Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Microbiological Methods 53 (2003) 97–106
Journal of Microbiological Methods
www.elsevier.com/locate/jmicmeth

Assessing technician effects when extracting quantities from microscope images

D. Webb[a,*], M.A. Hamilton[b], G.J. Harkin[b], S. Lawrence[b], A.K. Camper[b], Z. Lewandowski[b]

[a] *Department of Mathematics, Bemidji State University, HS-341, Box 23, 1500 Birchmont Dr NE, Bemidji, MN 56601-2699, USA*
[b] *The Center for Biofilm Engineering, Montana State University, Bozeman, MT 59717-3980, USA*

Received 18 September 2002; received in revised form 10 October 2002; accepted 17 October 2002

Abstract

Consider an experiment where the response is based on an image; e.g., an image captured to a computer file by a digital camera mounted on a microscope. Suppose relevant quantitative measures are extracted from the images so that results can be analyzed by conventional statistical methods. The steps involved in extracting the measures may require that the technicians, who are processing the images, perform some subjective manipulations. In this case, it is important to determine the bias and variability, if any, attributable to the technicians' decisions. This paper describes the experimental design and statistical analyses that are useful for those determinations. The design and analysis are illustrated by application to two biofilm research projects that involved quantitative image analysis. In one investigation, the technician was required to choose a threshold level, then the image analysis program automatically extracted relevant measures from the resulting black and white image. In the other investigation, the technician was required to choose fiducial points in each of two images collected on different microscopes; then the image analysis program registered the images by stretching, rotating, and overlaying them, so that their quantitative features could be correlated. These investigations elucidated the effects of the technicians' decisions, thereby helping us to assess properly the statistical uncertainties in the conclusions for the primary experiments.

Keywords: Bias; Image analysis; Microscope; Technician variability; Uncertainty

1. Introduction

The development of quantitative methods for image analysis has been an active area of research in microbiology. This is especially true in the study of biofilms where microscopes and surface science instruments are required to view both biofilm bacteria and the surface to which they attach (Condron et al., 1999; Heydorn et al., 2000; Hope et al., 2002; Xavier et al., 2001). The resulting images may be digitized so they can be viewed on a computer monitor and saved in a computer file for subsequent analysis. To compare and correlate images, it is important to find valid methods for converting the images into numerical quantities. It is not unusual for the quantitative image

* Corresponding author. Tel.: +1-218-755-2846; fax: +1-218-755-2822.
E-mail address: dwebb@bemidjistate.edu (D. Webb).

0167-7012/02/$ - see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0167-7012(02)00228-2

analysis protocol to require some subjective input by the technician. When this is the case, it is important to assess the extent to which the technicians' decisions can influence the end results of the analysis. If the technicians' subjective decisions greatly influence the conclusions, those conclusions may well not be convincing. The direct approach for assessing technicians' influence is a special set of experiments involving multiple technicians, each following the same image analysis protocol. A mixed model analysis of variance can be conducted to estimate the intra- and inter-technician components of variability. The experiments may also uncover bias caused by the image analysis protocol.

The purpose of this paper is to describe an experimental design and mixed model analysis of variance for elucidating the within- and between-technician contributions to statistical variability in quantitative image analysis. The approach is illustrated by two different examples. One example pertains to choosing a cut point on a gray-scale for purposes of thresholding (converting each pixel to either pure black or pure white), often a preliminary step when extracting numerical quantities from an image. The other example pertains to the registration (overlaying) of two different images for purposes of correlation analysis where the quantities in one image are correlated to quantities at the same spatial positions in the other image. In this example, the investigation showed that the registration protocol could lead to important bias. The methods of this paper are of quite general applicability and are not limited just to the examples included here. The design and analysis strategy could be applied to assess most microbiology protocols on which technician effects are potentially influential.

2. Materials and methods

2.1. Experimental design and statistical analysis

Two experiments were designed to quantify the variability and/or bias associated with subjective manipulations performed by technicians on digital images of biofilm systems. Both experiments emphasize randomization, replication, and blind assessment.

Replication of technicians' manipulations allowed assessment of the variability associated with the manipulations. Both experiments utilize replication at multiple levels. Technicians perform the same manipulation multiple times in order for the variability associated with each individual technician (intra-technician variability) to be quantified. Multiple technicians are also part of each experiment. This enables the quantifying of technician-to-technician (inter-technician) variability.

Blind assessment in the context of these experiments means that the technicians who perform the experiment do not know that they are manipulating the same set of digital images multiple times. Blind assessment reduces potential bias in replication that may result if a technician recalls a particular digital image from a previous experimental run and manipulates the image in a fashion similar to how they first manipulated the image.

2.2. Experiment A: image thresholding and areal porosity

A mixed species biofilm was grown in an open channel reactor using methods and protocols described elsewhere (Lewandowski et al., 1999). In brief, the open channel, flat plate reactor was made of polycarbonate, 4 cm deep, 4 cm wide, and 75 cm long with a total working volume of 420 ml. The lid of the reactor was sealed with silicon rubber to prevent contamination and the entire system was sterilized with 70% alcohol before each experiment. The reactor and components were rinsed with sterile water (autoclaved) until all the alcohol was removed. The reactor components (tubings, connectors, air filter, etc.) and growth medium were autoclaved at 1218 °C prior to experiments.

The nutrient solution was made of monobasic potassium phosphate KH_2PO_4 (0.69 mM), dibasic potassium phosphate K_2HPO_4 (1.5 mM), ammonium sulfate $(NH_4)SO_4$ (0.079 mM), magnesium sulfate $MgSO_4 \cdot 7H_2O$ (0.013mM), glucose (0.040 g/l), and yeast extract (0.031 g/l). The inoculum, 1 ml of frozen stock culture (from the Center for Biofilm Engineering), consisted of wild type strains of the three species: *Pseudomonas aeruginosa*, *Pseudomonas fluorescens*, and *Klebsiella pnenmoniae.* The reactor was operated as a batch for 12 h

and then switched to a continuous flow (flow velocity of 1.6 cm/s) for 4 to 5 days. The hydraulic retention time was less than 4 min to prevent growth of planktonic bacteria. The nutrient solution was aerated in the mixing chamber. Flow velocity was maintained by recycling the growth medium using peristaltic pumps (Cole-Palmer, Chicago, IL).

At the chosen sampling time, five images were captured through the bottom of the reactor on a an inverted confocal scanning laser microscope (CSLM; Bio-Rad, model MRC 600 CSLM associated with an Olympus BH-2 microscope), three using a 10 × objective and two using a 20 × objective, giving total magnifications of 100 × and 200 ×, respectively. All images were captured at a depth of approximately 30 μm from the surface of the biofilm at randomly chosen locations. The locations were randomly chosen by defining one corner of the flat plate reactor to be the origin. Random numbers between 0 and 75 cm and between 0 and 4 cm were then chosen. These random numbers established the (x,y) locations at which the images were captured. The procedure was repeated four more times. The images were electronically stored as TIFF files, 512 × 768 pixels in size, where each pixel takes on an eight bit gray-scale value from 0 to 255.

The images show light-shaded biofilm cell clusters against a background of dark-shaded void or bulk water areas. The investigators planned to convert the images into quantitative measures that describe the morphology and spatial pattern of the biofilm cell clusters. Subsequent experiments would show the extent to which changes in the experimental parameters (e.g., flow conditions, substrate concentration, and bacteria species) cause alterations in the quantities. We focused on a single parameter in this study, areal porosity, which is the ratio of void area to total area in an image. Areal porosity ranges in value from 0 to 1, with higher values representing more pores and interstitial space within the image (Lewandowski et al., 1999). In order to calculate the areal porosity, each pixel in the image was classified as either a cell cluster (white) or a void area (black) and areal porosity was the ratio of black pixels to total pixels. The initial gray scale image was converted into a black and white image using a process known as thresholding. For the eight bit gray scale images collected in this study, the level of gray was indicated by an integer between 0 and 255. To threshold the image, one chooses a threshold number between 0 and 255. All pixels having gray scale values less than the threshold level are defined to be biofilm material (pixel value set to 0) and all pixels having gray scale values greater than or equal to the threshold level are defined to be void or non-biofilm material (pixel value set to 255). The image analysis software (MARK) included a useful tool for choosing the threshold value. The technician moved a pointer to a position on the 0 to 255 scale and simultaneously viewed the original gray scale image and the thresholded image corresponding to the threshold value at the pointer. Because the areal density depended on thresholding and the thresholding included a key subjective step, it was important to conduct a special experiment designed to determine the influence of that subjective step.

Six technicians were chosen to participate in the experiment. All technicians had knowledge of biofilm appearance and structure and were accustomed to working with digital images captured by a CSLM. Each technician observed the same set of five images a total of three times and chose a threshold level for each image. Therefore, a total of 18 threshold levels were obtained for each image.

Before being presented to the technicians, the five images were electronically duplicated two times. The first duplicate was flipped along its vertical axis and the second duplicate was flipped along its horizontal axis. All 15 images were then presented to each technician one image at a time in a randomized order. The changed orientation adequately disguised the images and the technicians were unaware that they were choosing three independent threshold levels for each image. This assured blind assessment of the images by the technicians. The recorded data were the threshold levels and associate areal porosity for each of the 18 images for each technician.

2.3. Experiment B: image registration

Experiment B was part of a project conducted to correlate the positions of individual colonizing bacteria at the initial stage of biofilm formation with

the topographical features on the surface of a 316L stainless steel coupon embedded in a flow cell (Hamilton et al., 1995). The flow cell consisted of two parallel flat plates. The top plate allowed for a large glass coverslip that served as an observation window, and the bottom contained a recessed well for the coupon. Overall chamber dimensions were 40 mm long by 12 mm wide by 0.01 mm thick. Flow conditions were maintained at a Reynold's number of 6 and a shear stress at the coupon surface of 0.75 N m^{-2}. The entire flow cell was autoclaved before inserting the coupon, which had been sterilized using ultraviolet light prior to inoculation.

The test organism was a *P. aeruginosa* originally isolated from an industrial biofilm. It was grown in a chemostat with a residence time of 5 h. Sterile influent feed contained (per liter of distilled water) 0.1 g glucose, 0.7 g dibasic potassium phosphate (K_2HPO_4), 0.3 g monobasic potassium phosphate (KH_2PO_4), 0.1 g ammonium sulfate ($(NH_4)_2SO_4$), and 0.01 magnesium sulfate ($MgSO_4 \cdot 7H_2O$). Chemostat effluent, diluted with the same medium without glucose to provide 108 cells per milliliter, was pumped through the flow cell.

The adsorption of individual cells to the coupon surface within the field of view was observed nondestructively and in real time by reflected light imaging of 1000× magnification, using a BioRad MRC 600 CSLM associated with an Olympus BH-2 microscope. The topographic features were mapped using an atomic force microscope (AFM; Digital Instruments Nanoscope II AFM-STM; performed by Imaging Services, Santa Barbara, CA) and the results were saved as a computer image for which an eight bit gray scale indicated the altitude at each pixel location. The AFM image was saved in a 256 × 256 pixel array and the CSLM was saved in a 512 × 768 pixel array. On the computer monitor, the AFM image appeared smaller even though it showed the same field of view as the CSLM image (Fig. 1). The striations that are evident in the CSLM image are only a few nanometers deep.

Using a laser etching technique, four tiny, shallow holes were etched at the center of the coupon to mark the corners of a square field of view, 50 μm on a side. Both a CSLM image and an AFM image were taken of that field of view. The holes were the fiducial points (or landmarks) that the registration software used to register the AFM image to the CSLM image. Special software was written in-house (MARK) to accomplish the expansion, translation, and rotation necessary to register two different digital images of the same field of view.

At the CSLM magnification used to capture the whole field of view, there were 60–100 pixels per hole and the holes were not round or regular in shape. Image analysis software could not successfully recognize the fiducial points; therefore, the locations of fiducial points on each image had to be selected by the technician. To do this, the technician viewed the image on a monitor and used the computer mouse to move a cross-hair cursor to the chosen center of a fiducial point. When the mouse was clicked, the computer recorded the pixel location. It was apparent that a single technician did not necessarily choose the same center pixel for a given fiducial point if the process was repeated. Additionally, two technicians would make somewhat different choices. To assess the extent of variation in selecting each fiducial point's location, and the effect of that variation on the registration process, Experiment B was conducted.

Two laser-marked 316L stainless steel coupons were chosen. One coupon had an electropolished surface; it was called the "rough" coupon. The other coupon was electropolished and further hand-polished with 0.3-μm alumina grit to a mirror finish; it was called the "smooth" coupon. For this investigation, the coupons were sterile because registration variability and bias depend on marking the locations of fiducial points, a process independent of the presence or absence of bacteria.

Seven technicians were chosen to participate in the experiment; all technicians were experienced at choosing the centers of fiducial points and running the registration software (MARK). Prior to starting the experiment, the investigator altered each of the CSLM images by electronically selecting 13 "benchmarks," each having a regular shape and known center pixels. The benchmarks, which were arranged in a pattern that covered the field of view, were essentially transparent to the technicians and had no effect on their task of selecting the centers of the fiducial points. Some critical results depended on tracking the benchmark

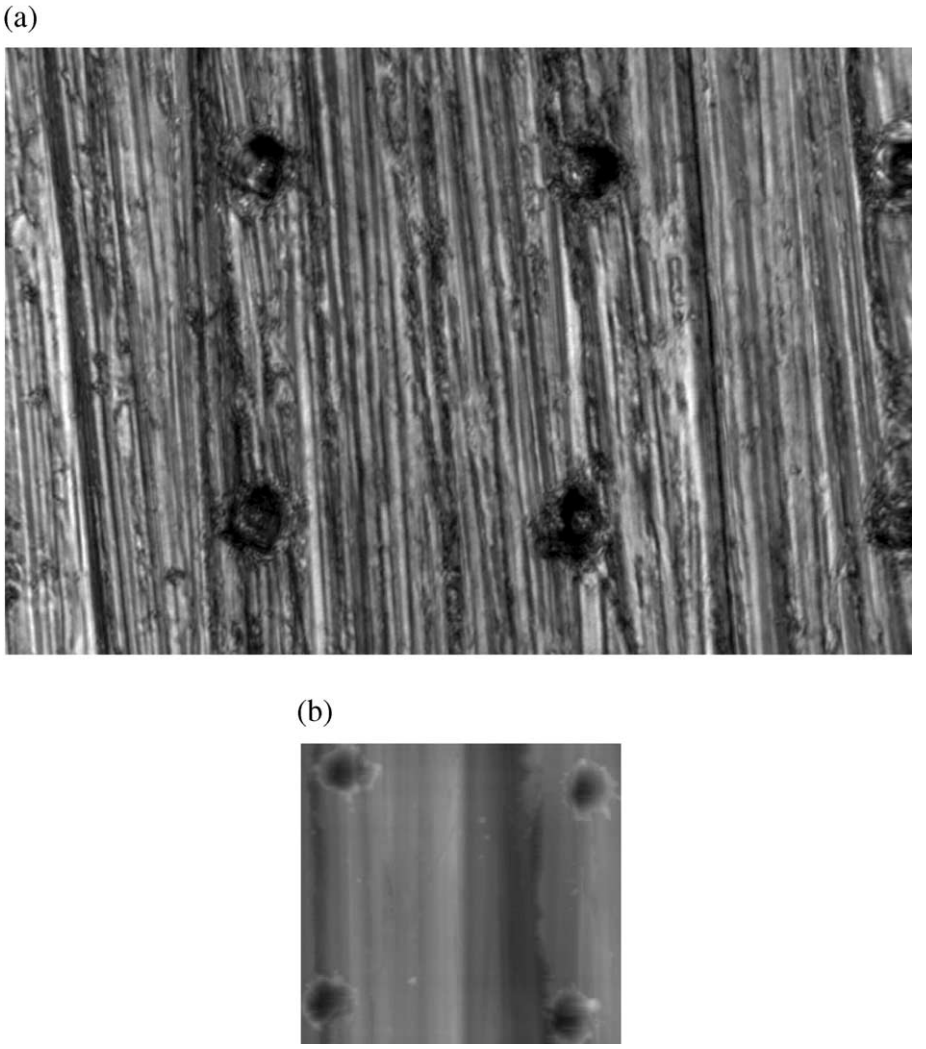

Fig. 1. Microscope images of the same area on the rough stainless steel coupon. (a) is the CSLM image and (b) is the AFM image. To register the CSLM image to the AFM image, the CSLM image needs to be rotated clockwise a few degrees, then resized. The fiducial points are the laser-etched holes that delineate a square field of interest approximately 50 μm per side.

positions as they were shifted by the registration process.

Each run of the experiment involved a technician following this two-step protocol (see Fig. 2):

1. *Registration step*—Mark each of the four fiducial points on the CSLM image and the corresponding four fiducial points on the associated AFM image. Direct the software to register the CSLM image (called CSLM1) to the AFM image. The result was a new CSLM image (called CSLM2) reduced to a 256×256 pixel array, rotated and resized to match the AFM image. The same mathematical transformation that accomplished the registration was applied to the 13 bench points. The centers of those benchmarks as they appear on CSLM2 were electronically recorded for analysis at the conclusion of the experiment.
2. *Reregistration step*—Mark each of the four fiducial points on CSLM2 and the four fiducial points on

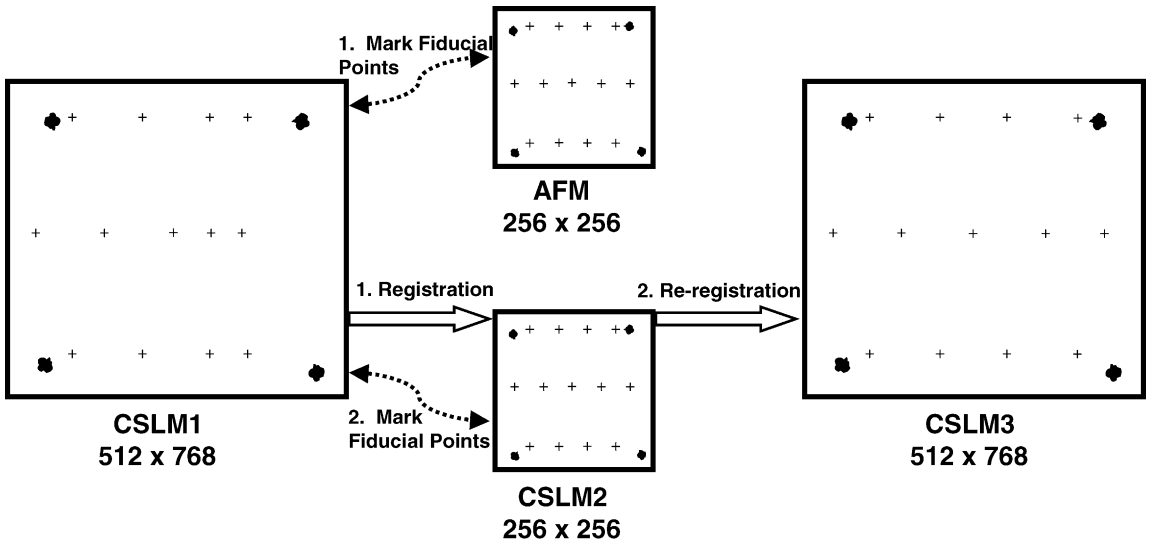

Fig. 2. Schematic showing the two-step registration–reregistration protocol in which a CSLM image is registered to a smaller AFM image's size and orientation, then reregistered back to the original size and orientation. Each square indicates an image. The four black, irregular circles in each square indicate the approximate positions of the fiducial points. The +'s indicate the approximate locations of the 13 benchmarks which were invisible to the technicians. The center-most benchmark is point 7. The fat arrow points from the initial image to the registered image. To accomplish a registration, the technician had to mark the locations of the four fiducial points in each of the two images connected by double arrows.

CSLM1. Direct the software to register CSLM2 to the original CSLM1. The resulting image, called CSLM3, will be CSLM2 after being rotated, resized, and converted to a 512×762 pixel array so it matches CSLM1. The centers of those benchmarks as they appear on CSLM3 were electronically recorded for analysis at the conclusion of the experiment.

Each of seven technicians repeated the protocol on three occasions for the same smooth and rough coupon images.

The analysis consisted of two components, one focusing on the variability in selecting individual fiducial point centers and the other on the shift of individual image features due to the registration–reregistration. The protocol required that the technician twice select the four fiducial point centers on the original CSLM image. For each experiment, the mean (x,y) pixel location for each fiducial point in the original CSLM image, averaged over all 42 selections (two per protocol for seven technicians and three trials), was subtracted from the individual center point selections to form x and y deviations from the mean. The deviations were converted from pixels to μm and converted to a Euclidean distance, D. The variabilities of the D responses were homogeneous across the four fiducial points and the D responses were submitted to analysis of variance using the two factor, mixed model described in the Statistical model and parameters section.

The second component of the analysis assessed the propagation of variability from the fiducial points to individual features of the image. If the registration–reregistration process was perfect, the centers of the 13 benchmarks in CSLM3 would be exactly at the same pixel locations as in CSLM1. By comparing the location of a benchmark in CSLM3 to the location in CSLM1, one can characterize the bias and variability attributable to the registration–reregistration steps. This part of the analysis focused on the difference of benchmark locations between CSLM3 and CSLM1.

For a specific benchmark point, let its center on CSLM1 be denoted by the pixel location $(x1, y1)$ and its center on CSLM3 be denoted by $(x3, y3)$. The

response variables are the differences $dx=(x3-x1)$ and $dy=(y3-y1)$, and the Euclidean distance $dE=[(x3-x1)^2+(y3-y1)^2]^{1/2}$, all quantities measured in μm. Analysis of the dE values was based on the mixed model analysis of variance presented in the Statistical model and parameters section. The dx and dy values were studied primarily to see if there was a consistent direction in which the points moved during the protocol.

2.4. Statistical model and parameters

For both experiments, the analysis was based the two factor, additive, mixed model (Searle et al., 1992) of Eq. (A1) where the meaning of the symbols depends on the experiment. The model is also known as a randomized block model where the technicians are the blocks.

$$Y_{ijk} = \mu + \beta_i + T_j + \varepsilon_{ijk}. \quad \text{(A1)}$$

For Experiment A, the two responses (threshold value and the areal porosity) were analyzed separately, and Y_{ijk} denotes the response for the kth time that image i was thresholded by technician j, and $i=1,\ldots,5$, $j=1,\ldots,6$, and $k=1, 2, 3$. The symbol μ denotes the overall mean across all five images, β_i denotes the fixed effect for image i.

For Experiment B, the analysis was performed separately for each image and each of the three responses, x, y, and d. The response Y_{ijk} denotes the benchmark error (x, y, or d) for point i the kth time that technician j performed the registration–reregistration protocol, where $j=1,\ldots,7$, $k=1, 2$. Here $i=1,\ldots,4$ for the responses (x,y) pertaining to the fiducial points, and $i=1,\ldots,13$ for the response d pertaining to the benchmark points. The symbol μ denotes the overall mean across all points, β_i denotes the fixed effect for point i.

For both experiments, T_j denotes the random effect due to technician j, and ε_{ijk} denotes the random deviation from the mean for technician j and image i. Let σ_T^2 denote the variance of T_j (the inter-technician variance) and let σ^2 denote the variance of ε_{ijk} (the intra-technician variance). The total variance for a randomly chosen response is $\sigma_Y^2=\sigma_T^2+\sigma^2$; that is, the total variance is the sum of two components, the inter- and intra-technician variances. The associated total standard deviation, which is called the repeatability standard deviation in the main body of the paper, is $\sigma_Y=[\sigma_T^2+\sigma^2]^{1/2}$. Because the experimental designs were balanced, the variance components σ_T^2 and σ^2 were estimated using the ANOVA method (Searle et al., 1992). For both experiments and all responses, preliminary statistical analyses showed that the additive model was appropriate.

For experiment A, the coefficient of variation (CV) was also calculated for each of the five images for both responses, threshold value and areal porosity. The coefficient of variation is a dimensionless quantity conventionally used to compare variability across different measurement scales; it is the repeatability standard deviation divided by the mean, expressed as a percentage.

3. Results

3.1. Experiment A: image thresholding

The results for both the threshold value and areal porosity are summarized in Table 1. The estimated standard deviation and inter-technician and intra-tech-

Table 1
Results for Experiment A, showing the standard deviation, which is the square root of the total variance, and the components that make up the total variance for the threshold values and the areal porosity

Image	Repeatability standard deviation	Coefficient of variation (%)	Variance components (%)	
			Inter-technician	Intra-technician
Threshold values				
1	5.4	6	55	45
2	7.7	15	84	16
3	8.4	19	64	36
4	3.4	5	77	23
5	9.7	19	20	80
Areal porosity				
1	0.11	19	71	29
2	0.08	12	87	13
3	0.03	6	62	38
4	0.08	17	74	26
5	0.02	4	17	83

nician components of variance are reported. The components of variance for threshold values from images 2, 3, and 4 were statistically significantly different from zero (all p-values < 0.01). The inter-technician variance from image 1 was marginally significant (p-value $= 0.013$) and from image 5 was non-significant (p-value $= 0.198$). The components of variance for porosity from images 1 through 4 were statistically significantly different from zero (all p-values < 0.01). The inter-technician variance from image 5 was non-significant (p-value $= 0.228$). The inter-technician and the intra-technician variance components differ greatly from image to image. For images 1 through 4 the inter-technician variance component is larger than the intra-technician variance component. But, for image 5 the opposite is true. This holds for both threshold and areal porosity. Image 5 had a statistically non-significant inter-technician variance and was the only image for which the intra-technician variance was larger than the inter-technician variance.

Areal porosity can only be calculated after a threshold value has been chosen for an image. An image with a high amount of variability associated with the threshold choice does not necessarily have a high amount of variability associated with the areal porosity calculation. For example, the CV for threshold for image 3 was 19% while the CV for areal porosity for image 3 was 6%. The opposite is also true, an image with a low amount of variability associated with the threshold choice may have a high amount of variability associated with the areal porosity calculation. Image 1 had a CV for threshold of 6% while image 1 had a CV for areal porosity of 19%.

3.2. Experiment B: image registration

The estimated standard deviation and inter-technician and intra-technician components of variance for the fiducial point response D are shown in Table 2. The components of variance were statistically significantly different from zero (all p-values < 0.01). The ranges of D responses were 0.0 to 1.4 μm for the smooth coupon and 0.1 to 3.4 μm for the rough coupon.

Fig. 3 shows the positions of the CSLM3 benchmark centers relative to the original (CSLM1) centers for benchmark number 7 which was at the approximate middle of the images. If the registration–reregistration had returned the benchmark to its original pixel location, the (dx,dy) point would be (0, 0). The distance of the observed (dx, dy) from (0, 0) indicates how much the location of a feature, such as a peak or valley that is located near the benchmark position, would be artificially shifted by the registration process.

Table 2
Results for Experiment B, showing the standard deviation, which is the square root of the total variance, and the components that make up the total variance for the difference D for fiducial points and the shift dE for benchmark points

Coupon	Repeatability standard deviation (μm)	Variance components (%)	
		Inter-technician	Intra-technician
D for fiducial points			
Smooth	0.3	28	72
Rough	0.5	36	64
dE for benchmark points			
Smooth	0.7	27	73
Rough	0.6	37	63

The 21 points plotted for each coupon correspond to the three repeated runs by each of the seven participants. The plots show a bias; that is, a consistent movement away from (0,0) in the positive dx, negative dy direction. For benchmark point 7, the estimated mean (x,y) differences were $(1.9, -2.5)$ for the smooth coupon and $(0.8, -1.1)$ for the rough coupon. The plots for the other benchmark points were nearly identical to Fig. 3, indicating that all features in the image were artificially moved about the same direction and distance.

The variability of benchmark shifts as measured by the Euclidean response dE was homogeneous across benchmark points and it is appropriate to pool variance estimates across benchmark points. The estimated standard deviation of a single, randomly chosen dE response, and the components of variance are presented in Table 2. The components of variance were statistically significantly different from zero (all p-values < 0.001). The mean dE response for benchmark point 7 was 3.1 μm for the smooth coupon and 1.7 μm for the rough

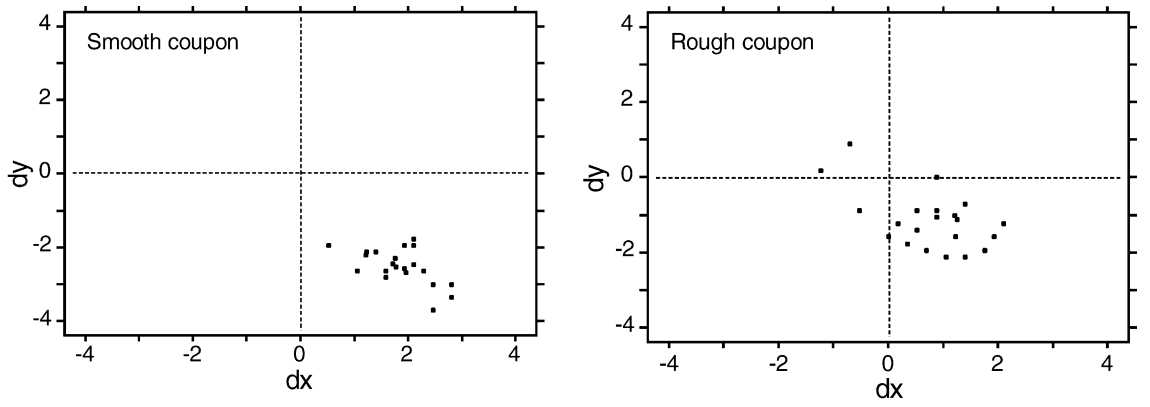

Fig. 3. Shift (μm) in benchmark point 7 on each coupon due to the register–reregister protocol. The 21 points pertain to the three trials by each of seven technicians.

coupon. The point 7 results were typical of results for the other 12 benchmark points.

4. Discussion

Experiment A comprised the steps of capturing an image, creating digital files, subjectively thresholding the image, then calculating the parameters of interest. As with any measurement process, the potential sources of measurement uncertainty are random variability or chance error and bias or systematic error (Wheeler and Lyday, 1989). Experiment A provided an estimate of the random error associated with thresholding but it was not designed to give an estimate of bias. There was no standard threshold value known to be correct and against which all the subjective threshold values could be compared.

If there was absolutely no intra-technician or inter-technician variability associated with thresholding an image, then there would also be no intra-technician or inter-technician variability associated with the areal porosity calculation. Variability in calculating areal porosity results from the propagation of variability in thresholding an image. This study demonstrated that error or variability does not propagate in a predictable fashion from the step of choosing a threshold of an image to the step of calculating an areal porosity of an image. The functional relationship between the variability in calculating areal porosity and the variability in thresholding an image is complex and beyond the scope of this investigation.

Depending on the characteristics of the images to be thresholded, inter-technician variability is the major contribution to the total variance for some images and the minor contribution for other images. For an image analysis protocol that requires subjective thresholding, it is important to evaluate both intra-technician and inter-technician variability when developing the protocol. This study indicates that inter-technician variability is of greater concern than intra-technician variability in thresholding an image and in computing areal porosity. In general, however, no generalization should be made about which source of variability is more important. In some images both sources may be important; e.g., image 3 for areal prorosity where the components split 60% and 40% attributable to inter- and intra-technician, respectively.

Experiment B illustrated an approach for estimating the bias due to subjective image analysis steps. A potential design deficiency is that observed shifts in the benchmark points due to registration–reregistration do not necessarily equal the bias and variability that would occur from the registration step alone. It is conceivable that either the reregistration error cancels out some of the registration error or the two steps produce double the bias of one step. In either case, the research team judged that the observed bias was cause for concern about the image analysis methodology.

The standard deviations for the registration–reregistration benchmark shifts were larger than for the fiducial point selection data. This could be because the combined effect of subjectively selecting all four fiducial points creates a large shift in a benchmark point. On the other hand, the greater standard deviation could simply be due to the two-step process used to measure the shift of a benchmark point.

The practical results of both experiments described here were that the inter-technician and intra-technician variances were too large for practical use of the original experimental methodologies. In the case of the image analysis method related to Experiment A, the research team sought an objective method for thresholding images (Yang et al., 2001), thereby eliminating the subjective component entirely. Experiment B indicated that the bacteria positions in a CSLM image might have been artificially shifted 2 μm or more when the CSLM image was registered to the AFM topography image. Such large errors in superimposing the bacteria on the AFM image prevented any reliable correlation between bacteria positions and topographic features. The research team decided to abandon their method and devised a different approach that used silicon surfaces instead of stainless steel and did not require registering two different images (Scheuerman et al., 1998).

In the absence of the knowledge produced by Experiments A and B, the research teams might well have wasted time trying to find statistically significant effects, when these important effects would have been masked by methodological variability. Our experience, illustrated by the two experiments of this paper, indicates that it is important to recognize and evaluate the uncertainty attributable to subjective components in an image analysis protocol. The examples presented in this paper show how one can perform such evaluations.

Acknowledgements

This work was supported by the Industrial Associates of the Center for Biofilm Engineering (CBE) and by a cooperative agreement EEC-8907039 between the National Science Foundation and Montana State University. The authors thank CBE technicians for their willing and productive participation in Experiments A and B.

References

Condron, P., McLoughlin, A.J., Upton, M., 1999. Quantitative determination of the spatial distribution of microbial growth kinetics within alginate beads using an image analysis technique. Biotechnol. Tech. 13, 927–930.

Hamilton, M., Johnson, K., Camper, A., Stoodley, P., Harkin, G., Gillis, R., Shope, P., 1995. Analysis of bacterial spatial patterns at the initial stage of biofilm formation. Biom. J. 37, 393–408.

Heydorn, A., Nielsen, A.T., Hentzer, M., Sternberg, C., Givskov, M., Ersbøll, B.K., Molin, S., 2000. Quantification of biofilm structures by the novel computer program COMSTAT. Microbiology 146, 2395–2407.

Hope, C.K., Clements, D., Wilson, M., 2002. Determining the spatial distribution of viable and nonviable bacteria in hydrated microcosm dental plaques by viability profiling. J. Appl. Microbiol. 93, 448–455.

Lewandowski, Z., Webb, D., Hamilton, M., Harkin, G., 1999. Quantifying biofilm structure. Water Sci. Technol. 39, 71–76.

Scheuerman, T.R., Camper, A.K., Hamilton, M.A., 1998. Effects of substratum topography on bacterial adhesion. J. Colloid Interface Sci. 208, 23–33.

Searle, S.R., Casella, G., McCulloch, C.E., 1992. Variance Components. Wiley, New York, NY, pp. 118–130.

Wheeler, D.J., Lyday, R.W., 1989. Evaluating the Measurement Process. SPC Press, Knoxville, TN, pp. 35–36.

Xavier, J.B., Schnell, A., Wuertz, S., Palmer, R., White, D.C., Almeida, J.S., 2001. Objective threshold selection procedure (OTS) for segmentation of scanning laser confocal microscope images. J. Microbiol. Methods 47, 169–180.

Yang, X., Beyenal, H., Harkin, G., Lewandowski, Z., 2001. Evaluation of biofilm thresholding methods. Water Res. 35, 1149–1158.