JOM
https://doi.org/10.1007/s11837-021-04703-0

Identification and Quantification of 3D Fiber Clusters in Fiber-Reinforced Composite Materials

MATHEW J. SCHEY ,[1,4] TIBOR BEKE,[1] LARS APPEL,[2] SIMON ZABLER,[3] SAGAR SHAH,[1] JIE HU,[1] FUQIANG LIU,[1] MARIANNA MAIARU,[1] and SCOTT STAPLETON[1]

1.—Department of Mechanical Engineering, The University of Massachusetts Lowell, Lowell, MA 01854, USA. 2.—Institute for Textile Technology, RWTH Aachen University, Aachen, Germany. 3.—Fraunhofer Institute for Integrated Circuits IIS, Würzburg, Germany. 4.—e-mail: Mathew_schey@student.uml.edu

Microscale computed tomography scans of fiber-reinforced composites reveal that fibers are most often not strictly parallel to each other but exhibit varying degrees of misalignment and entanglement. One characteristic of this entanglement is the degree to which fibers stay together as clusters. In this study, a method for identifying and isolating fiber clusters was established, and scans of two different composite microstructures were analyzed. To identify clusters, fiber center points of the first cross-section were triangulated, and the variation of the perimeter and area of triangles along the fiber direction was used to identify fiber triads which stay together. A filtering process eliminated fiber triads not part of a larger cluster. Geometric properties of the clusters such as cluster orientation, radius of gyration, cluster density, and volume fraction were calculated and compared. The metrics revealed fundamental differences between the two samples, suggesting that clusters have origins in manufacturing.

INTRODUCTION

One of the keys to understanding the performance of fiber-reinforced composites lies in understanding the organization and geometry of their microstructure. Proper quantification of the behavior of composites at the fiber level can provide critical information on quantities such as their stiffness and strength. It is extremely important, therefore, that any representation of the microstructure be as accurate as possible.

Early efforts at understanding the effects of fibers involved modeling perfect, parallel cylinders packed in a repeating hexagonal or square pattern. These studies showed that some elastic properties, especially those that are axially dominated, are nominally captured well by such ideal models.[1–6] However, they often fall short when predicting transversely dominated properties, and strength, as well as the variation of properties. Additionally, cross-sectional images of fiber microstructures reveal that ideal fiber packing models are not representative of the variation found in manufactured composites.

Recently, much work has been done to try to understand the effects of random placement of fibers. Newer representative volume elements (RVEs) have been generated by varying the two-dimensional (2D) locations of fibers randomly.[7–17] Constructing RVEs in this way allows localized effects such as stress concentrations to be captured.[18] The results of these efforts have shown that the random spatial variation of fibers has a strong effect on the transverse strength of composites. Numerous works have also shown that fiber misalignment has a strong effect on the compressive strength of composites.[19–22] The scatter on the values of properties of composites has thus been attributed in part to the apparently random positions of fibers in real composites.

One of the aspects of fiber morphology that is not necessarily represented by many of these models is the presence of lengthwise fiber behaviors such as entanglement. A number of studies have been

(Received December 30, 2020; accepted April 19, 2021)

Published online: 17 May 2021

carried out using micro computed tomography (CT) scans of composites to explore the morphology of fibers along the fiber direction.[23–29] These works have shown that the spatial variability in fiber positions is subject to change from cross-section to cross-section due to the degree of misalignment of the fibers. The degree of fiber misalignment and its effects on strength have also been studied extensively.[30,31] Some studies have also attempted to quantify the behavior of fibers using various metrics. These works have found that many composite microstructures are host to several three-dimensional (3D) features.

One feature of microstructures exhibiting significant entanglement is fiber clustering, where groups of neighboring fibers remain together throughout the volume of a composite. Defining and identifying fiber clustering may be a way to describe the features of the entanglement that may have a significant impact on local effects such as failure initiation and progression, and could serve towards homogenizing parts of the microstructure for more efficient microstructural representation. Few studies have been done on the detection of fiber clusters, while virtually no analyses have been done on cluster morphology or mechanical effects.[23,32] Therefore, there is a need to create a methodical definition of fiber clusters to study their behavior and eventual effects on composite properties.

In this study, a novel method for identifying and analyzing fiber clusters is presented. High resolution cross-sectional images from two different composite microstructures were analyzed. One scan set is from an aerospace-grade laminate, made from a prepreg in an autoclave using materials manufactured for higher performance. The second scan set comes from an automotive-grade heavy tow (48k fibers per tow) consolidated using vacuum-assisted resin transfer molding (VARTM), where the reduced cost of working with a larger tow results in more entanglement of fibers. Fiber paths are extracted using a custom-written MATLAB code. Clusters are detected using the extracted fiber paths. A series of metrics are then used to quantify the behavior of the clustered fibers. Finally, the results of the metrics for the two different microstructures are compared (Fig. 1).

EXPERIMENTAL PROCEDURES

Materials and Manufacturing

For this study, scans of two fiber-reinforced composites were compared, an aerospace- and automotive-grade composite. The aerospace-grade composite was constructed from an IM7/5320-1 unidirectional prepreg and processed using an autoclave. The specimen was a curved bracket geometry, quasi-isotropic layup. From the aerospace sample, layers 5, 9, and 13 were isolated and analyzed (Fig. 2a).

The automotive-grade composite was manufactured at the Institute for Textile Technology at RWTH University from Toho Tenax STS40 F13 48k tows. Each tow contained 48,000 fibers, which dereases the cost of the carbon fiber due to bulk manufacturing. The tows were infiltrated with Hexion Epikote RIMR 135 with an Epikure RIMH curing agent. The tow was infiltrated using VARTM at pressure of 1000 mbar.

Imaging

The automotive-grade specimens were scanned at the European Synchrotron Radiation Facility (Fig. 2a). Scans were taken at a resolution of 0.332 μm/voxel, with 2048 pixels in the fiber direction and 2098 pixels in the transverse directions. To reduce the computational load, every third image was analyzed in the stack of automotive scans, resulting in a total of 682 images with spacing of 0.996 μm between images. Compared with micro-CT scanners, synchrotrons have the advantage of being able to operate at significantly higher energy levels. Typical micrographs at submicron resolution often contain significant amounts of Gaussian pixel noise. This noise effectively blurs the boundaries between fiber and matrix, leading to errors in fiber detection in each scan. The scans of the automotive composite, however, exhibit high levels of contrast between fibers and matrix.

The scan of the aerospace-grade composite was taken at the Air Force Research Lab using their Robomet.3D robotic serial sectioning system (Fig. 2-b).[33] A robotic arm system polishes the surface of the specimen, takes an image, then continues polishing. One of the main benefits of optical microscopy serial sectioning (compared with

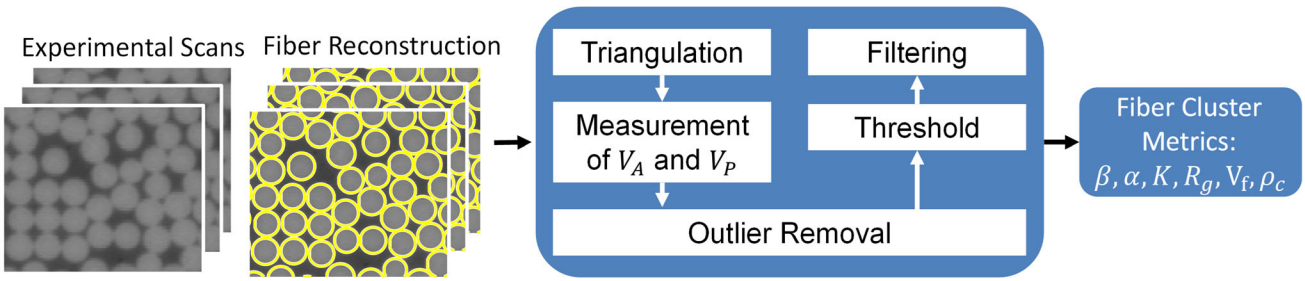

Fig. 1. Flowchart of clustering process.

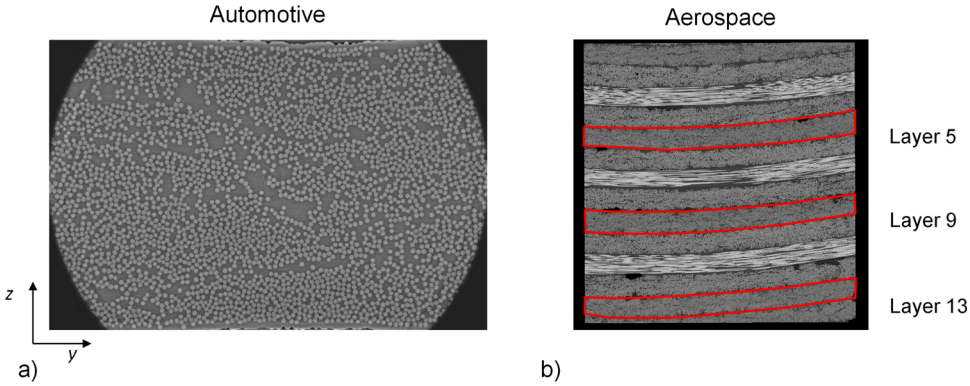

Fig. 2. Cross-sectional images of the (a) automotive-grade heavy tow and (b) aerospace-grade quasi-isotropic layup with layers highlighted.

volumetric methods such as x-ray CT for polymer-matrix composites) is that bright-field reflective optical images from composite materials provide a combination of high signal-to-noise and micron-scale resolution that greatly aids in the segmentation of carbon fibers from matrix. Note that this is not a "volumetric" reconstruction as one obtains from x-ray CT systems; rather, it is information from just the surface. The resolution of the images was 0.336 μm/pixel, and images were taken with an average spacing of 1.07 μm between images. A total of 208 images were taken with each image being 4844 $\times$ 5939 pixels. Three of the four 0° sections of the aerospace composite were cropped and used for analysis.

Microstructure Reconstruction

A script was used to reconstruct the fiber paths of both samples. First, fiber centers and radii were found in each cross-section transverse to the fiber direction (as in Fig. 2) using a circular Hough transformation. The fibers used in this study did not show significant cross-sectional eccentricity, and the standard deviation of the fiber radii was computed to be 0.104 μm for the aerospace samples and 0.201 μm for the automotive samples. These values are relatively low when compared with the mean radius of 2.8 μm for the aerospace samples and 3.38 μm for the automotive samples, thus the mean fiber radius for each sample was used for all subsequent calculations. After the fiber center points were found in each cross-section, the center points had to be linked between cross-sections to form the fiber. Due to the stiffness of the fibers and the relatively short spacing between images (compared with the fiber diameter), the fiber positions did not vary significantly between sequential cross-sections and a nearest-neighbor search algorithm from one cross-section to the next was thus used to link fibers.

Fibers located at the edge of each scan were often only partially within the image boundaries, effectively making them discontinuous. Many of the metrics used for analysis required fibers that were represented in all cross-sections, so all edge fibers were removed from the useable dataset.

Cluster Analysis

A formal definition for isolating clustered fibers from all other fibers is described below. After fiber clusters were identified, various quantities are presented to describe the features of individual clusters, as a means of characterizing the composite microstructures.

Cluster Identification

Cluster analysis is a method for describing the tendency of groups of fibers to remain together throughout the volume of fiber-reinforced composites. The literature contains several approaches to formalize the notion of "cluster"' or "fiber bundle." Most approaches are two dimensional in nature and rely on the distribution of distances between nearby fiber centers, whether measured in absolute units or via graph-theoretic notions of adjacency in terms of nearest neighbors. In the analysis presented here, fiber triads are found by triangulation of all fiber centers in the first cross-section of an image stack. Alpha shapes of the fibers are used to ensure the boundary fiber connectivity is restricted to the convex hull. The fiber triads found in the first cross-section were held constant for the remainder of the analysis. Triangulated fiber triads belonging to clusters will not vary as much in area or perimeter as those that do not belong to a cluster. Whether or not a triad belongs to a cluster should also be independent of fiber orientation. Clustered fibers should be permitted to meander as long as their neighbors tend to meander in parallel. The requirement that all fibers stay together also makes clusters independent of the degree of fiber packing. Therefore, the cluster metric was defined in terms of the total variation of certain geometric measurements of fiber triads. The total variation of the area, V_A^l, is used here to measure the 3D variation of triad areas within the sample.

$$V_A^l = \sum_{n=2}^{N} \left| A^l(x_n) - A^l(x_{n-1}) \right| \quad (1)$$

where N is the total number of cross-sections and A^l is the area of triad l at a particular cross-section (Fig. 3a). Initially the total variation of triad areas was the sole criteria used to determine whether fibers belonged to a cluster or not. It was discovered, however, that it was possible for three fibers to bound a relatively small area but still be spread out. To handle such cases, the total variation of the perimeter, V_P^l, was added as a second criterion.

$$V_P^l = \sum_{n=2}^{N} \left| P^l(x_n) - P^l(x_{n-1}) \right| \quad (2)$$

where V_P^l is the total variation of the perimeter, and P^l is the perimeter of triad l at a particular cross-section (Fig. 3a). The final cluster determination is made using the histogram distributions of both criteria. The analysis is complicated, however, by the presence of "stray fibers." These fibers travel at an angle to their neighbors that is considerably (by three standard deviations or more) larger than the variation in the angles between nearby fibers. The centers of cross-sections of stray fibers create triads whose areas are subject to total variation that is an order of magnitude greater than that between fibers belonging to the same cluster. Therefore, the histogram distributions for both criteria were trimmed to within the mean plus two standard deviations to remove outliers from the data. Otsu's method with three bins was then used to threshold V_A^l and V_P^l independently.[34] In this method, three bins are created, with bin boundaries set such that the variance within each bin is minimized. While it may seem that using three bins is an arbitrary selection, it relies on the assumption that there are generally three groups of fibers of interest: those that diverge from each other along the fiber direction, those that stay close together, and all the others. For this analysis, only the group which stayed together were selected for further analysis, but the group of fiber triads which diverge from each other could similarly be used to identify fiber groups which are highly entangled. Finally, using relative rather than static thresholds allows the cluster identification process to take place independent of the compaction of a specimen, since the threshold is specific to the

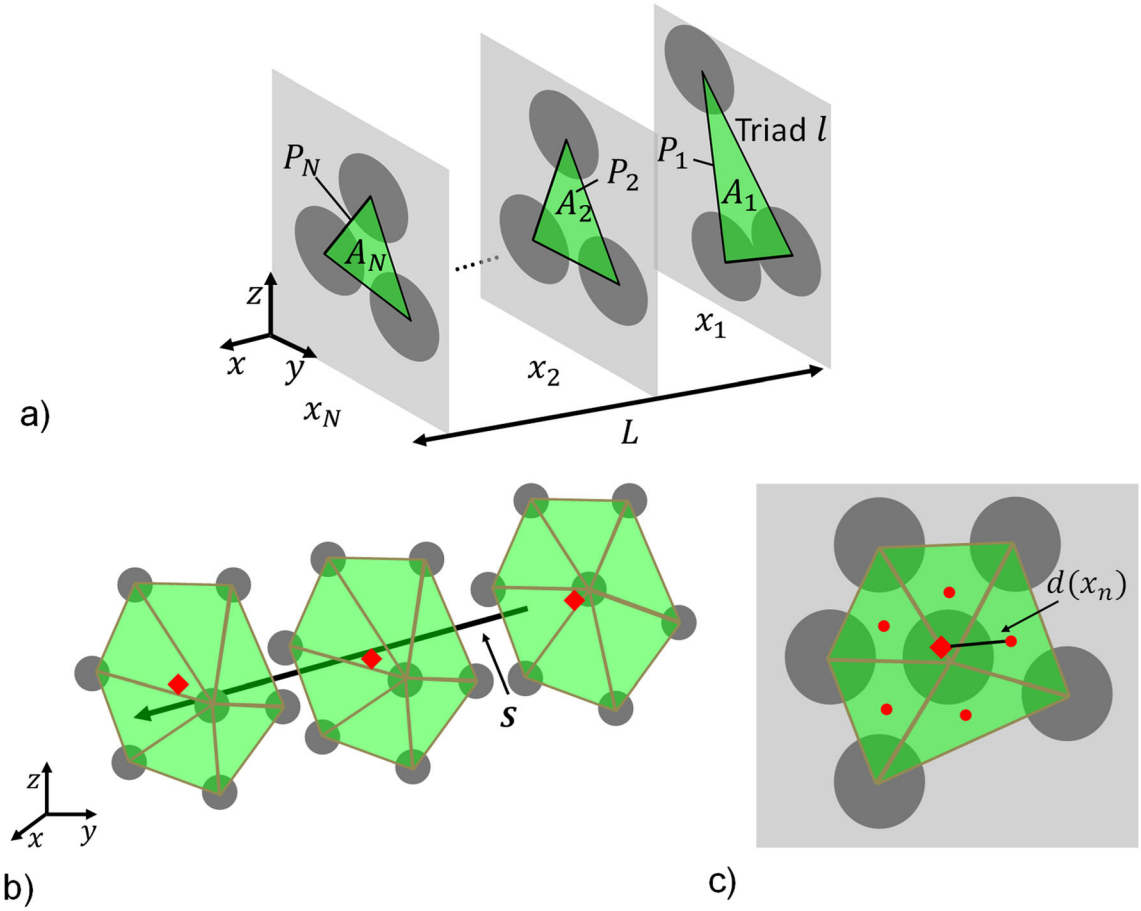

Fig. 3. Diagrams of (a) a typical triad l with perimeter P and area A through N cross-sections and length L, (b) a linearized cluster over length L with vector S connecting the cluster's centroids, and (c) a cross-section of a cluster showing the distance between a triad centroid (red dot) and the cluster centroid (red diamond) (Color figure online).

specimen. Additionally, despite the total variation values being sensitive to the spacing of the cross-sections, the cluster analysis itself is independent of the spacing. When the spacing between scans increases or decreases, the variation values for each triad in the sample change equally and the distribution of values are then shifted higher or lower. Finally, the determination of whether a triad is clustered or not depends solely on the triad's relative location in the distribution of all triads rather than its specific value, and the threshold is invariant of a length scale.

The lower thresholds for the area and perimeter variance are then combined via a Boolean operation. For each triad, the values of V_A^l and V_P^l must be below their respective thresholds to be deemed clusters. The triads in the two upper bins for both criteria, as determined by Otsu's method, were not studied in this analysis. There are two major drawbacks to this method of cluster detection. The first is that the thresholding method used can be sensitive to outlier triads. While these outliers were filtered out during the analysis, and filtered equally for both samples, it is difficult to remove them objectively. The second drawback is that the proposed method will always find fiber clusters in a sample, even if they do not appear to be present. Other methods can be applied to look at the averages of certain metrics of the sample as a whole and check for fibers whose values deviate. Metrics such as the neighbor change rate, for example, have been used in the past as a basis for the existence of clusters in a particular sample.[23] Within clusters, there are typically fibers whose neighbors do not change or change very little throughout the sample. Searching for these fibers, coupled with the approach mentioned in this work, could serve as an alternative approach to cluster detection. Details of the use of Otsu's method and the method for removing outliers using the histogram distributions are presented in Appendix Sect. 6.2.

The thresholding typically results in a noisy mix of threshold-passing triads, but clusters should be a unique subset of these triads that are aggregated together. To isolate the fiber clusters, a filtering algorithm was created. The intent of the filtering is to remove passed triads that are not part of a larger cluster but include fiber triads that are surrounded by a cluster but did not pass the thresholding. This results in a smoothing of fiber clusters, where the feature size can be controlled via the filtering method. The basic filtering operation checks whether the neighbors of each triad also passed the thresholding. The algorithm then changes the designation of each triad according to the designation of its neighbors. Details of the filtering operation can be found in Appendix Sect. 6.3. The entire clustering process is also summarized in Fig. 1.

Cluster metrics

A variety of measurements can be made on clusters to reveal details of their geometry and makeup. Two types of metrics were used to quantify the clusters (Fig. 4). The first type were 3D metrics, such as orientations, which described features of the clusters over their entire length. The second type were 2D metrics, such as the second moments of area and radius of gyration, which were calculated at each cross-section and averaged along the length of the clusters.

The centroid of the cluster at each cross-section can be found by finding the average of the centers of area for all triads that make up that cluster. These centers can be found in every cross-section of each sample and stitched together similarly to the fiber centers found in each scan. A straight line can be used to approximate the overall path of the centers of area for each cluster. A linear fit was used so that the orientation metrics could be formulated as global rather than local metrics. One of the main issues with local metrics is that they are sensitive to any deviations in position from cross-section to cross-section. Using these linear fits, the polar angle, γ^m, can be found for each cluster m by

$$\gamma^m = \operatorname{acos}\left(\frac{L}{\|\boldsymbol{S}\|}\right) \tag{3}$$

where L is the length of the entire scan volume in the fiber direction and $\boldsymbol{S}$ is the vector of the line of best fit through all centroids of a cluster in each cross-section (Fig. 3b). The polar angle is a useful quantity because if gives the magnitude of a cluster's deviation from the global fiber direction. While many clusters tend to consist of fibers that deviate very little from the global fiber direction, a number of clusters were seen to be made of fibers which were oriented some offset from the global fiber direction. Another useful angle measure is the pairwise angle between clusters m and n, β^{mn}, defined as

$$\beta^{mn} = \operatorname{acos}\left(\frac{V^m \cdot V^{n'}}{\|V^m V^{n'}\|}\right) \tag{4}$$

where V^m is the direction vector for a linearized cluster m and $V^{n'}$ is the direction vector for cluster n. Note that both direction vectors are the unit normal vectors of the vector $\boldsymbol{S}$ for both clusters. The angle between fibers provides an indication of how parallel the clusters are to each other.

The radius of gyration, $R_g^m(x_n)$, of each cluster m at cross-section x_n can be calculated about the center of area from each cluster using

$$R_g^m(x_n) = \sqrt{\frac{\sum_{l=1}^{N_t^m} d^l(x_n)}{N_t^m}} \tag{5}$$

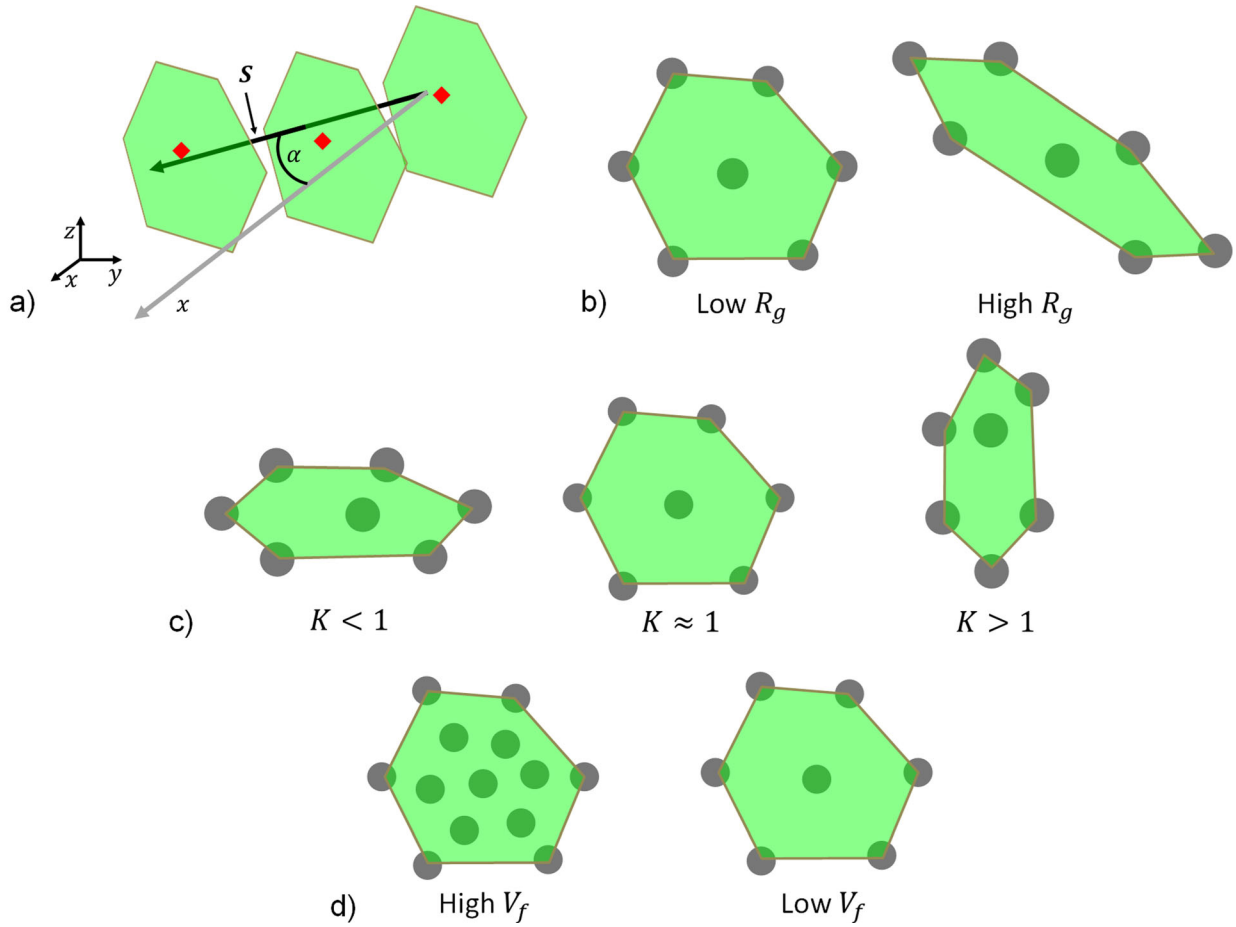

Fig. 4. Examples of cluster metrics including (a) polar angle from the *x*-axis, (b) contrasting R_g values, (c) contrasting K values, and (d) contrasting V_f values.

where N_t^m is the number of fiber triads in a cross-section of cluster m, and $d^l(x_n)$ is the distance from the centroid of the cluster to the center of triad l inside the cluster at cross-section x_n (Fig. 3c). While the value for R_g has many structural uses, it is used here as a measure for how spread out the triads, and subsequently the fibers, are inside a cluster (Fig. 4c).

The second moments of area, I_{yy} and I_{zz}, can also be calculated for each cross-section (x_n) of a cluster using

$$I_{ii}(x_n) = \sum_{l=1}^{N_t^m} A^l(x_n) d_{ii}^l \quad \text{where } i = y, z \tag{6}$$

d_{ii}^l is the distance from the cluster centroid to the center of the triads within the cluster in either the y or z direction. Like the radius of gyration, the moments of inertia are used to understand the geometry of clusters. The values of I_{yy} and I_{zz} provide insight into how long the clusters are along both the y and z axes, in the sense of the distribution of area along those dimensions (Fig. 4d).

The value $K(x_n)$ is used to determine how symmetric the clusters are about both axes, defined as

$$K(x_n) = \sqrt{\frac{I_{yy}(x_n)}{I_{zz}(x_n)}} \tag{7}$$

When the values for the moments of inertia are very similar, the value of K is close to 1, indicating that the clusters are fairly symmetric about the y and z axes. When I_{yy} is larger than I_{zz}, it means that the cluster is longer in the z than y direction, and vice versa. The values for K, therefore, give a sense of how symmetric the clusters are about both axes (Fig. 4a).

The average cluster volume fraction, $\overline{V}_{fc}^m(x_n)$ is a measure of the average fiber volume fraction in a cluster m at cross-section x_n.

$$\overline{V}_{fc}^m(x_n) = \frac{\pi R_f^2 \sum_{l=1}^{N_t^m} \frac{1}{A^l(x_n)}}{2N_t^m} \tag{8}$$

where R_f is the radius of the fibers in the sample. The assumption for all samples analyzed in this work is that the radius of the fibers is nearly constant and that all deviations from the mean

radius are inconsequential. While this may not be valid for many composites, the quality of fibers used in the composite specimens of this study permits this assumption. The average volume fraction for each cluster is a measure of how compacted the clusters are (Fig. 4b).

Finally, the cluster density, ρ_c, can be used to help compare samples of different sizes and fiber counts.

$$\rho_c = \frac{N_c}{A_{\text{total}}} \tag{9}$$

where N_c is the number of clusters per sample and A_{total} is the total area of cross-section of the scanned composite.

Additional metrics such as the linearity and global twist of the clusters could also be explored using this analysis. Little is known about which descriptors will correlate with structural properties such as the fatigue strength and transverse strength. Study is needed to quantify the effects of these descriptors and provide information on the effects of clusters.

RESULTS AND DISCUSSION

A method for detecting and isolating fiber clusters using cross-sectional images of fiber-reinforced composites was created. Various measurements and metrics were used to quantify the shape and behavior of the clusters. In this section, automotive-grade and aerospace-grade composites are analyzed, and the cluster metrics compared.

Cluster Analysis

Cluster Identification

The clustering process isolates groups of neighboring fibers whose spacing with respect to each other varies little throughout the sample (Fig. 5a and b). Cropped regions of the automotive sample and layer 13 are shown (Fig. 5a). The average intensity images, referred to in this work as "ghost plots," were created by averaging each cross-section together (Fig. 5b). The colored triangles show the magnitude of the values of V_A and V_P, while the grey triangles represent the outliers which were removed before calculating thresholds (Fig. 5c and d). The fiber triads that are below both thresholds were then found before clusters were isolated (Fig. 5e). The fiber clusters for both sections are then shown after rounds of filtering (Fig. 5f).

As seen in the ghost images of the automotive sample, the orientation of many of the clustered fibers deviated far from the intended fiber direction. In the aerospace sample, however, the degree of misalignment of the fibers was far less severe. Instead, the clusters seen here actually seem to be near hexagonally packed in many places. As expected, the total variation values for the two samples are significantly different (Fig. 5c and d). In the automotive sample, there are distinct resin-rich regions that contain low fiber counts and meandering fibers. These regions appear as grey triangles when the triads are colored based on their total variation value. In the aerospace samples, the holes, if any, appear as very small regions made up of only a few triangles. Effectively, the level of compaction in the aerospace samples forces the total variation values to decrease relative to the automotive sample. This makes the thresholds for the aerospace samples generally smaller than those for the automotive sample. The filtering process applied here removed all of the fiber triads whose neighbors were not below both thresholds. The filtering process was run until no more triads were removed, allowing the shapes of the clusters to converge.

The lower pressure used to manufacture the automotive sample suggests that the passed fiber triads in the automotive sample tend to be much more spread out on average than those of the aerospace samples; That is, the clusters in the aerospace samples tend to be denser with respect to fibers than the clusters in the automotive sample.

Plotting the total variation of fiber triad area and perimeter for both samples shows that the automotive triads are generally larger than the aerospace ones (Fig. 6a and b). The area thresholds for the automotive sample and layers 5, 9, and 13 of the aerospace sample were 68.3 μm^2, 7.34 μm^2, 8.61 μm^2, and 5.72 μm^2, respectively. The perimeter thresholds for the automotive sample and layers 5, 9, and 13 were 24.6 μm, 4.65 μm, 5.29 μm, and 3.35 μm, respectively. In both materials, triads with higher average volume fractions tend to contain more passed triads. In other words, there is a noticeable correlation between increasing volume fraction and the number of passed triads. This finding makes sense because a higher total variation in the area, for example, would indicate that the fibers bounding that area are relatively far from each other. Comparing across samples, however, reveals that clusters in the automotive material occur at lower average volume fraction values than in the aerospace material.

Plotting the two total variation values together provides clearer insight into the shapes of the fiber triads in the two composites (Fig. 6c). In general, the aerospace samples have more fiber triads that pass only one of the two thresholds. There is a definite positive correlation between the total variation of area and total variation of perimeter for each sample. The fibers in the triads of the aerospace samples, however, have a greater tendency to diverge from each other in such a way as to increase either the perimeter or the area bounded between them, but not both. This leads to the aerospace envelope being much wider than the automotive one. The aerospace fibers tend to slip past each other in shear bands as they meander, causing more obtuse triangles that vary more in perimeter than area.

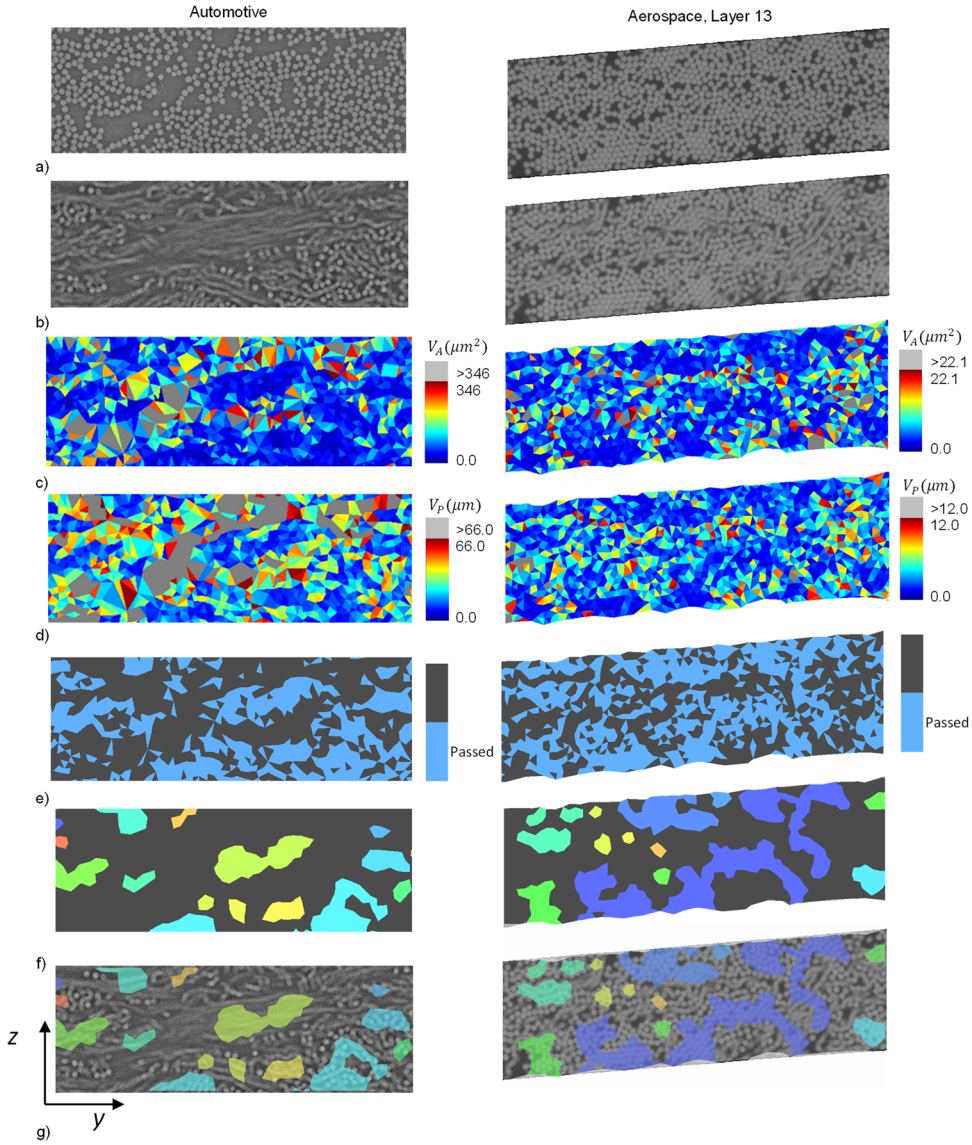

Fig. 5. Clustering process for two samples showing (a) cross-sections, (b) ghost plots, (c) triangulation colored by V_a values, (d) triangulation colored by V_P values, (e) triads that passed the variation thresholds, (f) final cluster result with each individual cluster colored uniquely, and (g) final cluster results overlayed on top of ghost plots.

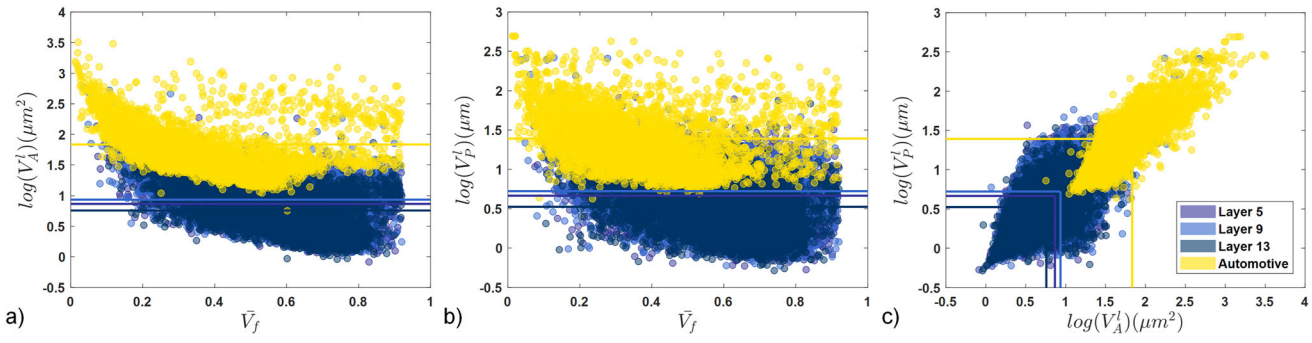

Fig. 6. Plots of (a) total variation of area versus average triad volume fraction, $\overline{V}_f$, (b) total variation of perimeter, V_A^l, versus average volume fraction, and (c) total variation of perimeter, V_P^l, versus total variation of area with thresholds for each sample shown as vertical and horizontal colored lines.

Cluster Metrics

For the automotive sample, the clustering analysis identified 32 clustered regions. The aerospace layers 5, 9, and 13 had 71, 64, and 73 clusters, respectively. One of the known details about both samples is that they were compacted vertically, which would cause the fibers in both samples to spread out horizontally. It was hypothesized, therefore, that the cluster shapes would be affected by this. Visually, it is apparent that the fiber clusters are longer horizontally than vertically (Fig. 7).

Various metrics were chosen to try to obtain a more complete picture of the geometry of clusters (Fig. 8). The average cluster area for each sample gives a general idea of the magnitude of the cluster sizes (Fig. 8a). The average polar angle is a measure of the cluster centroid's deviation from the fiber direction (Fig. 8b). The average angle between cluster centroids is a measure of how parallel the clusters are to one another in each sample (Fig. 8c). The average moments of inertia give a sense of how long the clusters are in each direction (Fig. 8d and e). The average value of K is the ratio of the moments of inertia and provides information on the symmetry of clusters (Fig. 8f). The average cluster volume fraction is a measure of the average fiber volume fraction in each cluster per cross-section (Fig. 8g). Finally, the cluster density is a normalized measure of how many clusters are present in a given area (Fig. 8h).

The average polar angle and average angle between cluster centroids for each cluster in each sample did not have standard deviations of more than 2° per sample. This means that the majority of clusters do not deviate very much from the intended fiber direction and remain relatively parallel to each other. The greatest difference between samples was seen in the angles between the cluster centroids. The automotive sample had higher averages than the aerospace samples in every metric except its K value. The cluster density, ρ_c, considers the number of clusters per area of the scan space for each sample. This metric provides a way to compare clusters in samples of different shapes. The cluster density for all aerospace samples was significantly larger than for the automotive samples, while the average area for all aerospace samples was lower than for the automotive samples.

Centerline plots of cluster centroids were generated for all four samples, where the view is directly into the fiber direction and the automotive sample has been trimmed to the size of the aerospace samples (Fig. 9a). The centerline plots provide a good way of viewing the orientation of the cluster centroids. The degree of misalignment of each centroid can be seen by its own length in a particular direction. Centroids which do not move very much appear very similar to the actual width of the line. Another benefit of the centerline plot is that it can show, qualitatively, how parallel the centroids are. The parallelism of clusters is also shown in the histogram of angles between cluster centroids (Fig. 9b).

The angles between fiber cluster centroids between the automotive and aerospace samples differ mainly in the magnitude of the probability density. Clusters belonging to the aerospace samples are more parallel to each other than those of the automotive sample in the sense that lower values of β occur more often. In both samples, however, the majority of cluster centroids were within 1° of each other. These results show the fiber clusters to be relatively straight and aligned with each other.

The histogram of all values for K for the four samples is shown in Fig. 10a. In a sense, K is a measure of the anisotropy of the clusters. Higher values of K would mean that the clusters were longer in the z direction, while lower values of K would mean the clusters are longer in the y direction. A K value of 1 would mean that the clusters are perfectly symmetric about both axes. The average cluster volume fraction, however, is a measure of the average compaction of the fibers inside each cluster in each cross-section (Fig. 10b). The volume fractions of all triads in each cluster were averaged in each cross-section. The areas of

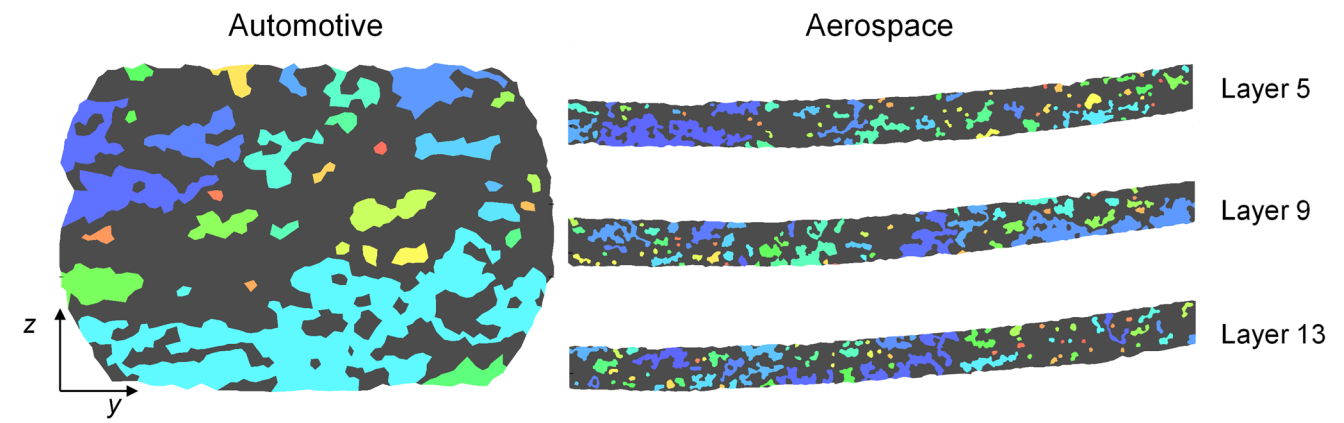

Fig. 7. Final results of cluster analysis for automotive sample (left) and four aerospace samples (right).

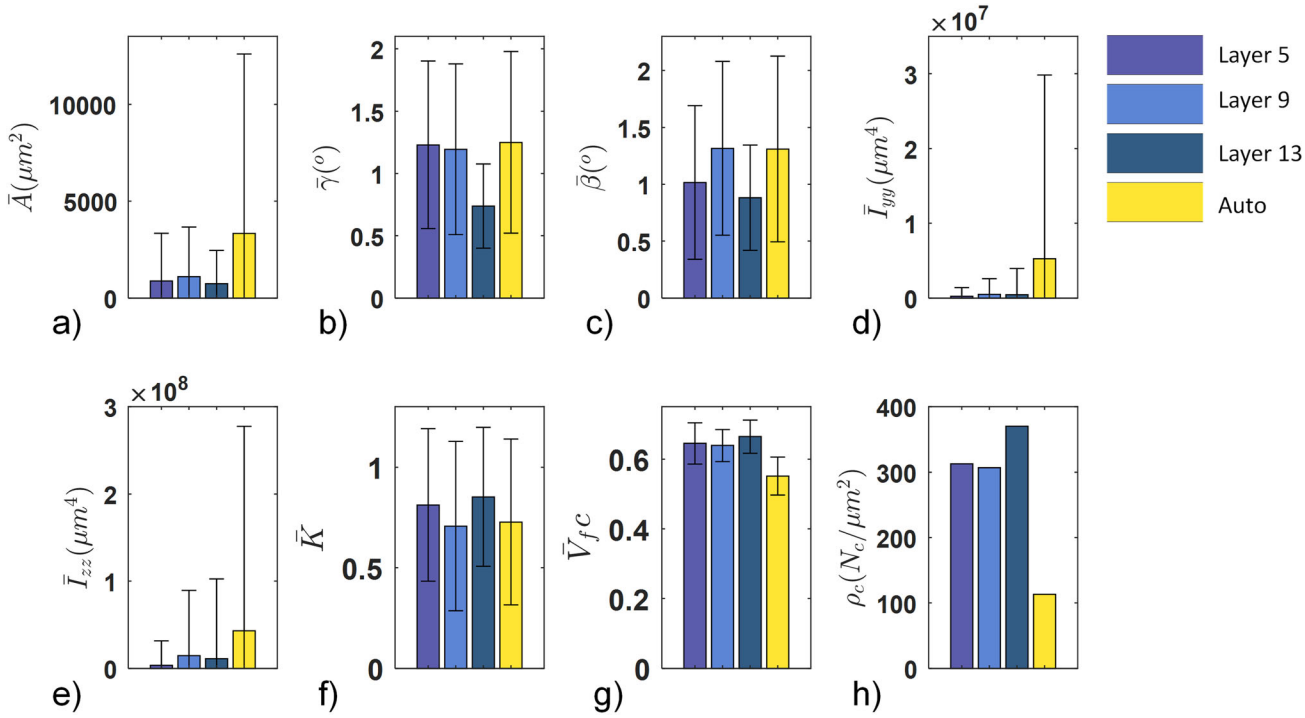

Fig. 8. Plots of average (a) cluster area, A, (b) polar angle of cluster centroids, γ, (c) angles between cluster centroids, β, (d) moment of inertia about the y-axis, I_{yy}, (e) moment of inertia about the z-axis, I_{zz}, (f) ratio of the square root of the moments of inertia, K, (g) cluster volume fraction, V_{fc}, and (h) cluster density, ρ_c, with standard deviations shown.

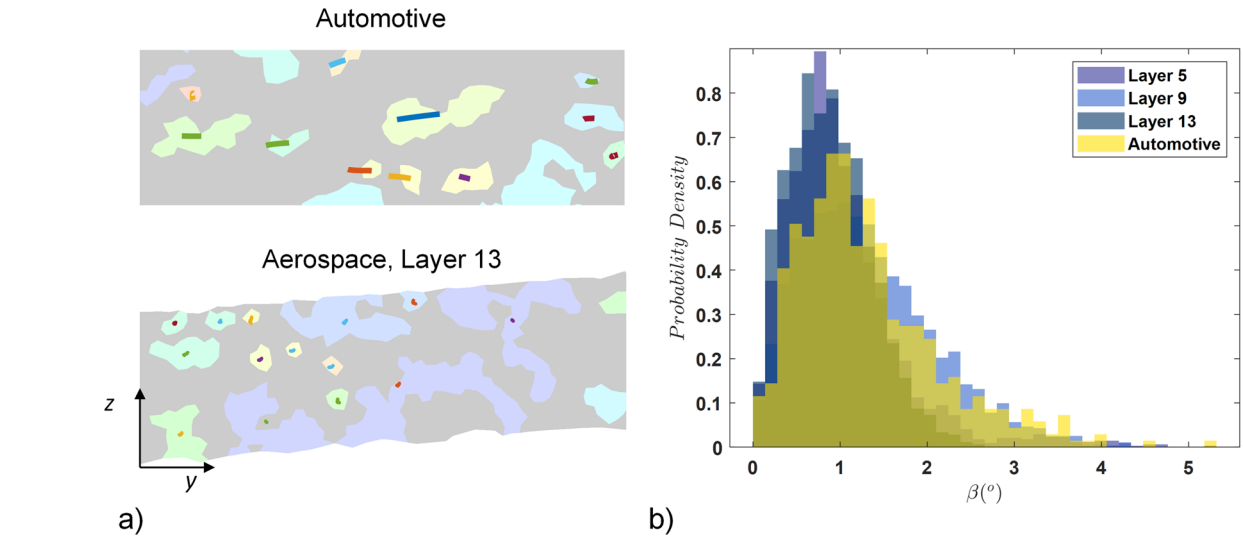

Fig. 9. (a) Centerline plots of the centroids for representative clusters in the automotive and aerospace, layer 13 samples. (b) Histogram of the angles between clusters, β.

each cluster were measured in each cross-section by summing up the areas of all triads in the clusters (Fig. 10c). The radius of gyration of each cluster was calculated in each cross-section by finding the root-mean-square distance of each cluster centroid to the centroid of the constituent triads (Fig. 10d).

The majority of the K values for both samples lay below 1, with the aerospace sample peaking just under. The distribution shows that the clusters in both samples tend heavily to be larger in the y direction, normal to compaction. Many of the aerospace clusters, however, lay just under a K value of 1. This means that aerospace samples each have clusters that are very symmetric about both axes, while the automotive clusters have a stronger preference for their direction.

The cluster volume fractions depict more of a difference between the two samples than the values for K. Due to the different manufacturing methods and the qualitative information obtained from the

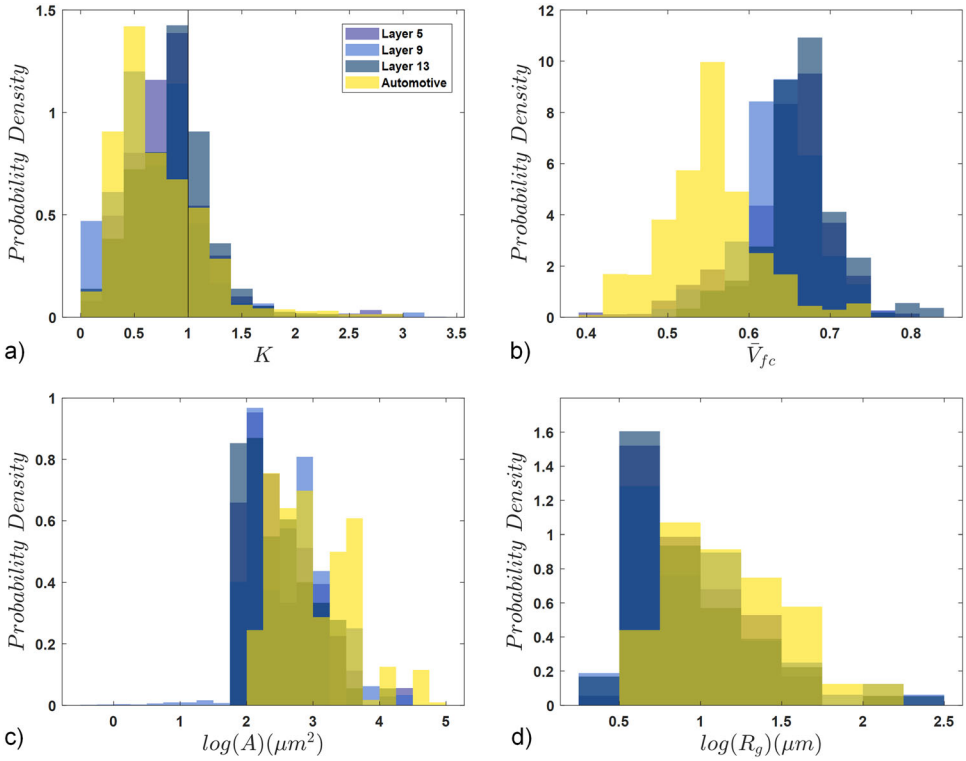

Fig. 10. Histograms of (a) ratio of the square roots of the moments of inertia, K, (b) cluster volume fraction, $\overline{V}_{fc}$, (c) area, A, of the clusters, and (d) the radius of gyration, R_g, of the clusters in each cross-section.

cross-sectional images, it is known that the levels of compaction between the two samples differed greatly. The story is the same when looking at the volume fractions of the clusters. The majority of the aerospace samples lay between cluster volume fraction values of 0.5 and 0.8, while the majority of automotive cluster volume fraction values were between 0.4 and 0.7. The peaks occurred at 0.65 for the aerospace samples and 0.55 for the automotive samples.

The data presented here characterize the clusters of the two samples as being oriented similarly with respect to angle, but different in the sense of geometry and density. The automotive clusters tended to be longer in the y direction and relatively large when compare with the aerospace samples. The aerospace samples were mostly oriented horizontally as well, but exhibited higher degrees of symmetry. The aerospace clusters were also found to be smaller, more numerous, and much more compacted than the automotive clusters. Further analysis is needed to fully understand the ramifications of these findings. Future work will include finite-element-based analyses of composite 3D composite RVEs that have been homogenized using the cluster formulation. The findings of this work and future analyses into other samples will serve as a blueprint for the design of clustered composite RVEs.

CONCLUSIONS

Composite microstructures are typically characterized by the overall fiber volume fraction and global fiber orientation, but these two measures alone may be insufficient for accurate representation. Different metrics to describe the structural distribution of fibers and their entanglement have been developed, but one higher-order structural feature that has been observed in many microstructures is the clustering of fibers, where groups of fibers remain together throughout the entire sample volume. While this can be recognized qualitatively, a method for identifying and isolating clusters was the focus of this study. Fiber clusters were identified by triangulating all fibers and studying the total variations of the area and perimeter of these triangles (or triads of fibers) in every cross-section. The triads were divided into three groups or bins based on the variation, and the thresholds between the bins was optimized to minimize the variance in each group. Using thresholds relative to the sample rather than absolute thresholds allows the identification of clusters regardless of global compaction levels and sample size. The triads in the bin with the lowest variance were selected as potentially part of a fiber cluster, and a filtering operation was applied to identify discrete clusters of fibers with low variance. The clusters were then isolated, and various metrics such as cluster

orientation, cluster symmetry, cluster fiber volume fraction, cluster density, and radius of gyration were calculated.

Microstructures from scans of two different composites, one from an automotive-grade and the other from an aerospace-grade composite, were analyzed for cluster metrics. Both analyzed samples had a uniform global fiber direction (ether unidirectional or constructed of unidirectional layers) but differed significantly in the level of compaction or global volume fraction, with the automotive composite being less compact than the aerospace composite. The analysis allowed a more subtle comparison of the two samples beyond global fiber volume fraction.

The metrics presented revealed some key differences and similarities between the automotive and aerospace samples. The aerospace samples had a larger density of clusters, smaller clusters, and higher volume fraction of clusters than the automotive sample. The automotive sample had a large variation of cluster sizes, with one large cluster pulling the average up. The orientation of fiber clusters was similar in both samples, with clusters varying less than 5° from each other. Additionally, metrics describing the cross-sectional shape of clusters showed equivalency between samples. The radius of gyration for both had similar values, and both showed anisotropy in fiber shape, with a tendency to be flatter in the direction of compaction (through-thickness). One would have to take more samples of each type of composite to thoroughly characterize the microstructure, but it was significant that the analysis showed differences in cluster geometry between the two samples nonetheless.

One major question which remains unanswered in this study is how far (in the fiber direction) the fiber clusters persist beyond the field of view of these particular scans. The scans were around 0.2 mm and 0.7 mm in the fiber direction, and longer scans would be needed to answer this question. It is currently unknown whether clusters eventually disperse, or stay together for extended lengths. Little is known to date regarding the morphology of individual fibers in the fiber direction, because continuous fibers would need to be scanned at much larger length scales, while still maintaining a high enough resolution transversely to recognize and isolate individual fibers.

One can speculate about the manufacturing origins of the clusters for the different samples. The automotive sample consisted of a heavy tow, which was processed in a bundle of around 48k fibers. During manufacturing, these tows are much wider than they are thick, which makes them susceptible to the tow folding in on itself. This mechanism may be one of the contributors of the larger, more distinct fiber clusters. The aerospace samples, on the other hand, were made from highly aligned prepregs compressed with a greater pressure during manufacturing. These fibers are generally more aligned, allowing greater compaction. The fiber clusters here may be more a product of either stray fiber individuals or shear planes of fibers shifted under high compaction pressures. Further work is needed to identify these causes, which could lead to tuning the processing towards achieving the desired fiber clustering.

While clusters exist within the microstructure of certain fiber-reinforced composites, further work needs to be done to understand how this feature affects macroscopic characteristics such as strength. While volume fraction and global orientation are well known to be influential metrics, the size, density, and alignment of fiber clusters may have a significant impact on the ability of a crack to travel transverse to the fibers. It may be difficult to determine such an effect experimentally, so it is expected that computational models will be pivotal.

The identification of fiber clusters could be used to perform a level of homogenization below typical composite mesoscales (tows/plies) but above the fiber microscale. This additional level could provide morphological context for local failure initiation, where perhaps failure initiation is not solely determined by the presence of a microscale flaw or stress riser, but where that initiation point is located in relation to larger features such as fiber clusters, matrix-rich regions, and larger voids. Additionally, these descriptors provide a more refined way to determine microstructural features and compare between processing conditions, as was done here. Identifying these morphological features such as fiber clusters and developing methodical ways to quantify them is an important link in the process–microstructure–properties chain, being one step closer to designing manufacturing processes to obtain desired material and part performance

ACKNOWLEDGEMENTS

This work is based upon work supported by the National Science Foundation and Air Force Office of Scientific Research under grant number IIP-1826232. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. The authors would like to thank Helga Krieger for her help in obtaining some of the micrographs and contributing to the early versions of the fiber tracking, and Jamal Husseini for his contributions to the cluster analysis code. The serial sections of the aerospace composite were provided by Mike Uchic and Craig Pryzbyla of the Air Force Research Laboratory at Wright-Patterson Air Force Base in Dayton, OH.

CONFLICT OF INTEREST

The authors declare they have no conflicts of interest.

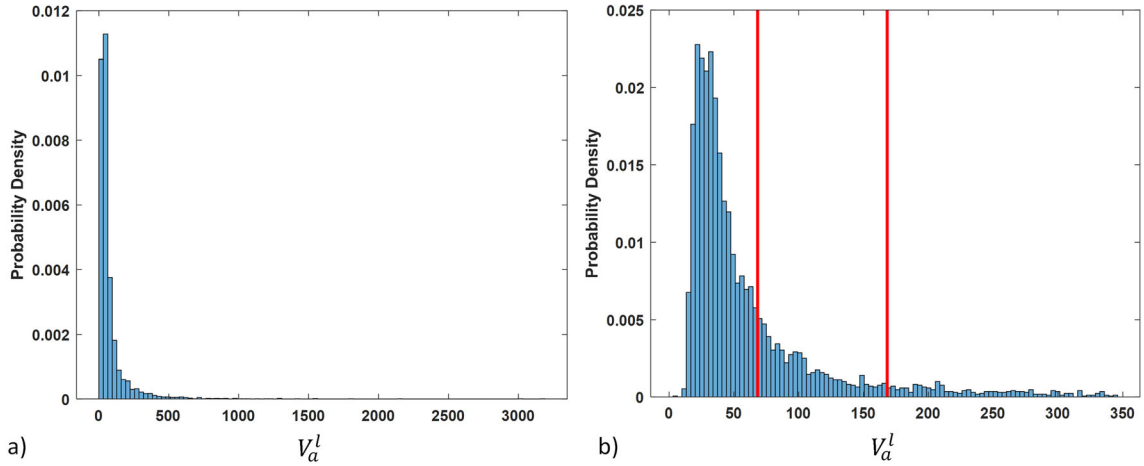

Fig. 11. Histogram plots of (a) untrimmed and (b) trimmed histograms with thresholds for total variation of area created using Otsu's method

APPENDIX

Fiber Path Extraction

To confirm the validity of the extracted fiber paths, a second method for fiber detection was used. The second method used a novel algorithm, coded in Mathematica, that mixed tools from mathematical morphology and traditional Gaussian filters. A median filter, with a disk structuring element, was applied to the images, followed by thresholding and Gaussian smoothing. The fiber centers were identified by finding the local maxima of the filtered image. The blending of image foreground and background made traditional gradient-based edge detection methods inapplicable. It was concluded that these optical artifacts were due to surface contamination since they did not persist from one cross-section to the next. The algorithm based on mathematical morphology is computationally more intensive but performs better on noisy images. These difficulties notwithstanding, the two methods were always within 1 pixel on the centers of the fiber cross-sections and agreed with each other on more than 90% of the disks.

Use of Otsu's Method

Otsu's method is typically used to threshold greyscale images for the purpose of, for example, separating background from foreground or isolating objects. It was recognized, however, that Otsu's method provides a reliable way of splitting histograms into specific regions. The implementation of Otsu's method used in this work is a MATLAB function, multithresh, which can split histograms into multiple regions based on the input number of thresholds.

To find the thresholds for V_A and V_P, histograms of both values were generated. For all samples, these histograms contained long, discontinuous tails filled with outliers. For each histogram, the average and standard deviation were calculated. The tails were then trimmed by filtering out all values greater than the average plus twice the standard deviation (Fig. 11).

The multithresh function was then used to split the total variation histograms into three bins. The multithresh function splits the distribution of values into k number of bins with the objective of minimizing the variation of values within each bin as expressed in

$$\arg\min_{S} \sum_{i=1}^{k} \sum_{x \in S_i} \|x - \mu_i\|^2 \tag{10}$$

where S is the set of partitioned values of the input data, x is a particular entry of S, also known as S_i, and μ_i is the mean of the points in x. This is also the function used in K-means clustering, a process in which groups of values are clustered together by sets of points in which the variance for those points is minimized.[35]

All triads with V_A and V_P values below the lower threshold were determined to be clustered triads. The triads in the two upper bins were not studied in this work. These bins are hypothesized to contain both intermediate triples, which contain relatively equal proportions of fiber and matrix, and matrix-rich triads. Future works will focus on what these bins truly represent.

Filtering

Filtering algorithms are applied to the clustered triads to remove extraneous fibers which do not appear to be clustered. The subtraction algorithm loops through each clustered triad and removes it from the cluster if two of its neighboring triads do not belong to a cluster. The addition algorithm loops through every nonclustered triad and adds it to a cluster if two of its neighbors belong to one. A generalized version of these two algorithms, where each triad becomes whatever two of its neighbor triads are, is used as a first step in the overall

filtering process. A final, subtractive filtering step is used to remove any remaining noise and allow the clusters to converge to their final shapes. The entire filtering process is then repeated until the number of clusters converges (Fig. 5e).

REFERENCES

1. F. H. Bhuiyan, S. H. R. Sanei, and R. S. Fertig, Compos. Struct. **237**, 111887 (2020).
2. B. Fiedler, M. Hojo, S. Ochiai, K. Schulte, and M. Ochi, *Compos. Sci. Tech.* 61, 95. (2001).
3. C.T. Sun, and R.S. Vaidya, *Compos. Sci. Tech.* 56, 171. (1996).
4. L. Bouaoune, Y. Brunet, A. El Moumen, T. Kanit, and H. Mazouz, *Compos. Part B* 103, 68. (2016).
5. J.C. Michel, H. Moulinec, and P. Suquet, *Comput. Method. Appl. M.* 172, 109. (1999).
6. O. van der Sluis, P.J.G. Schreurs, W.A.M. Brekelmans, and H.E.H. Meijer, *Mech. Mater.* 32, 449. (2000).
7. A. Wongsto, and S. Li, *Compos. A* 36, 1246. (2005).
8. S.H.R. Sanei, and R.S. Fertig, *Compos. Sci. Tech.* 117, 191. (2015).
9. P. Pineau, and F. Dau, *Comput. Method. Appl. M.* 241–244, 238. (2012).
10. X. Chen, and T.D. Papathanasiou, *Compos. Sci. Tech.* 64, 1101. (2004).
11. S. H. R. Sanei, E. Barsotti, D. Leonhardt, and R. Fertig III, J. Compos. Mater. **51**, (2016).
12. C. Fu and X. Wang, Compos. Struct. 113343 (2020).
13. M. Wang, P. Zhang, Q. Fei, and F. Guo, *Compos. Struct.* 227, 111287. (2019).
14. L. Wang, G. Nygren, R.L. Karkkainen, and Q. Yang, *Compos. Struct.* 230, 111462. (2019).
15. T.J. Vaughan, and C.T. McCarthy, *Compos. Part A* 42, 1217. (2011).
16. E. Totry, C. González, and J. Lorca, *Compos. Sci. Tech.* 68, 829. (2008).
17. F. Larsson, K. Runesson, S. Saroukhani, and R. Vafadari, *Comput. Method. Appl. M.* 200, 11. (2011).
18. Y. Swolfs, L. Gorbatikh, and I. Verpoest, *Compos. Sci. Tech.* 85, 10. (2013).
19. B. Fedulov, F. Antonov, A. Safonov, A. Ushakov, and S. Lomov, *J. Compos. Mater.* 49, 2887. (2015).
20. D. Wilhelmsson, D. Rikemanson, T. Bru, and L.E. Asp, *Compos. Struct.* 233, 111632. (2020).
21. T. Zheng, L. Guo, R. Sun, Z. Li, and H. Yu, *Compos. Part A* 143, 106295. (2021).
22. A. Li, J. Zhang, F. Zhang, L. Li, S. Zhu, and Y. Yang, *New Carbon Mater.* 35, 752. (2020).
23. T. Fast, A.E. Scott, H.A. Bale, and B.N. Cox, *J. Mater. Sci.* 50, 2370. (2015).
24. M.W. Czabaj, M.L. Riccio, and W.W. Whitacre, *Compos. Sci. Tech.* 105, 174. (2014).
25. P.J. Creveling, W.W. Whitacre, and M.W. Czabaj, *Compos. Part A* 126, 105606. (2019).
26. J. Weissenböck, A. Bhattacharya, B. Plank, C. Heinzl, and J. Kastner, *Case Stud. Nondestruct. Test. Eval.* 6, 39. (2016).
27. R.M. Sencu, Z. Yang, Y.C. Wang, P.J. Withers, C. Rau, A. Parson, and C. Soutis, *Compos. Part A* 91, 85. (2016).
28. A.E. Scott, I. Sinclair, S.M. Spearing, A. Thionnet, and A.R. Bunsell, *Compos. Part A* 43, 1514. (2012).
29. N.Q. Nguyen, M. Mehdikhani, I. Straumit, L. Gorbatikh, L. Lessard, and S.V. Lomov, *Compos. Part A* 104, 14. (2018).
30. D. Wilhelmsson, R. Gutkin, F. Edgren, and L.E. Asp, *Compos. Part A* 107, 665. (2018).
31. B. Larrañaga-Valsero, R.A. Smith, R.B. Tayong, A. Fernández-López, and A. Güemes, *Compos. Part A* 114, 225. (2018).
32. B.A. Bednarcyk, J. Aboudi, and S.M. Arnold, *Int. J. Solids Struct.* 69–70, 311. (2015).
33. J.E. Spowart, H.E. Mullens, and B.T. Puchala, *JOM* 55, 35. (2003).
34. N. Otsu, *IEEE Trans. Syst. Man Cybern.* 9, 62. (1979).
35. J. Macqueen, in *5th Berkeley Symposium on Mathematical Statistics and Probability* (1967), pp. 281–297.