International Journal of Production Research,
Vol. 44, No. 14, 15 July 2006, 2839–2851

Taylor & Francis
Taylor & Francis Group

Two-stage data mining for flaw identification in ceramics manufacture

ORHAN DENGIZ†, ALICE E. SMITH*† and IAN NETTLESHIP‡

†Department of Industrial and Systems Engineering, Auburn University, Auburn, AL 36849, USA
‡Department of Materials Science & Engineering, University of Pittsburgh, Pittsburgh, PA 15261, USA

(*Received July 2005*)

Advanced ceramics are commonly manufactured by sintering high-purity powders. The design of ceramic elements is governed by its fracture strength, which is greatly influenced by microstructural flaws. Three ceramic powder preparation methods for ceramics manufacturing are considered in this paper—uniaxial pressing followed by isostatic pressing, flocculated slip casting, and dispersed slip casting. Their effects on the growth and characteristics of microstructure flaws and damage on the ceramic surface are investigated using a two-stage data-mining approach. In the first stage, digital microstructural images are mined to characterize the flaws and surface damage. In the second stage, an extreme value probability distribution is fitted using the information from stage 1. The extreme value distribution estimates large flaws which are highly correlated with subsequent fractures. Results of the two-stage data mining show that ceramic production method significantly affects flaw characteristics that, in turn, determine the ceramics' fracture strength.

Keywords: Advanced ceramics; Data mining; Extreme value distribution; Image processing; Powder manufacturing; Sintering

1. Introduction

Advanced ceramics have superior physical and chemical characteristics that make them fit in a wide range of application areas (Reeber 1993). Although used today in many fields including electronics and chemical applications such as spark plugs, electro-mechanical sensors, and pipe seals, the structural uses are currently limited. Compared with most metallic materials, the uniaxial strength of ceramics is more sensitive to size, shape, and surface conditions. Since ceramics are highly brittle, failures are often sudden, with no prior sign. Because of this property, the fracture strength governs structural ceramic design, and large safety factors are required, which means increased weight and cost.

The fracture strength depends on complex interactions between material characteristics and processing variables. Volume flaws that are present in ceramics

*Corresponding author. Email: aesmith@eng.auburn.edu

International Journal of Production Research
ISSN 0020–7543 print/ISSN 1366–588X online © 2006 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/00207540500534454

can originate or propagate failures (Krell 1998). Fracture strength can be better controlled, or estimated if the flaw sizes are controlled. Improving the materials performance through microstructural control became a new interest of researchers following the better understanding of fracture dynamics (Reeber 1993). A design approach that involves statistical analysis to compute probabilities of fracture is necessary for ceramics (Evans 1984, Quinn and Morrell 1991).

Advanced ceramics are commonly manufactured by a process called sintering. Sintering is a powder-manufacturing process where a pre-formed powder is kept under high temperature levels for a certain period of time. The resulting product is a densified, solid ceramic. The ceramic powder can be prepared in different ways before the sintering process. The choice of the forming method depends on the shape of the manufactured piece and the experience of the manufacturer. Three methods—die pressing, slip casting with flocculated slips, and slip casting with dispersed slips—are most commonly used, and therefore these are the methods investigated in this paper. Even if two different ceramics are from the same raw material, the preparation and process methods have a significant impact on material properties (Quinn and Morrell 1991).

Internal flaws, in the form of dense heterogeneities, grow in ceramics at the microstructural level during sintering. These flaws, if large, can initiate or propagate fracture under stress, causing the ceramic to suddenly fail. With a high frequency of internal defects that are larger than machining-induced flaws, strength depends on the internal flaws and not on the machining defects (Krell 1998). To investigate the effects of manufacturing process on the final ceramic property, this paper uses a two-stage data-mining approach. In stage 1, an image analysis is conducted on the microstructural images of sintered samples to recover the heterogeneities (flaws) that are formed during sintering, both internally and on the polished surface. In stage 2, data analysis using a common extreme value probability distribution, the generalized Pareto distribution (GPD), establishes the relationship between powder preparation and the characteristics of the large flaws in the ceramic. GPD has been widely applied in natural sciences, especially for disastrous events whose effects are catastrophic beyond a certain magnitude but considered regular otherwise. Examples of such natural events are earthquakes, rainfall, and hurricanes. GPD has also been applied to modelling high risks in financial markets (Gilli and Kellezi 2005), earthquake magnitudes (Caers *et al*. 1999), heavy winds (Simui and Heckert 1995, Heckert *et al*. 1998), and particle size distributions (Dierickx *et al*. 2000). The only application of GPD similar to that in this paper was for the estimation of the size of the maximum inclusion in clean steels (Shi *et al*. 1999, 2001, Anderson *et al*. 2000, 2003).

Data mining is new to the area of materials manufacture. There have been a few distantly related works in the literature. Hofmann and Apostolakis (2003) use data mining to predict molecule arrangement in crystals. They consider a large database of entries concerning the known chemical properties of certain crystals. The data are first filtered by removing poor-quality observations. They then define problem-specific descriptors and train them using simple decoys generated by slight distortions of the observed data. The descriptors are trained to discriminate between the observed data and the decoys. Chen *et al*. (1998, 2000) develop a proprietary commercial software that uses data mining for materials manufacture. They use a neural network to develop processing relationships and a genetic algorithm

optimizer to find the best conditions for the preparation or production of materials, and give three short, high-level application possibilities in rubber and aluminium manufacture. Evans *et al*. (2001) use data mining with neural networks to select proper settings for ink-jet printing on ceramics. This was a work in progress and treats the data mining as a combinatorial optimization problem. None of these cited papers considers image processing or material microstructure.

The remainder of this paper is organized as follows; image processing for stage 1 data mining is described in detail in section 2. In section 3, the stage 2 data mining by statistical extreme-value analysis is described. Results and conclusions are given in sections 4 and 5, respectively.

2. Stage 1: data-mining image processing

2.1 *Acquiring digital images*

Grayscale digital microstructural images were acquired using an optical microscope with a digital camera attached to the eyepiece. Dense heterogeneities show up as brighter regions than their less dense surroundings within the viewed sample surface. The surface pull-out damages appear as very dark areas on the sample surface. However, a considerable amount of noise and imperfections is present in the images, which could not be eliminated due to the reflective properties of the ceramics. This rendered commonly available simple data-mining tools such as thresholding and density slicing ineffective. Thresholding is used when extracting parts of a digital image that are either darker or brighter than a certain level. Density slicing is used to extract parts of a digital image that fall within a certain predetermined interval of two greyscale values. It is a binary conversion tool where all pixels that fall within the interval are set to one binary value, while pixels outside the interval ('slice') are set to the other binary value. Since none of the conventional feature extraction algorithms performed satisfactorily alone, a more complex data-mining algorithm was needed to automatically enhance the images and set the best thresholding and density slicing parameters for every image to effectively recover the features of interest.

2.2 *Image-processing data mining*

As stated before, internal flaws appear as brighter regions than their surroundings, whereas the surface damage appears as darker regions, usually very close to black (figure 1). For this reason, extracting both types of flaws requires two different algorithms. The major difference between the two data-mining algorithms of stage 1 is the colour of the detected feature.

For each data-mining algorithm, sample images were taken randomly, and recovery of the features of interest was done by a human operator while recording the density slice parameter values. Out of the entire set of acquired images, 28 images of flaws were used to develop an automated algorithm to mine the flaw data, and 18 images of surface damage were used to develop an automated algorithm to mine the surface damage data. More images were needed for the flaw data-mining

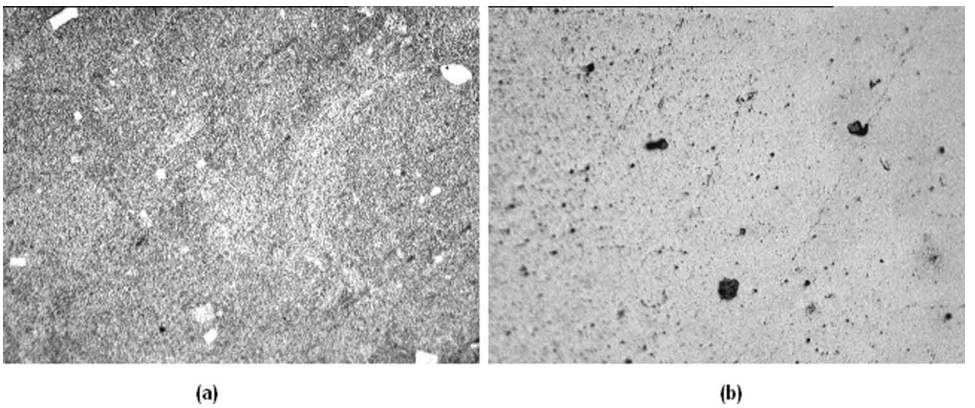

Figure 1. (a) Microstructural flaws and (b) surface-damage areas.

Table 1. Regression analysis results for equations (1)–(3).

		ANOVA					
Regression	R^2 (%)	Source	df	SS	MS	F	p
Equation (1)	87.1	Model	7	22090.2	3155.7	19.25	0.000
		Residual error	20	3277.9	163.9		
		Total	27	25368.1			
Equation (2)	79.2	Model	7	13496.3	1928.0	10.86	0.000
		Residual error	20	3551.8	177.6		
		Total	27	17048.1			
Equation (3)	78.0	Model	3	2970.5	990.2	16.57	0.000
		Residual error	14	836.5	59.8		
		Total	17	3807.0			

algorithm because flaw images have a greater variation in brightness and contrast among them than do the surface damage images. The manual recovery of the features was followed by a regression analysis to develop an estimation function for the best density slicing intervals using the predictor variables average, standard deviation, minimum and maximum grey values, and dependent variables of the density slice parameters. The statistical software package Minitab was used to conduct the regression analyses. The results of the regression analyses are equations (1)–(3). Minitab outputs for the analysis of variance and R^2 values are provided in table 1. It is important to note that these regression equations are constructed using samples from the data collected during this study. Therefore, although the procedures can be applied in general, the given equations will perform satisfactorily only for images acquired in a similar manner. Images with different colour and brightness characteristics will require their own regression analyses to form equations (1)–(3).

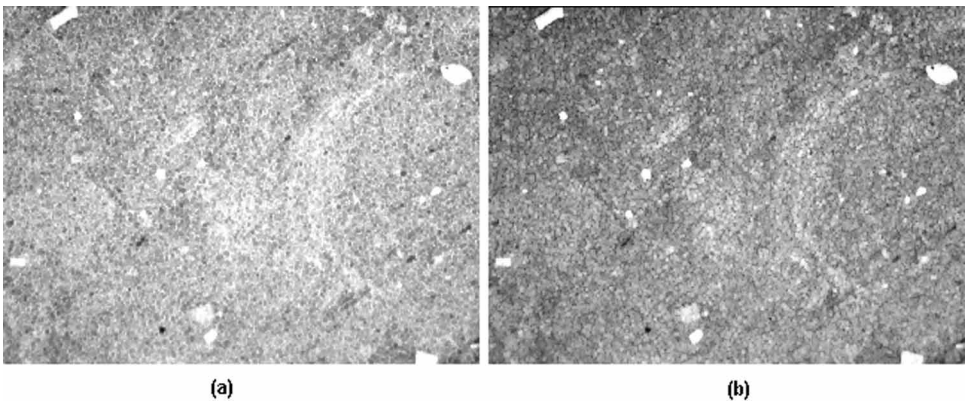

Figure 2. Image of figure 1(a) using the min–max filter (a) and max–min filter (b).

Processing with the algorithms is as follows: the images are enhanced by min and max rank filters. These are standard operators found in Scion Image software (http://www.scioncorp.com). Notice that the ordering of these two filters is reversed and gives different results. Figure 2 shows figure 1(a) with min–max on the left (figure a) and max–min on the right (figure b). After image enhancement, the best density slice interval is calculated by the regression equation, and density slicing is applied to recover the features. Feature recovery is followed by noise reduction, the standard 'Erode' and 'Dilate' functions, both standard Scion Image operators. Erode takes the minimum pixel value of a moving window (3×3 in this application) and replaces the value of the centre pixel with that minimum value. Dilate does the same thing but uses a maximum pixel value. Density slicing is the main operator to extract features, whereas others are required to enhance image quality and eliminate noise.

2.2.1 Algorithm for the extraction of the microstructure flaws.

(1) Set two copies of the original image to Orig_1 and Orig_2.
(2) Apply max and min filters to Orig_1, min and max filters to Orig_2.
(3) Set Mean_1 and Mean_2 as the average grey value of Orig_1 and Orig_2.
(4) Set St Dev_1 and StDev_2 as the standard deviation of the grey values of Orig_1 and Orig_2.
(5) Set Min_1 and Min_2 to the minimum grey value in Orig_1 and Orig_2.
(6) Calculate DSup_1 as the upper density slice level of Orig_1 from regression equation (1).
(7) If ($\text{DSup}_1 > 255$) or ($\text{DSup}_1 < 0$) Then ($\text{DSup}_1 = \text{Min}_1$).
(8) Calculate DSup_2 as the upper density slice level of Orig_2 from regression equation (2).
(9) If ($\text{DSup}_2 > 255$) or ($\text{DSup}_2 < 0$) Then ($\text{DSup}_2 = \text{Min}_2$).
(10) Density slice Orig_1 with levels (0, DSup_1) to get Bin_1 and Orig_2 with levels (0, DSup_2) to get Bin_2.

(11) Apply image math to combine Bin_1 and Bin_2 by OR operator, as BinCombine.
(12) Apply Erode and Dilate binary operators respectively to BinCombine.

$$\begin{aligned} DSup_1 = {} & -102 + 2.42 \cdot Mean_1 - 2.5 \cdot StDev_1 + 1.18 \cdot Min_1 - 0.00084 \cdot Mean_1^2 \\ & + 0.163 \cdot StDev_1^2 + 0.0111 \cdot Min_1^2 - 0.0593 \cdot Mean_1 \cdot StDev_1 \\ & - 0.0198 \cdot Mean_1 \cdot Min_1 + 0.0761 \cdot StDev_1 \cdot Min_1 \end{aligned} \quad (1)$$

$$\begin{aligned} DSup_2 = {} & -356 + 2.97 \cdot Mean_2 + 17.1 \cdot StDev_2 + 9.65 \cdot Min_2 - 0.00303 \cdot Mean_2^2 \\ & - 0.220 \cdot StDev_2^2 - 0.202 \cdot Min_2^2 - 0.0719 \cdot Mean_2 \cdot StDev_2. \end{aligned} \quad (2)$$

2.2.2 Algorithm for the extraction of the ground and polished surface damage.

(1) Set the original image as Orig.
(2) Set Mean as the average grey value of Orig.
(3) Set StDev as the standard deviation of the grey values of Orig.
(4) Set Max and Min as the maximum and minimum grey values in Orig.
(5) Calculate DSlow as the lower density slice level of Orig from regression equation (3).
(6) Apply density slicing with levels (DSlow, Max) to Orig to get Bin.
(7) Apply Erode and Dilate binary operators respectively to Bin.

$$DSlow = 77.0 + 0.789 \cdot Mean + 1.57 \cdot StDev - 0.591 \cdot Min. \quad (3)$$

2.2.3 Results of the image-processing data mining. The results of the two image-processing algorithms using the microstructure images in figure 1(a) and (b) are shown in figure 3(a) and (b), respectively.

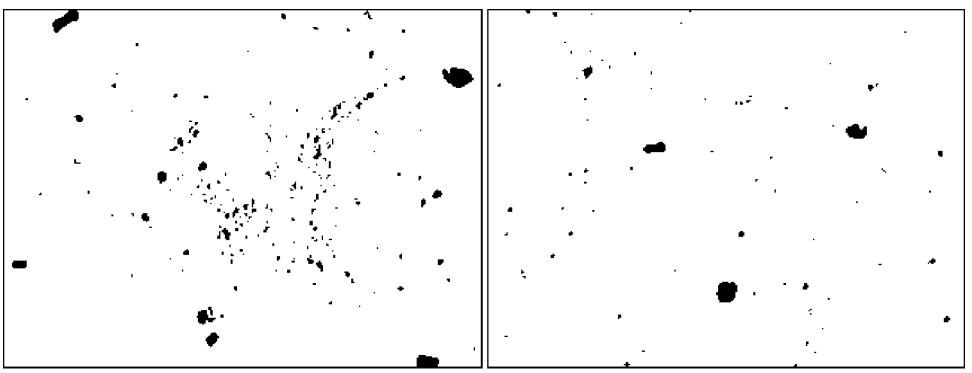

Figure 3. (a) Image in figure 1(a) after processing. (b) Image in figure 1(b) after stage one data mining.

3. Stage 2: fitting an extreme value distribution

Once the images have been mined for flaws or surface damages, there remains the task to estimate the largest such imperfections that might occur throughout the material. Because each image is a tiny part of the material manufactured and because the largest imperfections are those which reduce fracture strength, an extreme value probability distribution was chosen for the second stage of the data mining. Size distributions in nature generally show a power law relationship, where small occurrences are common, but large occurrences are rare. The Pareto law with a shape parameter γ, a scale parameter s, and a location parameter ν can be stated as:

$$1 - F_x(x) = P(X > x) = \left(\frac{x - \nu}{s}\right)^{-1/\gamma}, \quad \gamma > 0,\ x \geq \nu. \tag{4}$$

A flexible and widely applied limit distribution is the Generalized Pareto Distribution (GPD), which is a generalized form of equation (4) (Pickands III 1975). Dargahi-Noubary (1989) recommends GPD for use as the distribution of the excess of observed values over a certain threshold, as the threshold increases toward the right-hand tail. Gilli and Kellezi (2005) provide detailed information on the GPD, from which equations (5)–(9) can be derived.

$$F_u(y) \sim G_{\xi,\sigma}(y),\ u \to \infty$$

$$P(Y \leq y | Y > 0) = G_{\xi,\sigma}(y) = \begin{cases} 1 - \left(1 + \dfrac{\xi}{\sigma} y\right)^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - e^{-y/\sigma} & \text{if } \xi = 0, \end{cases} \tag{5}$$

for $0 \leq y \leq x_{\max} - u$, is the Generalized Pareto Distribution (GPD). If x is defined as $x = u + y$, the GPD can be expressed as a function of x as:

$$P(X < x | X > u) = G_{\xi,\sigma}(x) = 1 - \left(1 + \xi \frac{x - u}{\sigma}\right)^{-1/\xi}. \tag{6}$$

The cases where $\xi > 0$, $\xi = 0$, and $\xi < 0$ correspond to Frechet, Gumbel, and reverse Weibull distributions, respectively. If there is any evidence that there is a certain upper bound for X, then $\xi < 0$ is used, else $\xi > 0$ is used. Equations (5) and (6) are the conditional cumulative distribution of excess $Y = X - u$ given $Y > 0$ or the cumulative distribution of X given $X > u$, respectively, for a sufficiently large threshold u. From the conditional probability,

$$F_u(x) = \frac{F(x) - F(u)}{1 - F(u)}. \tag{7}$$

Then, solving for $F(x)$ yields

$$\begin{aligned} F(x) &= P(X \leq x) = F_u(x) \cdot (1 - F(u)) + F(u) \\ &= \left(1 - \left(1 + \xi \frac{x - u}{\sigma}\right)^{-1/\xi}\right) \cdot \frac{N_u}{n} + \left(1 - \frac{N_u}{n}\right), \end{aligned} \tag{8}$$

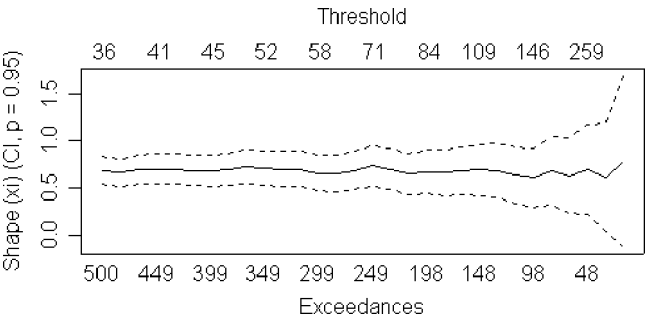

Figure 4. Behaviour of the shape parameter (ξ) with different threshold values.

which can be further simplified as

$$F(x) = 1 - \frac{N_u}{n}\left(1 + \xi\frac{x-u}{\sigma}\right)^{-1/\xi}, \tag{9}$$

where, N_u is the number of data points larger than the threshold u, and n is the size of the sample.

To fit a GPD to data, a specific threshold value, u, is needed. One approach is to plot the fitted GPD parameters, along with their confidence intervals, for different threshold values and locate the largest threshold value after which the fitted GPD parameters start to vary or deviate significantly. This is shown in figure 4 for the pressed sample flaws. A threshold value of 90 was found to be suitable, and similar analyses were done for the other production methods.

The estimation of the GPD parameters is done by maximum likelihood. The likelihood function is given by equation (10)

$$L = \prod_{i=1}^{N_u}\frac{1}{\sigma}\left[1 + \frac{\xi(x_i - u)}{\sigma}\right]^{(-1/\xi)-1}. \tag{10}$$

Then, $\log(L)$ can be written as

$$\log(L) = -N_u\log(\sigma) - \left(1 + \frac{1}{\xi}\right)\sum_{i=1}^{N_u}\log\left(1 + \frac{\xi(x_i - u)}{\sigma}\right). \tag{11}$$

The values of σ and ξ that maximize the above function are the estimates of the GPD parameters. The statistical software package, R-project, (available at http://www.r-project.org) is used for all distribution fitting and confidence interval calculations, with the special toolbox called 'evir'. The evir toolbox estimates the values of ξ and σ.

4. Results

Each pixel^2 area in the digital microstructure images corresponds to 1.10803 micron^2 of flaw or damage area. 0.1 and 0.5 h sintered samples are used for the flaw

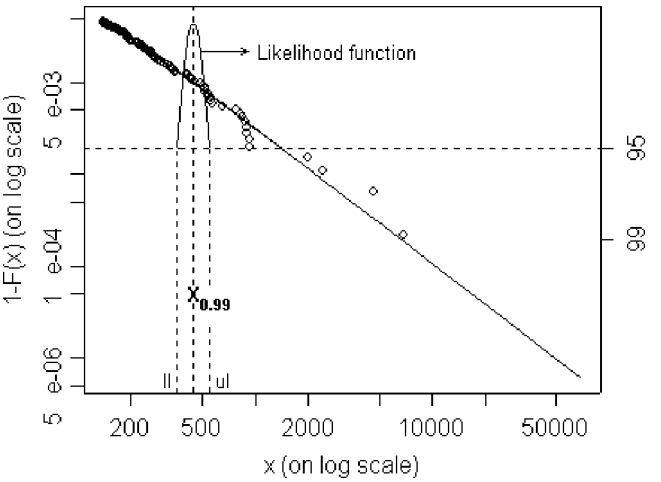

Figure 5. 99th percentile, $X_{0.99}$, and 95% CI for the 99th percentile, [*ll*, *ul*].

analyses, and 5.0 and 10.0 h sintered samples are used for damage analyses. Stage 1 data mining is used, and the extracted areas greater than five pixel^2 are recorded as the flaw or damaged areas. The reason for this size threshold is to eliminate any noise still present in the images. Stage 2 data mining, fitting the extreme value distributions, then took place, followed by the calculation of 40 different percentiles ranging from 0.960 to 0.999, along with their 90% CIs. (Since the confidence intervals are fairly wide, 90% was used for comparison.) When the percentiles for the samples with different preparation methods are plotted on the same graph, differences in the tail distributions are readily perceived. While pairwise comparisons are shown, simultaneous considerations of the three methods might also have been considered using a multiple comparison procedure such as Tukey's or Bonferroni's comparison. Furthermore, observed differences that are of practical importance may not always be shown as statistically significant.

A representative graph of pressed sample dense heterogeneities is given in figure 5. In figure 5, the curve used to calculate the 95% CI of the 99th percentile, $X_{0.99}$, is superimposed on the tail probability plot. The peak shows the maximum likelihood estimate, and *ll* and *ul* on the *x*-axis show the lower and the upper limits of the 95% CI. The distance between the peak and the horizontal dashed line, in terms of the maximum likelihood function value, is $\chi^2_{1-\alpha,1}/2 = 1.92$ when the Chi-square has one degree of freedom and $\alpha = 0.05$. Additional information on calculations for an extreme value distribution can be found in Shi *et al.* (1999) and Gilli and Kellezi (2005).

Figure 6 shows the comparisons of the percentile plots of the pressed vs. flocculated slip cast, pressed vs. dispersed slip cast, and the flocculated slip cast vs. dispersed slip cast flaws, respectively. From figure 6, 99.5% of the flaws using the pressed production method are expected to be smaller than 730 pixel^2, whereas 99.5% of the flaws using the flocculated slip cast production methods are expected to be smaller than 285 pixel^2. In other words, pressing tends to cause much larger microstructural flaws than either slip cast production, whereas there is no statistical difference in the largest flaws for the slip cast methods.

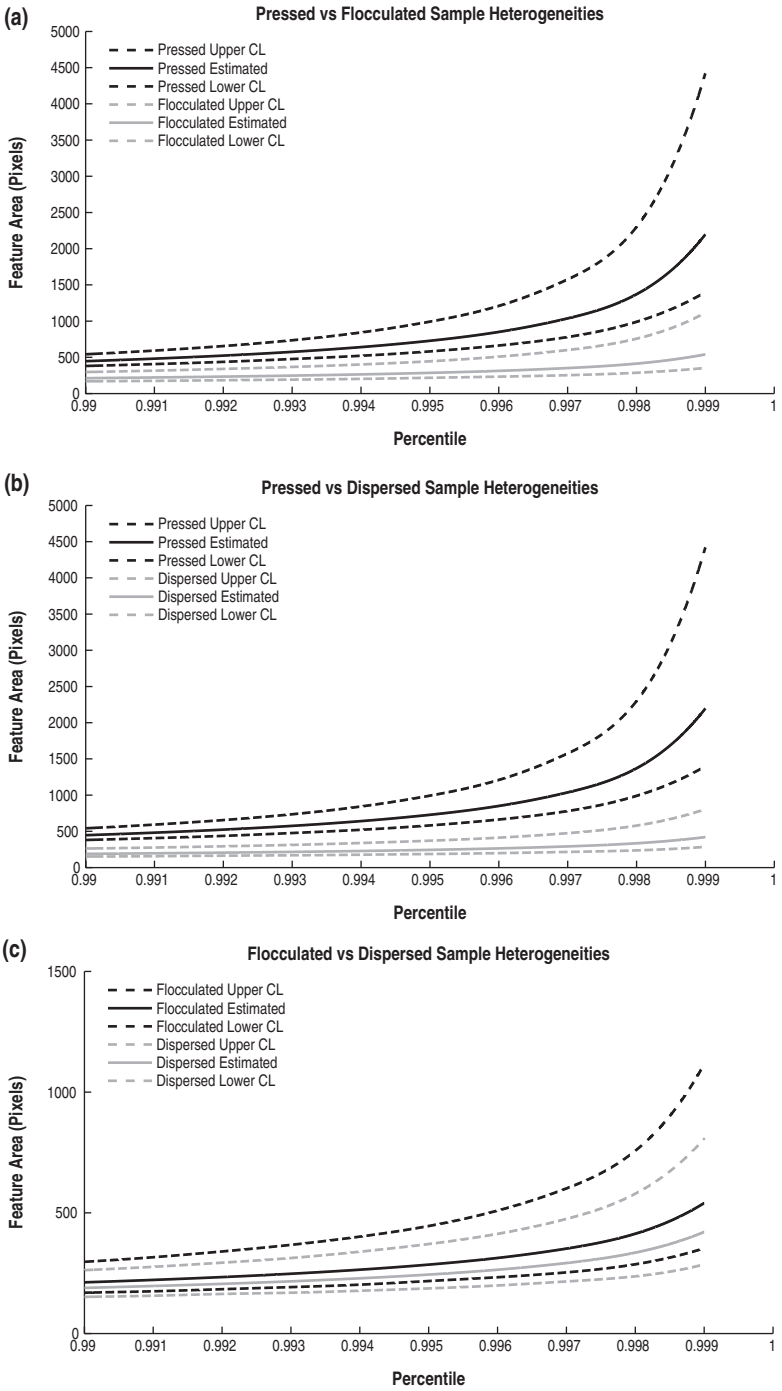

Figure 6. GPD percentile plots for flaw populations (estimate and 90% upper and lower limits).

Similarly, figure 7 shows the comparisons of the percentile plots of the pressed vs. flocculated slip cast, pressed vs. dispersed slip cast, and flocculated slip cast vs. dispersed slip cast sample damages, respectively. The pressed–dispersed, and the flocculated–dispersed pairs have significant differences, but the pressed–flocculated

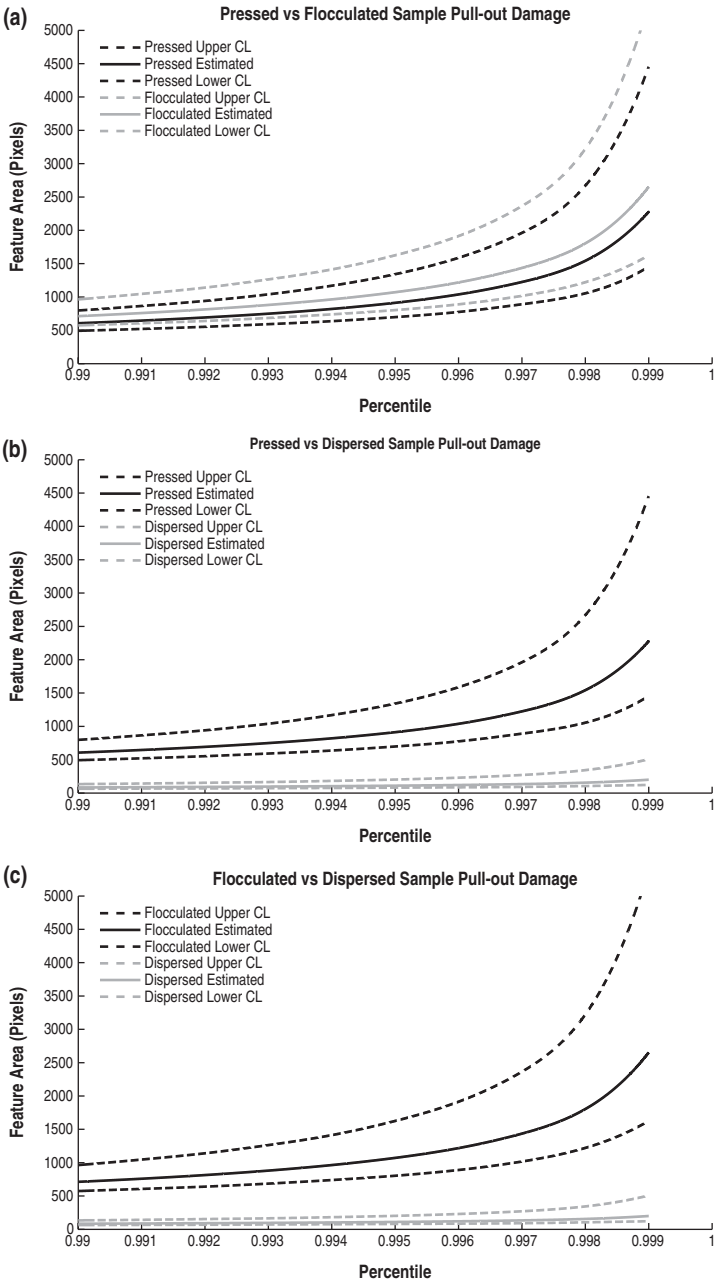

Figure 7. GPD percentile plots for damage populations (estimate and 90% upper and lower limits).

pair does not show any significant difference when the expected damage sizes at different probability values are compared. The dispersed slip cast production method shows significantly smaller damaged areas in the tail distribution than either pressing or flocculation slip cast (these two are statistically identical). Therefore, using this two-stage data-mining approach, the production method which tends to produce smaller flaws and surface damage areas is dispersed slip casting. Further, the results show the distribution of sizes of flaws and damaged areas for each production method based on sampling of images.

5. Conclusions

Even though large ceramic flaws are infrequent, such flaws can cause structural failures in advanced ceramics. A two-stage data-mining approach to accurately estimate flaw distribution at the size extreme has been developed and tested on three different processing methods. The first stage of data mining is image processing—identifying and recovering large flaws from noisy microscopic images automatically. This is done by a sequence of standard algorithms combined with adaptive density slicing. The second is the use of the flaw information to fit an extreme value probability distribution. This allows manufacturers and designers to estimate the number and distribution of large flaws for a given processing method using a sample of images taken from the material during the sintering process.

The results herein would indicate that dispersed slip casting would be preferred as the production method, since it tends to produce both smaller large microstructural flaws during sintering and smaller large damaged areas upon surface grinding and polishing. These indicate a ceramic with better strength properties. This two-stage approach could be used, with modifications, for other materials and manufacturing methods where the salient information comes from images and occurs at a physical extreme (e.g. length, volume, perimeter).

Acknowledgement

This project was supported by US National Science Foundation grant DMI-0301273.

References

Anderson, C.W., Shi, G., Atkinson, H.V. and Sellars, C.M., The precision of methods using the statistics of extremes for the estimation of the maximum size of inclusions in clean steels. *Acta Mater.*, 2000, **48**, 4235–4246.

Anderson, C.W., Shi, G., Atkinson, H.V., Sellars, C.M. and Yates, J.R., Interrelationship between statistical methods for estimating the size of the maximum inclusion in clean steels. *Acta Mater.*, 2003, **51**, 2331–2343.

Caers, J., Beirlant, J. and Maes, M.A., Statistics for modeling heavy tailed distributions in geology: Part I methodology. *Math. Geol.*, 1999, **31**, 391–410.

Chen, N., Li, C. and Qin, P., KDPAG expert system applied to materials design and manufacture. *Eng. Applic. Artif. Intell.*, 1998, **11**, 669–674.

Chen, N., Zhu, D.D. and Wang, W., Intelligent materials processing by hyperspace data mining. *Eng. Applic. Artif. Intell.*, 2000, **13**, 527–532.

Dargahi-Noubary, G.R., On tail estimation: an improved method. *Math. Geol.*, 1989, **21**, 829–842.

Dierickx, D., Basu, B., Vleugels, J. and Van der Biest, O., Statistical extreme value modeling of particle size distributions: experimental grain size distribution type estimation and parameterization of sintered zirconia. *Mater. Character.*, 2000, **45**, 61–70.

Evans, A.G., Aspects of the reliability of ceramics. in *Materials Research Society Symposia Proceedings*, 1984, **24**, 63–80.

Evans, J.R.G., Edirisinghe, M.J., Coveney, P.V. and Eames, J., Combinatorial searches of inorganic materials using the ink-jet printer: science, philosophy and technology. *J. Eur. Ceram. Soc.*, 2001, **21**, 2291–2299.

Gilli, M., Kellezi E., An application of extreme value theory for measuring financial risk. Research paper. Preprint submitted to *Elsevier Science*, 2 May 2005, Available online at: http://www.unige.ch/ses/metri/gilli/evtrm/GilliKelleziEVT.pdf (accessed 14 July 2005).

Heckert, N.A., Simiu, E. and Whalen, T., Estimates of hurricane wind speeds by 'peaks over threshold' method. *J. Struct. Eng.*, 1998, **124**, 445–449.

Hofmann, D.W.M. and Apostolakis, J., Crystal structure prediction by data mining. *J. Molec. Struct.*, 2003, **647**, 17–39.

Krell, A., Fracture origin and strength in advanced presureless-sintered alumina. *J. Am. Ceram. Soc.*, 1998, **81**, 1900–1906.

Pickands, J. III, Statistical inference using extreme order statistics. *Ann. Stat.*, 1975, **3**, 119–131.

Reeber, R.R., Surface engineering of structural ceramics. *J. Am. Ceram. Soc.*, 1993, **76**, 261–268.

Quinn, G.D. and Morrell, R., Design data for engineering ceramics: a review of fracture test. *J. Am. Ceram. Soc.*, 1991, **74**, 2037–2066.

Shi, G., Atkinson, H.V., Sellars, C.M. and Anderson, C.W., Application of the generalized Pareto distribution to the estimation of the size of the maximum inclusion in clean steels. *Acta Mater.*, 1999, **47**, 1455–1468.

Shi, G., Atkinson, H.V., Sellars, C.M., Anderson, C.W. and Yates, J.R., Computer simulation of the estimation of the maximum inclusion size in clean steels by the generalized Pareto distribution method. *Acta Mater.*, 2001, **49**, 1813–1820.

Simui, E. and Heckert, N.A., Extreme wind distribution tails: a 'peaks over threshold' approach. *National Institute of Standards and Technology Building Science Series 174*, *Coden: NBSSES*, March 1995.