

RESEARCH

Open Access

Novel microstructure quantification framework for databasing, visualization, and analysis of microstructure data

Stephen R Niezgoda^{1,2*}, Anand K Kanjarla³ and Surya R Kalidindi⁴

*Correspondence:

niezgoda.s@gmail.com

¹Materials Science and Technology Division, Los Alamos National Laboratory, Los Alamos NM 87545, USA

²Present Address: Department of Materials Science and Engineering, Department of Mechanical and Aerospace Engineering, The Ohio State University, Columbus, OH 43218, USA

Full list of author information is available at the end of the article

Abstract

The study of microstructure and its relation to properties and performance is the defining concept in the field of materials science and engineering. Despite the paramount importance of microstructure to the field, a rigorous systematic framework for the quantitative comparison of microstructures from different material classes has yet to be adopted. In this paper, the authors develop and present a novel microstructure quantification framework that facilitates the visualization of complex microstructure relationships, both within a material class and across multiple material classes. This framework, based on the stochastic process representation of microstructure, serves as a natural environment for developing relational statistical analyses, for establishing quantitative microstructure descriptors. In addition, it will be shown that this new framework can be used to link microstructure visualizations with properties to develop reduced-order microstructure-property linkages and performance models.

Keywords: Microstructure; Two-point correlations; Microstructure database; Structure-property relationships; Reduced-order models

Background

It is well understood that advances in the development of materials with enhanced performance characteristics have been critical in the successful development of advanced technology, and are important drivers for continued economic prosperity. It is also widely recognized that modern materials cannot be understood through their chemistry and bulk processing alone, but that the key to continued material development lies in understanding and optimizing the myriad details of the hierarchical three-dimensional (3-D) internal structure, or microstructure, which spans several disparate length scales (from the electronic to the macroscale). While developing microstructure/processing/property linkages is the central theme in materials science, as a field we are only beginning to develop the tools necessary to truly explore and harness multi-scale microstructure sensitive design and manufacturing. Given the complexity of microstructure in virtually all engineering and natural materials, the traditional approaches to materials development, relying on combinatoric experimentation guided by engineering principles and physical intuition, have only categorized and exploited a small handful of the readily accessible material microstructures.

In recent years, impressive advances have been made in materials characterization as well as the development of sophisticated physics-based multi-scale modeling and simulation tools. Through these advents, materials science is undergoing a transition from a data-limited field to a data-driven but analysis-limited field. It is now possible to automate acquisition of large experimental and simulation datasets, and our ability to generate data is rapidly outstripping our ability to process it. Radically new approaches are required to integrate the looming deluge of data from advanced simulation and characterization techniques into *useful materials knowledge*. While this data crisis is a significant technical challenge, it is also an opportunity to explore completely novel inverse approaches to material design and deployment while at the same time reducing our dependency on slow combinatoric experimental approaches. This realization has been highlighted via the Materials Genome Initiative for Global Competiveness [1], the DOE Needs Reports on Computational Materials Science and Chemistry [2] and the continued growth of Integrated Computational Materials Engineering (ICME) [3]. It is widely realized that the community must develop a Materials Innovation Infrastructure (MII) or Materials Innovation Ecosystem [1,2] to exploit fully advanced simulation and coupled experiments, improve predictive capabilities, and provide the design, certification and monitoring tools for rapid and holistic materials development and deployment.

While the need for and end goals of a practical MII are well articulated, its requirements and components are largely undefined. It is envisioned that a successful MII framework will accelerate the development and deployment of materials by reducing/replacing expensive and evolutionary empirical experimentation with revolutionary optimization of structure and processing through simulations that are verified and validated with key experiments. The absence of a general framework for categorizing and visualizing the space spanned by collected microstructure data, understanding the statistical properties and variability inherent in materials, and efficiently integrating digital materials data with the disparate components of multi-scale modeling frameworks is a significant barrier to the development of a practical MII and to the overall goals of the Materials Genome Initiative.

The creation and curation of large scale materials databases has been widely cited as a critical required component for the acceleration of materials development and deployment [1,2] and has been a recognized need by the materials community since the 1970's [4]. Prior efforts at the creation and maintenance of large scale materials databases, such as the efforts of ASTM Committee E-49 for the Computerization of Materials Property Data [5] and the National Materials Property Data Network [6] (which was operated commercially from the mid 1980's until 1995), were largely focused on chemical composition, average properties and effective material response. A shortcoming of these efforts, largely due to the limited computing power of the times, was that details of the materials internal structure or microstructure was not considered and captured alongside the composition and properties. Understanding the relationship between structure, properties and processing is the central theme in materials science and engineering, and it is critical that future materials databases effectively capture all three legs of the triangle if such databases are to be successfully applied to materials design applications. While a significant amount of development, across the entire materials community, is required to even outline the structure and specific requirements of such a database, this paper focuses on

one critical area; namely, the definition and description of a computational framework and tools for the visualization, analysis, and quantitative comparison of microstructure datasets (collected across several different “materials”) consisting of ensembles of 2D or 3D micrographs collected for each material.

The work addresses several aspects necessary for the development and deployment of successful materials databases, or more exactly microstructure databases. In particular, we seek to address the following questions: 1) What is the appropriate way to represent the microstructure for inclusion into the database? 2) How do we describe the inherent structural variability in a material, and how do we connect this variability to variability in properties? 3) How do we quantitatively compare ensembles of material volumes from different samples or material systems? 4) How do we visualize the relationship between materials and the span of collected microstructure data? 5) How do we determine if additional characterization would add to our knowledge base or if it would be redundant? In this work we demonstrate a reduced-order microstructure visualization space, built upon a stochastic process representation of microstructure. It is envisioned that this visualization space will serve as the interface between the large scale database of microstructure data and the user, and will provide tools that will allow the user to interact with microstructure data in new ways that go far beyond those accomplished traditionally by visual inspection of micrographs or three-dimensional data sets.

Previous work from the authors and others laid the groundwork for reduced-order representation and categorization of microstructure data, on which many of the ideas presented in the manuscript build. Sundararaghavan and Zabaras proposed the construction of a material library from a reduced-order representation of digital micrographs, and demonstrated the application of supervised learning techniques such as support vector machines for material classification [7,8]. Ganapathysubramanian and Zabaras explored more advanced non-linear dimensionality reduction schemes and used these representations to construct inputs to stochastic multi-scale models [9,10]. Niezgoda [11] and Niezgoda, Yabansu and Kalidindi [12] formalized the description of microstructure via statistical metrics into a stochastic process interpretation, and developed the basic theory for the microstructure visualization space presented in this work. Additionally, they developed relationships between the observed variance in microstructure for a material ensemble and the corresponding ensemble variance in properties/performance and proposed applications of the microstructure visualization space in materials design and manufacturing quality control and process monitoring.

While the previous work largely focused on developing the microstructure space and visualizations for ensembles of a single material or microstructure class, this work is focused on the extension of the above concepts to the simultaneous categorization, analysis and visualization of multiple materials, the development of descriptive and relational statistics that facilitate the exploration of microstructure relationships between different materials, and the exploration of microstructure/property relationships in the microstructure visualization space. The main purpose of this manuscript is to present the philosophy of the visualization space and the mathematical formulation that underlies its construction, followed by some simple examples of the type of microstructure analyses and property explorations that can be performed in the space. The examples are chosen to show the utility of the proposed space, rather than to suggest a specific analysis procedure or regime. The main idea is that the proposed visualization space is a platform upon

which a large suite of analysis tools can be built, and where the analysis can be tailored for a specific material system and design problem.

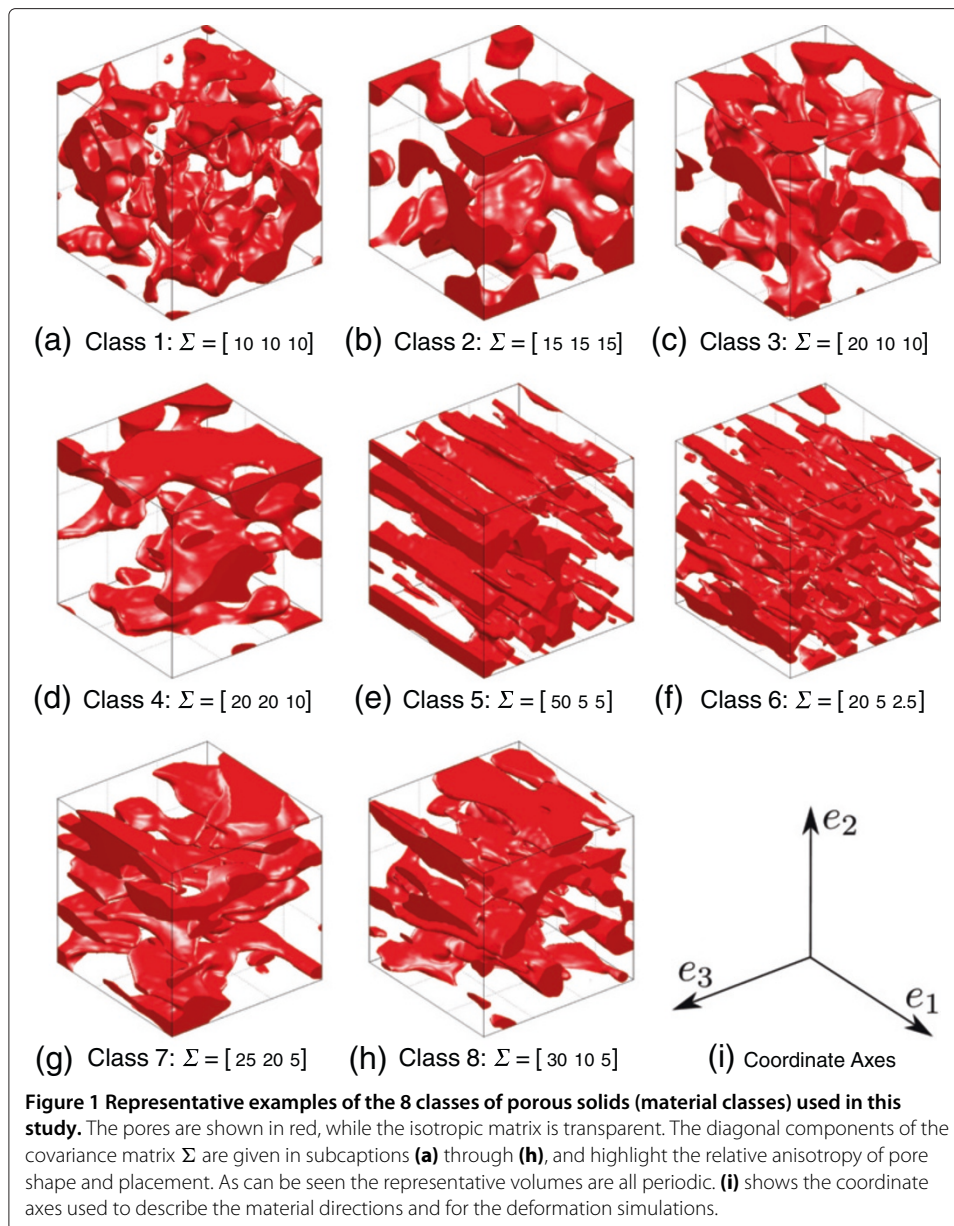
The following key ideas and features behind the proposed microstructure visualization space will be explored in detail throughout the remainder of this paper: 1) The understanding that microstructure possesses an inherent randomness suggests that the mathematics of stochastic processes is a natural description of the multi-scale variability of materials. 2) While powerful, such an approach leads to significant abstraction of the data and the resulting data-set lies on the surface of a complex high-dimensional manifold which is not intuitive to many materials, design, or manufacturing practitioners. In order to simplify the presentation of the data a reduced-order representation of the microstructure data is required. Here we apply principal component analysis to construct the reduced order representation. 3) The reduced order representation forms the basis for the visualization space. Each sample is represented by a single point in the space. Points that are near each other correspond to samples with similar structures, disparate points correspond to samples with drastically different structures. 4) The space spanned by all the characterized samples from a material or material class gives an indication of the microstructure variability within that material. This scatter in the visualization space can be directly tied to the scatter in properties or response of the material. 5) Comparison tools based on relational statistics can be easily developed for further analysis in the space. Additional tasks such as automated classification or clustering analysis are naturally formulated in the visualization space. 6) Explicit mapping of properties or performance characteristics into the visualization space can be performed to formulate invertible structure/property linkages.

Methods

Microstructure data generation and modeling methodology

For this study ensembles of 8 different material classes were generated by thresholding random fields. The 8 material classes were all porous composites with varying degrees of anisotropy in pore shape, connectivity, and spatial distribution. In order to prevent variations in porosity between classes from exerting an overwhelming influence on property calculations and to highlight the higher-order effects of the spatial distribution of pores, the volume fraction of pores was controlled to be nominally 25% and normally distributed with a standard deviation of $\sigma = 8.7\%$ across all material classes. $64 \times 64 \times 64$ arrays were populated by sampling from a uniform distribution $\mathcal{U}(0, 1)$, the field was then locally averaged by circular convolution with an anisotropic 3-D Gaussian filter with a diagonal covariance matrix, Σ . The resulting periodic random field was thresholded, with the threshold value sampled from a normal distribution, $\mathcal{N}(\mu = 0.25, \sigma^2 = 0.0075)$. Representative realizations of the 8 material classes are shown in Figure 1. For each class two ensembles of 50 volumes were generated. The first ensemble was used to create the microstructure database and calibrate the classification and property models, while the remainder were reserved for validation.

The mechanical response of each volume element was evaluated using an image based fast Fourier transform (FFT) elastic-viscoplastic model, that directly uses the three dimensional voxel material volumes as input. This type of model was first developed by Suquet and collaborators [13], for nonlinear composites, and further developed by Lebensohn and collaborators for polycrystals [14]. For this work we adopted



the elasto-viscoplastic formulation of Lebensohn and colleagues [15,16], to generalized power-law composites. The FFT-based algorithm computes a compatible strain-rate field that minimizes the average work rate under the constraints of the constitutive relation and stress equilibrium. The FFT method is predicated on the observation that the mechanical response of a heterogeneous non-linear medium can be calculated as a convolution between a linear reference material and a polarization field, which describes the non-linearity and heterogeneity of the actual composite material response. The FFT name comes from the fact that the resulting stress equilibrium equations take a computationally convenient form when cast in terms of Fourier transforms.

The matrix material is considered as elastically and plastically isotropic with a Young's modulus of 40GPa and a Poisson's ratio of 0.3. The matrix had an initial yield strength,

σ_0 , of 45 MPa. A visco-plastic constitutive model was adapted to relate the stress to the plastic strain rate as

$$\dot{\varepsilon}_{ij}^p = \frac{3}{2} \dot{\varepsilon}_0 \left(\frac{\sigma_{ef}}{\sigma_0} \right)^n \frac{\sigma'_{ij}}{\sigma_{ef}} = C^* (\sigma_{ef})^{n-1} \sigma'_{ij} \quad (1)$$

where $\dot{\varepsilon}_{ij}^p$ is the viscoplastic strain rate, σ'_{ij} is the deviatoric stress tensor, $\dot{\varepsilon}_0 = 1$ is a reference strain rate, and σ_{ef} is the scalar Von Mises effective stress. C^* can be considered as an effective plastic modulus given by $C^* = \frac{3}{2} \dot{\varepsilon}_0 (\sigma_0)^{-n}$.

Stochastic process representation of microstructure

The mathematical formalism of stochastic processes has been demonstrated to be a powerful tool for the quantification of microstructure variability and variance [12]. Here we present a brief overview of the theory as required for the applications presented in this manuscript; for a more complete overview of the theory of stochastic processes and the application to microstructure modeling please see [11,12]. Consider a probability space defined by the ordered triplet $(\Omega, \mathcal{F}, \mathcal{P})$. The first element Ω is termed the sample space and is a non-empty set of experimental outcomes or observations, individually denoted as ω . \mathcal{F} is the set of all theoretically possible events and is formally defined as a Borel σ -algebra [17], and \mathcal{P} denotes the standard probability measure on \mathcal{F} . A random variable \mathbf{x} is then defined as a function, with domain \mathcal{F} , which maps to each experimental outcome ω a number $\mathbf{x}(\omega)$ such that the axioms of probability are satisfied. The probability distribution function (pdf), $f(\mathbf{x})$, associates each possible experimental outcome with a probability such that $\mathcal{P}\{x_1 \leq \mathbf{x} < x_2\} = \int_{x_1}^{x_2} f_{\mathbf{x}}(x) dx$.

A stochastic process, $\mathbf{x}(t)$, is by extension a set of rules which assign a function $x(t, \omega)$ to every experimental outcome ω of experiment Ω . These rules take the form of associated probability distributions. To completely determine the statistical properties of a stochastic process the n^{th} order distribution function $f(x_1, \dots, x_n; t_1, \dots, t_n)$ must be known for all t_i , x_i , and n . For virtually all processes (and microstructures) this is impossible. Instead, attention is usually restricted to lower order statistical measures such as the mean and autocorrelation [12,18].

The microstructure of virtually all materials exhibit rich details that span several length scales. A microstructure constituent that can be assigned a distinct local structure can be considered to possess a distinct local state. The local state is denoted h and can be considered an element of a local state space, H , that identifies the complete set of local states that could theoretically be encountered. For example, the local state at point x in a polycrystalline metal sample may be defined by the thermodynamic phase ρ , the local lattice orientation g , the state of dislocation α etc. In this way the local state may be considered a random vector $\mathbf{h}(x) = [\rho, \mathbf{g}, \alpha, \dots]$, and assuming a stationary process the local state distribution $f_{\mathbf{h}}(h)$ gives the volume density of material points in a volume with local state h . A common example of local state distribution is the well known orientation distribution function used to quantify preferred crystallographic orientations in polycrystalline materials [19].

In terms of the probability space defined above, the set of all possible events \mathcal{F} is the set of all possible spatial arrangements of local states $h \in H$. The sample space Ω consists of an ensemble of microstructure realizations, where each microstructure realization is considered an experimental outcome ω . The microstructure can then be interpreted as

a stochastic process $\mathbf{m}(x, h)$, sometimes termed the microstructure function, that assigns a local state distribution field $m(x, h, \omega)$ to every realization. It is important to note that the stochastic process $\mathbf{m}(x, h)$ cannot be observed directly. The microstructure function is best understood as a series of higher order probability distributions which describe how the local states are placed in a material relative to each other. Instead, the local state fields $m(x, h, \omega)$ can be observed for an ensemble, and can be used to estimate the microstructure statistics.

Statistics of the microstructure

Consider a material system where the local state is defined by a combination of k features of interest, thus the local state is described by the random vector $\mathbf{h} = [\beta_1, \dots, \beta_k]$. The probability of finding local state \mathbf{h} at material point x is in a region of local state space \mathcal{H} is given by

$$\begin{aligned} \mathcal{P}\{\mathbf{h}(x) \in \mathcal{H}\} &= \int_{\mathcal{H}} f(h, x) dh \\ &= \int_{\mathcal{H}} f(\beta_1, \beta_2, \dots, \beta_k, x) d\beta_1 d\beta_2 \dots d\beta_k \end{aligned} \quad (2)$$

where the PDF $f(h, x)$ is referred to as the first order density of the microstructure function, and can be interpreted as the spatially resolved volume fraction of local state $\mathbf{h} = h$.

Estimation of 2nd order PDFs from characterized microstructure realizations, is often impractical and estimation of higher order PDFs ($n \geq 2$) is often impossible. Fortunately, a framework of hierarchical spatial statistics of the microstructure is available in the literature in the form of n -point correlation functions [20-25]. Local state distributions, $f(h)$, are often termed one-point statistics, as they reflect the probability density of finding a specific local state at a randomly selected point in the microstructure.

Expanding on this idea, the two-point correlation is the joint density of occurrence of local state h_1 and h_2 at material points separated by the vector r

$$f_2(h_1, h_2|r) = E\{\mathbf{m}(x, h_1)\mathbf{m}(x+r, h_2)\} \quad (3)$$

The microstructure statistics must be estimated from an ensemble of characterized realizations. Consider an ensemble of P volumetric realizations $\Omega = (\omega_1, \dots, \omega_p)$, where the p^{th} realization has an associated local state field $m(x, h, \omega_p)$. The two point correlation for the microstructure can be estimated as

$$\begin{aligned} f_2(h_1, h_2|r) &\approx \widehat{f}_2(h_1, h_2|r) = \langle f_2(h_1, h_2|r, \omega_p) \rangle \\ &= \frac{1}{P} \left[\sum_{p=1}^P \frac{1}{\text{vol}(\omega_p|r)} \int_{x \in \omega_p|r} m(x, h_1, \omega_p) m(x+r, h_2, \omega_p) dx \right] \end{aligned} \quad (4)$$

where $\omega_p|r = \{x|x \in \omega_p \cap x+r \in \omega_p\}$. $f_2(h_1, h_2|r, \omega_p)$ is the estimate of the two-point correlation obtained from a single realization or volume. Equation (4) is readily identified as a convolution integral which is readily computed via Fourier transform techniques [22, 25]. Three-point and higher order correlations are defined analogously, however, for the remainder of this paper the discussion will be limited to 2-point correlations.

Results and discussion

The microstructure space

Principal component representation

The set of physically meaningful n -point correlation functions lie in a very high dimensional space that is not amenable for efficient computation, visualization, analysis, or other design purposes [25]. For even the simple material systems considered here, where the local state is either pore or solid, and restricting the statistical description to two-point correlations, some form of reduced order representation of the microstructure statistics is required. While there are numerous approaches to dimensionality reduction including spectral methods, manifold learning approaches, metric multidimensional scaling, among others [26], the requirements of the proposed microstructure space places strong constraints on the type and complexity of the dimensionality reduction scheme. In order to determine an appropriate low dimensional representation the following factors were considered: 1) The approach must be computationally robust and insensitive to the type or structure of the underlying data. 2) The representation must be able to be updated in real time as new information or datasets are added to the system 3) The representation must be invertible in real time, so that the microstructure space can be used to explore new microstructures in both an interpolative or extrapolative manner. 4) In order to facilitate the computation of descriptive and relational statistics the representation should be an orthogonal decomposition of the data so that each dimension in the reduced frame can be considered as independent variables. Based on these requirements and considerations of computational complexity, and the excellent performance in previous work [12], dimensionality reduction via principal component analysis (PCA) was chosen for this study.

PCA is a linear approach to dimensionality reduction which can be understood as a coordinate transformation that maps a set of (possibly) correlated variables onto a new set of orthogonal (independent) variables. This is most easily understood as projection of a high dimensional dataset onto a new orthogonal coordinate frame where the axes are defined by the directions of highest variance [27]. PCA has been likened to a shadow of the data cast from its most informative projection. One key strength of PCA is that it forms a natural basis for the exploration of microstructure variance as the principal directions align with the directions of highest variance in the microstructure data [12]. A significant drawback to using PCA is that, as the microstructure data (2-point correlations) is inherently non-linear (meaning the data naturally lies on an embedded curved manifold in the high dimensional space rather than on a hyper-plane), the resulting representation will not necessarily be the most compact or efficient. However it will be shown that when the different materials are of the same “family”, such as the various classes of porous composites explored here, a suitably compact representation can be developed and useful visualizations and maps can be developed in just a few dimensions. If the range of structures were to be expanded to include multiple material “families” then more advanced dimensionality reduction approaches would be required. However, in practice, for most materials design applications the materials system is fixed as a design constraint and the goal is optimizing processing or chemistry to achieve a target microstructure or properties. For these cases the PCA representation is expected to perform quite well.

Consider an ensemble of P realizations or volumes from a single material class (in multiple volumes cut from a single large sample or multiple samples with the same nominal

processing history) denoted $\omega_1, \omega_2, \dots, \omega_p$ each with associated two-point correlation estimate $f_2(h_1, h_2|\omega_p)$. The PCA representation of the correlations measured from the p^{th} member of the ensemble can be written as

$$f_2(h_1, h_2|r, \omega_p) = \sum_{j=1}^{P-1} \alpha_j^p \phi_j + \widehat{f}_2(h_1, h_2|r) \quad (5)$$

where $\widehat{f}_2(h_1, h_2|r)$ is the ensemble average of the overall microstructure statistics (see Equation (4)). ϕ_j are the orthogonal principal component vectors and α_j^p are the corresponding weights or the PCA representation of the p^{th} member of the ensemble. Mathematically the decomposition consists of a few basic steps

1. Mean center the data

$$\Phi^p = f_2(h_1, h_2|r, \omega_p) - \widehat{f}_2(h_1, h_2|r) \quad (6)$$

2. Compute the covariance matrix of the mean centered data

$$C = \frac{1}{P} \sum_{p=1}^P \Phi^p (\Phi^p)^T \quad (7)$$

3. Perform an eigenvalue decomposition

$$C\phi_j = b_j\phi_j \quad (8)$$

4. Project Φ^p onto the eigenvectors

$$\alpha_j^p = (\phi_j)^T \Phi^p \quad (9)$$

In practice, it is not necessary to explicitly construct the covariance matrix C . Instead, for large datasets, an algorithm called the method of snapshots is used to reduce the computational burden [28]. Additionally, it is possible to incrementally build the basis (realization by realization) by taking advantage of the understanding that any portion of new data not in the span of the eigenvectors must be orthogonal to the current basis [29]. Building the basis in this manner has the advantage that addition of new realizations does not require the re-computation from scratch. Assuming no linear dependencies, the rank of C is $P - 1$, implying that the maximum number of parameters necessary to represent the data is approximately the number of members in the ensemble. The eigenvalues of the decomposition are an indication of the significance of that principal component (degree of variance in the data). By taking only the components with the highest eigenvalues it is often possible to approximate the data in a handful of parameters.

For this study, two PCA decompositions were performed to generate a two tiered representation. First an intra-class decomposition was performed on the members of each material class. The intra-class PCA representation will be used to explore microstructure variance within a material system, and to identify interesting or representative members of a specific class. Then an inter-class PCA decomposition was performed over all the realization of all classes. The inter-class representation will be used to explore relationships between the individual classes and to develop property relationships across multiple material systems. The inter-class representation was truncated to 50 principal components, primarily for computational reasons. However, it will be shown that for this example fewer than 20 components are sufficient for the classification and property modeling examples presented here. 50 realizations from each materials class were used to

build the inter-class representation, the additional 50 members were reserved to validate the representation and microstructure classification scheme.

Visualizations in the microstructure space

The inter-class PCA weights, α_j^p , are an explicit representation of the individual microstructure realizations in an “optimal” orthogonal reference frame defined by the eigenvectors of the data covariance matrix. This reference frame spans the space of the collected microstructure datasets (up to the error introduced by truncation), and is a natural setting in which to visualize and explore the range of data collected, the relationships between datasets and the variability within material classes. By projecting the microstructure data onto the first three eigenvectors, microstructure maps can be constructed. Other operations such as descriptive and relational statistical analysis and the development of homogenized property relationships can be performed in this microstructure map space. As such, these maps and visualizations are a central result of this work. The projection of the first eight microstructure classes onto the inter-class PCA basis is shown in Figure 2. As a reminder, each data point in the figure corresponds to a single material realization. Convex hulls bounding the realizations of each microstructure class are plotted to help visualize the differences between the volume of space occupied by the members of each class.

An advantage of the PCA representation is that the axes are naturally ordered by significance in that the first principal component is the direction of highest variance and the 2nd is the direction of highest variance orthogonal to the first and so on. By choosing the first three components or directions for visualizing the projected microstructure data we are selecting our vantage point so that we are capturing the view which captures the most variability in the collected microstructure data. This does not, however, guarantee that the view is optimal for separating the different realizations by microstructure class. As can be seen in Figure 2 the different microstructure classes are indeed heavily overlapped in three dimensions, indicating that more terms are needed to delineate the different microstructure classes. It will be shown for this particular example that 15 principal components are sufficient to build a robust classification model for these eight material classes (see Section “Support vector machines for classification of microstructures”).

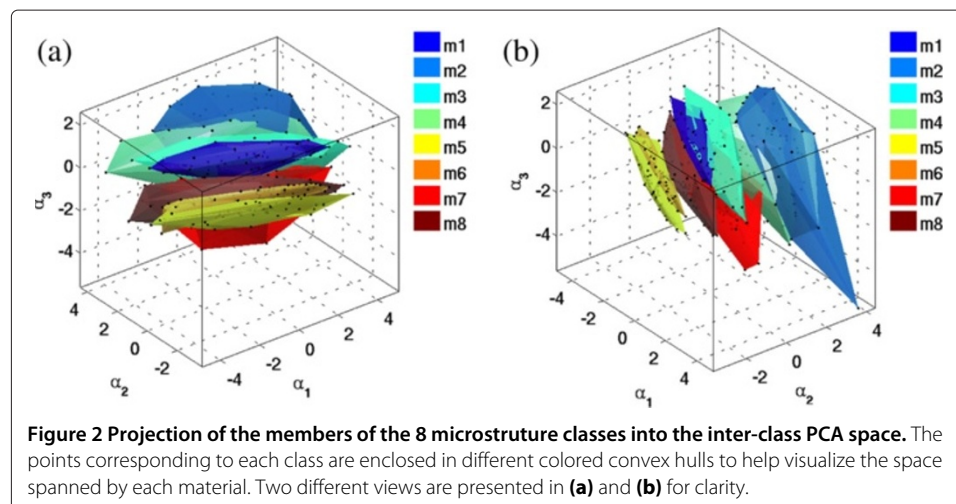
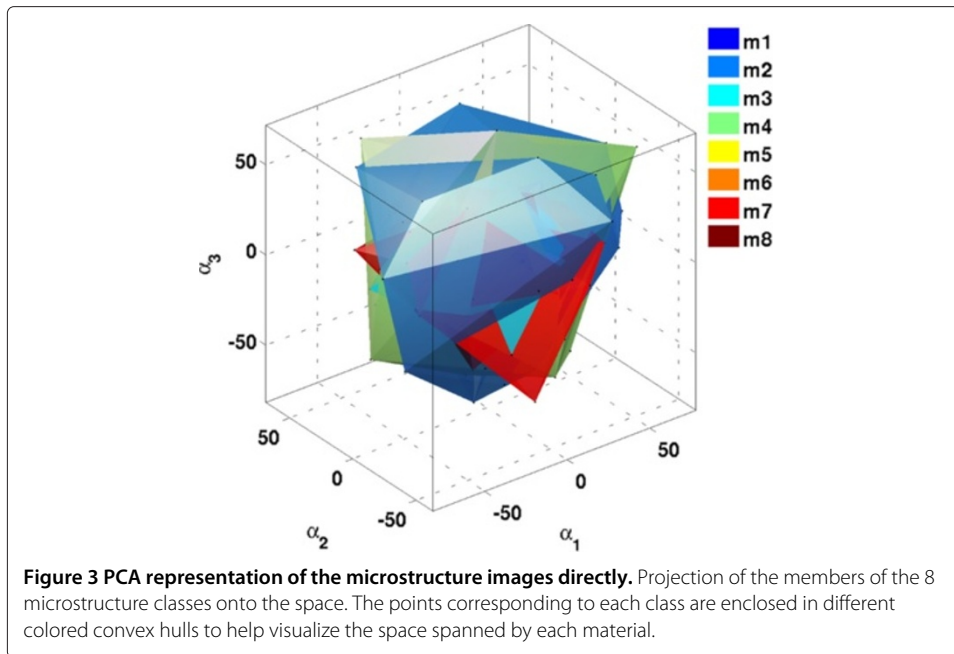


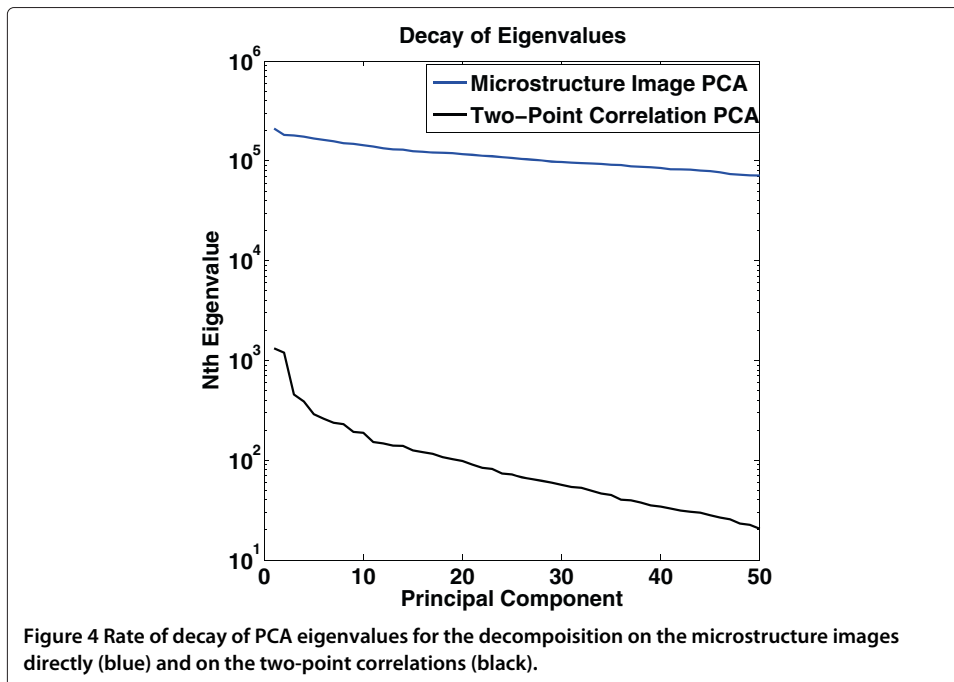
Figure 2 is a key result of this work. At a glance it shows the relationships between the different microstructure groups. Distances in this space (defined below) serve as an indication of how similar or different structures are. Material classes whose bounding hulls are close (or overlapping) share more common features than those classes whose bounding hulls are well separated. This can be seen in Figure 2(b) which shows the hulls bounding classes 5 and 6, which share a highly anisotropic elongated pore shape, are overlapping and are well offset from the classes 1 and 2 which have an isotropic pore shape. The center of mass of the cloud of points for each class, serves as an estimate of the average or representative material volume for that class. The volume of the hull bounding a microstructure class also serves as a qualitative measure of the inherent structural variability of the members of a material class relative to the other classes. Again, the figure shows that the classes 5 and 6 exhibit less structural variability than the other material classes. Microstructural clustering or the uneven dispersion of points within the hull, can be also readily observed. While not the focus of this paper, the scatter in microstructure realizations collected for a single class allows the ready identification of structural outliers or realizations expected to exhibit performance or property values far from the mean. Previous work from the authors focused on intra-class variance and the development of tools for analysis of multiple ensembles of a single microstructure class [11,12].

Sundararaghavan and Zabarar also used PCA for reduced order representation for the construction of microstructure libraries [7,8]. In that work they chose to perform the decomposition on the characterized microstructures directly, rather than working with statistical descriptors as we do here. While the motivations behind this work and theirs are substantially different, it is worthwhile briefly discussing some of the advantages and disadvantages of each approach. PCA is a linear transformation and thus effective dimensionality reduction can only be accomplished if the data can be approximately fitted to an embedded linear manifold (hyperplane) in the high dimensional space. Unfortunately, as the ongoing work of Zabarar group has demonstrated, the underlying data in microstructure datasets often lies on an embedded highly non-linear surface and large numbers of principal components must be kept for reasonable representations or non-linear data mappings must be employed [10]. Some basic properties of the n -point correlations, including having a natural origin at $r = 0$ and translation invariance, greatly reduce this non-linearity and adequate representation can be produced with only a few principal components. For comparison, a PCA decomposition was done directly on the 3-D microstructure datasets and the resulting projection into the PCA space is shown in Figure 3. As can be seen in the figure, the hulls bounding the microstructure classes are heavily overlapped and the relationships between the classes is not readily apparent from a view of the first three dimensions of the space. The eigenvalues of the decomposition give an indication of how significant each principal component is to the representation of the data-set, and the rate of decay of the eigenvalues gives an indication of how many terms must be kept for accurate representation. Figure 4 shows the eigenvalues for the first 50 principal components for the inter-class PCA, performed on the spatial correlations (described earlier), and the PCA performed on the microstructure images. For the inter-class representation constructed on 2-point correlations there is a sharp drop in the eigenvalues within the first 10 principal components while there is no such drop in the representation constructed from the images themselves. The slow decay is highlighted by the ratio between eigenvalues, for the decomposition on the structures



$\alpha_1/\alpha_{50} \approx 3$, while for the decomposition on the statistics $\alpha_1/\alpha_{50} \approx 65$. These results indicate a much stronger concentration of the representational power in the first few principal components when the microstructure statistics are used to construct the reduced order representation.

Working in the correlation space does, however, add a significant abstraction. When working directly with the microstructure realizations, once the basis is defined, new realizations can be created for any point in the spanned space by simple linear combinations



of the eigenvectors. When working with the correlations each point in the spanned space represents a set of potential microstructure correlations, and new microstructure realizations must be reconstructed from the statistics. Reconstruction from statistics is an active area of research, and significant advances have been made in recent years [30,31]. In the opinion of the authors the benefits of a more compact linear (PCA) representation and a cohesive framework based on the formalism of stochastic processes outweigh the added abstraction.

Relational statistics in the microstructure space

In order to effectively utilize the PCA space for microstructure quantification or comparison of structures, a measure of distance in the space must be introduced. For intra-group comparison between individual points (i.e. how similar two points are) the Euclidean distance is adequate. However in this work we are interested in comparisons between groups of points and for the application of microstructure classification we are interested in the question of how close are the clouds of points representing two material classes or how close is a new realization from an unknown class to the cloud of points from a known class. For this we apply the Mahalanobis distance, which gauges the similarity between an unknown sample (or sample set) to a known or classified sample set [32]. Consider a multivariate vector $x = (x_1, \dots, x_n)^T$, the Mahalanobis distance between x and a distribution of vectors $f(y)$ with mean μ_y and covariance matrix C_y is defined as

$$D_M(x, f(y)) = \sqrt{(x - \mu_y)^T C_y^{-1} (x - \mu_y)} \quad (10)$$

The feature of the Mahalanobis distance is that it takes the shape of the known probability distribution into account, in that all points that lie on the same iso-probability surface of the known distribution will have the same Mahalanobis distance to that distribution regardless of the shape of the distribution^a. The Mahalanobis distance can be generalized to a similarity measure between two multivariate vectors x and y , where y comes from a distribution $f(y)$ with covariance C_y as

$$d(x, y) = \sqrt{(x - y)^T C_y^{-1} (x - y)} \quad (11)$$

When computing distances in the PCA space the individual components of x and y , x_i and y_i are independent (orthogonal) variables and the covariance matrix is by definition diagonal. In this case $d(x, y)$ becomes the normalized Euclidean distance, $d(x, y) = \sqrt{\sum_i (x_i - y_i)^2 / \sigma_i^2}$, where σ_i is the standard deviation of $f(y_i)$. In the limit that the covariance matrix is the identity matrix (e.g. for the standard multivariate normal distribution) the Mahalanobis distance reduces to the standard Euclidean distance.

When working with ensembles of characterized material volumes in the design space, the first natural question to ask is whether two samples (material volume ensembles) truly possess different microstructures or have the same microstructure but different structural variance. In other words, how likely is it that the two ensembles share the same mean or that same realization can serve as a representative volume for both ensembles. For quality control and process monitoring applications it is important to have a quantitative measure of drift in the manufacturing process or be able to compare material from different processing lines. It is a natural question to ask if process paths 1 and 2 are indeed producing material with the same microstructure, and if not how different are they? If they are

confirmed to possess the same microstructure, then analysis and comparison of the variability of the ensembles may be carried out to determine the likely effects on properties or performance (see [12]).

The standard tool for hypothesis testing on the means of multivariate data is the 1-way multivariate analysis of variance (1-way MANOVA). The 1-way MANOVA tests the null-hypothesis that the means of each group lie in the same d -dimensional subspace. Choosing $d = 0$ tests that the means are the same or that the ensemble possesses the same microstructure. By extension, choosing $d = 1$ tests that the means lie on the same line through the PCA space, etc. For a full discussion of the 1-way MANOVA see standard multivariate statistical texts such as [32]. The major assumptions made in performing a MANOVA analysis is that the different classes have the same covariance matrix, which is clearly not the case in our data. However, MANOVA analysis has been shown to be robust to a violation of this assumption provided the groups are all well behaved unimodal distributions [32]. The basic notion of the MANOVA analysis is to transform the original n -dimensional dataset into a new variable on a d -dimension subspace, termed the canonical variable, that maximizes the separation between the classes, then to perform a statistical test on the ratio of within group scatter of the canonical variable to the total scatter across all groups.

If we consider a dataset of K material volume ensembles where $\Omega_k = \{\omega_1^k, \dots, \omega_p^k\}$ denotes the P_k members of the k^{th} ensemble. By projecting onto the PCA space each ω_p^k is represented by the J weights ${}^k\alpha_j^p$. For compactness, let A_{pk} denote the J dimensional column vector of PCA weights for the p^{th} member of the k^{th} ensemble. Further let \bar{A}_k indicate the mean vector of PCA weights over all P_k members of ensemble k . The within group scatter is characterized by the intra-class sum of squares and cross product (SSCP) matrix, W , defined as

$$W = \sum_{k=1}^K \sum_{p=1}^{P_k} (A_{pk} - \bar{A}_k)(A_{pk} - \bar{A}_k)^T \quad (12)$$

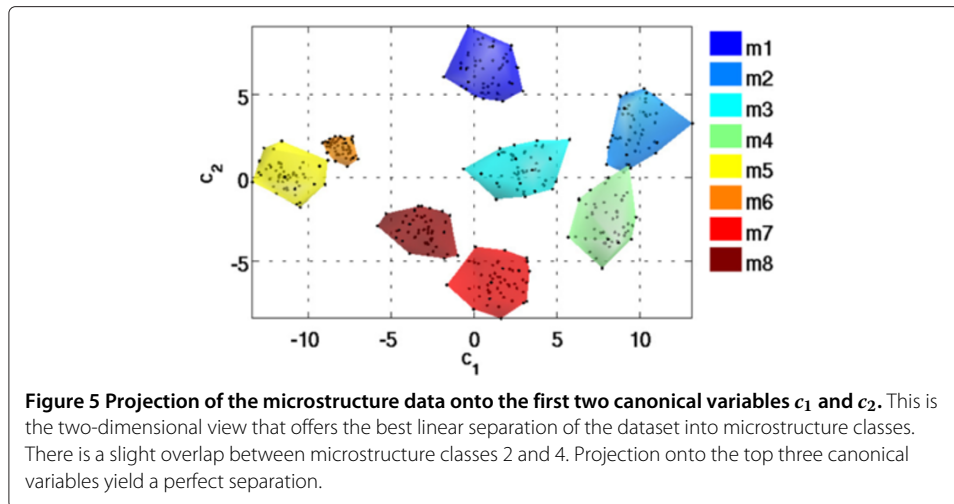
The between group scatter is characterized by the inter-class SSCP matrix defined as

$$B = \sum_{k=1}^K P_k (\bar{A}_k - \bar{A})(\bar{A}_k - \bar{A})^T = \sum_{k=1}^K P_k \bar{A}_k \bar{A}_k^T \quad (13)$$

where \bar{A} is the mean vector of PCA weights across all K ensembles which is by definition the zero vector. The total scatter is simply $T = W + B$. The test statistic for the 1-way MANOVA can be chosen as Wilk's lambda $\Lambda = \frac{|W|}{|T|}$ or Pillai's trace $Tr(WT^{-1})$. Wilk's lambda is a multivariate extension of the F test, while Pillai's trace (also F distributed) is more robust to violations of the assumptions (such as our case).

The representation of the data in terms of the canonical variables can be found by performing an eigenvalue decomposition of $W^{-1}B$ then projecting the PCA representation onto the eigenvectors. Let $L = \text{eigenvectors}(W^{-1}B)$, then the canonical representation can then be found as $c_{pk} = LA_{pk}^T$.

The canonical representation is useful for visualizing clustering between the groups. Figure 5 shows that microstructure classes can be easily separated by projection onto the first two canonical variables. In other words there exists a plane in the 50 dimensional PCA space, that when the microstructure data is projected onto it there is virtually no

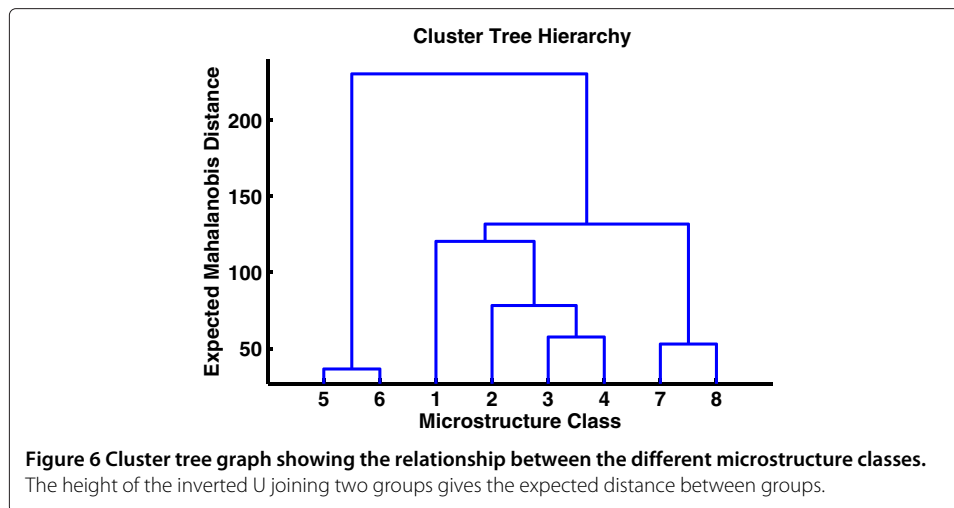


overlap between the classes. Projection onto the first three canonical variables yields a perfect separation, for this dataset.

That a perfect separation can be achieved by only three canonical variables does not imply that the data lies on a hyperplane in the 50 dimensional PCA space. The results from the 1-way MANOVA indicate that at 5% significance we can reject the null hypothesis that the means of the 8 microstructure classes lie in a 6 or lower dimensional subspace ($p = 0.048$ for $d = 7$), but that they may lie in a 7 dimensional space. Performing the test pairwise between the groups indicates that we can state with statistical certainty that the means of the 8 classes are all different or that the microstructures of the 8 ensembles are all distinct ($p \approx 0$ for $d = 0$ for all pairwise tests).

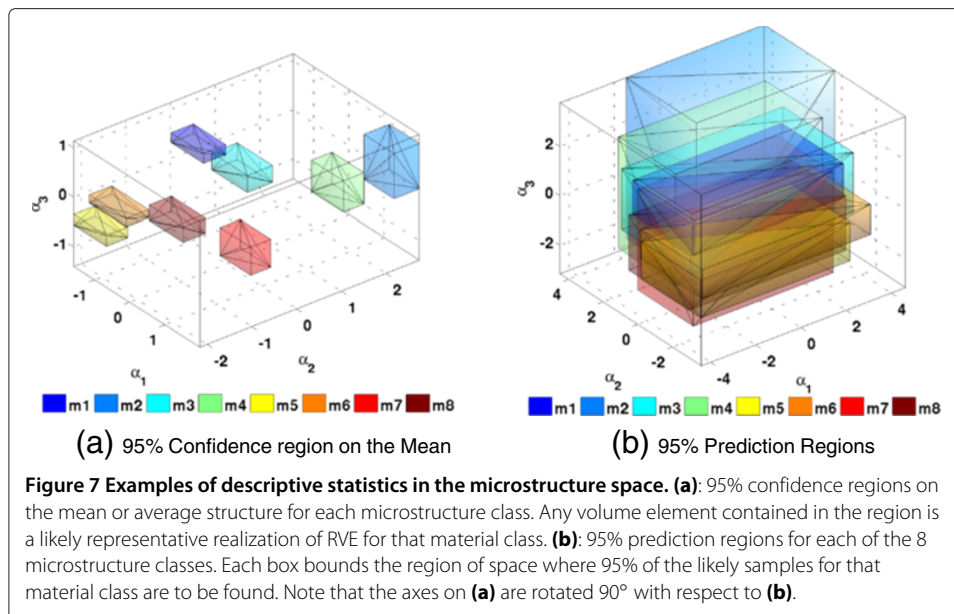
This notion of Mahalanobis distance can be used to further explore the relationships between the various groups. A key question to ask is what is the expected distance between a randomly selected member of class a to class b . By computing this expectation for all combinations of classes the entire dataset (8 material classes) can be broken up into groups and subgroups of classes that share similar features by a hierarchical cluster analysis. Hierarchical cluster analysis is a technique to build a multi-level hierarchy of groups and subgroups that are easily represented by a tree-like graph. The analysis is performed by first computing the distance between the different groups, in the case the distance measure is the expected Mahalanobis distance between a randomly selected member of class a and the cloud of points corresponding to class b , $\bar{D}_M(a, b) = E\{D_M(x \in a, f(y \in b))\}$. Then the two closest groups are combined into a larger group. The procedure is then repeated until all the smaller groups have been combined into a single large class.

The results of the hierarchical cluster analysis are shown in Figure 6 as a cluster tree. The cluster tree is interpreted by moving up on the y -axis, which shows the expected Mahalanobis distance, the height of the inverted U connecting two groups is the expected distance between those groups. Note that $\bar{D}_M(a, b) \neq \bar{D}_M(b, a)$ due to the differences in covariance between the groups, for plotting the cluster tree the minimum of the two expectations is taken. Groups that are linked lower down are more similar, than groups that are linked higher up. For example classes 5 and 6 are more similar than 7 and 8, and the expected distance between a group formed from classes 1–4 and the group consisting



of classes 7 and 8 is approximately 135. While the cluster tree graph is a useful visualization of the relationships between the microstructure classes, it is up to the end user to decide on the number of clusters or groups in the dataset or where to cut the tree to define clusters. For this example a likely pruning scheme is to break the data into three groups: the first consisting of classes 5 and 6, the second containing classes 1–4, and the third classes 7 and 8. It is interesting to note that this grouping naturally follows the degree of anisotropy in the pore shape as seen in Figure 1. Classes 1–4 all have a low degree of anisotropy with a covariance ratio of 2 or less between the highest and lowest direction, while the other two groups are more highly anisotropic with 5 and 6 having highly elongated pores along 1 primary axis, while classes 7 and 8 have pores elongated along two directions.

The PCA or visualization space serves as a natural setting for a wide range of different statistical analyses on the microstructure data. As mentioned above the center of mass for each class can be thought of as the average or representative realization for that class. A natural question to ask is “How well is the mean known given the amount of characterization for this class?” or equivalently “If the experiment were repeated and a different ensemble collected how different would our estimate of the average structure be?” Given the orthogonality of the principal component vectors, common tools of statistical inference can be applied treating each principal component as independent. This allows us to use readily available univariate statistical tools and descriptors without concerning ourselves with correlations between the different directions. For example the questions asked above can be readily answered by computing confidence regions about the mean of each microstructure class. 95% Confidence regions are shown in Figure 7(a). This gives us a formal probabilistic approach to selecting realizations to serve as representative volume elements (RVE) for the ensemble. The confidence region is interpreted not as a probability on the true microstructure mean (which is unknowable) but rather on the process by which we are estimating the average structure. If we were to validate the confidence regions by collecting a very large number of ensembles and estimating the ensemble mean, 95% of the time the calculated mean would lie within this box. Under this interpretation any material volume lying in our confidence region is a potential RVE for that



microstructure class. The RVEs shown in Figure 1 were chosen as the material realization closest to the center of the confidence region.

In a similar manner, we may wish to have an estimate on the range of each microstructure class or given the variability observed in a class how likely are additional samples to fall within some region. Prediction bounds and related measures can also be readily calculated in the PCA space. Examples of 95% prediction regions are shown in Figure 7(b). Simple analysis such as this give a way of sampling representative volumes from a fuller range of likely structures for modeling and simulation [33] or more advanced intra-class statistical analysis.

Support vector machines for classification of microstructures

Projection onto the canonical variables showed that the microstructure classes can be well separated in a low dimensional subspace, therefore it should also be possible to develop a robust classification scheme to sort new microstructure realizations into their appropriate classes. For the example classes presented here where there are clear microstructural differences between the classes, classification could be accomplished by simply assigning a new realization to the nearest class, the class which minimizes the Mahalanobis distance. While this simple approach will work for the “toy” examples given in the paper, we wish to apply a more formal probabilistic approach to classification that is able to handle the general case when the data may not be linearly separable. Following Sundararaghavan and Zabaras, we adopt a kernel support vector machine (SVM) classification scheme [7,8]. Unfortunately, a mathematical treatment of SVM classification, particularly for the multi-class examples of interest here are beyond the scope of this manuscript. For the interested reader a full treatment of SVM classification and regression can be found in standard references such as [34], here we will provide a descriptive overview of the technique and a classification example.

If we consider the projection of the microstructure classes onto the canonical variables (see Figure 5), and we wish to develop a linear categorization rule that assigns a

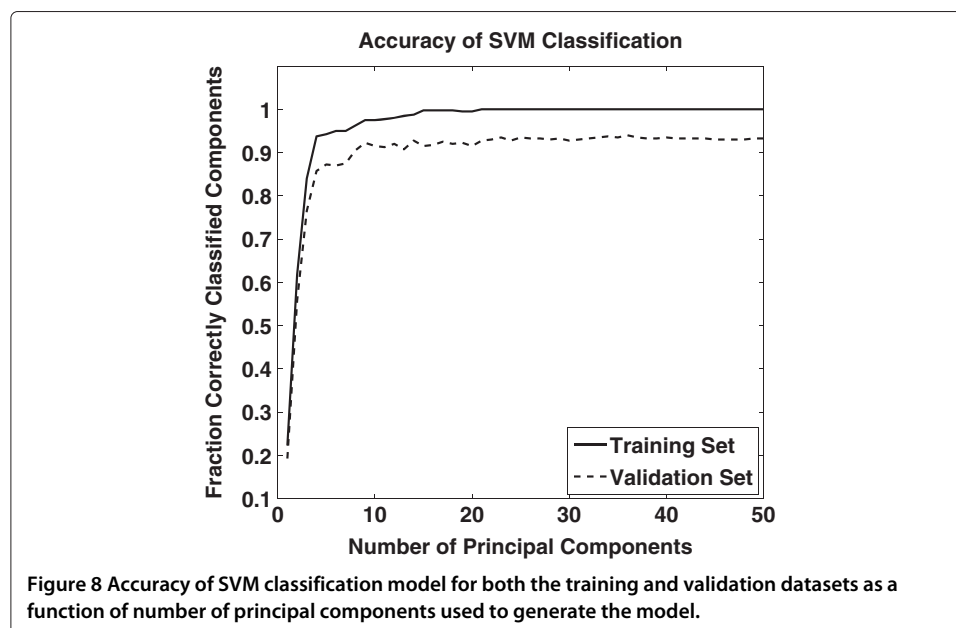
new point to either microstructure class 1 or 2 it is obvious that there are numerous possible lines that could be drawn which would partition the space between these two classes. The intuitive approach to finding an “optimal” partition would be to imagine a line connecting the centers of mass of each class and draw the partition normal to this line at the midpoint. In performing this operation we would be finding a partition that approximately maximizes the distance from each class to the partition. SVMs are supervised machine learning models for binary linear classification, that implement, in a more theoretically rigorous manner, the basic approach just described. They attempt to find the separating hyperplane through the data such that the distance from it to the nearest data points on each side (the support vectors) is maximized. Such a plane, provided it exists, is called the maximum-margin hyperplane. There are however many cases where such a plane does not exist, such as when the data is non-linearly separable or when there is noise in the data and the outliers from each class overlap. In the case of outliers where the data is nearly linearly separable, a penalty is introduced on points on the wrong side of the margin boundary that increases with the distance from it. This is known as a soft margin SVM and was introduced in 1995 by Cortes and Vapnik, and revolutionized the machine learning field [35]. Previous to soft margin approaches the options for nearly linear or fuzzy classification were limited to neural networks or other unsupervised techniques that have a strong tendency to overfit the data. The soft margin concept allowed for an optimal classification without overfitting, in the sense that the margin is maximized, even when a clean linear separating hyperplane did not exist. In instances where the data is not linearly separable, such as a quadratic separating surface, an approach termed kernel SVM introduced by 1992 by Boser, Guyon and Vapnik, is used to map the data points to a higher dimensional space where they are linearly separable [36]. The approach relies on the so called kernel trick, where it was realized that if the optimization was formulated properly the mapping to the higher dimensional space only appears as a dot product, implying that the mapping never needs to be explicitly performed thus keeping the dimensionality of the optimization equation low. For multi-class problems such as our example, the problem is reduced to multiple binary classification problems by developing pairwise classification for all combinations (1-against-1) or by considering one class against the remainder of the data set (1-against-rest) [37].

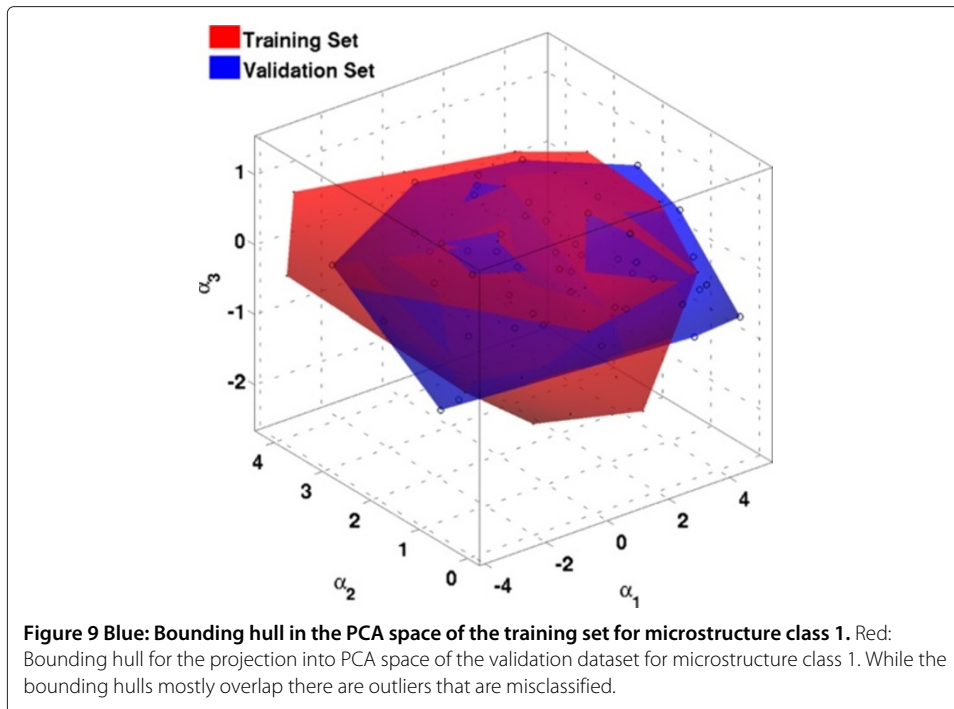
The open-source software library LIBSVM was used to create a classification on the data in the PCA space [38]. Performing the classification in terms of the canonical variables would have been more efficient as the data is 100% linearly separable in three dimensions. However for this example we are interested in exploring how many principal components are really necessary to capture the range of microstructures in the dataset. The 50 realizations from each of the 8 classes that were used to construct the PCA representation were once again used as the training dataset. As we know *a priori* that the data is linearly separable the applied kernel is simply the identity matrix. The 1-against-1 approach was found to perform better as the individual classifications remained linear rather than needing to apply a non-linear kernel for 1-against-rest. Validation was performed by projecting the 50 members of each set held in reserve into the PCA space (the PCA basis was not updated, the weights of the new data given the current basis were computed) and exercising the classification model assuming no prior knowledge of correct class membership. Classification models were constructed for increasing numbers of

principal components. The accuracy of the classification model for both the training and validation sets is shown in Figure 8.

As can be seen in the figure, the accuracy of the classification rapidly increases with additional principal components, by 15 components the accuracy of the training set is 100% at the accuracy of the validation set saturates around 92%. That the classification accuracy of the validation dataset does not approach 100% implies that 1) there are outliers in the validation dataset that are being misclassified and 2) the additional material realizations contain some useful microstructure information that is absent in the original dataset and is not captured in the original PCA decomposition. By examining the PCA representation of the training data and the projection into this space of the validation set, it can be seen that the hulls bounding the points do not entirely overlap and indeed there are outliers which skew the shape of the bounding convex hull. An example of this projection for microstructure class 1 is shown in Figure 9.

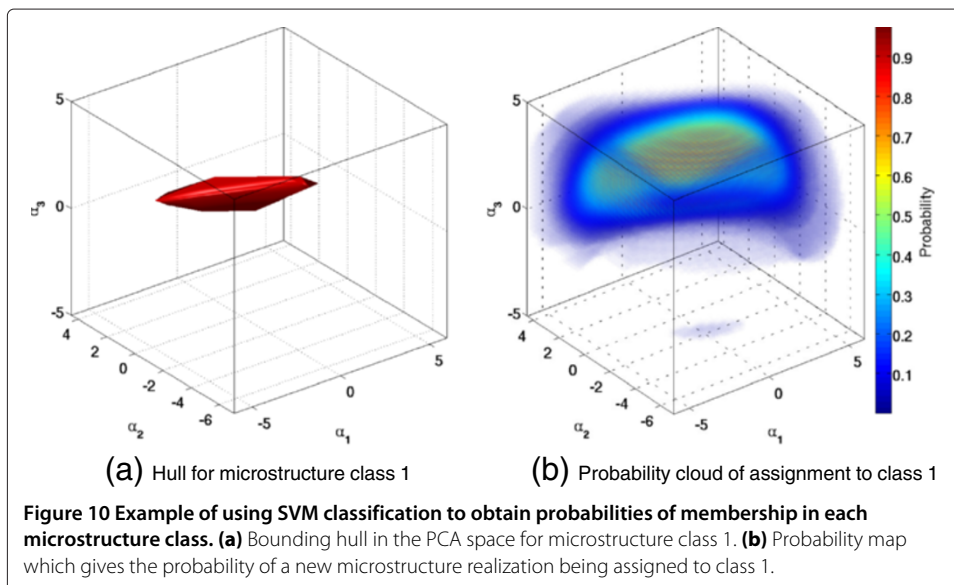
A key advantage of using SVMs for classification is that when new points are classified they are assigned a probability of membership for each existing class. This probability can be used to construct visualizations and maps in the design space on class membership. Such a map giving the probability of class membership over a region in the PCA space is shown in Figure 10. The probability of class assignment gives additional information concerning the state of the classification model and the sufficiency of the microstructure data collected. Virtually all of the misclassified points in the validation dataset were assigned a significant probability of belonging to two or more microstructure classes. When adding a new ensemble to the system the classification probabilities are a powerful tool in determining if the ensemble belongs to a currently defined class or a new class. While not discussed in detail here, the probabilities can also be used as weights when assigning average properties or other metrics to the various classes. In effect this is one method of handling the outliers of a microstructure class or of estimating the tails of the performance distribution, given the current state of the microstructure knowledge.





Connection with properties and performance

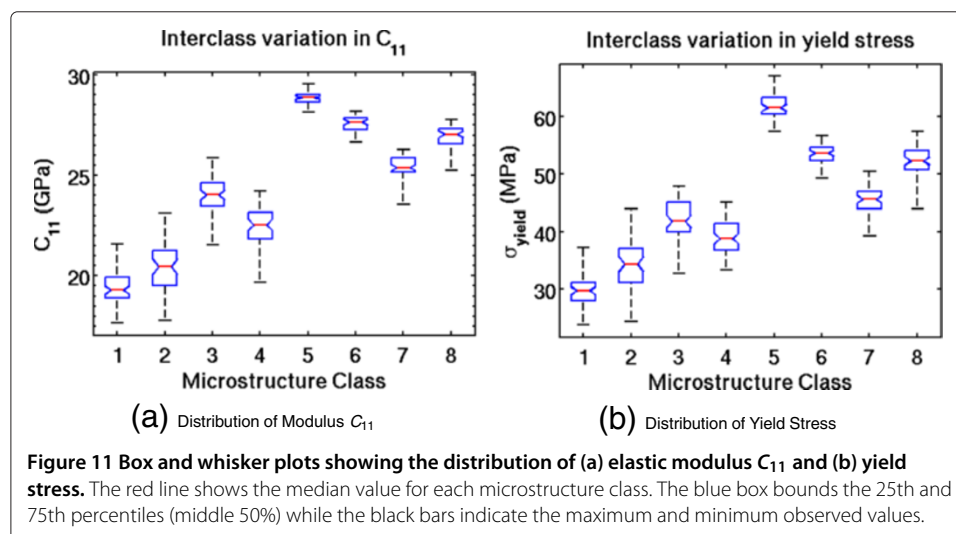
The above discussion was focused solely on the exploration and visualization of the microstructural relationships between multiple materials classes in an interactive materials design space. In this section we link the materials design space with properties. Being able to connect the interactive materials data with properties and response data is a critical task if the materials design space concept is to be applied in a practical manner. The idea of generating invertible mappings between a microstructure space and a property space is not new. The Microstructure Sensitive Design framework [22,39],



for example, was predicated on delineating a microstructure design space termed the microstructure hull that bounded the complete set of theoretically feasible microstructures of a given class and then projecting microstructures from the design space to a property space, typically in the form of homogenization relationships for the delineation of property closures. The main difficulty of such an approach is the construction of the microstructure hull, typically only first order microstructure descriptors such as volume fraction or crystallographic texture could be used. As was described above the space of two point correlations for even the simple porous composite material system here, renders the problem intractable for higher order microstructure descriptors [25]. Here we are restricting the problem from the complete set of theoretically possible structures to the space of microstructures available through either characterization or digital simulation. Instead of delineating property closures, or strict bounds on theoretical property values, we are exploring and mapping the property space that is currently achievable.

The successful production of property models directly in terms of distributions of the microstructure process or microstructure function is important if such structural models are to be incorporated by industry and the larger ICME or materials design community. In the long term, it is hoped that the microstructure design space described here will find applications in quality control and design certification, where current decisions are made on measured property or experimental response values rather than on the microstructure details. Moving toward microstructure based acceptance criteria will require the widespread construction and validation of the types of linkages presented here.

As described in Section “Microstructure data generation and modeling methodology”, the mechanical response of each realization for the 8 microstructure classes was simulated using elastic visco-plastic FFT modeling. The resulting distributions in yield stress and effective elastic modulus are shown in Figure 11. In previous work Niezgoda et al. observed a nearly perfect linear relationship between a measure of microstructure variance and the observed variance in properties for multiple ensembles of the same material class [11,12,40]. In that work the material system was a low to moderate contrast two phase composite, and while the ensembles all had large differences in structural variance,



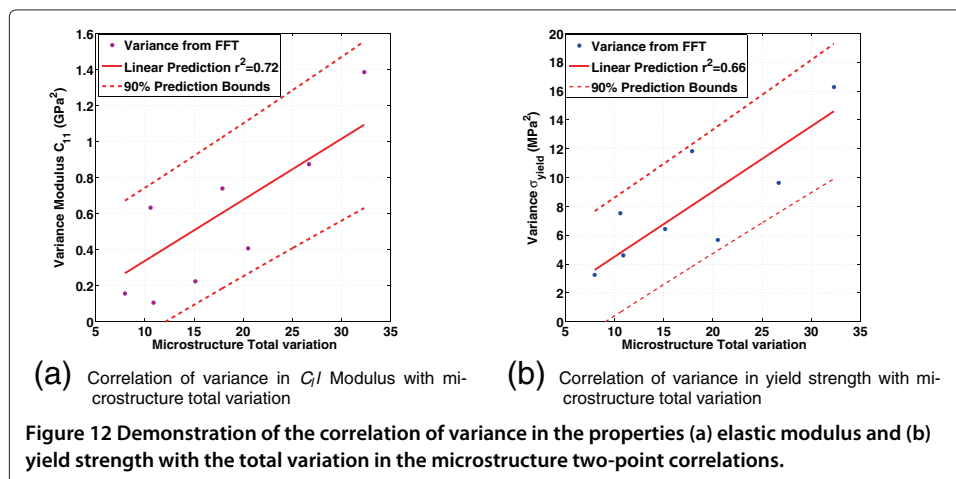
the multiple ensembles all had the same mean or average structure. Here we extend that observation to effectively infinite contrast porous solids and show that the relationship between structural variance and property variance shows the expected linear trend across multiple material classes.

The descriptor of microstructural variance found to be most useful was the total variation defined as

$$VAR_{tot}^p = b_1^p + b_2^p + \dots + b_j^p \quad (14)$$

where b_j^p are the eigenvalues of the intra-class PCA decomposition for the p^{th} microstructure class. The total variation can also be approximated for a given class from the inter-class PCA representation by $VAR_{tot} \approx \sum_{j=1}^J var(\alpha_j^p)$ where the variance is calculated over the members of the p^{th} specific class. For each class the total variation was computed from the intra-class PCA representation, and the relationship with property variance is shown in Figure 12. As can be seen in the figure there is a strong correlation between microstructure variance and property variance. The observed effect is not as strong as seen in previous cases, however this example was deliberately set up as a worst case in that the contrast between matrix and pore was effectively infinite and that the microstructures spanned a wide range of anisotropy in pore shape and distribution.

Ashby charts or maps are a well known design tool for visualizing relationships between properties, performance and design constraints such as cost. Such maps have proved invaluable in guiding material selection based on macroscale properties or performance criteria. An ongoing goal of the authors is to link microstructure and microstructure variance into similar type maps to extend the utility to the realm of microstructure sensitive design and design of materials/microstructures. In order to develop such tools, properties and performance must be described as a function of the microstructure, or in our case the PCA representation of microstructure. Here we we perform a proof of concept example, by developing linear homogenization relationships in the microstructure design space. A linear model was fit between effective modulus and yield strength and the PCA weights of each realization using a weighted least squares approach. The eigenvalues from the inter-class PCA representation were chosen as the least squares weights. Weighting the least squares regression in this manner enforced the assumption that the significance of



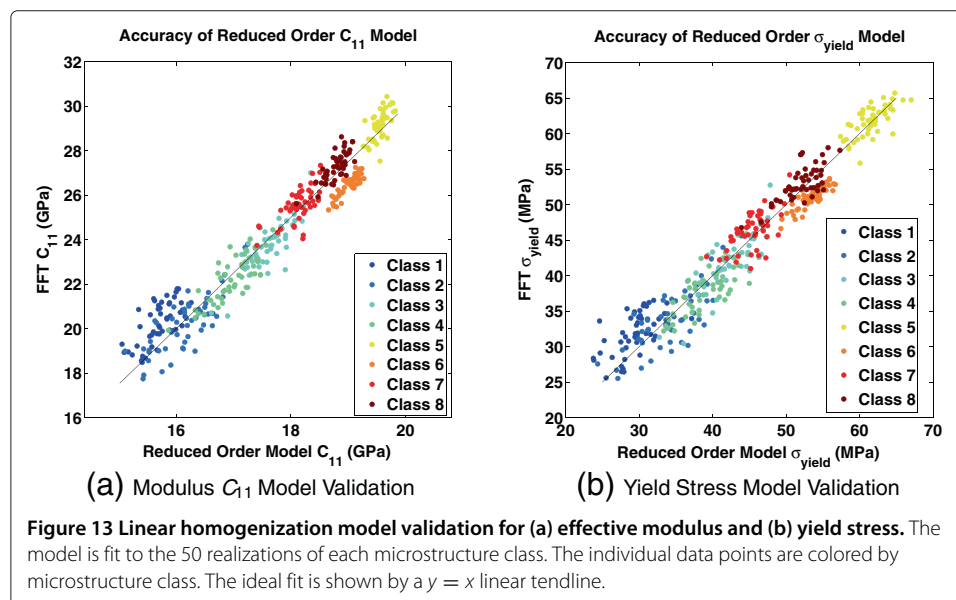
each successive principal component on properties will decrease, and add a physical basis for the simple linear model. In effect we are assuming the directions of greatest variability in the microstructure will have the largest influence on the change in properties. The results of the fit are shown in Figure 13, which shows that the simple linear model does an excellent job of capturing properties across the complete range of material classes.

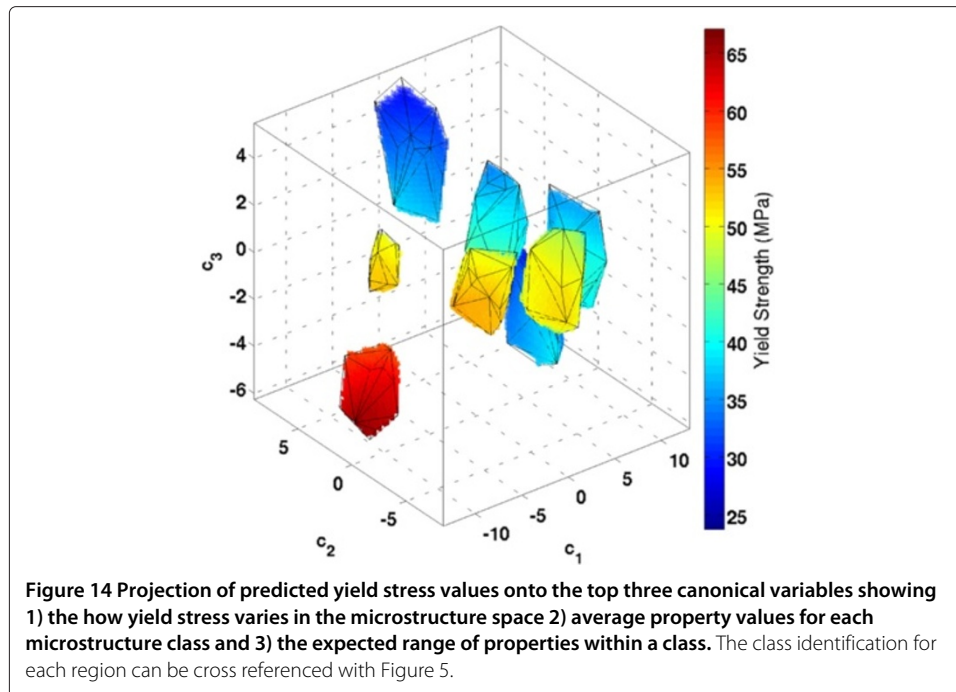
Once an appropriate homogenization scheme is developed, property maps can be constructed by projecting the property values into the PCA space or onto the canonical representation. The projection of properties into the canonical microstructure representation is shown in Figure 14. The canonical representation was chosen as the cloud of realizations for each microstructure class are significantly overlapped in the PCA space but are well separated in the canonical representation. In contrast to Ashby charts and similar maps, constructing property maps in the microstructure space allows for the visualization of the range of average properties for a microstructure class but also to see the expected range of properties for individual microstructure realizations within a given class. Such maps offer a quick but powerful visualization of the effect of structure on properties and gives the user a means of rapid identification of microstructure realizations of interest.

Conclusions

This work was largely driven by the need, within the materials community, for effective and efficient microstructure databases and microstructure based design and development tools. In this paper, we presented a novel framework for the visualization and analysis of microstructure data which fills several of the critical technological gaps limiting the development of large scale microstructure design libraries. This work directly targets the following key technological questions:

1. What measures of the material microstructure best capture its salient features?
2. How do we describe the inherent variability in a material and how do we link this variability with scatter in properties or performance?





3. How do we quantitatively compare materials from different classes or material systems?
4. How do we place metrics on the quality or sufficiency of the microstructure data collected, or how do we place measures on how well we have characterized a particular material?
5. What objective data-driven reduced-order measures of the material microstructure are best suited for establishing the invertible structure-property relationships needed for materials design?

While a concerted effort from the larger materials community is required if such databases are to ever come to fruition, the authors believe this work represents a critical first step in their development.

In this work we developed several proof-of-concept tools and visualizations for the analysis of microstructure data. While the case studies presented were largely “toy-problems” or highly simplified examples they demonstrate the main features of the microstructure space and the range of tools that can be developed and the breadth of analyses that can be performed. While it is still in its infancy, we believe that this work represents a completely new approach to handle the analysis and visualization of microstructure data. In particular this work offers for the first time, an objective framework for the quantitative comparison of microstructure between differing material classes. These simple results potentially open completely new avenues for the exploration of structure-processing-property relationships. A key difficulty in incorporating microstructure evolution into design frameworks such as Microstructure Sensitive Design has been the lack of a suitably defined microstructure space for higher order statistical microstructure descriptors. In future work, the authors hope to explicitly incorporate microstructure evolution maps as processing path-lines or flow vectors through the microstructure space. Such maps

could find great utility in process monitoring and material quality control applications, in addition to providing novel structure-processing-property visualizations to the design community.

The work presented here also has immediate relevance to the ongoing efforts in multi-scale modeling and characterization. If useful multi-scale structure/property/processing knowledge are to be developed, an automated analysis built on a rigorous mathematical formalism is necessary. Such formalism should handle a wide variety of data types across multiple length-scales. Such an approach will enable scientists and engineers to model the intrinsic randomness of microstructures, relate it to property variance, and treat it as a materials design parameter for verification and validation. Given the vast separation between the macro-, meso-, and atomic length scales of interest in most materials design problems, it should be clear that microstructure datasets collected at differing length scales are best integrated in a stochastic manner. Using this formalism a natural multi-scale representation of microstructure emerges where the structure at lower length scales is described in terms of conditional probability densities given the structure at higher length scales. Such a method aligns naturally with emerging multi-scale modeling techniques based around stochastic partial differential equations [41-43] and provides an elegant framework for the inclusion of materials uncertainty into material model uncertainty quantification and into component level design validation and verification.

The need to move beyond design based on average or effective properties is driving the integration of simulation and experiment within frameworks such as ICME and is a central theme of the Materials Genome Initiative. As time goes on this need is likely to become more acute. In the opinion of the authors, the integration of the disparate techniques and data paths required for ICME and the MGI necessitates a corresponding evolution of the way we think about microstructure data. It is the hope of the authors that this work will help jump-start discussions within the field of materials science and engineering, and pave the way for future developments.

Availability of supporting data

Computer code to create the simulated digital micrographs and to reproduce all of the case studies presented in this manuscript are available directly from SRN (Niezgoda.s@gmail.com).

Endnote

^aThis is only strictly true for Gaussian or near-Gaussian “well behaved” unimodal distributions. However, this will be the case for all microstructure classes examined here, and is expected to hold for virtually all real material systems.

Competing interests

The authors declare no competing interests.

Authors' contributions

SRN developed the case studies, performed the analysis and composed the manuscript text. AKK performed the FFT micromechanical simulations and discussed the results and manuscript drafts. SRK discussed the results and manuscript drafts. All authors read and approved the final manuscript.

Acknowledgements

S.R.N. acknowledges partial funding for this work from the Los Alamos National Laboratory LDRD program, Project 20110602ER. LANL is operated by LANS, LLC, for the National Nuclear Security Administration of the U.S. DOE under contract DE-AC52-06NA25396. S.R.K acknowledges funding from ONR award N00014-11-1-0759 (Dr. William M. Mullins, program manager).

Author details

¹Materials Science and Technology Division, Los Alamos National Laboratory, Los Alamos NM 87545, USA. ²Present Address: Department of Materials Science and Engineering, Department of Mechanical and Aerospace Engineering, The Ohio State University, Columbus, OH 43218, USA. ³Department of Metallurgical and Materials Engineering, Indian Institute of Technology Madras, Chennai, 600036, India. ⁴George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA.

Received: 29 March 2013 Accepted: 10 May 2013

Published: 2 July 2013

References

1. Materials Genome Initiative for Global Competitiveness. Tech. rep., National Science and Technology Council (2011)
2. Crabtree G, Glotzer S, McCrady B, Roberto J (2010) Computational materials science and chemistry: accelerating discovery and innovation through simulation-based engineering and science. Tech. rep. U.S. Department of Energy, Office of Science
3. Allison J (2011) Integrated computational materials engineering: A perspective on progress and future steps. *JOM J Minerals, Metals Mater Soc* 63(4): 15–18
4. National Academy of Sciences (US) Committee on the survey of materials science and engineering (1975) *Materials and man's needs: materials science and engineering – Volume II, The Needs, Priorities, and Opportunities for Materials Research*. The National Academies Press, Washington
5. Rumble J (1991) Standards for Materials Databases: ASTM Committee E49. In: Kaufman J, Glazman J (eds) *Computerization and networking of materials databases, Volume 2*. ASTM International, pp 73–83
6. Freiman S, Madsen LD, Rumble J (2011) Perspective on materials databases. *Am Ceramics Soc Bull* 90(2): 28–32
7. Sundararaghavan V, Zabarar N (2004) A dynamic material library for the representation of single-phase polyhedral microstructures. *Acta Materialia* 52(14): 4111–4119
8. Sundararaghavan V, Zabarar N (2005) Classification and reconstruction of three-dimensional microstructures using support vector machines. *Comput Mater Sci* 32(2): 223–239
9. Ganapathysubramanian B, Zabarar N (2007) Sparse grid collocation schemes for stochastic natural convection problems. *J Comput Phys* 225: 652–685
10. Ganapathysubramanian B, Zabarar N (2008) A non-linear dimension reduction methodology for generating data-driven stochastic input models. *J Comput Phys* 227(13): 6612–6637
11. Niezgoda SR (2010) Stochastic representation of microstructure via higher-order statistics: theory and application. PhD thesis, Drexel University
12. Niezgoda SR, Yabansu YC, Kalidindi SR (2011) Understanding and visualizing microstructure and microstructure variance as a stochastic process. *Acta Materialia* 59(16): 6387–6400
13. Moulinec H, Suquet P (1998) A numerical method for computing the overall response of nonlinear composites with complex microstructure. *Comput Methods Appl Mech Eng* 157: 69–94
14. Lebensohn RA (2001) N-site modeling of a 3D viscoplastic polycrystal using fast Fourier transform. *Acta Materialia* 49(14): 2723–2737
15. Lebensohn RA, Kanjarla AK, Eisenlohr P (2012) An elasto-viscoplastic formulation based on fast Fourier transforms for the prediction of micromechanical fields in polycrystalline materials. *Int J Plast* 0: 59–69
16. Kanjarla AK, Lebensohn RA, Balogh L, Tomé C N (2012) Study of internal lattice strain distributions in stainless steel using a full-field elasto-viscoplastic formulation based on fast Fourier transforms. *Acta Materialia* 60(6–7): 3094–3106
17. Billingsley P (2012) *Probability and measure*, 3 edition. Wiley, New York
18. Papoulis A, Pillai SU (2002) *Probability, random variables and stochastic processes*. McGraw-Hill Education, Boston
19. Bunge HJ (1982) *Texture analysis in materials science: mathematical methods*. Butterworths, London
20. Tewari A, Gokhale AM, Spowart JE, Miracle DB (2004) Quantitative characterization of spatial clustering in three-dimensional microstructures using two-point correlation functions. *Acta Materialia* 52(2): 307–319
21. Torquato S (2001) *Random heterogeneous materials: microstructure and macroscopic properties*. Springer, New York
22. Fullwood DT, Niezgoda SR, Adams BL, Kalidindi SR (2010) Microstructure sensitive design for performance optimization. *Prog Mater Sci* 55(6): 477–562
23. Huang M (2005) The n-point orientation correlation function and its application. *Int J Solids Struct* 42(5–6): 1425–1441
24. Adams BL, Morris PR, Wang TT, Willden KS, Wright SI (1987) Description of orientation coherence in polycrystalline materials. *Acta Metallurgica* 35(12): 2935–2946
25. Niezgoda SR, Fullwood DT, Kalidindi SR (2008) Delineation of the space of 2-point correlations in a composite material system. *Acta Materialia* 56(18): 5285–5292
26. Lee JA, Verleysen M (2007) *Nonlinear dimensionality reduction*. Springer, New York
27. Fukunaga K (1990) *Introduction to statistical pattern recognition*. Academic Press, Boston
28. Turk M, Pentland A (1991) Eigenfaces for recognition. *J Cogn Neurosci* 3: 71–86
29. Li Y (2004) On incremental and robust subspace learning. *Pattern Recognit* 37(7): 1509–1518
30. Yeong CLY, Torquato S (1998) Reconstructing random media. *Phys Rev E* 57: 495
31. Fullwood DT, Niezgoda SR, Kalidindi SR (2008) Microstructure reconstructions from 2-point statistics using phase-recovery algorithms. *Acta Materialia* 56(5): 942–948
32. Stevens JP (2001) *Applied multivariate statistics for the social sciences*. Lawrence Erlbaum, Mahwah
33. Qidwai SM, Turner DM, Niezgoda SR, Lewis AC, Geltmacher AB, Rowenhorst DJ, Kalidindi SR (2012) Estimating the response of polycrystalline materials using sets of weighted statistical volume elements. *Acta Materialia* 60(13): 5284–5299
34. Steinwart I, Christmann A (2008) *Support vector machines*. Springer, New York
35. Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3): 273–297

36. Boser BE, Guyon IM, Vapnik VN (1992) A training algorithm for optimal margin classifiers In: Proceedings of the fifth annual workshop on Computational learning theory, 5. ACM, Pittsburgh, pp 144–152
37. Hsu CW, Lin CJ (2002) A comparison of methods for multiclass support vector machines. *IEEE Trans Neural Netw* 13(2): 415–425
38. Chang CC, Lin CJ (2011) LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol* 2(3): 27
39. Adams BL, Henrie A, Henrie B, Lyon M, Kalidindi SR, Garmestani H (2001) Microstructure-sensitive design of a compliant beam. *J Mech Phys Solids* 49(8): 1639–1663
40. Niezgoda SR, Turner DM, Fullwood DT, Kalidindi SR (2010) Optimized structure based representative volume element sets reflecting the ensemble-averaged 2-point statistics. *Acta Materialia* 58(13): 4432–4445
41. Karakasidis T, Charitidis C (2007) Multiscale modeling in nanomaterials science. *Mater Sci Eng: C* 27(5): 1082–1089
42. Xu X (2007) A multiscale stochastic finite element method on elliptic problems involving uncertainties. *Comput Methods Appl Mech Eng* 196(25): 2723–2736
43. Li Z, Wen B, Zabaras N (2010) Computing mechanical response variability of polycrystalline microstructures through dimensionality reduction techniques. *Comput Mater Sci* 49(3): 568–581

doi:10.1186/2193-9772-2-3

Cite this article as: Niezgoda et al.: Novel microstructure quantification framework for databasing, visualization, and analysis of microstructure data. *Integrating Materials and Manufacturing Innovation* 2013 **2**:3.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
