinity of fibers in contact. This suggests that cracks emanating from fiber/coating/matrix interfaces may favor high-stress regions of clustered fibers. Recognizing the significance of fiber distribution to CMC performance, Yang<sup>12</sup> developed random volume elements (RVE) that characterized the random spatial arrangement of fibers with the intent of incorporating them into micro-mechanical damage models. However, fibers in the RVE contained no coatings and the results of their micro-mechanical modeling were not reported.

The analytical models discussed above used statistical distributions of an input variable, such as fiber strength, to predict a global response in the composite, such as cracking. We sought an alternate approach by experimentally measuring all input and response variables, and fitting statistical models to this data. Several variables may influence the fracture characteristics of CMCs, including residual stresses in the fibers and matrix, fiber size, coating thickness, fiber distribution, inclusions and defects in the matrix, and the local stress state. However, this work focused on factors pertaining to fiber/coating size and distribution to develop statistical damage models for CMCs, as these parameters are important, measureable with a high degree of accuracy, and controllable.

In this work, cross-ply CMCs comprised of silicon carbide fibers clad with boron nitride and encased in a silicon carbide matrix (abbreviated SiC/SiC) were loaded in tension in-situ in a scanning electron microscope (SEM). High-magnification micrographs of the damaged CMCs were captured and digital image correlation (DIC) was used to compute full-field strains. Microstructural and deformation data were cataloged on a fiber-by-fiber basis to include the following:

1. Principal strains in the fiber coatings oriented transverse to the loading direction
2. Fiber and coating sizes, proportions, and spatial distributions
3. Fiber proximity to matrix cracks

Unsupervised statistical methods (ie, statistical calculations generated without a response variable) were used to group the variables into categories with unique traits. Microstructural parameters pertaining to coating area and fiber distribution constituted a majority of the first principal component, indicating that they shared a common attribute in the overall pool of data. It was later found, through supervised statistical methods, that the common attribute was an affinity for accruing highly localized strains. Covariance matrices were used to assess relationships

between variable pairs. This analysis indicated that fiber distribution and coating area influenced the location of matrix cracks in transverse plies, but did not distinguish whether these relationships were causal or correlated by association with another variable.

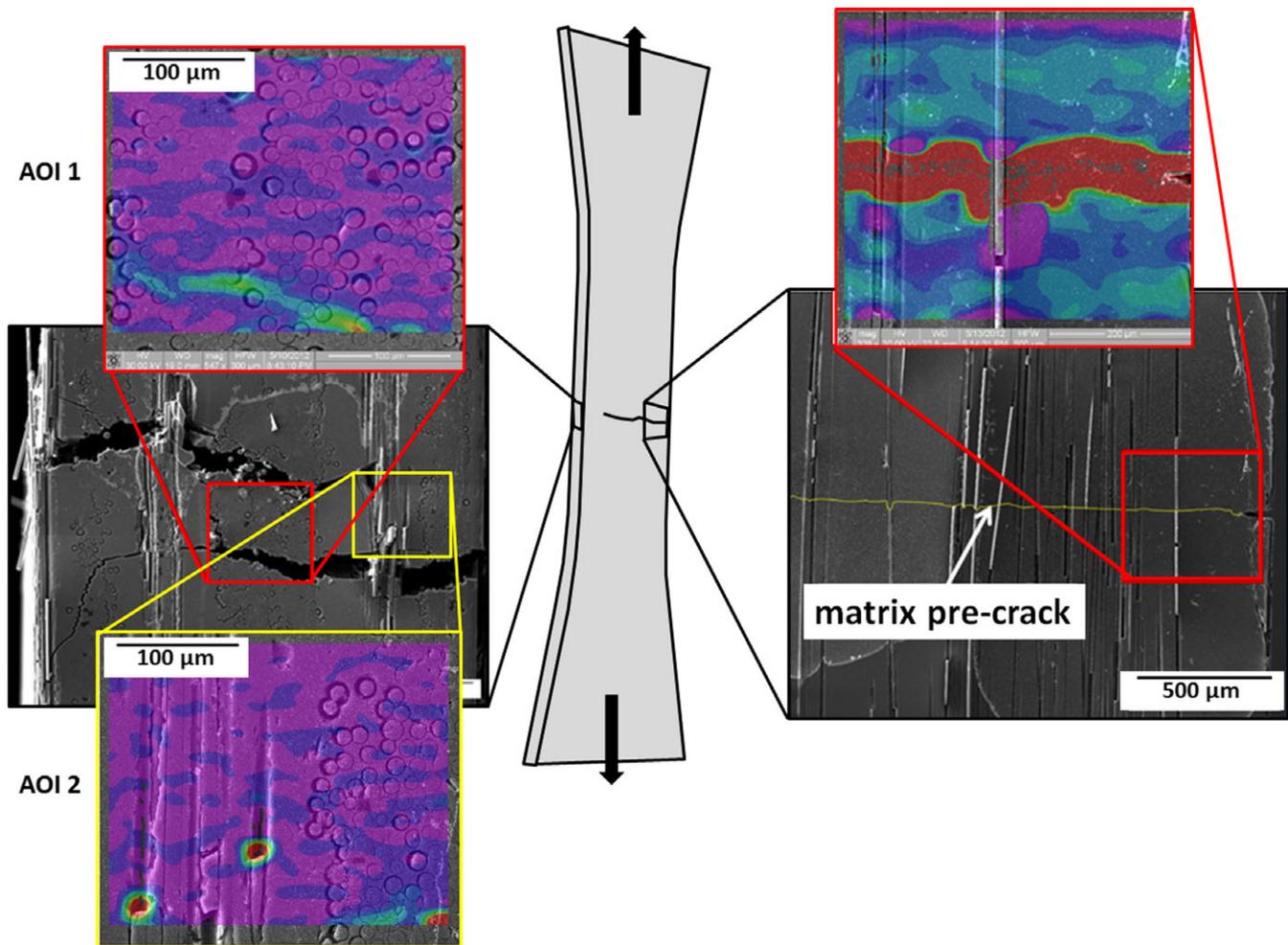
Linear and logistic models were developed to refine the findings from covariance analyses by identifying which parameters were causal to strain localization and crack propagation. Two variable selection algorithms, best-subsets and lasso, were used to determine significant predictor variables for incorporation into statistical models. The linear model predicted that the magnitude of maximum principal strains in transverse fiber coatings increased as (i) the fiber coating area increased and (ii) the length of matrix ligament between nearest-neighbor fibers decreased. The logistic model predicted the probability of a transverse fiber residing on a matrix crack path. It was found that this likelihood increased as the (i) ratio of coating to filament decreased, (ii) nearest-neighbor distance between fibers decreased, or (iii) coating area increased. However, logistic model findings were interpreted with caution, as they under-predicted the amount of fibers that resided on a crack path upon failure. Additionally, there were inherent errors associated with the experimental data, and methods to improve the accuracy of all models are discussed.

## 2 | EXPERIMENTAL METHODS

### 2.1 | Microscale data collection

All data were collected from cross-ply SiC/SiC CMC coupons loaded in uniaxial tension and imaged in-situ in a FEI Quanta 3D SEM. Three in-situ tensile tests were conducted using two edge-notched specimens and one smooth (ie, unnotched) specimen. The two edge-notched tests were on pre-cracked specimens with 300  $\mu\text{m}$  fields of view (FOV): one test yielded two sets of image data (Figure 1), whereas the other yielded one set. The smooth test specimen had no pre-crack and yielded one data set at a 200  $\mu\text{m}$  FOV. Specimens were loaded in SEM using a miniature tension/compression stage (Kammrath & Weiss) and were oriented to image the edge of the gage section. For the edge-notched tensile coupons, the areas of interest (AOIs) were situated along the edge opposite the notch, such that matrix cracks propagating through the gage width manifested in the AOI prior to final fracture.

All tensile coupons were [0/90/0/90]<sub>s</sub> with approximately 200  $\mu\text{m}$  thick individual lamina. The two edge-notched coupons were pre-loaded to initiate through thickness matrix cracks. For these two samples, 300  $\mu\text{m}$  FOV AOIs were positioned within transverse laminae. FOVs of this size were small enough to permit the use of DIC subset sizes similar to the fiber diameters, yet large



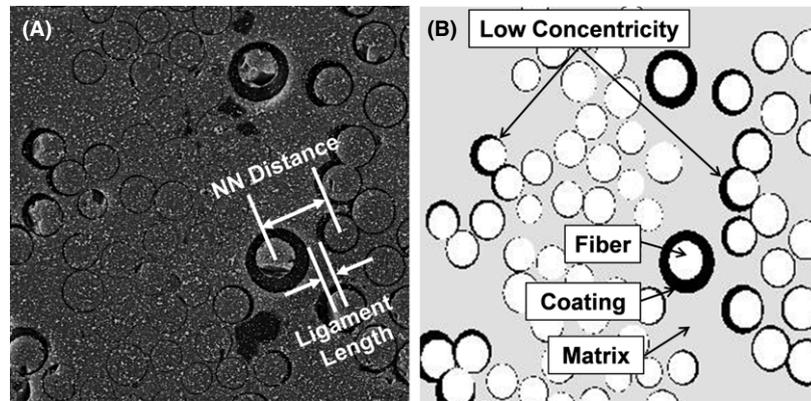
**FIGURE 1** A matrix pre-crack was initiated from a laser ablated edge-notch (right-hand side of diagram) and allowed to propagate partially through the gage width of the tensile coupon. After crack initiation, the sample was oriented for SEM imaging of the coupon edge opposite the notch (left-hand side). Images were captured at incremental loads in two locations along the coupon edge (AOI-1 and AOI-2), both in transverse plies. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

enough to sample a sufficient number of fibers for statistical analysis. In the absence of an automated system for capturing images along the entirety of the edge of the gage section, notches were needed to localize matrix crack propagation and improve the likelihood of capturing propagating cracks in a predetermined AOI. An illustration of data collection on a notched tensile coupon is provided in Figure 1. Two areas of interest (AOI-1 and AOI-2) were selected for image capture during loading, where AOI-1 captured a larger portion of the propagating matrix crack. A 200  $\mu\text{m}$  FOV was used for the smooth bar coupon, which sampled fewer fibers but provided greater spatial resolution in full-field deformation measurements.

SEM images were taken at discrete load increments. The tests were run in displacement control, which caused the samples to relax as much as 15 N. Consequently, image capture, which itself took approximately 90 seconds, began after there was no further drop in load (between 30 and

60 seconds). Two images were collected at each load increment; the latter images were used for correcting spatial and temporal image distortions inherent in SEM imaging.<sup>13-16</sup> Note that imaging the entire coupon edge is ideal as it provides the largest possible pool of statistical observations, but was impractical due to the rastering nature of SEM imaging. The notched specimen tensile test depicted in Figure 1 fortunately had matrix cracks manifest in both AOIs selected for imaging, whereas only one AOI intercepted a matrix crack during the test conducted on the second notched coupon.

In each of the four data sets, fibers and their coatings were segmented, binned, and indexed. Principal, longitudinal, transverse, and shear strains were calculated pixel-by-pixel in each micrograph using DIC software.<sup>17</sup> The strains in each fiber and coating were averaged and indexed. Additionally, the following microstructural variables were measured and indexed; an illustrative example of each variable is provided in Figure 2:



**FIGURE 2** Microstructural parameters used for statistical modeling. (A) Nearest-neighbor distance was measured between fiber centroids. Ligament length was a measure of the amount of matrix between nearest-neighbor fibers. (B) Constituent bins and examples of low concentricity fibers are indicated by arrows.

1. Location of fiber and coating centroids
2. Area of fiber and coating
3. Ratio of coating area to fiber area
4. Fiber-coating concentricity (ie, the amount of collinearity between fiber and coating centroids)
5. Fiber-to-fiber nearest-neighbor distance (measured from centroids)
6. Length of matrix ligament between nearest-neighbor fibers
7. Whether or not any part of the fiber or its coating resided on a crack path upon final fracture (note: these data were only collected for the pre-cracked coupons)
8. Averaged strains in fiber and coating (minimum and maximum principal, longitudinal, transverse, and shear strains)

To ensure that each data set was a representative sample of the total population of transverse fibers, the distribution of fiber data collected from multiple fiber tows along a transverse lamina was compared with those of each data set. The data sets in Figure 1 were comprised of 183 fibers in AOI-1 and 89 fibers in AOI-2. Each fiber constituted an experimental observation for statistical modeling purposes. The data set for the second edge-notched coupon contained 101 observations while that of the smooth bar sample contained 86 observations. Within the gage section of a tensile coupon, however, there were thousands of transverse fibers. As shown in Figure 3, the individual data sets contained similar distributions to the lamina.

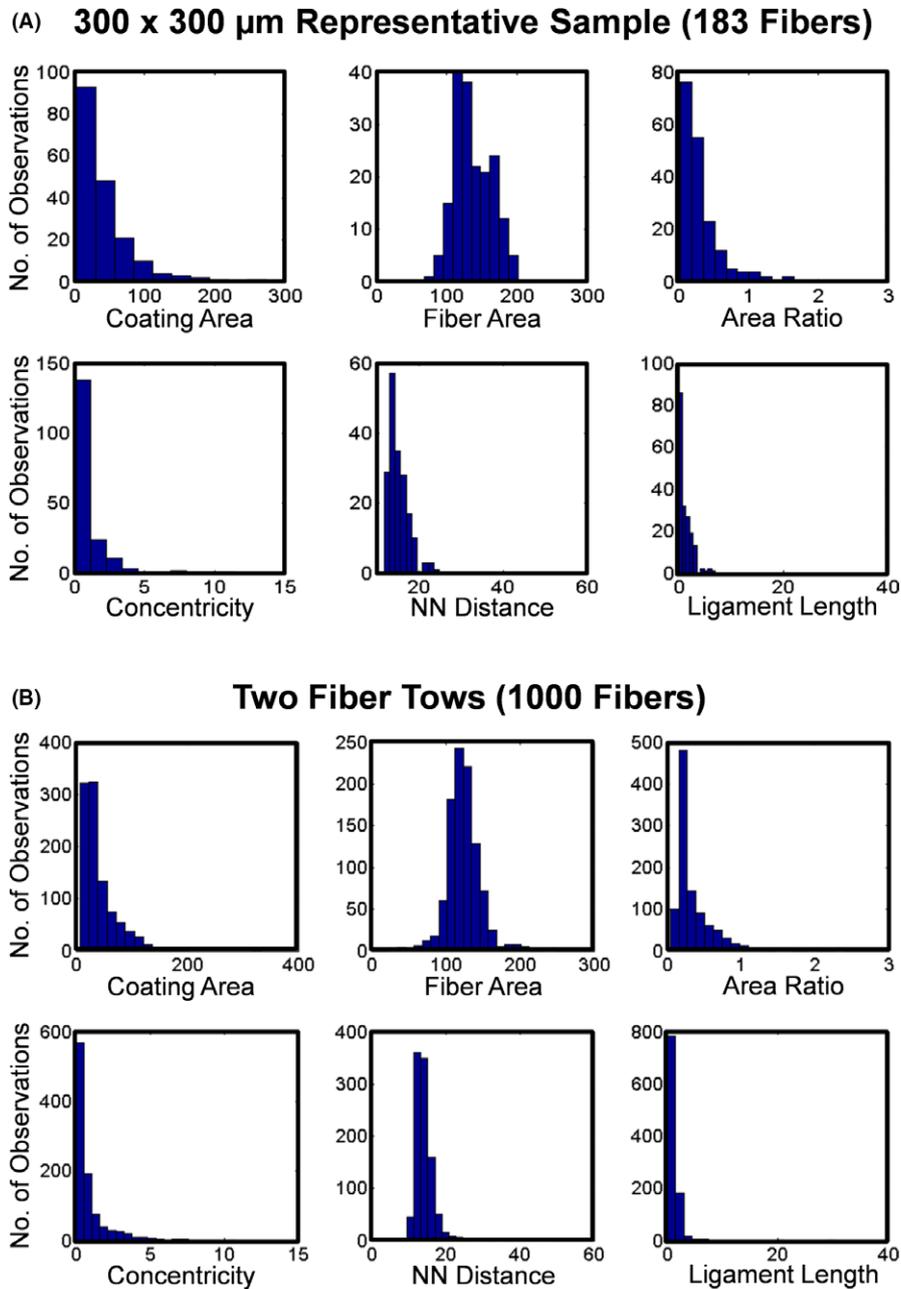
## 2.2 | Unsupervised statistical learning

Scatter plots of microstructural variables, collected from the notched-coupon test of Figure 1, provide a broad-based graphical depiction of correlations between variable pairs. This is evident in Figure 4, where for example, coating

area (“coatarea”) and area ratio (“ratio”) have a clear and expected positive correlation. Other correlations, such as that between area ratio (“ratio”) and nearest-neighbor distance (“nndist”), are not as clear.

Covariance matrices were used to determine the magnitude and direction of correlations between all microstructural and mechanical variables. Covariance matrices provide a quantitative measure of correlations between variable pairs, summarized in a single metric; an example and further explanation of covariance matrices is provided in Appendix 1. As previously mentioned, the data for statistical modeling were sampled fiber-by-fiber. For the covariance analysis, rather than sample data fiber-by-fiber, data were sampled point-by-point where each point corresponded to a DIC data node. This was so that the analysis would incorporate deformation data in the SiC matrix in addition to the fibers and their coatings. Covariance matrix input data were binned according to the type of constituent from which it was collected. For example, for any data point residing within a SiC carbide fiber, the corresponding predictor and response values for that point were consolidated into a data matrix that contained data only for SiC fibers. The same concept was applied for the fiber coatings and the SiC matrix. Covariance matrices were constructed for each load increment over which DIC data were collected.

When a data set contains more than one predictor, visualizing the contribution of each predictor to a response becomes difficult. Principal component analysis (PCA) provides a means of graphically representing a data set that contains several dimensions (such as the microstructural variables identified in Figure 4) in fewer dimensions. In PCA, eigenvalue decomposition is used to transform a data set into a space comprised of principal components, where the number of components is equal to the number of predictors in the data set, and each component is a linear

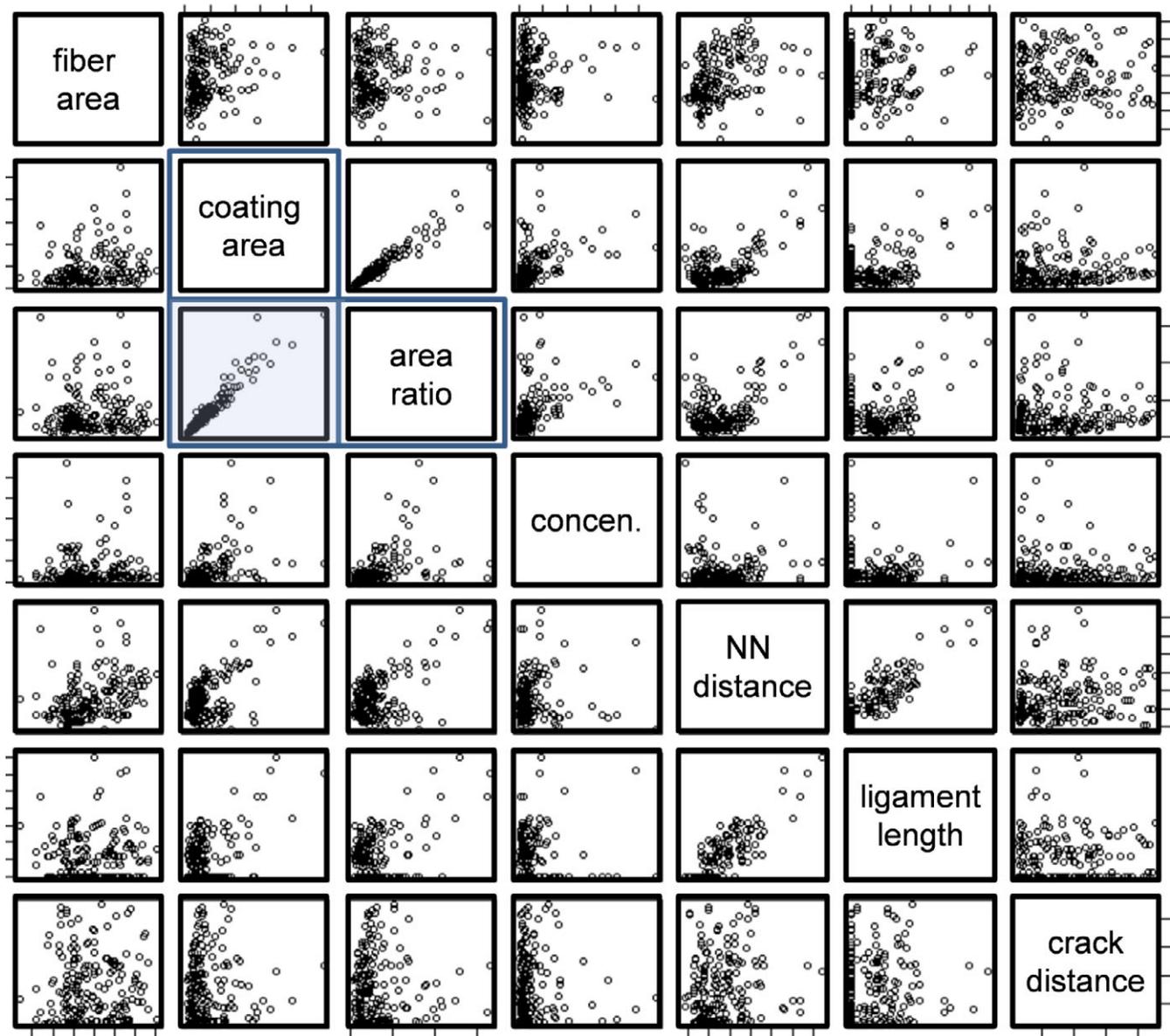


**FIGURE 3** Distribution of microstructural parameters for the cases of (A) a  $300 \times 300 \mu\text{m}$  representative area element and (B) two fiber tows (each tow containing roughly 500 fibers). The representative area contains an equivalent distribution of fiber parameters as that of multiple tows. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

combination of all the predictors. The first component accounts for the greatest amount of variability in the data set. Each subsequent component accounts for a lesser amount of variability in the data set and is a linear combination of variables that are uncorrelated with the first principal component. Further details of method can be found in.<sup>18</sup>

A majority of data variance was accounted for in just a few principal components. Using the data set from AOI 1 in Figure 1 as an example, given an input of six microstructural predictors, the first three principal

components accounted for approximately 90% of the variability in the data set. When plotted in principal space, the predictors clustered into unique groups. Binary plots of the first three principal components aided in grouping variables as illustrated in Figure 5. The red crosses represent scores for the respective principal components (indicated by the axes) of each fiber. The blue lines represent the principal component loading for each predictor-type, as projected onto the two principal components indicated by the axes. Looking along the 1<sup>st</sup> principal component axis in Figure 5(A), the variables weighted most heavily are coating



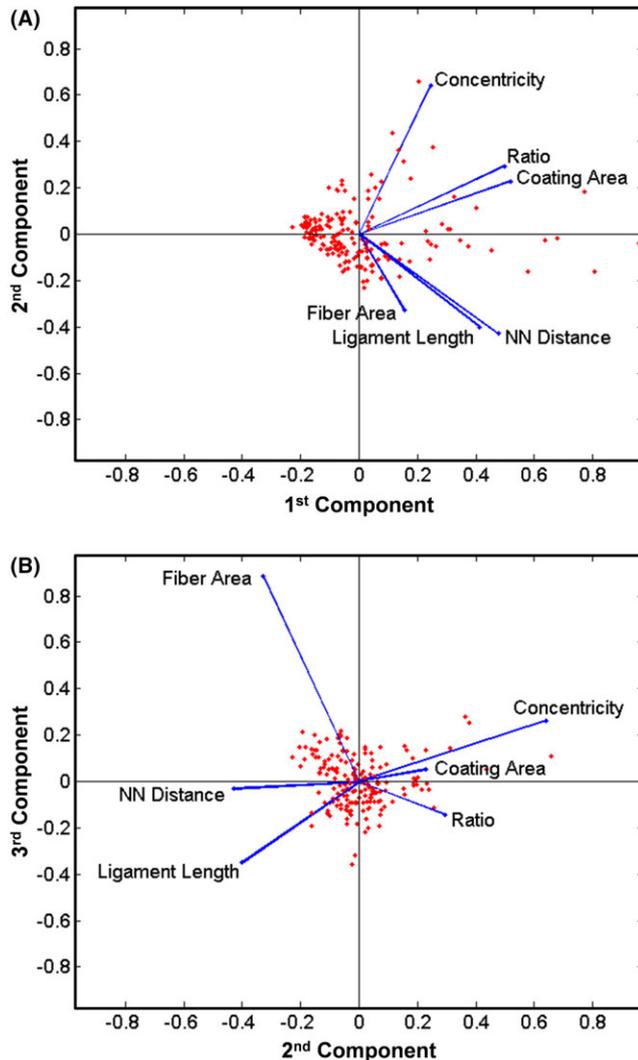
**FIGURE 4** Example of scatter plots of microstructural variables. Plot of coating area and ratio (highlighted in a blue box) shows a clear positive correlation. The trend is not as clear for many other variable pairs, hence the need for covariance analysis. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

area, coating-fiber area ratio, nearest-neighbor distance, and nearest-neighbor ligament length. Thus, the first component accounts for coating size and fiber separation. It also accounts for 50% of the variability in the data set. As these fibers are clumped together, they are correlated with each other. Observing the 2nd and 3rd component axes in Figure 5(B), these components correspond to measures of fiber-coating concentricity and fiber area, respectively. Note that while the PCA biplots do not represent correlations between individual variables, they do indicate a common attribute that accounts for the greatest variability in the data set, that is, there is something about these variables that is relevant. The nature of their relevancy must be determined through more pointed statistical methods such as supervised

modeling. The supervised statistical models of Section 3 will show that these variables likely influence localized damage in the composite.

### 2.3 | Statistical models

Linear and logistic models were developed to assess the influence of microstructural parameters of CMCs on (i) localized strain in fiber coatings and (ii) the probability of transverse fibers residing on a matrix cracks. The linear models were developed using all four data sets. The smooth bar data set (with 86 observations) contained the least noisy DIC data and was used for training the model. The logistic models did not incorporate DIC data, so the



**FIGURE 5** Biplots of principal component analysis of predictor variables. The first two principal components are plotted against each other in (A). The second and third components are plotted in (B). Red markers indicate scores for individual observations. Blue vectors indicate position of individual predictors in principal space. Area ratio, coating area, ligament length, and nearest-neighbor distance have the largest values along the first principal component. This indicates that they share a common attribute that accounts for the greatest variability in the data set. Further statistical analysis (see Section 2.3) is needed to identify the nature of the common attribute. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

data set from AOI 1 of Figure 1 (with 183 observations) was used for training.

The linear regression models constituted the least squares solution to equations of the form:<sup>18</sup>

$$y(x) = c_0 + c_1x_1 + c_2x_2 + \dots + c_px_p,$$

where  $p$  is the number of predictors,  $x_p$ ,  $y$  is the predicted response (maximum principal strain in the fiber coating), and  $c_p$  are coefficients. Logistic regressions were of the form:<sup>18</sup>

$$P(x) = \frac{e^{c_0+c_1x_1+\dots+c_px_p}}{1 + e^{c_0+c_1x_1+\dots+c_px_p}}$$

where  $P(x)$  represents the probability of an observation residing on a crack as a function of its predictor values.

Best-subsets selection and lasso shrinkage algorithms adopted from<sup>18</sup> were used to filter out unnecessary predictors. The best-subsets algorithm proceeded as follows. Given a data set of  $p$  predictors:

1. For all  $k=1:p$  combinations of predictors, fit a regression to the appropriate response such that  $2^p$  models are generated. For logistic regressions, the response was binary; if a fiber resided on a crack path, its response value was designated as 1, otherwise the response was designated as 0. For linear regression, the response was the maximum principal strain in a fiber coating (averaged over all DIC data points within the coating).
2. Among all models consisting of  $k$  predictors, identify the model with the largest  $R^2$  value (for linear regressions) or lowest deviance (for logistic regressions).
3. Among the  $p$  models that remain, select the model with the largest adjusted  $R^2$  value. As more predictors are added to a model,  $R^2$  and deviance may improve only as a consequence of adding more predictors. The adjusted  $R^2$  criterion adds a small penalty for having more predictors so that the largest predictor subset will not necessarily be the best subset.

The algorithm was run using a six-predictor data set comprised of microstructural variables (fiber area, coating area, coating-to-fiber area ratio, concentricity, centroid-to-centroid fiber nearest-neighbor distance, and nearest-neighbor matrix ligament length) and their inverses. For both the linear and logistic models, the most significant subsets were associated with coating properties and fiber separation, coinciding with the 1st and 2nd principal component predictors of Section 2.2. Higher-order variables (eg, quadratic terms and inverses) were excluded from the best subsets for two reasons: (i) their inclusion resulted in impractical processing times and (ii) plots of predicted responses vs residuals did not indicate that higher-order terms would improve the fits. The influence of quadratic terms, however, was investigated using an alternative means of subset selection—the lasso method.

Lasso variable selection suggested that only the variables associated with fiber separation (ie, nearest-neighbor distance and matrix ligament length) were relevant in predicting the response. Like best subsets, the lasso method performs variable selection. However, it does so by fitting a model with all variables, then shrinking the coefficients of insignificant variables to zero using a tuning

parameter.<sup>18</sup> Significant variables are those that have non-zero coefficients after the tuning parameter has been optimized to produce a fit of minimum deviance. A shortfall of this method is that it often selects highly collinear variables. When this occurred, the variables with largest collinearity, as measured via the variance inflation factor (VIF), were removed from the subset. This resulted in degradation to the  $R^2$  and  $P$  values of the model. Results are discussed in Section 3.

### 3 | RESULTS

#### 3.1 | Results of unsupervised models

Consolidating covariance data from all bins (fiber, coating, and matrix) for each of the four data sets resulted in the following findings:

1. Fibers with thick coatings exhibited low fiber-filament concentricity;
2. Maximum principal strains were largest in fibers that were closely spaced and exhibited low concentricity;
3. Crack paths favored fibers that were closely spaced, but fibers with thicker coatings on the crack path often spread out rather than clustered.

When interpreting these results, two important points must be noted: (i) covariance matrices do not account for interactions between multiple variables (ie, a positive correlation between Variable 1 and Variable 2 may be an indirect artifact of Variable 3, but covariance matrices alone do not provide that information) and (ii) conservative constraints were applied to draw each conclusion; if an observation was not consistent from one bin to another and from one test to another, then it was not included in the list of conclusions. In order to better understand how microstructural variables influenced fracture, linear and logistic statistical models were developed. They are presented in Sections 3.2 and 3.3.

#### 3.2 | Linear model results

A linear statistical model was developed to determine the influence of microstructural parameters on strain localization. For the stiff components of the CMC (ie, the fiber filaments and matrix), it was assumed that high-strain DIC measurements indicated cracking. Using maximum principal strain to indirectly observe cracks in transverse plies is a reasonable assumption and has been previously demonstrated.<sup>19</sup> The strain data were modestly filtered (45-pixel strain filter window given a 29-pixel subset size and 1-pixel step size) to reduce noise. The impact of noise on model results is discussed in Section 4.

**TABLE 1** Linear model results

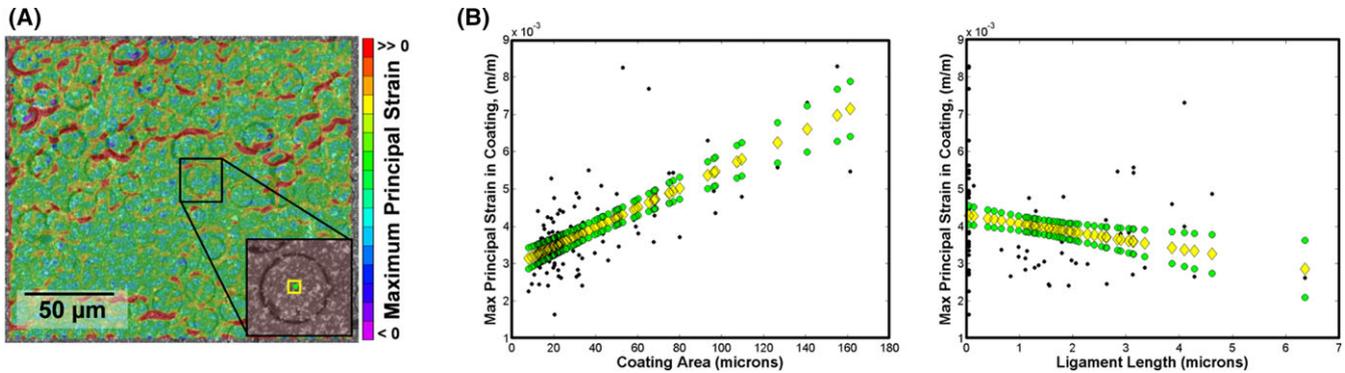
Model	Best subsets	Lasso shrinkage
$R$ -squared	.49	.51
Adjusted $R$ -squared	.48	.47
$P$ value	0	0
RMS error	908 $\mu\epsilon$	913 $\mu\epsilon$

	Variable	$P$ value	Variable	$P$ value
Significant variables	Coating area	.00	Coating area	.10
	Ligament length	.00	Area ratio [coating:fiber]	.84
			Ligament length	.29
			Inverse coating area	.39
			Inverse ligament length	.31
			Inverse squared coating area	.28

Best-subsets and lasso algorithms were used to reduce the number of predictors used for modeling and prevent overfitting of the data. Both methods selected variables pertaining to coating thickness and fiber dispersion. Recall from Section 2.3 that the best-subsets method selected variables by omitting those that degraded the adjusted  $R^2$  value of the linear fit, while the lasso method fit a model to all variables but reduced the weights of insignificant variables to zero. It was determined that a linear fit of first-order terms is sufficient. The results of both methods are listed in Table 1. The best-subsets model only included first-order terms, which resulted in a goodness of fit (adjusted  $R^2$ ) of .48. The lasso model incorporated inverse and higher-order terms, but this did not improve the goodness of fit (adjusted  $R^2$  was .47). The scatter plot of residuals vs fitted values for the best-subsets model showed no distinct trend to suggest anything other than a linear fit to the data points.

Coating area and fiber distribution impact strain localization and should be considered critical factors in optimizing a CMC microstructure. Plots of the best-subsets linear model applied to training data are shown in Figure 6. To examine the effect of each variable, maximum principal strain (the failure criterion for matrix cracking chosen in this study) was plotted as a function of each predictor and the other predictor was held constant at its mean value. Though the fracture strain of the bulk siliconized-silicon carbide matrix is not known, for reference the fracture strain of reaction bonded silicon carbide is  $\sim 1200 \mu\epsilon$ .<sup>20</sup> Figure 6(A) shows an overlay of the principal strain fields on the microstructure. Inset is a yellow square indicating the DIC subset size relative to a fiber. Figure 6(B) shows the plots of each best-subset predictor. The red squares are



**FIGURE 6** Statistical model developed using strain fields from smooth bar tensile coupon as training data suggest high-strain localizations are prevalent in fibers with thick coatings spaced close together. Deformation fields are plotted on top of micrograph in (A). Predicted strain responses as a function of predictor values are shown in (B). Fitted strains responses are regressed against best-subset predictor (yellow diamonds) holding other predictor constant. 95% confidence interval is plotted in green circles. Black dots indicate experimentally observed response. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

experimental data points. Black diamonds indicate the fitted response and are bound by magenta diamonds depicting the 95% confidence interval. The linear model suggests that localized maximum principal strains (ie, cracking in the CMC) increase as the fiber coating area increases and/or the matrix ligament length between nearest neighbors decreases (which may be indicative of clustering). Results for the test data are shown in Figures 7 and 8, which indicate that the trends observed all three sets of test data match with that of the training data.

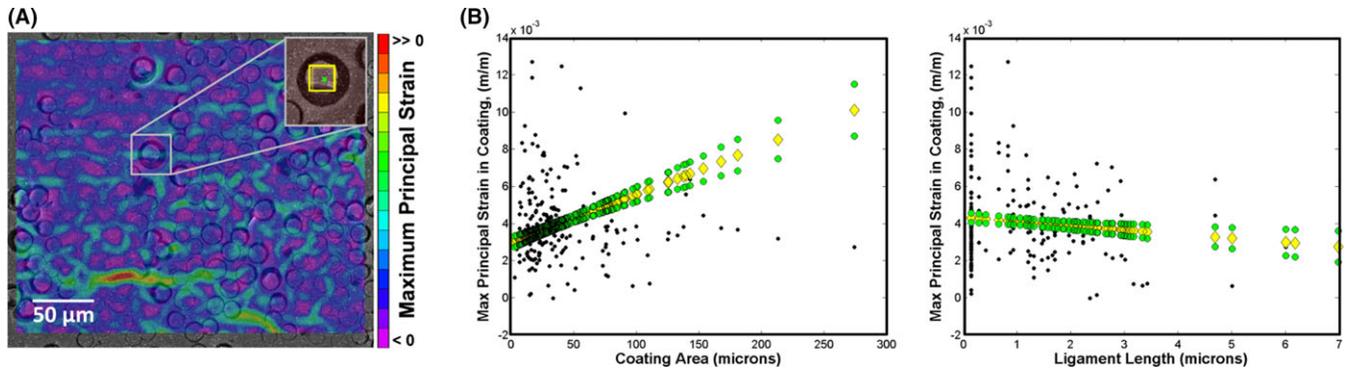
### 3.3 | Logistic model results

The influence of fiber properties on the likelihood of a fiber residing on a crack path was assessed using a logistic statistical model. First, best-subsets selection and lasso shrinkage were used to identify microstructural variables relevant to the response. Because the smooth bar tensile coupon was not tested to final fracture, its data set was not included in the logistic model. The model was trained with the data from AOI 1 of Figure 1, as it had the greatest number of observations. Results of each model are presented in Table 2. Both the best-subsets and lasso shrinkage algorithms selected variable subsets pertaining to coating area and fiber separation, as was the case for the linear regression model. Both models had low training and test error rates (ie they misclassified predictions 18.5% and 15.8% of the time, respectively). These low error rates indicate that the logistic models reliably predict the likelihood of a crack developing on a fiber if the microstructural parameters of the fiber meet the criteria determined by the model. However, note that the models have low  $R^2$  values and grossly underestimate when a fiber resides on a crack path. The best performing model (best-subsets) only predicted 6% of crack-path fibers correctly in the training data

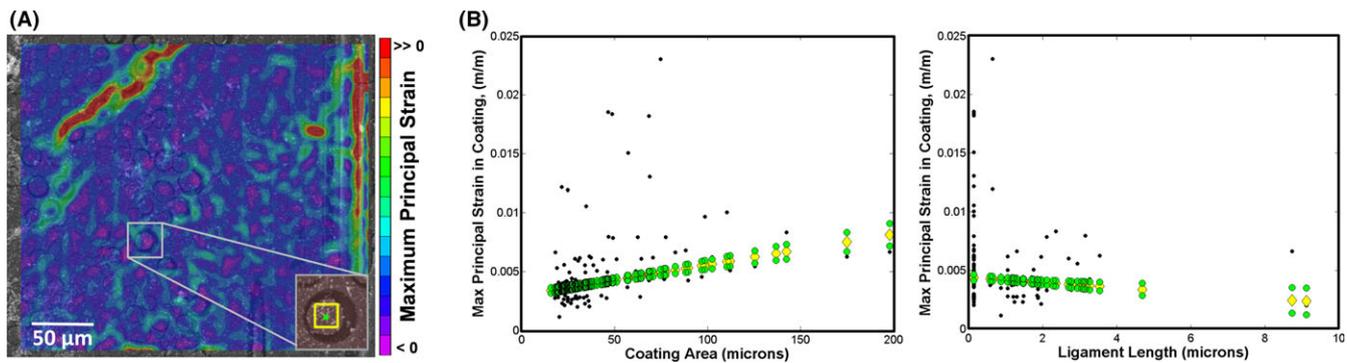
and 3% in the test data. This renders the results of the model impractical for use in optimizing the microstructure of fibrous CMCs. Thus, ensemble methods were pursued to improve the predictive capability.

Ensemble methods were used to improve the crack classification error rates of the best-subsets model and bolster the fits of the regression models. Three ensemble algorithms were investigated—bootstrap aggregation, boosting, and subspace nearest-neighbor classification—and it was found that boosting improved the classification error rates. All ensemble algorithms were implemented using MATLAB's Statistics Toolbox.<sup>21</sup> In brief, ensembles incorporate a combination of multiple fits to improve poor model diagnostics. A randomly selected subset of training data is used to grow a decision tree. New decision trees are grown sequentially from the original tree using the training data and information from the previously grown tree. The final tree used for predictions is built from all of its combined successors. Specifics on the individual ensemble techniques can be found in.<sup>18,21</sup>

The judicious use of boosting ensembles, as described below, significantly improved crack prediction. However, the error rate was still too high to accurately implement the binomial logistic (or classification) model in the prediction of microstructure influence on crack propagation in CMC transverse plies. Using AOI 1 as training data and performing 10-fold cross-validation, boosting ensembles were evaluated using the best-subsets predictor selection. Best-subsets were chosen because it produced a lower deviance and larger adjusted  $R^2$  than the lasso predictor set. The test error rate fell to 16.9% and the percentage of cracks correctly predicted rose to 37.5% when applied to the data set from AOI 2. The test error rate was 24.8% with only 25% correctly predicted cracks for the data set from the second notched-coupon tensile test.



**FIGURE 7** Application of model to test data (Test 1, AOI 1). Deformation fields are plotted on top of micrograph in (A). Predicted strain responses as a function of predictor values are shown in (B). Fitted strains responses are regressed against best-subset predictor (yellow diamonds) holding other predictor constant. 95% confidence interval is plotted in green circles. Black dots indicate experimentally observed response. Test response is in agreement with training response. [Color figure can be viewed at [wileyonlinelibrary.com](#)]



**FIGURE 8** Application of model to test data (Test 1, AOI 2). Deformation fields are plotted on top of micrograph in (A). Predicted strain responses as a function of predictor values are shown in (B). Fitted strains responses are regressed against best-subset predictor (yellow diamonds) holding other predictor constant. 95% confidence interval is plotted in green circles. Black dots indicate experimentally observed response. Test response is in agreement with training response. [Color figure can be viewed at [wileyonlinelibrary.com](#)]

## 4 | DISCUSSION

There are significant challenges associated with using DIC data from brittle materials like CMCs to statistically model damage. Two major challenges are that (i) cracks evolve over large swaths of the material during loading, sometimes closing as new cracks nucleate; and (ii) DIC measurements in CMCs are noisy, but removing noise comes at the expense of mitigating or even eliminating crack-induced strain signals. Regarding challenge (i), cracks that relax as a result of adjacent crack formation can present dampened strain signals that fall below the strain to noise ratio. Fortunately, most of the cracks captured in our data sets appeared to continually open as indicated by strain signals that intensified with load.

Recall from Section 3.3 that the logistic model underestimated the number of fibers residing on cracks (meaning that either a fiber filament or its coating intercepted a crack). This is likely a result of measuring the binomial

response (ie, does or does not a fiber reside on a crack path) after final fracture, as some matrix cracks could have closed. If possible, the entire edge of the gage section should be scanned at each increment, using the smallest FOV available. This would provide more observations for linear modeling, and strain signals could be converted to a binomial response for logistic modeling. As DIC data are near real time, the issue of crack closure would be resolved. This would also permit modeling crack initiation (rather than a given damage state) as a function of microstructural parameters.

Regarding the second point, DIC data often contain unwanted artifacts, namely noise and bias. Combating noise is particularly challenging in the SEM, where specimen charging, chamber contamination, and equipment vibration contribute to image contamination.<sup>22</sup> Noise and bias amplitudes can be reduced through data filtering, but this comes at the expense of spatial resolution. For quantitative statistical modeling, high spatial resolution is desired

TABLE 2 Logistic model results

Model	Best subsets	Lasso shrinkage
Deviance	162.15	166.57
Adjusted <i>R</i> -squared	.06	.03
Training error rate	18.5%	18.5%
Cracks correctly classified	6%	0
Test error rate	15.8%	15.8%
Cracks correctly classified	3%	0

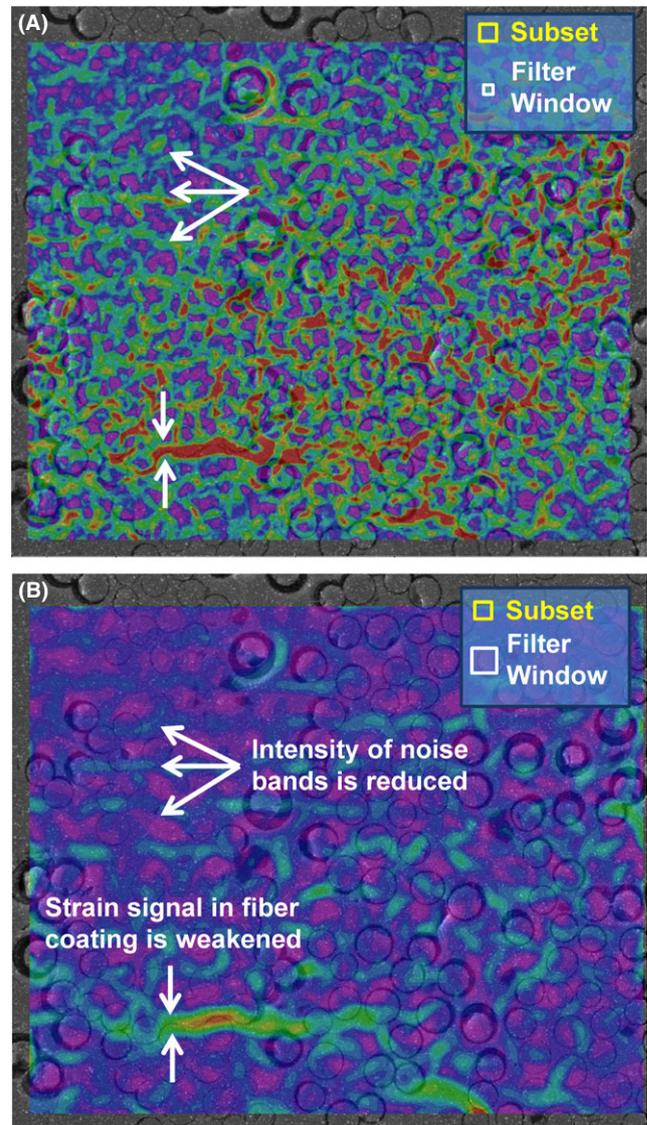
  

	Variable	<i>P</i> value	Variable	<i>P</i> value
Significant variables	Coating area	.00	Inverse nearest-neighbor distance	.00
	Area ratio	.16	Inverse squared ligament length	.53
	Nearest-neighbor distance	0		

to minimize smearing of deformation data at constituent interfaces. This is illustrated in Figure 8, where maximum principal strain fields with different amounts of filtering are overlaid on a micrograph of the loaded, cracked CMC.

Noise filtering presents an unavoidable conundrum: larger strain filters applied to the DIC data not only reduce periodic noise, but also reduce the intensity of real strain measurements. In Figure 9(A), strains are computed from full-field displacements, then smoothed with a  $5 \times 5$  pixel averaging filter window. The 29-pixel subset is depicted by the yellow box for reference. In Figure 9(B), the strain filter is increased to 45 by 45 pixels. As depicted by the trio of white arrows, this reduces the periodic bands of noise. However, as depicted by the double arrows, filtering reduces the intensity of strain measurements, in this case around the thin coating at the top of a fiber. The thickness of the band of localized strain around the coating remains about the same; however, the band in (B) is nearly uniform in intensity. This is a serious deficiency because using DIC data to detect cracks requires a clear distinction between regions that contain cracks and those that do not. For this purpose, the data of (A) are more suitable, but it is corrupted with noise that can be misconstrued as localized strain/cracking. Noise will present less of an issue in more advanced microscopes. However, once present, the low strain signal-to-noise ratio that results from running DIC on a stiff ceramic precludes filtering without a loss in spatial resolution.

The strength of the statistical modeling approach is that it may be used to quantify the synergistic influences of microstructural variables on damage progression that are difficult to reliably incorporate into phenomenological



**FIGURE 9** Maximum principal strain fields overlaid on microstructure for CMC loaded in tension. In (A), a five-node filter (1/6th of DIC subset size) has been applied to strain data. In (B), a 45-node filter (1.5 $\times$  DIC subset size) has been applied to strain data. Relative size of filter window to subset is inset in upper-right of both figures. Triple arrows point to bands of noise, which are reduced when filtering is increased. Double arrows point to a matrix crack, the signal of which is also reduced as a result of filtering. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

models. A limitation is that critical values of any given variable (to predict cracking events) may be evaluated only by fixing all other variables. However, provided noise limitations are addressed, this type of modeling provides designers a metric to identify likely cracking events based on tunable inputs of microstructural parameters. To improve accuracy of the model predictions, future efforts should address as many variables as are reliably measurable (eg, matrix defects, residual stresses, sub-surface damage).

## 5 | CONCLUSION

Unsupervised and supervised learning techniques were used to analyze the effect of fiber attributes on cracking in SiC/SiC CMCs. It was found that the amounts of coating surrounding fibers and their dispersion within the matrix were most likely to influence where cracks evolved in transverse plies. Covariance analysis and linear and logistic statistical models all implied that as coating area or the distance between adjacent fibers increased, the presence of localized maximum principal strain increased, and therefore the likelihood of a fiber residing on a crack path increased. The linear model showed a reasonable goodness-of-fit ( $R^2=0.49$ ) and low root-mean-square error ( $RMS=\pm 908 \mu\epsilon$ ).

There were several limitations in the statistical modeling techniques presented, and results be interpreted with caution. Subsets of the microstructure were used for modeling after being determined as representative elements of the overall transverse microstructure. Under this assumption, the distribution of fiber attributes and matrix cracks within the random subsets of the material are assumed to be similar. Without the capability to comprehensively image the entire gage section, this could not be guaranteed. Time and resources permitting a similar statistical analysis of the entire microstructure would be preferred. Furthermore, the SEM-DIC deformation data collected from SiC/SiC CMCs at FOVs on the order of hundreds of microns are inherently noisy. High-quality speckle patterns can help minimize noise, but some degree of filtering is necessary. Comprehensive imaging of the entire gage section at the smallest FOVs attainable is the best way to remedy this problem. Finally, a thorough investigation of the influence of microstructure on cracking should incorporate variables such as flaws, inclusions, and residual stresses. In spite of these limitations, this investigation suggests that coating thickness and fiber dispersion are significant factors in optimizing the microstructure of CMCs.

## REFERENCES

1. McDonald KR, Dryden JR, Zok FW. Effects of matrix cracks on the thermal diffusivity of a fiber-reinforced ceramic composite. *J Am Ceram Soc.* 2001;84:2015–2021.
2. Morscher GN. Stress-dependent matrix cracking in 2D woven SiC-fiber reinforced melt-infiltrated SiC matrix composites. *Compos Sci Technol.* 2004;64:1311–1319.
3. Morscher GN, Singh M, Kiser JD, Freedman M, Bhatt R. Modeling stress-dependent matrix cracking and stress – strain behavior in 2D woven SiC fiber reinforced CVI SiC composites. *Compos Sci Technol.* 2007;67:1009–1017.
4. Morscher GN, Dicarolo JA, Kiser JD. Effects of fiber architecture on matrix cracking for melt-infiltrated SiC/SiC composites. *Int J Appl Ceram Technol.* 2010;7:276–290.
5. Chateau C, Gélébart L, Bornert M, Crépin J. Journal of the mechanics and physics of solids modeling of damage in unidirectional ceramic matrix composites and multi-scale experimental validation on third generation SiC/SiC minicomposites. *J Mech Phys Solids.* 2014;63:298–319.
6. Tracy J, Daly S, Sevener K. Multiscale damage characterization in continuous fiber ceramic matrix composites using digital image correlation. *J Mater Sci.* 2015;50:5286–5299.
7. Curtin WA. Theory of mechanical properties of ceramic-matrix composites. *J Am Ceram Soc.* 1991;45:2837–2845.
8. Calard V, Lamon J. Probabilistic-statistical approach to matrix cracking in ceramic matrix composites (CMCs). *Key Eng Mater.* 1999;164–165:205–208.
9. Calard V, Lamon J. A probabilistic-statistical approach to the ultimate failure of ceramic-matrix composites — Part I: experimental investigation of 2D woven SiC/SiC composites. *Compos Sci Technol.* 2002;62:385–393.
10. Calard V, Lamon J. A probabilistic-statistical approach to the ultimate failure of ceramic-matrix composites — Part II: macroscopic model. *Compos Sci Technol.* 2002;62:395–399.
11. Sorenson B, Talreja R. Effects of nonuniformity of fiber distribution on thermally-induced residual stresses and cracking in ceramic matrix composites. *Mech Mater.* 1993;16:351–363.
12. Yang S, Tewari A, Gokhale AM. Modeling of non-uniform spatial arrangement of fibers in a ceramic matrix composite. *Acta Mater.* 1997;45:3059–3069.
13. Sutton MA, Li N, Garcia D, et al. Metrology in a scanning electron microscope: theoretical developments and experimental validation. *Meas Sci Technol.* 2006;2613:2613–2622.
14. Sutton MA, Li N, Joy DC, Reynolds AP, Li X. Scanning electron microscopy for quantitative small and large deformation measurements Part I: SEM imaging at magnifications from 200 to 10,000. *Exp Mech.* 2007;47:775–787.
15. Sutton MA, Li N, Garcia D, Cornille N. Scanning electron microscopy for quantitative small and large deformation measurements Part II: experimental validation for magnifications. *Exp Mech.* 2007;47:789–804.
16. Kammers AD, Daly S. Small-scale patterning methods for digital image correlation under scanning electron microscopy. *Meas Sci Technol.* 2011;22:125501.
17. VIC-2d, *Correlated Solutions*; 2009.
18. James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning*. New York, NY: Springer; 2013.
19. Tracy J, Daly S. Characterization of fracture in CMCs at the microstructural length scale. In: *38th Annu. Conf. Compos. Mater. Struct.* Cocoa Beach, FL; 2014.
20. CoorsTek. *Advanced Silicon Carbide for Critical Components*; 2016.
21. Mathworks. *MATLAB and Statistics Toolbox*; 2014.
22. Tracy J, Waas A, Daly S. A new experimental approach for in situ damage assessment in fibrous ceramic matrix composites at high temperature. *J Am Ceram Soc.* 2015;9:1–9.

**How to cite this article:** Tracy J, and Daly S. Statistical analysis of the influence of microstructure on damage in fibrous ceramic matrix composites. *Int J Appl Ceram Technol.* 2017;14:354–366. doi:10.1111/ijac.12646.

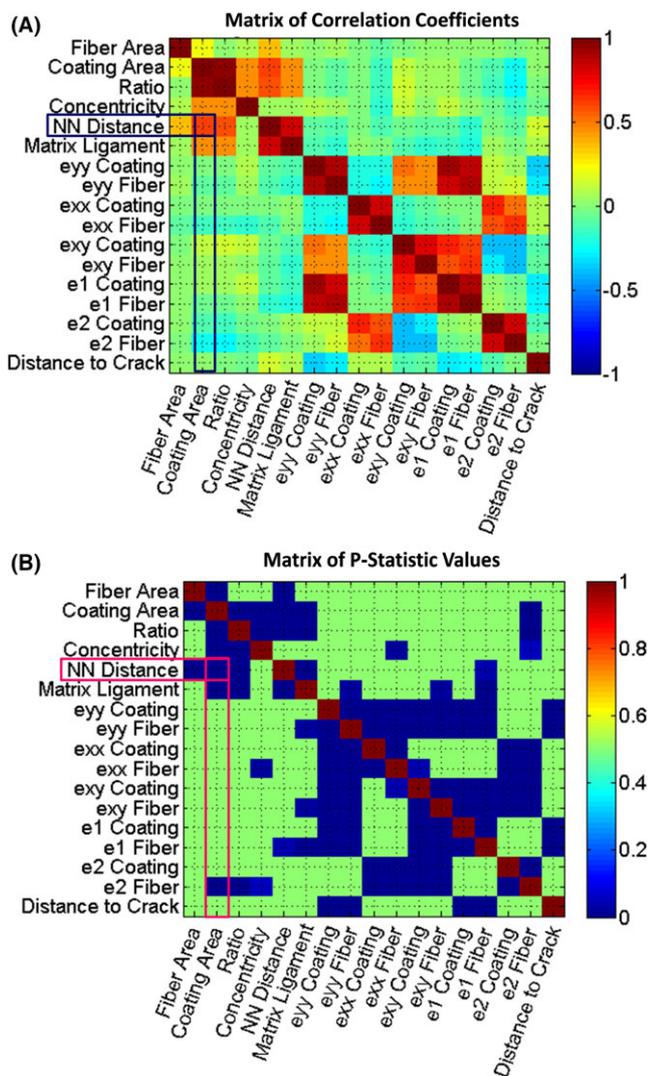
## APPENDIX 1

### Supplementary Examples of Statistical Methods

An example covariance matrix is illustrated in Figure A1 (A). The horizontal and vertical axes contain variable names. The intersection of any combination of variables contains a correlation coefficient which indicates the magnitude and direction of the correlation between the variable pair. A positive value implies a positive correlation; as the value of one of the values increases, so does the other. A value of 1 indicates a perfect positive correlation, whereas

a value of 0 indicates no correlation. The diagonals contain correlations between identical variables, and hence should always be positive 1.

In Figure A1(A), the intersection of coating area and nearest neighbor (NN) distant indicates a strong, positive correlation; as the amount of coating around a fiber increases, so does the centroid-to-centroid distance between it and its nearest neighbor. Figure A1(B) contains  $p$ -statistics for each correlation coefficient from Figure A1 (A).  $p$ -Statistic values less than or equal to .05 indicate strong enough confidence to consider the correlation statistically significant. All correlations with  $p$ -statistics greater than .05 (ie, those colored green in Figure A1(B)) were disregarded.



**FIGURE A1** Examples of (A) covariance and (B)  $p$ -statistic matrices from fiber bin from AOI 1 of Figure Anaheim. In (A), nearest-neighbor distance and coating area show a strong positive correlation. The corresponding  $P$ -value in (B) is low, indicating that the correlation is significant