An average enumeration method of hyperspectral imaging data for quantitative evaluation of medical device surface contamination

Hanh N. D. Le,1,2 Moon S. Kim,3 Jeeseong Hwang,4 Yi Yang,1 Paweena U-Thainual,1 Jin U. Kang,2 and Do-Hyun Kim1,*

1Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, M.D. 20993, USA
2Department of Electrical and Computer Engineering, Johns Hopkins University, 3400 North Charles Street, Baltimore, M.D. 21218, USA
3United States Department of Agriculture, 1400 Independence Ave., S.W, Washington, D.C. 20250, USA
4National Institute of Standards and Technology, 325 Broadway St., Boulder, C.O. 80305, USA

*do-hyun.kim@fda.hhs.gov

Abstract: We propose a quantification method called Mapped Average Principal component analysis Score (MAPS) to enumerate the contamination coverage on common medical device surfaces. The method was adapted from conventional Principal Component Analysis (PCA) on non-overlapped regions of a full frame hyperspectral image to resolve the percentage of contamination from the substrate. The concept was proven by using a controlled contamination sample with artificial test soil and color simulating organic mixture, and was further validated using a bacterial system including biofilm on stainless steel surface. We also validate the results of MAPS with other statistical spectral analysis including Spectral Angle Mapper (SAM). The proposed method provides an alternative quantification method for hyperspectral imaging data, which can be easily implemented by basic PCA analysis.

©2014 Optical Society of America

OCIS codes: (110.4234) Multispectral and hyperspectral imaging; (170.3880) Medical and biological imaging.

References and links

1. M. G. Paraje, “Antimicrobial Resistance in Biofilms,” in Science against microbial pathogens: communicating current research and technological advances, A. Mendez Vilas, Ed., pp. 736–744, Formatex Research Center (2011).
2. M. Kalmokoff, P. Lanthier, T.-L. Tremblay, M. Foss, P. C. Lau, G. Sanders, J. Austin, J. Kelly, and C. M. Szymanski, “Proteomic Analysis of Campylobacter jejuni 11168 Biofilms Reveals a Role for the Motility Complex in Biofilm Formation,” J. Bacteriol. 188(12), 4312–4320 (2006).
3. D. Davies, “Understanding biofilm resistance to antibacterial agents,” Nat. Rev. Drug Discov. 2(2), 114–122 (2003).
4. M. A. De Groote, “Disease resulting from contaminated equipment and invasive procedures,” in Pathogenic Mycobacteria in Water: A Guide to Public Health Consequences, Monitoring and Management, Ed. S. Pedley, J. Bartram, G. Rees, A. Dufour and J. Cotruvo, Ed., pp. 132–142, IWA Publishing, London (2004).
5. Department of Health and Human Services, Immunology of biofilms: http://grants.nih.gov/grants/guide/pa-files/P4-07-288.html
6. T. F. Mah and G. A. O’Toole, “Mechanisms of biofilm resistance to antimicrobial agents,” Trends Microbiol. 9(1), 34–39 (2001).
7. I. I. Raad, M. F. Sabbagh, K. H. Rand, and R. J. Sherertz, “Quantitative tip culture methods and the diagnosis of central venous catheter-related infections,” Diagn. Microbiol. Infect. Dis. 15(1), 13–20 (1992).
8. M. L. Corona, S. G. Peters, B. J. Narr, and R. L. Thompson, “Subspecialty clinics: critical care medicine. Infections related to central venous catheters,” Mayo Clin. Proc. 65(7), 979–986 (1990).
1. Introduction

Bacteria exist in both free floating planktonic form and biofilm form; the latter consists of organized communities of bacteria living in polymeric forms. Biofilms can form on various surfaces, and are capable of harboring more than 99.9% of bacteria found in aquatic ecosystems [1, 2]. The growth of biofilm on surfaces can be categorized in three stages:
adsorption, adherence and ultimately the formation of extracellular polymeric substances (EPS) [2]. The EPS enhances the bacteria’s resistance towards desiccation, antimicrobials, and host defense [2, 3].

Studies have found that cases of pathogenic mycobacterial disease, including the formation of biofilm on the surface and interface of such devices, introduced by contaminated medical devices are on the increase [4]. According to a National Institute of Health estimation, 80% of human infections can be related to biofilm [5]. Medical devices and implants such as catheters, orthopedic devices, and contact lenses, are all subjected to biofilm contamination [6]. For instance, biofilm growth on a catheter tip can introduce bloodstream infections to patients [7–9]. Biofilms developed on mechanical heart valves cause conditions known as prosthetic valve condition [10]. Urinary catheters inserted inside patients can develop biofilm, and this accumulation of biofilms will lead to urinary tract infections [11]. Development of a biofilm surveillance method for medical devices is vital to the interest of public health [12]. The complex biofilm maturation process requires different treatments for each stage of development. For instance, antibiotic therapy on matured biofilm kills bacteria only in the superficial layer, which forms a protective layer for the bacteria underneath and allows it to remain dormant but alive after the treatment [13]. Assessing the effectiveness of antibiotic therapy requires better surveillance on the amount of biomass and percent surface coverage by biofilm [3].

Conventional contamination quantification methods, including spectrophotometric methods [14] have been reported, which add absorptive dyes such as crystal violet to the samples [15–17], and then indirectly measure the number of attached bacteria by determining the absorption extinction at 590 nm between clean polystyrene and one covered with biofilm [17]. The roll-plate technique is an application specific method developed to detect and quantify biofilms on catheter tips. This method first removes the tip of the catheter and then rolls the tip over the surface of nonselective medium. Biofilm on the tip is assessed from the number of transferred organisms recovered from the contacting surface [7, 12]. While effective, these methods are often laborious and time consuming, and lack the ability to extract dimensional information on the distribution of organisms on the tip [16]. Other optical modalities to image biofilm growth, such as confocal laser scanning microscopy [18], optical coherence tomography [16, 19], and optical interferometry have been demonstrated [20]. A comparative evaluation of the effectiveness of these promising techniques on medical device materials is yet to be presented.

Hyper-Spectral Imaging (HSI) has emerged as a quick and effective detection method to obtain a high order of spectral information of certain substance within a wide view [21]. Similar to spectroscopy, HSI acquires spectral information of the object of interest. However, instead of measuring point spectra, HSI data consist of spatially distributed spectra. Therefore, both spectral and spatial information are acquired at the same time, enabling the HSI technique to be optimized for complex object with mixture of different materials. Public health benefit using this technique ranges from quality and safety inspection of agriculture products [21–23] to tissue oxygenation monitoring in human body [24, 25].

Recently, the feasibility of utilizing HSI to analyze bacterial biofilm on stainless steel was reported [26]. The authors were able to obtain fluorescence images of two biofilm samples on stainless steel surfaces. The authors detected the distinct spatial features of concentrated biofilm growth, leading to biofilm quantification [26]. Recent work also studied biofilm attachment on the surfaces of materials such as metal (stainless-steel, titanium, and titanium-alloys), glass, polycarbonate, and polytetrafluoroethylene, all of which are materials commonly found in medical devices [27, 28]. These investigations demonstrate the potential of HSI for detecting and monitoring biofilm attachment on medical device surfaces. In this paper, we propose a rapid quantification method called Mapped Average Principal component analysis Score (MAPS) to quantitatively analyze medical device surface contamination from controlled materials of artificial test soil, color simulating organic mixture, and biofilm. The
motivation for exploring the average principal component (PC) score is that contaminants on medical device surfaces are commonly either microbial or chemical, and thus produce a very weak optical spectral signature especially in the visible spectral range. Therefore, compared to applying classification on the spectrum from one pixel, averaging over multiple pixels from a finite region of interest will produce signals with a larger signal-to-noise ratio. Throughout this paper, several terms will be used to provide reference to a certain physical object or quantity:

- Sample: a physical object which is comprised of a substrate and contaminants.
- Contaminant: the material that simulates real-world contaminating chemical or biological substance such as biofilm or soil.
- Substrate: the base substance on which contaminants are intentionally introduced.
- Reference spectra: a spectroscopic signature from the material of interest, which is identified by comparison to isolated spectrum of same material.
- Background spectra: spectroscopic signature from the substrate.

2. Method

2.1 Hyperspectral Imaging system

The schematic diagram of our HSI setup is illustrated in Fig. 1. The system consists of an Electron Multiplying Charge Coupled Device (EMCCD) equipped with an f/2.8 objective lens (F-stop 1.4/23, Schneider – Kreuznach Xenoplan) to capture an area of $10 \times 10$ cm$^2$ on the sample stage. To the camera was added an imaging spectrograph (Headwall Hyperspec VIS, Fitchburg, Massachusetts) with a fixed line slit of $25\mu m \times 18$ mm for line view restriction and a grating to disperse the collected light over a spectral range from 380 nm to 825 nm. The sample is illuminated by two 150 W quartz halogen lamps (Fiber-LiteA-241, Dolan-Jenner Industries, Boxborough, Massachusetts) with adjustable intensity for reflectance measurement and two UV-A lamps (Spectroline XX-15A 365 nm, Spectronics Corporation, Westbury, New York) for fluorescence measurements. The sample is linearly translated by a stepper motor stage (XSlide, VELMEX, Inc., Bloomfield, New York) to enable line-by-line imaging collection.

![Schematic of Hyperspectral Imaging System](image)

Fig. 1. Schematic of Hyperspectral Imaging System.

2.2 Sample preparation

Common medical device materials include metals such as stainless steel and titanium, ceramics, and plastics [29]. In our study, we chose multipurpose 304 stainless steel as a substrate to mimic medical device surface and applied our MAPS technique to the following two samples.
Sample 1: Artificial Test Soil (ATS, Healthmark Industries, Fraser, Michigan) and Color Simulating Organic Mixture (CSOM) were pipetted on an unpolished piece of stainless steel (SS, Mill finish, Multipurpose 304 stainless steel, Finish Selector Pack, McMaster-Carr, Robbinsville, New Jersey) substrate. ATS and CSOM had similar color under a visible illumination, and were chosen to test MAPS’s ability to distinguish spectrally similar substances. A photograph of droplets of these two materials is illustrated in Fig. 2.

Sample 2: Biofilm sample was grown from micro-organism *Pseudomonas aeruginosa* ATCC 9027 (American Type Culture Collection, Manassas, Virginia) on a satin finished stainless steel substrate (brushed #4 Satin, Multipurpose 304 stainless steel, Finish Selector Pack, McMaster-Carr, Robbinsville, New Jersey). Before biofilm culturing, the coupon was sterilized with germicidal ultraviolet-C light (Spectroline XX-15G 254 nm, Spectronics Corporation, Westbury, New York) on both sides at 244.8 mJ/cm² per side. The biofilm culturing was prepared from incubating bacterial stock in Tryptic Soy Broth (TSB) in an incubator shaker at 37°C for 24 hours. A serial dilution with phosphate buffered saline without magnesium and without calcium were consequently implemented for plating. The recorded concentration was $5.02 \times 10^7$ (colony forming unit/mL). 5 mL of bacterial stock was placed on the SS coupon, which was then placed inside a 100 mm × 15 mm sterile petri dish. A 1 mL volume of TSB was gently pipetted onto each coupon sample after 24 hours in incubation to prevent dehydration. The biofilm sample taken out of the incubator after 48 hours was transferred in a sterile petri dish. The sample was placed inside the HSI systems for reflectance and fluorescence spectra after it was held vertically for 30 seconds for non-biofilm elimination. The waiting time was determined empirically.

2.3 Endmember spectral characteristic

Reflectance and fluorescence spectra of the above samples were collected using HSI. The spectral range covers from 401 nm to 995 nm for reflectance measurement and 411 nm to 697 nm for fluorescence measurement. Both data sets have spectral resolution of 4.7 nm. To quantify the coverage area of interested contaminant and identify the contaminant origin within a sample, a reference spectra list was used. The reference spectra list included spectra of ATS and CSOM (Sample 1), and biofilm (Sample 2). Each reference spectrum of each contaminant is an average of spectra from interior regions within that contaminant. The collected spectra, illustrated in Fig. 3 are reference spectra that were subsequently used for spectral similarity analysis using MAPS.

Displayed in Fig. 3 are the original and normalized spectra of each material in reflectance and fluorescence mode. The original spectra are dominated by the light source spectra reflected from the background surface. To discriminate the subtle differences between the spectra, we normalized the signal with its respective minimum and maximum and unique signatures of each sample are resolved within the rescaled range from 0 to 1, respectively. These normalized spectra are shown for illustration of different spectra only, original spectra are used for the analysis.
From the rescaled spectra in Fig. 3, the spectral signatures indicate the distinguished peaks across the whole spectral range in fluorescence mode for all materials. The reflectance spectrum of SS indicates that the SS spectrum is basically that of halogen light source. Fluorescence spectrum of SS indicates that UV-A component is not present at the measured spectral range, and that SS shows negligible fluorescence.

2.4 Current quantification methods for hyperspectral analysis

Hyperspectral analysis identifies unique spectral signatures, which are often referred as endmembers, and their contribution factor within a sample. Some algorithms for endmember extraction from a mixing sample include the Pixel Purity Index method [30], the N-finder algorithm [31], and the Automated Morphological Endmember Extraction method [32]. These methods return a constrained linear mixing of the key endmembers to produce a spatial abundance map of each endmember. To avoid the unnecessary overloading analysis from linearly proportional spectra, Principal Component Analysis (PCA) is commonly used as an initial processing method for HSI analysis [21, 24].

PCA is a form of multidimensional scaling, which linearly transforms possibly correlated original variables into independent or uncorrelated variables retaining larger amount of information of the variables. This new transformation projects the variables into a set of bases called PC bases. This new presentation of the variable is named PC score, or score, for short.
2.5 Spectral similarity analysis using MAPS

We propose a new quantification method called Mapped Average PCA Score (MAPS) to quantitatively analyze surface contamination associated with a specific spectral signature. The distinguished contaminant is named an endmember. Instead of processing a full image in the fashion of pixel-by-pixel analysis to identify precise pixel locations of the contamination, MAPS evaluates the approximate coverage contribution of contaminants in relation to the entire sample area by calculating an average PC score within each ROI. As shown in Fig. 4, the hyperspectral image across the wavelength range is first segmented into smaller non-overlapping ROI uniformly distributed without spacing \( d_0 \) (Fig. 4(a)) and with spacing \( d_1 \) (Fig. 4(b)) between ROIs. All ROIs in both samples share identical dimensions to resolve a certain contamination degree. Through data binning in the spatial domain, the original pixel values within a region are replaced by the average intensity of that region. The process produces a smaller datacube and reduces the computation time compared to a dense data set. In addition, by averaging pixels within an ROI, signal to noise level is improved, which minimizes the impact of unwanted fluctuations from the background surface. For contamination detection, identifying precise microscopic locations of contaminants is not always necessary. It is more useful to approximate the coverage of contaminants in relation to the entire sample area. Thus, by taking PCA from segmented ROIs, average data over enumerating contaminant covered areas can be collected.

As a proof of concept, we initially analyzed the controlled samples of ATS and CSOM on a SS substrate. The analysis result yielding endmember classifications is based on its score density distribution in the first two PCs. The contaminant coverage on each contaminant is subsequently computed. A work flow of MAPS is constructed in Fig. 5.

**Step 1**: Binning the input datacube \( X_{\text{in}} \) such that pixel intensities within an ROI are averaged and create a pixel of \( X_{\text{binned}} \). The binning process is repeated for all non-overlapping ROIs.
ROIs of $X_{in}$ for both spacing scenarios using $d_n$ and $d_l$ for neighboring and sparse ROI configurations. ROI spacing selection is illustrated in Fig. 4 and was implemented empirically to cover certain contamination level.

**Step 2:** Implementing PCA on a combined matrix of ROIs and the reference spectra. The outcome is score values of each ROI of $X_{binned}$ datacube in the first two PCs, namely PC1 and PC2. Only the first two PCs are utilized because of their most significant contribution to the quantification among other PC values.

**Step 3:** Visualizing score scatter plot in PC1 and PC2 axes.

**Step 4:** Calculating score distance from each score values of individual ROI to the reference spectra using Eq. (1).

$$
\text{Score distance} = \sqrt{(\text{Score distance}_1)^2 + (\text{Score distance}_2)^2},
$$

$$
= \sqrt{(\text{Score }_{1\text{ROI}} - \text{Score }_{1\text{ref}})^2 + (\text{Score }_{2\text{ROI}} - \text{Score }_{2\text{ref}})^2}
$$

where Score 1 and Score 2 are the score values of PC1 and PC2, respectively.

The score distance defines the spectral similarity degree for each ROI compared to the reference. Non-weighted score distance calculation is used in this paper. Only the contributions of the first two components are taken into account because the effect of PC scores higher than PC2 is negligible in the score distance computation if they are not weighted. For more complicated analysis scenario in the future, a weighted distance calculation with compensated weighting factor of all components will be used.

For visualization of spectral similarity in an intensity map, the score distance is normalized into its quantified brightness level from “0” to “1”, respectively indicating the perfect match and complete mismatch between ROI and reference spectra. “0” indicates contaminants appeared in dark color and “1” indicates area free of contaminants as is illustrated in bright color (Figs. 6 and 7).

**Step 5:** Quantifying the contaminant coverage using adaptive Gaussian kernel density estimation [33]. The method constructs an approximation function to characterize score value density and creates a quantified density range, namely contour. For coverage percentage, each enclosed group of score values within a density range is called a population. Among the multi-scale contours displayed in PC1 and PC2 domain, only one contour is chosen to group populations of similar score values so that the contour embodies the closest score value of interested contaminant reference and isolates the contaminant population other neighbored populations. The enclosed population is assigned to be related to the reference contaminant. The ratio between counts of bounded scores within the chosen contour relative to the overall score counts in the scattering plot subsequently computed to provide the coverage percentage of the reference contaminant.

**2.6 Comparison to SAM**

In this study, the percentage coverage obtained from a spectral angle mapper (SAM) is used as a validation of the MAPS results. SAM has been used as an effective spectral similarity analysis method and widely utilized in hyperspectral analysis. Both reflectance and fluorescence data from Sample 1 and Sample 2 are used in SAM validation. SAM is a supervised spectral analysis technique that specifies spectral similarity of each pixel to the reference by calculating the angle between two spectra, treating the spectra as vectors in an $n$-dimensional space defined by $n$ reference spectral bands [23, 33]. The smaller the angle, the closer the pixel spectrum matches the reference spectra.
\[
\alpha = \cos^{-1}\left(\frac{\sum_{i=1}^{n} \text{pxl}_i \cdot \text{ref}_i}{\sqrt{\sum_{i=1}^{n} \text{pxl}_i^2 \cdot \sum_{i=1}^{n} \text{ref}_i^2}}\right),
\]

where \( n \) is the wavelength number, and \( \text{pxl} \) and \( \text{ref} \) are the pixel and reference vectors.

The threshold used to determine a pixel endmember depends on the definition of maximum angle, which is the maximum acceptable angle that a pixel is verified as an endmember. The parameters for SAM are chosen to yield up to 2 or 3 endmembers for Samples 1 and 2, respectively, and the default maximum angle is kept at 0.1 radian (5.7\(^\circ\)) for full classification (i.e., no unclassified pixels). Global threshold is used in this study to maintain the consistency of the SAM results on same materials within a coupon, the effect of local threshold for each material in complex structure will be studied in the future. The reference endmember collection used for angle computation was imported from the predefined ROI described in Section 2.3. The outcome of SAM is an image with color-coded pixels where the number of colors represents the number of endmembers. The contaminant coverage percentage is the ratio of specific color-coded pixel to the total number of pixels.

3. Results and discussion

In order to indicate coverage results of contaminants, we compared PCA score images using MAPS with SAM abundance images. As illustrated in Fig. 6 and Fig. 7, SAM abundance images indicate color-coded portions of each contaminant in Sample 1 and Sample 2, respectively. In comparison, MAPS results in a score image in copper colormap of two ROI uniform distributions in relation with the corresponding reference spectra. SAM details the segmented boundaries between the endmembers, while MAPS sectors the endmember distribution based on the average intensity of each ROI. The most dominant signal from each ROI strongly affects coverage quantification. Uniform distribution implies an assigned equal distance between each ROI in the spatial domain of the hypercube. The two ROI distributions are collected with spacings \( d_i \) and \( d_j \) between ROIs. Bright field gray scale images in Fig. 6 and Fig. 7 represent reflectance and fluorescence images of interested sample corresponding to 650 nm and 500 nm for demonstration purpose only. Entire spectral range of both measurement modes are used for the analysis. Identical ROI dimensions of 5-by-5 pixels are used to resolve the degree of contamination. For a bigger sample area, input data for MAPS need to be compressed or truncated to a smaller data set to reduce the computation time. To demonstrate this scenario, we analyzed a smaller hypercube (less ROIs) using a uniform ROI distribution with spacing \( d_i \) between ROIs. The results provide spatial and quantitative information on the distribution of contaminants with a lower-resolution boundary, however, with higher time efficiency. For instance, with original datacube of 46 pixel-by-170 pixel-by-128 wavelengths, calculation time for all scores were 23 minutes on a computer equipped with Intel Core i7-2630QM CPU, 6GB RAM, 64-bit Operating System. With the same spectral range of 128 wavelengths, by using MAPS for binned data set of 306 ROIs (\( d_i \)) and 85 ROIs (\( d_j \)), computation time reduces to 69 seconds and 20 seconds, respectively. This example demonstrates that MAPS method not only averages the data in ROIs for better SNR, but it also saves computing cost while maintaining quantitative analysis result that is comparable to pixel-based method. The colorbar (in copper color) used in MAPS score image (Fig. 6 and Fig. 7) indicates the score distance from 0 to 1, where 0 is for the maximum similarity between the spectra of the ROI and the reference (the contaminant portion) and 1 is for the least similarity (the substrate portion). The contaminant coverage comparison between
SAM and MAPS is discussed using control samples in session 3.1 and biological sample in session 3.2.

3.1 Application to well-controlled samples ATS and CSOM

The SAM abundance image as illustrated in Fig. 6 indicates contaminant portions of ATS in red, CSOM in green and SS in blue. The white bar indicates 10 mm length and the black line in the middle of the sample (Fig. 6(a)) is the gap between two identical SS coupons. Figure 6 shows that MAPS signifies the strongest sensitivity in detecting CSOM with a clear boundary from the SS background, showing an agreement between SAM and MAPS. However, MAPS showed a lower sensitivity in distinguishing weak signal of ATS fluorescence compared to SAM. CSOM was clearly distinguished in Fig. 6(b), row 2, while ATS was buried in the background as can be seen in Fig. 6(b), ATS score distance. To overcome this issue, normalized score image was tested as shown in Fig. 6(b), ATS score distance normalized to SS score distance. Normalization was performed by dividing the score image of SS background from score image of the ATS reference. After the normalization, the boundary of ATS sample which is previously unclear (in Fig. 6(b), ATS score distance) could now be visualized. The copper colorbars from 0 to 1, indicating contaminant portion to non-contaminant portion of the sample in both measurement modes are applied to all score distance images.

![Fig. 6. Comparison of sample coverage between SAM and MAPS for (a) reflectance and (b) fluorescence hyperspectral data at two degrees of spacing. The white bars indicate distance of 10 mm length. The bright field images are representative reflectance or fluorescence image at 650 nm and 500 nm, respectively, for demonstration purpose.]

3.2 Application to biofilm sample

MAPS was tested on a biofilm covered SS substrate sample (Sample 1) for validation purpose. Biofilm contaminants are harder to distinguish from substrate in visual inspection compared to ATS. Similar agreement between SAM and MAPS in CSOM and ATS contaminant was observed from biofilm contaminant in Sample 2. The biofilm sample shows distinguished signals in both reflectance and fluorescence modes. In the SAM image, biofilm contaminant is in green and SS is in blue. In the reflectance sample (Fig. 7(a)), small portions of biofilm contaminant are illuminated at the thin-layer (red box). However, these weak signals were not recognized under fluorescence mode (Fig. 7(b)). Agreement between SAM and MAPS was apparent. A comparison between two ROI distributions (d₀ and d₁) signifies clearer contaminant segmentation when a higher number of non-overlapping ROIs are used.
3.3 Coverage threshold

Because differing contaminations are composed of different spectral components, establishing a threshold requires an educated initial guess of coverage. Depending on the threshold boundary, all or partial ROI population will be classified as a contaminant in the reference list. To measure the percentage of contaminant on a sample, we applied the kernel density approximation to localize the contaminant populations.

Fig. 8. Reflectance score scattering plots with density contours of (a) ATS and CSOM populations (b) the expanded view from two population windows in (a). The red dots are score values of each ROI. Solid red, green and blue circles indicate ATS, CSOM and SS population, respectively. Coverage boundary is illustrated with thick solid contour. The SS (background) population is indicated by black arrow in (a). The color bar indicates the population density level.
Fig. 9. Fluorescence score scattering plot with density contours of (a) ATS and CSOM populations (b) the expanded view from two population windows in (a). The red dots are score values of each ROI. Solid red, green and blue circles indicate ATS, CSOM and SS population, respectively. Coverage boundary is illustrated with thick solid contour. The SS (background) population is indicated by black arrow in (a). The color bar indicates the population density level.

Figures 8, 9 and 10 are the score value scattering plots with kernel density contours for each contaminant collection. Each score value represents a PCA transformation of an average intensity within an ROI of a sample. The data in Figs. 8, 9 and 10 represent data selection with ROI spacing \( d_x \). The grouping boundary is assisted by the first closed density boundary that separates the populations within a sample. The boundary also encloses populations that are closest to the reference score value (red, green or blue circle). The more enclosed score values, the higher the contaminant density becomes. The population of background substrate signal illustrated by highest density values is the most dominant substance in the sample (indicated by black arrows in Figs. 8(a) and 9(a)). In order to exclude its unwanted density effect from our focus in contaminant analysis of CSOM and ATS, we limited the scattering window to smaller density scales (Figs. 8(b) and 9(b)) to analyze interested populations of ATS and CSOM only. Since the concerned populations are now limited, the kernel density distribution is rearranged to adapt the new population dimension. When adjusting the windows, the density contours are chosen for optimized display in Figs. 8(a) and 9(a) but for analysis in Figs. 8(b) and 9(b). Due to strong signals in both reflectance and fluorescence mode, CSOM provides a clear density boundary between its population and that of the neighboring population, indicating strong separation from ATS and SS populations. The small impact from PC2 score values with the addition of small PC1 values as shown in Fig. 9 indicates the small separation between ATS and SS signals, suggesting lower specificity of HSI in visible spectrum for ATS samples. Density boundaries illustrated in thick contours in Figs. 8(b) and 9(b) are established for coverage quantification, which can be compared to SAM results. The coverage ratio is subsequently computed and is summarized in Table 1.
Fig. 10. Score scattering plot with density contours of *Ps. aeruginosa* biofilm populations in (a) reflectance and (b) fluorescence mode. The red dots are score values of each ROI. Solid green and blue circles indicate biofilm and SS population, respectively. Coverage boundary is illustrated with thick solid contour. The color bar indicates the population density level.

Figure 10 demonstrates the selection of *Ps. aeruginosa* biofilm contaminant population. A higher density of score value population is detected to be nearer to the biofilm reference than the SS reference, indicates the most coverage of this sample is dominant from biofilm contaminant. The chosen density contour illustrated in thick blue line (Fig. 10) is chosen to ensure the separation of the interested biofilm population from the SS substrate population. The coverage ratio of biofilm is computed via the ratio of count number of enclosed score values and the total score counts. Two thick contours are signified for demonstration in both reflectance and fluorescence data because of the same density contour value. However, only the boundary enclosing the biofilm population is taken into account for coverage ratio calculation.

### 3.4 Analysis of error in thresholds

In order to further indicate the clear separation between populations in both Samples 1 and 2, error plots of the computed average and standard deviation of the score values both in PC1 and PC2 are displayed in Figs. 11 and 12 for $d_0$ and $d_1$, respectively. Color strips are used for demonstration of chosen materials in each sample. If there is overlap of average score values between different materials in a sample (ATS and CSOM in Sample 1 or biofilm and SS in Sample 2), it is concluded that the materials are spectrally similar as related to that PC. For both $d_0$ and $d_1$ spacing, there is significant overlap in mean and standard deviation of both PCs in each sample. This indicates the strong similarities between the coverage analyses between the two ROI spacing scenarios. Moreover, it can be clearly seen that PC1 embodies more significant impact in differences between both materials in each sample, while PC2 scores were somewhat similar. Therefore, our purpose of using MAPS method to quantitatively distinguish target spectra from background was verified. Each population of different materials is separated, contributing to the reliability of the MAPS method.
Fig. 11. Computed average score value and standard deviation of PC1 and PC2 for each contaminant of interest. ROI spacing of $d_0$ is used for the represented reflectance signals.

Fig. 12. Computed average score value and standard deviation of PC1 and PC2 for each contaminant of interest. ROI spacing of $d_1$ is used for the represented fluorescence signals.

Table 1. Comparison between contaminant coverage using SAM and MAPS with selective threshold.

| Material | Reflection mode | Fluorescence mode |
|----------|-----------------|-------------------|
|          | ATS | CSOM | Biofilm | ATS | CSOM | Biofilm |
| SAM coverage (%) | 20.28 | 29.46 | 71.52 | 21.83 | 30.53 | 74.85 |
| MAPS coverage (%) | $d_0$ | 20.26 | 29.74 | 72.55 | 17.32 | 28.10 | 71.86 |
|               | $d_1$ | 21.18 | 30.58 | 73.63 | 16.47 | 29.41 | 73.63 |

Table 1 summarizes the contaminant coverage results in both reflectance and fluorescence data using the MAPS and SAM methods. Comparable results occur in both controlled and...
biological samples. Although a lower contrast between contaminants and background was obtained for \( d_i \) compared to \( d_o \), a similar coverage percentage of each material was achieved, justifying the possibility of using a smaller data set in contaminant coverage quantification when calculation of larger data sets are required for large area surveillance. The method accounts for larger populations than SAM when signals are low, such as with the ATS signal in fluorescence mode. However, MAPS does not require a set maximum angle threshold as in SAM method, thus MAPS can avoid the unclassified pixel situation.

4. Conclusion

We explored a quantification method for hyperspectral image data and applied the method to contaminations on medical device surfaces. By reducing computational cost while obtaining reliable results based on an averaged signal, MAPS provides a tool to quantify contaminant coverage on substrate surfaces with non-negligible spectral signatures. Contaminant materials and tested biofilms with known spectra were attached to the stainless-steel surfaces to collect spectral data cubes under reflectance and fluorescence hyperspectral modalities. The resulting spectral analysis justifies the use of a uniform distribution of non-overlapping ROIs as an effective way to quantify biofilm coverage. We also demonstrated the effect of varied threshold based on kernel density distribution to the quantification of the coverage. With a proper selection of the threshold, MAPS and SAM quantification result were comparable.

Acknowledgments

The authors thank Victoria M. Hitchins for the help in preparing biofilm samples. This work was supported by the U.S. Food and Drug Administration Critical Path Initiative. Disclaimer: The mention of commercial products, their sources, or their use in connection with material reported here is not to be construed as either an actual or implied endorsement of such products by the Department of Health and Human Services and National Institute of Standards and Technology (NIST).