Nondestructive evaluation of aircraft coatings with infrared diffuse reflectance spectra

Hans G. Korth[a], Kody A. Wilson[b], Kevin C. Gross[b],
Michael R. Hawks[b,c] and Timothy W.C. Zens[d]

[a]Air Force Research Laboratory, Munitions Directorate, Eglin AFB, FL 32542;
[b]Air Force Institute of Technology, Wright-Patterson AFB, OH 45433;
[c]Oak Ridge Institute for Science and Education, Oak Ridge, TN 37831;
[d]Air Command and Staff College, Air University, Maxwell AFB, AL 36112

ABSTRACT

Aircraft coatings degrade over time, but aging can be difficult to detect before failure and delamination. We present a method to evaluate aircraft coatings *in situ* using infrared diffuse reflectance spectra. This method can detect and classify coating degradation much earlier than visual inspection. The method has been tested on two different types of coatings that were artificially aged in an autoclave. Spectra were measured using a hand-held diffuse reflectance infrared Fourier transform spectrometer (DRIFTS). One set of 72 samples can be classified as either aged or unaged with 100% accuracy. A second sample set contained samples that had been artificially aged for 0, 24, 48 or 96 hours. Several classification methods are compared, with accuracy better than 98% possible.

Keywords: Nondestructive testing, Infrared spectra, Reflectance spectra, Polymer degradation

1. INTRODUCTION

One of the most important functions of a coating material is to protect against corrosion of the underlying panel. When the coating fails, moisture and air are able to penetrate to the interface between the coating and substrate where it reacts with both the coating and underlying panel.[1,2] Corrosion costs the United States Air Force (USAF) billions of dollars per year, making corrosion prevention and control a priority for the USAF and congress.[3]

The current inspection method for coating failure is a visual inspection. Visual inspections are inexpensive and do not require extensive training or certification of the inspector. The inspector walks around and on the aircraft looking for visible indications of coating failure. These indications are typically cracking, wrinkling, blistering, and chipping. Visual indications are ambiguous and qualitative and often present themselves too late for efficient maintenance scheduling.

There are laboratory characterization tools that are capable of characterizing the degradation present in the coating material. However, these tools are impractical due to both their destructive nature and the requirement that samples be removed from the aircraft and sent to the lab for evaluation. The USAF is in need of a nondestructive field inspection method that is capable of characterizing the current state of coating degradation and estimating the remaining service life of the coating.

The objective of this work was to investigate whether or not a handheld FTIR device is suitable as a non-destructive inspection tool that can be used in the field. The portable and nondestructive nature of a handheld FTIR device makes it ideal for this application. Specifically, Diffuse Reflectance Infrared Fourier Transform spectroscopy (DRIFTs) was used to characterize the state of degradation of polyurethane aircraft coatings.

Send correspondence to M.R.H.: E-mail: mhawks@afit.edu, Telephone: 1 937 255 3636 x4828

Table 1. Sample naming convention and treatment conditions for samples used in this study

Set A: REC/REC combination 95°C, 130% relative humidity		**Set B:** Top coat/REC combination 101°C, 130% relative humidity	
Sample	Exposure Time	Sample	Exposure Time
A1	0 hrs	B1	0 hrs
A2	0 hrs	B2	24 hrs
A3	144 hrs	B3	48 hrs
A4	144 hrs	B4	96 hrs

2. MATERIALS AND METHODS

2.1 Sample Preparation

The sample coating materials and layer thicknesses were representative of coating layers applied to fielded aircraft. All the samples used for this work were subjected to a combination of heat and moisture in an autoclave to generate the degradation. Sample set A was the first sample set prepared to determine if the DRIFTs spectrum could distinguish between an as-cured sample and a level of visual degradation. With that goal in mind, it contained two as-cured samples and two samples aged to visual degradation. Sample set B was created to determine if intermediate levels of degradation could be distinguished, so it contained an as-cured sample and one each aged for 24 hours, 48 hours, and 96 hours with the 96 hour sample showing visual degradation. The two different coating materials used for this investigation are labeled top coat and rain-erosion coat (REC). Table 1 summarizes the treatment conditions for the samples.

2.2 Equipment

The hand-held FITR device used for this study was an Agilent 4100 Exoscan containing a Michelson interferometer with a 4 cm^{-1} resolution and a sampling range of 4,000-650 cm^{-1}. The external reflectance sampling head was used for all data collection and a gold diffuse reflectance reference was used for calibration. The Exoscan uses a ZnSe beam splitter and a deuterated triglycine sulfate (dTGS) pyroelectric detector.

2.3 Data Collection

Spectra were collected from several locations across the sample surface to accurately represent variance in the data from non-uniform degradation. Each location was interrogated once per trial, with three trials conducted per day. The device was recalibrated between each trial with the gold diffuse reflectance reference standard provided with the device. Sample trials were conducted on sample set A for two consecutive days immediately after the aging regimen. Sample set B data was collected over the course of six weeks.

2.4 Data Analysis

The FTIR spectra were analyzed using singular value decomposition (SVD) and the samples were classified using linear discriminant analysis (LDA). SVD is commonly used as a pattern recognition technique, making it an ideal tool to search for a trend in the degradation levels of the samples.[4,5] LDA is a simple data classification algorithm that made sample classification objective and quantitative rather than subjective based on qualitative analysis.[6,7] It has been shown that when the two are combined, the predictive accuracy of discriminant analysis can be improved.[8]

3. THEORY

3.1 FTIR of Polyurethane

Generally, all polymer materials are susceptible to the same degradation mechanisms that include photo-degradation and thermal degradation.[9−17] In the case of aircraft coatings, exposure to heat, moisture, and UV radiation are the primary causes of degradation.[14−17] There are changes in the chemical composition and

polymer chain structure associated with these degradation mechanisms that should be detectable by FTIR spectroscopy.

There are two important regions of the IR spectrum of an organic polymer. The first region is called the functional region because it allows for the identification of functional groups present in the material. Depending on the source, the functional region is located in the 4,000-1,500 cm^{-1} (2.5-6.7 μm) to 4,000-1,300 cm^{-1} (2.5-7.7 μm) range of the IR spectrum.[18–20] The absorption bands seen in this region are typically due to stretching vibrations from single-, double-, and triple-bonded molecules.

The other region is known as the fingerprint region, which is of particular importance in a polymer's IR spectrum. The precise location of the fingerprint region is generally considered to begin where the functional region ends (1,500-1,300 cm^{-1}) and ends around 910-600 cm^{-1}(11-16.7 μm).[18–20] The fingerprint region is aptly named because it contains absorption frequencies arising from the complex interactions of molecules in the material and is unique to the material being examined. Molecular vibrations seen in this region include wagging, twisting, scissoring, and rocking as well as interactions between the functional groups.[20] Many polymer spectra will look very similar in the functional region, but will have a unique fingerprint region.

3.2 Pattern Recognition

Singular value decomposition (SVD) is a matrix decomposition method with many applications in image processing, data compression, pattern recognition.[5,21] For any $M \times N$ data matrix, D, there exists a factorization

$$D = U\Sigma V^T, \tag{1}$$

such that U and V are unitary matrices and Σ is a diagonal matrix whose elements are known as singular values. The rows of U and V are eigenvectors of DD^T and D^TD, respectively. Because the matrices are unitary, the rows can be used to form an orthonomal basis set.

For this data, the rows of D are mean-removed spectra, each of length N spectral channels chosen to use only the fingerprint region (850-1,220 cm^{-1}). Each of the M sample spectra can be characterized by weights $a_i, i \in \{1, 2...K < M\}$, which represent coordinates in a K-dimensional subspace of the basis set formed by V. These values are found by projecting the spectrum onto the basis set of eigenvectors, or

$$a_i = \vec{d} \cdot \vec{v_i}, \tag{2}$$

where $\vec{d}$ is a spectrum (row of D) and $\vec{v_i}$ is the i^{th} column of V. More generally,

$$A = DV, \tag{3}$$

so that a_{ij} represents the j^{th} coordinate of the i^{th} spectrum.

We should note here that the mean-removed spectra forming the rows of D can be computed in two different ways. For purely independent, uncorrelated samples, it would be appropriate to remove the average intensity over all bands, then divide by the standard deviation over all bands. This produces a zero-mean, unit area spectrum, equivalent to the z-score in statistics. The z-scored spectra are then

$$\vec{z_i} = \frac{\vec{d_i} - \bar{d_i}}{\sigma(\vec{d_i} - \bar{d_i})}, \tag{4}$$

where the operator $\sigma_i(\_)$ indicates the standard deviation and $\bar{d}$ represents the mean over the band (fingerprint region, in this case).

Alternatively, one could note that the differences between the sample spectra are relatively small. This suggests subtracting the sample mean (average spectrum over all samples) instead of the spectral mean in order to emphasize the small spectral differences between samples. If we denote these sample mean-removed spectra as $\vec{s}$,

$$\vec{s_i} = \vec{d_i} - \langle\vec{d}\rangle, \tag{5}$$

where $\langle \vec{d} \rangle$ represents the mean spectrum over all samples.

Henceforth, we use the notation a_i to indicate weighting coefficients (or coordinates) computed from the z-scored spectra, and b_i when the spectra in equation 2 have had the sample mean removed. The choice of spectra will, of course also affect the eigenspectra. In other words,

$$Z = U\Sigma V_z^T \Rightarrow A = ZV_z \qquad \text{and} \qquad S = U\Sigma V_s^T \Rightarrow B = SV_s, \tag{6}$$

where S is the data matrix composed of row-spectra $\vec{s_i}$ and Z is composed of rows $\vec{z_i}$.

3.3 Sample Classification

Discriminant analysis is a statistical approach to data classification first developed by Sir Ronald A. Fisher.[7] Commonly, either PCA/SVD or discriminant analysis is used independently to classify data.[4,6,21] PCA or SVD are more commonly used when all the variables are quantitative and discriminant analysis is more appropriate when the dependent variable is categorical and the independent variables are quantitative. However, if the principal components (or singular values) show separation or clustering, discriminant analysis can be applied to the principal components to improve classification accuracy.[8]

Much like SVD, the application of discriminant analysis to a data set is straightforward in commercial software. The general approach consists of collecting truth data to generate the discriminant function(s) and determining the prior probability that an observation belongs to a group. In this work, the prior probability for each sample is equal because there is the same number of samples in each group. The computational software will then calculate the discriminant score function(s) based on the sample mean vectors and variance-covariance matrices. After the discriminant function(s) are developed based on the truth data, the accuracy of the rule is checked using a method such as the holdout method or cross-validation. The classification model is then applied to new data with unknown classification for prediction purposes.

For illustration purposes, consider classifying samples into two classes based on only the first two singular values. Linear discriminant analysis (LDA) results in drawing a decision line that divides the $a_1 - a_2$ plane into two classes. For higher dimensional data (more than two principle components), the discriminant function will be a hyperplane rather than a line.

4. RESULTS AND DISCUSSION

Figure 1 is a plot of the mean reflectance collected the first week on sample set B in the 1,800-800 cm^{-1} range. The spectra were examined for immediately evident trends such as peak broadening or changes in peak height, but none were apparent. Several different bands of the sampling spectrum were analyzed using SVD, but the 1,220-850 cm^{-1} region of the spectrum provided the highest level of LDA discrimination. That result is expected since the 1,220-850 cm^{-1} band falls within the published bounds of the fingerprint region.

Figure 2 is a plot of the first two SVD coefficients for the 1,220-850 cm-1 band for sample set B. The left plot shows the a coefficients (computed using z-scored spectra), and the right plot shows the b coefficients computed by subtracting the mean spectrum (over all samples) as in equation 5. In both plots, you can see clear separation of the classes, although the z-scoring method does show significant overlap of the 24-hour samples (B2) with both 0 and 48-hour samples (B1 and B3). The separation is enhanced by removing the sample mean first, as shown in the right plot. Samples in the z-scored plot appear to show a smooth trajectory as the age increases, indicating that this method may be more useful for predicting different ages (e.g. a new 72-hour sample set), but the overlap between sets shows this will be less useful for classification of the present sample set.

Table 2 contains the confusion matrix and error rates of the classification rules generated using both Z and S spectra. When the first two coefficients were used to generate the classifier, the error rate was only 3% using sample mean subtration. The misclassifications occurred primarily between the 24 and 48 hour samples. When the third coefficient was included in the generation of the classifier, the error rate becomes less than 2% with the misclassifications being between the 24 and 48 hour samples. Classification accuracy in the z-scored spectra is lower, as expected from figure 2.

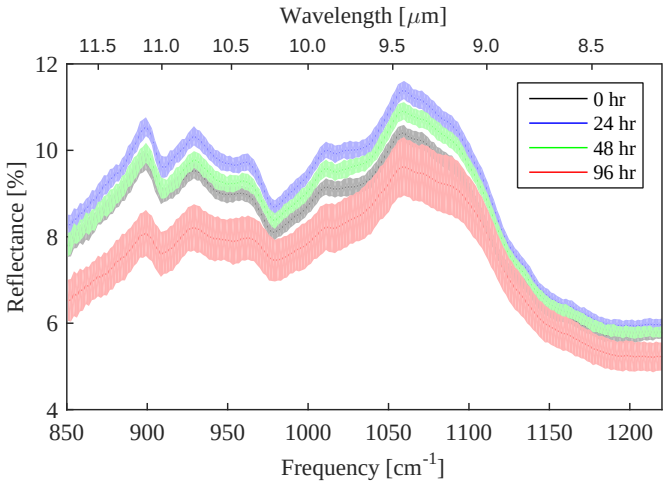

Figure 1. The long-wavelength end of the sampling spectrum, including the fingerprint region, for samples B1-B4 collected the first week. The solid lines show the mean spectrum over 75 samples and the shaded regions show ± one standard deviation.

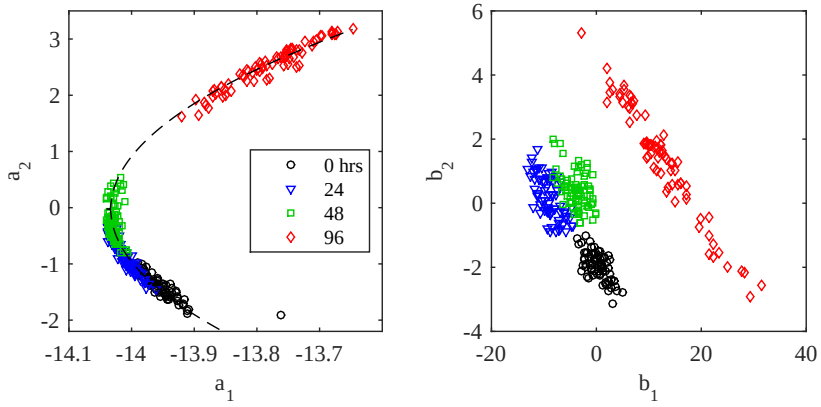

Figure 2. Plot of the first two coefficients in the 850-1,220 cm^{-1} band for sample set B. There is significant clustering of the samples that makes each sample clearly distinguishable from each other. The figure on the left was calculated from spectra in Z, while the figure on the right uses S. The dashed line on the left is a parabolic fit.

Table 2. Classification results for sample set B in the 850-1,220 cm^{-1} band. The confusion matrices are for 2-coefficient classification; only total error rate is shown for higher dimensional results.

Spectra	True Age [hours]	Predicted Age Age [hours]				Number of Coeff. used	Error [%]
		0	24	48	96		
Z	0	65	10	0	0	2	16.3
	24	8	49	18	0	3	15.3
	48	0	13	62	0	4	13.3
	96	0	0	0	75	5	13.0
S	0	75	0	0	0	2	3.0
	24	0	70	5	0	3	1.7
	48	3	1	71	0	4	1.7
	96	0	0	0	75	5	1.7

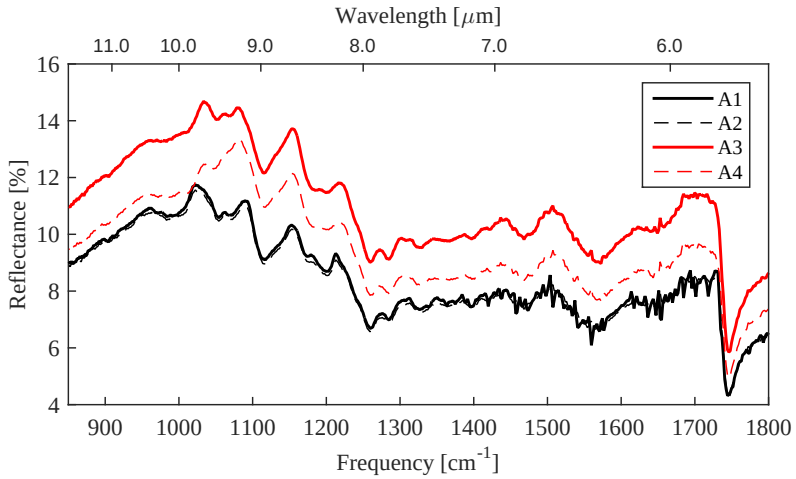

Figure 3. Mean reflectance of the four samples in set A. The two untreated samples (A1 & A2) lie almost on top of each other while the treated samples (A3 & A4) have a higher mean reflectance over the entire spectrum. Error bars have been omitted here for clarity, but are similar to those shown in Figure 1.

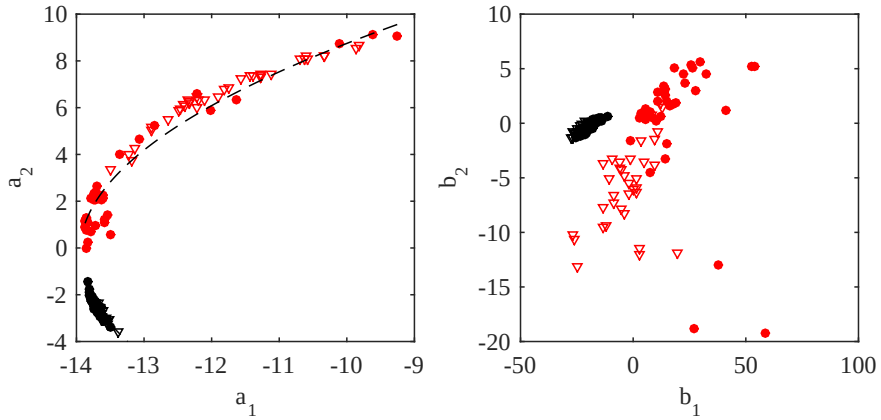

Figure 4. The first two coefficients generated from the fingerprint region (850-1,220 cm−1) for both the training set (•) and the test data set (▽) for prediction. The untreated samples are in black and the treated samples in red. As in Figure 2, the left plot shows results from z-scored spectra (Z) with a parabolic fit line, while the right plot is from spectra with the sample mean subtracted (S).

After verifying that different degradation levels are detectable, the final step in demonstrating the potential of this approach is classifying new sample data correctly. Sample set A is comprised of two untreated samples and two samples treated to a visibly degraded level. There are no intermediate levels of degradation to attempt to classify. This makes it ideal for a rudimentary test of the approach and its ability to correctly classify new samples. One untreated sample and one visibly degraded sample were used to generate the classifier and the other two samples used as blind samples. As with sample set B, the first step is an examination of the mean reflectance of the four samples in set A. Figure 3 reveals that the two untreated sample spectra are nearly identical and the two treated samples are distinctly different from the untreated samples and each other.

The two samples used as the training set to generate the classification rule were samples A1 and A3. Samples A2 and A4 were the blind samples for classification. Notice the clear separation between the two treated samples. The plan was to use A3 to train a rule that would be tested against A4, but clearly there is a lot of variability between sample sets that were prepared in the same way. It is not yet known what caused this variation, or whether this represents the variability between actual naturally aged samples. A much larger set of samples is currently being prepared which will help answer this question.

Figure 4 is a plot of the first two SVD coefficients of the training set and the unknown samples generated from

Table 3. Confusion matrices for 2-coefficient classification of sample sets A2 and A4, using only samples A1 and A3 to train the classifier.

Spectra	Method	Num of Coeff. Used	True Cond. [hours]	Predicted Condition		Error [%]
				Untreated	Treated	
Z	LDA	2	Untreated	36	0	0
			Treated	0	36	
S	LDA	2	Untreated	36	0	36.1
			Treated	26	10	
	LDA	3	Untreated	36	0	48.6
			Treated	35	1	
	QDA	2	Untreated	36	0	0
			Treated	0	36	

the fingerprint region. As in Figure 2, the analysis is repeated, once with z-scored spectra and once where the sample mean spectrum has been subtracted. In this case, the sample mean is calculated from only the training data set, then subtracted from both training and test sample spectra.

Based on figure 4, it was expected that the blind samples would classify well because there is a clear separation between the treated and untreated samples. Notice, however, the variance of the treated samples is much larger than the untreated samples (seen as the large spread of points in 4). LDA assumes identical covariances, which in this case leads to some misclassification, evident in the confusion matrices in table 3.

Table 3 shows the classification accuracy in this sample set is higher when using the z-scored spectra rather than the mean-removed spectra, S. Using the first two coefficients for classification, the rule correctly assigned all 36 untreated samples; however, only 10 of the treated samples were correctly classified for an accuracy of 64%. When the first three coefficients were used for classification, accuracy drops to 51%, indicating singular values above 2 are noise.

Figure 5 shows why results for LDA are so poor in sample set A. This is a plot of the first two SVD coefficients for the samples used to develop the discriminant function, along with the resulting discriminant function (a straight line). It shows why the discriminant function misclassifies so many of the treated samples. This results from the separation between the two treated samples in 4. The classification function derived from one treated sample does not adequately represent the other treated sample.

One way of improving the accuracy of a discriminant classification model is to adjust the cost of misclassification, as demonstrated in figure 5. In this example, we assume the cost of misclassifying a treated sample as untreated is 100× higher than vice versa. The adjusted classification rule still correctly assigned all 36 untreated samples, but now correctly assigns 28 of the unknown treated samples for an accuracy of 78%. This is much improved over the 28% accuracy before adjusting the cost; however, the classifier still does not achieve 100% accuracy as expected from an examination of figure 5. These results show that a larger sample set is necessary to ensure the training data set adequately represents the test data.

Another alternative is to use quadratic discriminant analysis (QDA), which results in the classification line being a quadratic curve. LDA does not consider differences in the covariance of the two classes in training data, so the classification line tends to be the linear bisector of the line connecting the centroids of the two point clouds. QDA, on the other hand, takes into account the larger spread of points in the treated class. The result is a classification ellipse that encloses the tightly grouped points of the untreated class. This increases the classification accuracy to 100% for both the training data and the blind samples, using either two or three coefficients.

5. CONCLUSIONS AND FUTURE WORK

This study shows there is strong evidence to support the further investigation of DRIFTs for use as a nondestructive inspection tool for aircraft coating characterization. The fingerprint region of the IR spectrum showed

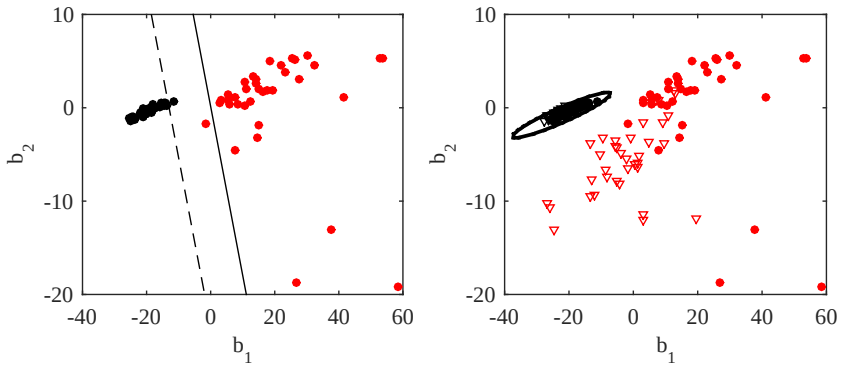

Figure 5. Classification results from LDA (left) and QDA (right). LDA trained with samples A1 and A3 (black and red •) produces the classification line (solid black line). This can be translated by adjusting the cost of misclassification, resulting in the weighted classification line (dashed black line). QDA trained with the same sets produces the classification ellipse (thick black line), which is also effective in classifying the test samples A2 and A4 (black and red ▽). Black points show untreated samples; red points are treated.

great potential to discern degraded material from non-degraded material as well as the ability to distinguish between intermediate levels of degradation. The data showed that multi-class samples could be correctly classified with 98% accuracy. When the approach was applied to classify samples with unknown degradation, the classification rule was 100% accurate. Note, however, that different normalization methods were used for the two different sample sets. The z-scoring method produced 100% accuracy in sample set A, where subtracting the mean spectrum (over all samples) was more accurate in sample set B. More research, using larger sample sets, is needed to determine the most robust classification method. Samples generated from many different coating configurations and fielded aircraft panels will be examined in future work. In order to extend the application to predicting remaining service life, the accelerated degradation applied to lab samples must be correlated to real on-aircraft degradation. While the combination of SVD and LDA revealed the potential of DRIFTs to detect coating degradation, there are many other classification techniques available, which will be compared in future work.

ACKNOWLEDGMENTS

The authors would like to thank the Materials and Manufacturing Directorate of the Air Force Research Laboratory and Dr. Adam Cooney for sponsoring this work and for the use of the equipment, lab space, and samples during this effort. The views expressed in this paper are those of the authors and do not necessarily reflect the official policy of the U.S. Air Force, the Department of Defense, or the U.S. Government.

REFERENCES

[1] Kosik, W. E., "Mechanisms of military coatings degradation: Final technical report," tech. rep., Air Force Research Laboratory, Aberdeen Proving Grounds, MD (2003).

[2] Grundmeier, G. and Stratmann, M., "Adhesion and de-adhesion mechanisms at polymer/metal interfaces: Mechanistic understanding based on in situ studies of buried interfaces," *Annu. Rev. Mater. Res.* **35**(1), 571–615 (2005).

[3] Forman, D., Baty, R., Herzberg, E., Kelly, A., Kumaran, M., and O'Meara, N., "The annual cost of corrosion for air force aircraft and missile equipment," tech. rep., LMI Government Consulting (2009).

[4] Mariey, L., Signolle, J. P., Amiel, C., and Travert, J., "Discrimination, classification, identification of microorganisms using FTIR spectroscopy and chemometrics," *Vib. Spectrosc.* **26**(2), 151–159 (2001).

[5] Wall, M., Rechtsteiner, A., and Rocha, L., [*Singular Value Decomposition and Principal Component Analysis*], 91–109, Springer, Berlin (2003).

[6] Yang, H., Irudayaraj, J., and Paradkar, M. M., "Discriminant analysis of edible oils and fats by FTIR, FT-NIR and FT-Raman spectroscopy," *Food Chem.* **93**(1), 25–32 (2005).

[7] Fisher, R. A., "The use of multiple measurements in taxonomic problems," *Ann. Eugen.* **7**(2), 179–188 (1936).

[8] Zhao, W., Krishnaswamy, A., Chellappa, R., Swets, D., and Weng, J., "Discriminant analysis of principal components for face recognition," in [*Face Recognition*], H. Wechsler, e. a., ed., Springer, Berlin (1998).

[9] Perera, D. Y., "Effect of thermal and hygroscopic history on physical aging of organic coatings," *Prog. Org. Coatings* **44**(1), 55–62 (2002).

[10] Perera, D. Y., "Physical aging of organic coatings," *Prog. Org. Coatings* **47**(1), 61–76 (2003).

[11] Perera, D. Y., "Effect of pigmentation on organic coating characteristics," *Prog. Org. Coatings* **50**(4), 247–262 (2004).

[12] Yilgor, I., Yilgor, E., Guler, I. G., Ward, T. C., and Wilkes, G. L., "FTIR investigation of the influence of diisocyanate symmetry on the morphology development in model segmented polyurethanes," *Polymer (Guildf).* **47**(11), 4105–4114 (2006).

[13] Croll, S. G., Shi, X., and Fernando, B. M. D., "The interplay of physical aging and degradation during weathering for two crosslinked coatings," *Prog. Org. Coatings* **61**, 136–144 (2008).

[14] Yang, X. F., Vang, C., Tallman, D. E., Bierwagen, G. P., Croll, S. G., and Rohlik, S., "Weathering degradation of a polyurethane coating," *Polym. Degrad. Stab.* **74**(2), 341–351 (2001).

[15] Yang, X. F., Vang, C., Tallman, D. E., Bierwagen, G. P., Croll, S. G., and Rohlik, S., "Blistering and degradation of polyurethane coatings under different accelerated weathering tests," *Polym. Degrad. Stab.* **77**(1), 103–109 (2002).

[16] Yang, X. F., Tallman, D. E., Croll, S. G., and Bierwagen, G. P., "Morphological changes in polyurethane coatings on exposure to water," *Polym. Degrad. Stab.* **77**(3), 391–396 (2002).

[17] Yang, X. F., Li, J., Croll, S. G., Tallman, D. E., and Bierwagen, G., "Degradation of low gloss polyurethane aircraft coatings under UV and prohesion alternating exposures," *Polym. Degrad. Stab.* **80**(1), 51–58 (2003).

[18] Settle, F. A., [*Handbook of instrumental techniques for analytical chemistry*], Prentice-Hall, Upper Saddle River (1997).

[19] Hsu, C., "Infrared spectroscopy," in [*Handbook of Instrumental Techniques for Analytical Chemistry*], Settle, F., ed., 247–284, Prentic-Hall, Upper Saddle River (1997).

[20] Leng, Y., [*Materials Characterization: Introduction to Microscopic and Spectroscopic Methods, 1st ed.*], John Wiley & Sons (2008).

[21] Wold, S., Esbensen, K., and Geladi, P., "Principal component analysis," *Chemom. Intell. Lab. Syst.* **2**, 37–52 (1987).