

Full length article

Parsing abnormal grain growth

A. Lawrence^a, J.M. Rickman^{a,*}, M.P. Harmer^a, A.D. Rollett^b^a Department of Materials Science and Engineering, Lehigh University, Bethlehem, PA 18015, USA^b Department of Materials Science and Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA

ARTICLE INFO

Article history:

Received 10 July 2015

Received in revised form 19 September 2015

Accepted 22 October 2015

Available online 18 November 2015

Keywords:

Grain growth

Microstructure

Statistics

Processing

ABSTRACT

Abnormal grain growth, the enlargement of a minority of grains in a polycrystal at the expense of the surrounding grains, occurs in both metallic and ceramic materials and can have a profound impact on their mechanical and electrical properties. Somewhat surprisingly then, there is little consensus as to which specific microstructural features provide a signature of abnormal growth. Indeed, some workers describe this phenomenon in terms of a bimodal grain size distribution, often without justification, while others focus on very few, elongated grains. Using specialty alumina (i.e., high-purity aluminum oxide with tailored impurity composition) as our prototype, we describe here a set of practical maps and metrics that are useful in quantifying various microstructural features that are associated with abnormal grain growth. These maps provide a visual “fingerprint” of abnormal growth, while the metrics permit the design of processing routes to obtain desired microstructures. We then present an application of correlation analysis that illustrates the efficacy of data analytics in quantifying which input (i.e., processing) variables exert the strongest influence on abnormal grain growth. Finally, we outline the use of this methodology to examine correlations among processing variables and the thermomechanical and kinetic properties of materials (e.g., strength, hardness, thermal conductivity, etc.).

© 2015 Acta Materialia Inc. Published by Elsevier Ltd. All rights reserved.

1. Introduction

When polycrystalline materials are held at sufficiently high temperatures, the grain structure coarsens in order to decrease the excess free energy associated with the grain boundaries [1]. Since this process determines important microstructural features over time, and therefore key mechanical and electrical properties of materials, it has been the subject of intense investigation over the years [2]. These investigations have focused on the driving forces for grain growth and the associated kinetics of microstructural evolution [3]. In describing the temporal evolution of a distribution of grains, it is useful to divide grain kinetics broadly into two categories, namely normal and abnormal grain growth. Normal grain growth (NGG), which occurs in a wide variety of metallic and ceramic systems, is characterized by a relatively narrow range in grain sizes and shapes and a grain-size distribution that obeys a simple scaling relation [1].

By contrast, abnormal grain growth (AGG) occurs in a polycrystalline material when a minority of grains having anisotropic boundary energies or mobilities becomes large and grows into the surrounding “normal” matrix [4]. AGG is ubiquitous in many systems, especially thin films, in which an anisotropy in grain-

boundary properties leads to the fast growth of a subset of grains relative to others in the microstructure [5]. This fast growth may occur, in some cases, due to the presence of impurity excesses, such as Ca or Si in Al_2O_3 [6]. Recent studies have also linked the presence of AGG with grain-boundary complexion transitions in which boundary structure and chemistry change to produce a new interfacial state [7]. In many cases AGG is undesirable as the resulting heterogeneous microstructure leads, for example, to a degradation in mechanical properties (e.g., hardness) [8]. In some situations, however, AGG is, in fact, beneficial as elongated grains can facilitate crack-tip bridging and thereby improve fracture toughness [9,10]. Moreover, there is the well-known use of AGG to generate large grain sizes in electrical steels in order to optimize their magnetic properties [11].

Despite the important implications of this phenomenon for material properties in polycrystals, there is little to no consensus as to which specific microstructural features indicate that AGG has occurred. For example, while some workers may focus on the presence of a single relatively large grain, others may interpret AGG in terms of a subset of large, non-equiaxed grains. Indeed, different researchers have adopted varying in-house heuristic criteria, which often involve only limited grain-size information. In reality, the situation is much more complex, and it is clear that both the grain-size and grain-shape distributions are critical to characterize abnormality and to correlate features of AGG with materials parameters, such as composition, temperature, etc.

* Corresponding author.

E-mail address: jmr6@lehigh.edu (J.M. Rickman).

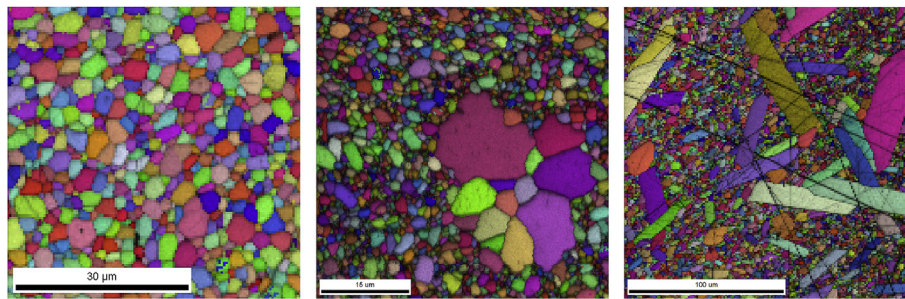


Fig. 1. Three illustrative microstructures for specialty alumina samples having different microstructures and processed under different conditions. a.) (Left) The system was annealed at 1750° C for 2 h. Each system contained the following dopants in different amounts: MgO, CaO, Na₂ O and SiO₂. The main dopant was MgO. b.) (Middle) The system was spark-plasma sintered at 1300° C for 1 h. The main dopants were MgO and CaO. c.) (Right) The system was annealed at 1750° C for 2 h. The main dopants were SiO₂ and CaO.

In this paper we propose a set of microstructural maps and metrics that are useful in quantifying microstructural features that are associated with abnormal grain growth. We then present an application of correlation analysis that highlights relationships between the independent (i.e., processing) and dependent (i.e., microstructural metrics) variables. This methodology illustrates the utility of data analytics in quantifying which inputs exert the strongest influence on abnormal grain growth. In so doing it exemplifies the development of the techniques that are required for the implementation of the materials genome [12].

2. Methodology

2.1. Metrics

Given the large amount of microstructural data that must be processed in characterizing grain growth, it is useful to develop tools that will permit both a visual and a quantitative description of microstructural features associated with abnormality. For this purpose, it is convenient to construct first a map of the joint probability density function (pdf) of grain size, \hat{G} , and aspect ratio, \hat{a} , for a given microstructure. Consider three representative microstructures that are associated with specialty alumina having different chemistries and processing routes, as described below. Several pdf maps for these three microstructures, as shown in Fig. 1a–c, are presented in Fig. 2a–c. For the purpose of comparison, we employ the quantities, \hat{G} and \hat{a} , which are the grain size and the aspect ratio, respectively, normalized by their microstructural averages. (In the latter case, this normalization of a dimensionless quantity is useful to relate it to its mean value, which typically differs from unity.) The dotted lines indicate the corresponding cutoffs for the grain size and the aspect ratio, G_c and a_c , used to identify “extreme” grains, as dis-

cussed below. While Fig. 1a is consistent with “normal” grain growth, Fig. 1b and c shows relatively large and non-equiaxed grains that are often taken as signatures of abnormality. Clearly, these pdf maps for disparate structures provide distinctive fingerprints that can be used to assess the degree of abnormality in terms of the number and distribution of outlying points.

To provide a quantitative description of AGG, it is essential to identify metrics that highlight specific features of the maps associated with abnormality. These metrics will necessarily focus on grains in the tail of the distribution. In our development, we borrow from the analysis of financial markets and risk analysis [13,14] to detect and quantify departures from grain size and shape distributions from that which is predicted by grain growth theory. In choosing quantities from the risk analysis and financial mathematics literature, we exploit an analogy between extreme events in a financial setting (e.g., unusually large monetary losses) and corresponding rare events in a microstructure (i.e., abnormal grains). In the former case, it is useful for planning purposes to have a measure of risk that reflects extraordinary occurrences, such as the mean value of the worse losses. Given that extreme events are of importance here, the corresponding quantities should be conditioned on being in the tail of the relevant distribution. As will be seen below, similar expressions based on conditional probabilities can be formulated for microstructural analysis. We note that, in this work, we focus on the moments of a distribution associated with finite sample sizes rather than on determining the limiting forms of statistics associated with very large sample sizes. This approach is appropriate in this context given the sample sizes considered and the fact that, in some cases, we are estimating multivariate probability densities with unknown limiting distributions.

In particular, we adapt to the current context the conditional probability $P(\mathbf{x}|\mathbf{x}>\mathbf{x}_q)$ that a random variable \mathbf{x} with associated pdf

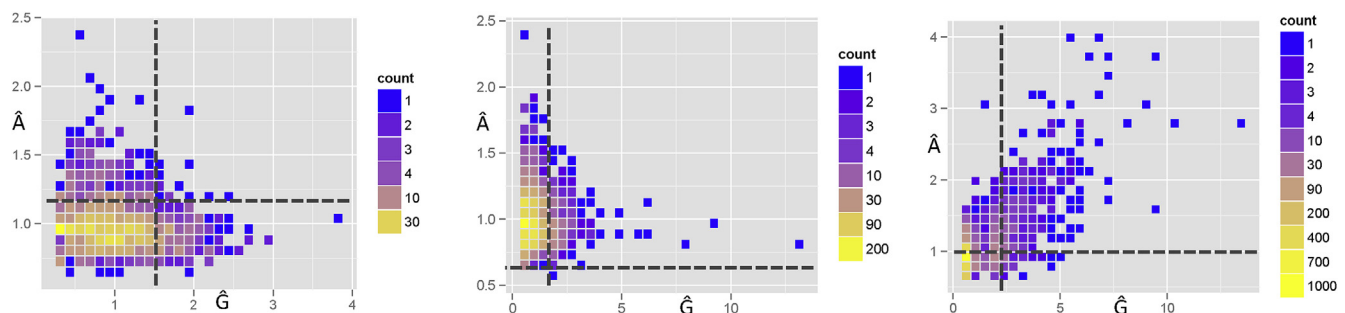


Fig. 2. The joint probability density function (pdf) maps corresponding to the microstructures shown in Fig. 1. These pdf maps summarize the density of normalized grain size, \hat{G} , and aspect ratio, \hat{a} , for the given microstructure, where each quantity is divided by its microstructural average. The dotted lines indicate the cutoffs G_c and a_c , used to identify “extreme” grains.

$p(x)$ exceeds a threshold value x_q , known as an exceedance (denoted exc) [15]. One can then write

$$P(\mathbf{x}|\mathbf{x})x_q = \int_{x_q}^{\infty} dx p(x). \quad (1)$$

The conditional tail expectation (CTE) [16] of a random variable \mathbf{x} is given by

$$E[\mathbf{x}|\mathbf{x})x_q] = \frac{\int_{x_q}^{\infty} dx xp(x)}{\int_{x_q}^{\infty} dx p(x)}, \quad (2)$$

where x_q is the value of \mathbf{x} for the q -th quantile. Thus, the CTE of a function $f_i(\mathbf{x})$ is given by $E[f_i(\mathbf{x})|\mathbf{x})x_q]$. This quantity can be interpreted as the mean of a function of the random variable given an extreme case scenario. In our analysis, we simply extend these definitions to a distribution of grain size and aspect ratio and prescribe a threshold (i.e., G_c and a_c) in this two-dimensional space. In addition to the exceedance, we define three metric functions that characterize important, yet distinct shape features of the tail, namely $f_1 = G$, $f_2 = G^2 - \bar{G}^2$ and $f_3 = G^2 + a^2$, where the overbar denotes the CTE of \mathbf{G} . Finally, we add to this list a fifth metric, namely the correlation between \mathbf{G} and \mathbf{a} (denoted by corr) [17].

In this context the CTE can be written in terms of the joint probability density function of the grain size and aspect ratio. More specifically, taking \mathbf{G} and \mathbf{a} as continuous random variables with the associated joint pdf $p(G, a)$, one can define the CTE of the microstructural function $f_i(G, a)$ as

$$E[f_i(\mathbf{G}, \mathbf{a})|\mathbf{G} > G_c, \mathbf{a} > a_c] = \frac{\int_{G_c}^{\infty} \int_{a_c}^{\infty} dG da f_i(G, a) p(G, a)}{\int_{G_c}^{\infty} \int_{a_c}^{\infty} dG da p(G, a)}, \quad (3)$$

where G_c (a_c) is a grain-size (aspect ratio) cutoff. One can similarly define the exceedance [15], the fraction of grains in the tail, in terms of the pdf as

$$\text{exc}(G_c, a_c) = \int_{G_c}^{\infty} \int_{a_c}^{\infty} dG da p(G, a), \quad (4)$$

In practice, the cutoffs G_c and a_c can be determined in various ways including, for example, fixing them to specific quantiles of the distribution. We chose to determine the cutoffs by calculating an effective radius of gyration of probability mass points for each sample considered here. In any case, the sensitivity of the analysis to the cutoff values can be assessed, as described below.

To interpret these metrics, we compare them with the corresponding quantities defined for normal grain growth (NGG). There has been considerable debate in the literature as to the appropriate form of the probability density function that describes NGG. For example, Palmer et al. [18] investigated normal and secondary grain growth in Ge thin films, fitting the acquired grain-size distribution to the lognormal, Louat [19] and Hillert [20] distributions, and it was found that the lognormal density provided the best fit at small grain sizes. In addition, Barmak et al. [21] investigated grain growth kinetics in CoPt alloy films and found that the grain-size distribution is lognormal at stagnation. More recently, Rickman et al. [22] calculated approximate, analytical corrections to the lognormal pdf of grain diameter for microstructures associated with polycrystalline, high-purity Al_2O_3 samples. Despite the success of a lognormal description, it should be noted that Mullins [23] has formulated a Fokker-Planck equation describing grain growth, and has used this approach to argue that the aforementioned stochastic models appear to have no theoretical

justification. In short, while Donegan et al. [13] have shown that grain-size distributions vary to some extent from the commonly accepted lognormal distribution in almost all cases, there does not exist an analytical description for this distribution that is applicable in general.

With this in mind, we make the *ansatz* here that for NGG \mathbf{G} and \mathbf{a} are independent random variables with lognormal and normal probability densities, $p^{\log}(G)$ and $p^{\text{gauss}}(a)$, respectively. The joint probability density is then given by

$$p^{\text{NGG}}(G, a) = p^{\log}(G) p^{\text{gauss}}(a), \quad (5)$$

where the lognormal density with parameters $\bar{\mu}$ and $\bar{\sigma}$ is given by

$$p_{\bar{\mu}, \bar{\sigma}}^{\log}(G) = \frac{1}{\sqrt{2\pi}\bar{\sigma}G} \exp\left[-\frac{(\ln G - \bar{\mu})^2}{2\bar{\sigma}^2}\right], \quad (6)$$

and the normal (Gaussian) density with parameters μ and σ is given by

$$p_{\mu, \sigma}^{\text{gauss}}(G) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(G - \mu)^2}{2\sigma^2}\right]. \quad (7)$$

The parameters $\bar{\mu}$ and $\bar{\sigma}$ were chosen so that the location and height of the peak of the lognormal density matches that of the marginal density $\int_0^{\infty} dG p(G, a)$. Thus, if the peak location and the peak height are given by G' and p' , respectively, then one finds that

$$p'G' = \frac{1}{\sqrt{2\pi}\bar{\sigma}} \exp\left(-\frac{\bar{\sigma}^2}{2}\right), \quad (8)$$

$$G' = \exp(\bar{\mu} - \bar{\sigma}^2).$$

Similarly, the parameters μ and σ were chosen so that the location (μ) and height ($1/\sqrt{2\pi}\sigma$) of the peak of the normal density matches that of the marginal density $\int_0^{\infty} da p(G, a)$.

With this prescription for the joint probability density for NGG, we then define the metrics

$$\phi_i = \frac{E[f_i(\mathbf{G}, \mathbf{a})|\mathbf{G} > G_c, \mathbf{a} > a_c]}{E^{\text{NGG}}[f_i(\mathbf{G}, \mathbf{a})|\mathbf{G} > G_c, \mathbf{a} > a_c]}, \quad (9)$$

for $i = 1, 2$ and 3 and

$$\phi_4 = \frac{\text{exc}(G_c, a_c)}{\text{exc}^{\text{NGG}}(G_c, a_c)}, \quad (10)$$

$$\phi_5 = 1 + \text{corr}(\mathbf{G}, \mathbf{a}),$$

where the superscript NGG denotes quantities calculated using the assumptions of normal grain growth. Given these definitions, ϕ_i is unity for NGG. We note that, given this *ansatz*, our working definition of a microstructure exhibiting AGG is one whose joint probability density $p(G, a)$ differs from that characterizing NGG. The aforementioned metrics simply permit a quantification of some salient differences.

2.2. Canonical correlation analysis

Our task is to quantify correlations to highlight those processing variables most responsible for AGG. For this purpose, we employ canonical correlation analysis (CCA) [24,25]. CCA is a generic, linear parametric model that is employed in the statistical analysis of data when there are many interrelated or interdependent variables. In this context, it is used to identify the linear combinations of processing variables (inputs) and metrics (outputs), known as canonical

variates, that maximize their cross-correlation. These input, \bar{X} , and output, \bar{Y} , variables can be summarized as follows.

$$\begin{aligned}\bar{X} &= \{c_{\text{MgO}}, c_{\text{CaO}}, c_{\text{Na}_2\text{O}}, c_{\text{SiO}_2}, r, T, t\}, \\ \bar{Y} &= \{\phi_1, \phi_2, \phi_3, \phi_4, \phi_5\}\end{aligned}\quad (11)$$

An extended discussion of CCA and its relationship to other dimensionality reduction techniques is given in the [Appendix](#).

In brief, a CCA begins with the construction of a covariance matrix (or the associated correlation matrix) Σ between the variables for the 33 samples considered here [24]. For the set of input and output variables, X_i ($i = 1, 2, \dots, 7$) and Y_i ($i = 1, 2, \dots, 5$) (see Eq. 11), one can define the corresponding components of the matrices \mathbf{X} and \mathbf{Y} , namely X_{ij} and Y_{ij} , respectively, where $j = 1, 2, \dots, 33$. Then, Σ may be written in block form as

$$\Sigma = \begin{bmatrix} \Sigma_{\bar{X}\bar{X}} & \Sigma_{\bar{X}\bar{Y}} \\ \Sigma_{\bar{Y}\bar{X}} & \Sigma_{\bar{Y}\bar{Y}} \end{bmatrix}. \quad (12)$$

Next, take the weighting coefficients α_i and β_i as the components of a 7- and a 5-dimensional vector, respectively. These quantities are the components of the eigenvectors of the product matrices,

$$\begin{aligned}\Sigma_{\bar{X}\bar{X}}^{-1} \Sigma_{\bar{X}\bar{Y}} \Sigma_{\bar{Y}\bar{Y}}^{-1} \Sigma_{\bar{Y}\bar{X}}, \\ \Sigma_{\bar{Y}\bar{Y}}^{-1} \Sigma_{\bar{Y}\bar{X}} \Sigma_{\bar{X}\bar{X}}^{-1} \Sigma_{\bar{X}\bar{Y}},\end{aligned}\quad (13)$$

respectively. The square root of the corresponding eigenvalues are the canonical correlations.

The corresponding eigenvectors may be used to determine the canonical variates, $V = \sum_i \alpha_i X_i$ and $W = \sum_i \beta_i Y_i$, where α_i and β_i are the associated canonical weights. To determine the correlation between these variates and the original variables, one can also calculate the canonical loadings from the correlation matrix. The output for a typical CCA analysis consists then of the canonical correlations, and the canonical weights and loadings for the input and output variables. The weights quantify the relative importance of input variables in obtaining a maximum correlation with output variables, while the loadings are the correlation coefficients between the canonical variates and the input/output variables from which they are constructed [24,27]. Canonical loadings therefore reflect the relative contribution of each variable to a given canonical variate [28]. These results are supplemented by tests to ascertain their significance, such as the calculation of the Wilks lambda [29], that focuses on the aforementioned eigenvalues. More specifically, significance is determined via a hypothesis test in which one tests the null hypothesis that a given pair of variates is uncorrelated. If there is contrary evidence of a correlation at a certain level of significance (usually corresponding to a p-value of less than 0.01), then this pair of variates is accepted for analysis, and the process is repeated for subsequent pairs of variates. In addition, a redundancy coefficient may be calculated that reflects the amount of the variance explained in one canonical variate by another, corresponding canonical variate. In other words, this coefficient is a measure of the ability of, for example, the input variables to explain observed variations in the output variables [27].

We note that the identification of a subset of relevant eigenvectors results in an effective reduction in the dimensionality of the parameter space that describes the full range of input and output variables. This reduction in the size of the multidimensional parameter space is critical for identifying the relevant linear combination of variables that are important for our analysis. It should be noted that one limitation of the CCA is that it provides only a linear combination of variables that maximizes input/output cor-

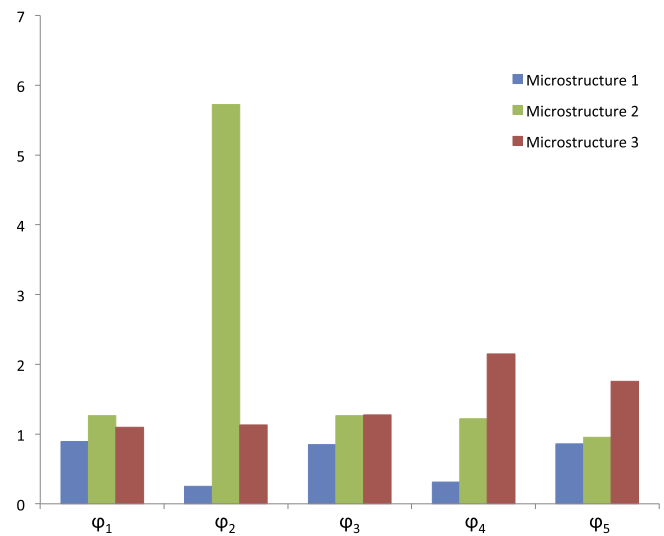


Fig. 3. A summary of the values of the metrics for the three microstructures shown in Fig. 1. As is evident from the figure, several metrics, including ϕ_2 , ϕ_4 and ϕ_5 , differ significantly from sample to sample.

relations. However, as will be seen below, this linear combination provides very useful information.

3. Results

These five abnormality metrics were tabulated for the microstructures shown in Fig. 1a–c and are summarized in the bar chart in Fig. 3. As is evident from the chart, differences in the values of the ϕ_i provide a means to distinguish among the populations of extreme grains in these microstructures. In particular, several metrics, including ϕ_2 , ϕ_4 and ϕ_5 , differ significantly from sample to sample. They capture distinct and important aspects of AGG. From a practical point of view, we can use these metrics to investigate correlations between sample processing variables, such as composition and temperature, and the presence of abnormal grains. To demonstrate the utility of this approach, we compiled statistics for microstructures associated with grain growth taken from 33 independent specialty alumina samples, each having a unique combination of processing variables, with grain populations in the range of approximately 1000–9000 grains. More specifically, powders with 11 different compositions, c , were employed through a combination of process-related chemical alterations of MgO, CaO, Na₂O, and SiO₂ and 9 different heat treatments. All powders were spark-plasma sintered and, in some cases, underwent additional heat treatments to promote grain growth. In addition to these four composition variables, we augment the list of processing variables with a ratio of composition variables, $r = c_{\text{MgO}}/(c_{\text{CaO}} + c_{\text{SiO}_2})$, as well as temperature, T , and time, t .

We find from our data that only the first canonical variate for the processing variables and the metrics, with associated canonical correlation 0.92, is required to adequately characterize input/output correlations. This finding was based on the outcome of the hypothesis test of significance described above. The results of a CCA analysis of our data are summarized in Table 1 and graphically in Fig. 4. Focusing on Fig. 4, the absolute values of the canonical weights, namely the α_i (blue circles) and β_i (red circles), are plotted versus the absolute values of the loadings for both the input and output variables. This figure can be understood by first examining the canonical weights, where it is evident that the canonical variate for the processing variables is dominated by the composition ratio, r , as the associated (negative) weight is substantially larger than that

Table 1

The canonical weights and loadings, α_i ($i = 1, 2, \dots, 7$) and β_j ($j = 1, 2, \dots, 5$), for the processing variables and the metrics, respectively.

α_i / β_j	Weights	Loadings
C_{MgO}	0.002	-0.65
C_{CaO}	-0.005	-0.34
C_{Na_2O}	-0.0006	-0.30
C_{SiO_2}	-0.005	-0.15
r	-0.98	-0.81
T	-0.0006	0.01
t	0.025	0.42
ϕ_1	0.039	0.52
ϕ_2	0.0008	0.80
ϕ_3	1.05	0.89
ϕ_4	-0.005	0.14
ϕ_5	2.92	0.90

for the other input variables. Using the same criterion, the corresponding variate for the metrics is therefore essentially a weighted average between ϕ_3 and ϕ_5 , with the weighting coefficients in the ratio of approximately 1–3. That is, $V \approx \alpha_5 X_5$ and $W \approx \beta_3 Y_3 + \beta_5 Y_5$.

The interpretation of these results requires some care. Given the strong correlation between the canonical variates, one interpretation is that large values of the composition ratio produce less abnormal microstructures, as characterized by ϕ_3 and ϕ_5 ($\alpha_5 < 0$ while β_3 and β_5 are positive). This finding is consistent with previous work that examined specifically the effects of MgO in relation to other impurities in alumina [6]. The canonical loadings are also summarized in Fig. 4, and the relative size of the circles indicates the fraction of the variance that is explained by a given variable. As is evident from the figure, the composition ratio explains a fair fraction of the processing parameter variate while ϕ_2 , ϕ_3 and ϕ_5 explain a significant fraction of the properties variate. In addition, we find from a redundancy analysis that approximately 51% of the variance in the metrics is explained by the input (i.e. processing) variables.

Thus, while the canonical loadings are consistent with the weights in many respects, there are indications that other variables, such as the composition of MgO (C_{MgO}) and the conditional tail variance of $G(\phi_2)$, may be important factors in the set of processing and metric variables, respectively. To examine these relationships in more detail, it is expected that we will need grain size and aspect ratio information from additional samples. Furthermore, as we have data for only a limited range of processing time, it will be especially important to obtain additional, time-dependent microstructural information. Nevertheless, more information can be obtained by performing a sensitivity test with the existing data. In particular, since disparities between canonical weights and variates may be an indication of variable collinearity, we repeated the CCA after eliminating some of those independent variables and found that the analysis was robust. In short, while the number of samples employed here is not especially large given that a relatively large number of grains must be analyzed per sample, we believe that the conclusions drawn here are significant and describe important trends in the data.

4. Discussion and conclusions

In this paper we construct microstructural maps reflecting the joint probability distribution of grain area and aspect ratio and associated metrics to quantify various microstructural features associated with abnormal grain growth. These maps provide a visual “fingerprint” of abnormal growth and, by examining correlations among the metrics and processing variables, we identify those variables that control observed microstructural features. One additional benefit of our approach is that it provides a working definition of AGG in terms of the joint probability density of grain size and aspect ratio. More specifically, we take any difference between this density and that prescribed for NGG to be a signature of abnormality. Our methodology was validated using high-purity aluminum oxide doped

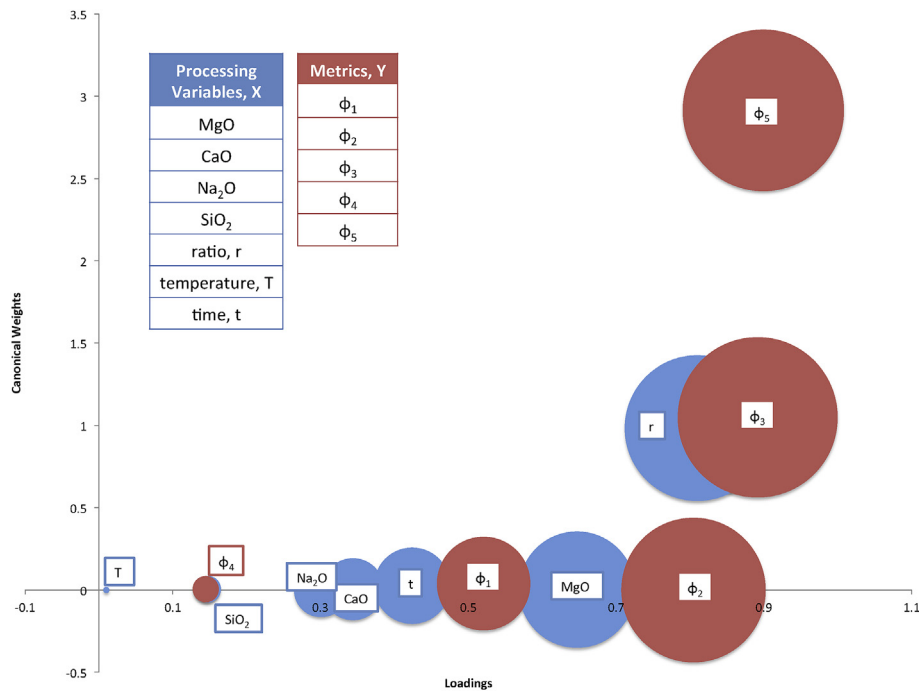


Fig. 4. A schematic showing the absolute values of the weights and loadings from the canonical correlation analysis. Blue circles denote processing variables while red circles denote metrics. The larger values of the weights associated with the composition ratio and the metrics ϕ_3 and ϕ_5 indicate that they comprise large fractions of the canonical variates. The relative size of the circles indicates the fraction of the variance associated with each variable. The table summarizes the processing variables (inputs) and the metrics (outputs). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

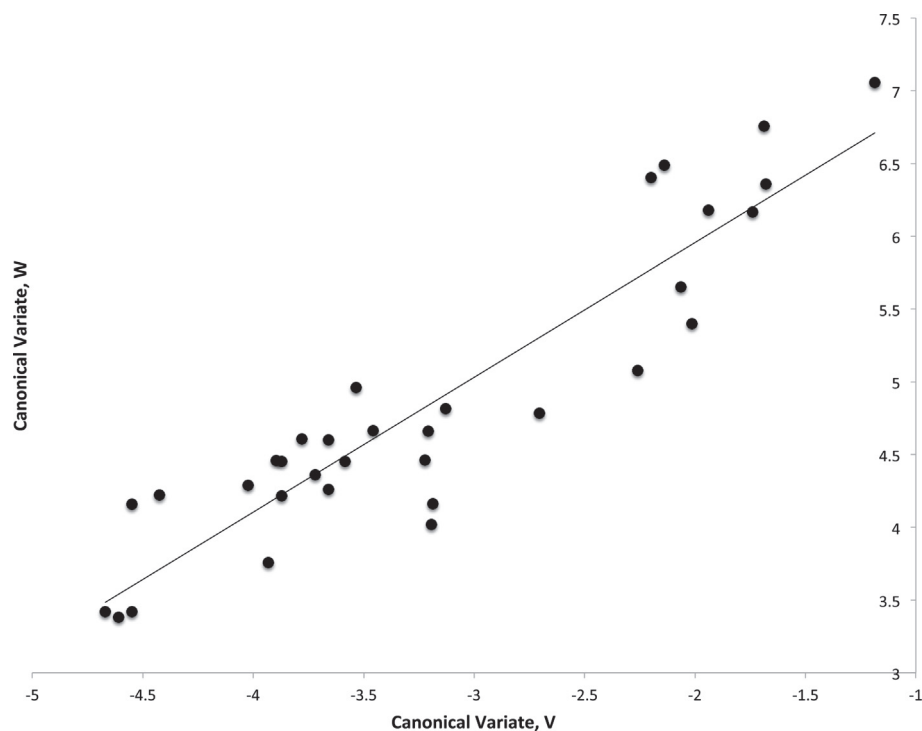


Fig. 5. A correlation plot showing the metric variate W versus the processing variate V for the 33 samples of specialty alumina considered here. The line is a least-squares fit to this data.

with selected impurities, but it can obviously be applied to other systems and to other sets of processing variables.

The methodology outlined here suggests a useful and robust strategy for materials analysis and design. In particular, the canonical correlation analysis of processing variables and metrics provides guidance as to which combination of inputs will produce microstructures with specific attributes. These attributes are linked, in turn, with desirable properties. To illustrate this capability, the correlation plot shown in Fig. 5 displays the metric variate W versus the processing variate V for the 33 samples of specialty alumina considered here. It should be noted that, while there is scatter in the data, there are no obvious outliers that will significantly affect the analysis. With regard to the fit shown in the figure, we note that the associated canonical correlation coefficient is 0.92. The coefficient of determination (i.e., r^2) is, therefore, 0.85, implying a very satisfactory goodness-of-fit despite some scatter. This assessment is strengthened by hypothesis testing using Wilks lambda, which tests the null hypothesis that the correlations between canonical variates is not very different from zero. The resulting p-value was about 0.001, permitting the rejection of the null hypothesis and confirming a significant correlation.

The information in Fig. 5 can be used for planning experiments. For example, one can obtain a value, W' , for a new value of the processing variate, V' , from the least-squares fit for a list of hypothetical processing variables. Alternatively, one can identify those processing variables that will produce a microstructure having certain target metrics. In addition, one can assess the impact of different cutoffs (G_c and a_c) on the analysis, and the sensitivity of the results to changes in either the list of processing or metric variables.

In conclusion, we note that many aspects of materials performance, such as degradation and service lifetime, depend not on mean values, but on rare events in microstructures. Abnormal grain growth is an exemplar of such a rare event, and the analysis that we offer here is likely to be a useful tool for understanding better these rare events. Our approach has the virtue that either variable set may be

augmented based on the accumulation of additional data, or in response to particular design specifications. Moreover, as a dimensional reduction strategy, the correlation analysis described here reduces a complex, multidimensional parameter space and thereby simplifies experimental planning. This approach can be generalized to other problems involving statistical correlations among input and output variables and, hence, is applicable to a wide spectrum of problems. For example, it can be used to examine correlations among processing variables and the thermomechanical and kinetic properties of materials (e.g., strength, hardness, thermal conductivity, etc.) by including these properties in the list of output variables. The application of this methodology to other systems and other metric sets is the subject of ongoing work.

Acknowledgments

The authors wish to thank Almatix, Inc. for kindly supplying the alumina samples for our microstructural analysis and for their guidance during this project. The authors also acknowledge support from the Office of Naval Research under grant N00014-11-1-0678.

Appendix.

While some materials researchers are aware of powerful dimensional reduction techniques such as principal-component analysis (PCA) [26], in general familiarity with such methods and their variants is minimal. We therefore provide here a somewhat more detailed discussion of CCA. As noted by Knapp, CCA is a general methodology for investigating relationships between two sets of variables, and most parametric tests of significance can be regarded as special cases of CCA [30]. More specifically, given two paired data sets that may be of unequal length, a CCA determines paired directions such that the projection of the first data set along the first direction is maximally correlated with the projection of the second data set along the second direction [31]. By ranking each pair of directions in terms of its associated correlation coefficient, one can

identify a subspace that best represents the data and achieve a reduction in dimensionality. Given these capabilities, the CCA is well-suited to investigate correlations among the processing and output variables in the problem studied here.

It is useful to compare the CCA with other correlation techniques. For example, a PCA involves a *single* data set, represented by \bar{x}_i in a d -dimensional space, for which one wishes to find a direction, \bar{v} , such that the projection $\bar{v}\bar{x}_i$ best represents the set [31]. Both PCA and CCA involve the calculation of spectra of operators that reflect variable correlations and the interpretation of the associated eigenvectors as special directions in a high-dimensional space. Moreover, as noted by Burges, both PCA and CCA projections decorrelate individual data sets; however, CCA also leads to vanishing cross-correlations and the identification of conjugate directions [31]. We note that a general multiple regression analysis is a special case of the CCA in which one data set contains only a single variable. Similarly, a one-way analysis of variance, as it is equivalent to a multiple regression for certain input/output variable number relationships, is also a special case [30,32].

References

- [1] H.V. Atkinson, *Acta metall.* 36 (1988) 469–491.
- [2] S.K. Kurtz, F.M.A. Carpay, Microstructure and normal grain-growth in metals and ceramics. I. theory, *J. Appl. Phys.* 51 (1980) 5725–5744.
- [3] A.M. Deus, M.A. Fortes, P.J. Ferreira, J.B. Vander Sande, A general approach to grain growth driven by energy density differences, *Acta Mater.* 50 (2002) 3317–3330.
- [4] A.D. Rollett, D.J. Srolovitz, M.P. Anderson, Simulation and theory of abnormal grain growth - anisotropic grain boundary energies and mobilities, *Acta Metall.* 37 (1989) 1127–1240.
- [5] W.W. Mullins, J. Vñals, Linear bubble model of abnormal grain growth, *Acta Mater.* 50 (2002) 2945–2954.
- [6] C.A. Handwerker, P.A. Morris, R.L. Coble, Effects of chemical inhomogeneities on grain growth and microstructure in Al_2O_3 , *J. Am. Ceram. Soc.* 72 (1989) 130–136.
- [7] P.R. Cantwell, M. Tang, S.J. Dillon, J. Luo, G.S. Rohrer, M.P. Harmer, Grain boundary complexions, *Acta Mater.* 62 (2014) 1–48.
- [8] D.A.H. Hanaor, W. Xu, M. Ferry, C.C. Sorrell, Abnormal grain growth of rutile TiO_2 induced by ZrSiO_4 , *J. Cryst. Growth* 359 (2012) 83–91.
- [9] N.P. Padture, B.R. Lawn, Toughness properties of a silicon carbide with an *in-situ* induced heterogeneous grain structure, *J. Am. Ceram. Soc.* 77 (1994) 2518–2522.
- [10] M. Mulla, V. Krstic, Pressureless sintering of β -SiC with Al_{12}O_3 additions, *J. Mater. Sci.* 29 (1994) 934–938.
- [11] J. May, D. Turnbull, Secondary recrystallization in silicon iron, *Trans. Metall. Soc. AIME* 212 (1958) 769–781.
- [12] K. Rajan, Materials informatics: the materials “Gene” and big data, *Annu. Rev. Mater. Res.* 45 (2015) 153–169.
- [13] S.P. Donegan, J.C. Tucker, A.D. Rollett, K. Barmak, M. Graeber, Extreme value analysis of tail departure from log-normality in experimental and simulated grain size distributions, *Acta Mater.* 61 (2013) 5595–5604.
- [14] A. Necir, A. Rassoul, R. Zitikis, Estimating the conditional tail expectation in the case of heavy-tailed losses, *J. Prob. Stat.* (2010) 596839.
- [15] E. Castillo, A. Hadi, N. Balakrishnan, J.M. Sarabia, *Extreme Value and Related Models with Applications in Engineering and Science*, John Wiley and Sons, Inc., Hoboken, NJ, 2005.
- [16] Z. Landsman, A. Tsanakas, Tail conditional expectations for exponential dispersion models, *ASTIN Bull.* 35 (2005) 189–209.
- [17] A. Hayter, *Probability and Statistics for Engineers and Scientists*, fourth ed., Brooks/Cole, Boston, MA, 2012.
- [18] J.E. Palmer, C.V. Thompson, H.I. Smith, Grain growth and grain size distributions in thin germanium films, *J. Appl. Phys.* 62 (1987) 2492–2497.
- [19] N.P. Louat, On the theory of normal grain growth, *Acta Metall.* 22 (1974) 721–724.
- [20] M. Hillert, On the theory of normal and abnormal grain growth, *Acta Metall.* 13 (1965) 227–238.
- [21] K. Barmak, R.A. Ristau, K.R. Coffey, M.A. Parker, J.K. Howard, Grain growth and ordering kinetics in Co-Pt thin films, *J. Appl. Phys.* 79 (1996) 5330–5332.
- [22] J.M. Rickman, A. Lawrence, A.D. Rollett, M.P. Harmer, Calculating probability distributions associated with grain-size distributions, *Comp. Mat. Sci.* 101 (2015) 211–215.
- [23] W.W. Mullins, Grain growth of uniform boundaries with scaling, *Acta Mater.* 46 (1998) 6219.
- [24] J.D. Jobson, *Applied Multivariate Data Analysis*, Vol. II, Springer-Verlag, New York, 1992.
- [25] R. Gittins, *Canonical Analysis: A Review with Applications in Ecology*, Springer-Verlag, Berlin, 1985.
- [26] S. Samudrala, K. Rajan, B. Ganapathysubramanian, Data dimensionality reduction in materials science, in: K. Rajan (Ed.), *Informatics for Materials Science and Engineering*, Elsevier, New York, 2013.
- [27] A.A.A. Kuylen, T.M.M. Verhallen, The use of canonical analysis, *J. Econ. Psych* 1 (1981) 217–237.
- [28] J.F. Hair, W.C. Black, B.J. Babin, R.E. Anderson, R.L. Tathan, *Canonical Correlation Analysis - Supplement to: Multivariate Data Analysis*, sixth ed., Pearson Prentice Hall, Upper Saddle River, NJ, 2006.
- [29] R.M. Thorndike, *Correlational Procedures for Research*, Gardner Press, New York, 1978.
- [30] T.R. Knapp, Canonical correlation analysis: a general parametric significance-testing system, *Psych. Bull.* 85 (1978) 410–416.
- [31] C.J.C. Burges, Dimension reduction: a guided tour, *Found. Trends Mach. Learn.* 2 (2009) 275–365.
- [32] J. Cohen, Multiple regression as a general data-analytic system, *Psych. Bull.* 70 (1968) 426–443.