Non-destructive Techniques for Classifying Aircraft Coating Degradation

THESIS

Kody A. Wilson, 1LT, USAF

AFIT-ENP-MS-15-M-88

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

DISTRIBUTION STATEMENT A
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

The views expressed in this document are those of the author and do not reflect the official policy or position of the United States Air Force, the United States Department of Defense or the United States Government. This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

AFIT-ENP-MS-15-M-88

NON-DESTRUCTIVE TECHNIQUES FOR CLASSIFYING AIRCRAFT COATING DEGRADATION

THESIS

Presented to the Faculty

Department of Engineering Physics

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of Master of Science in Applied Physics

Kody A. Wilson, B.S.

1LT, USAF

March 12, 2015

DISTRIBUTION STATEMENT A
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AFIT-ENP-MS-15-M-88

NON-DESTRUCTIVE TECHNIQUES FOR CLASSIFYING AIRCRAFT COATING DEGRADATION

THESIS

Kody A. Wilson, B.S.
1LT, USAF

Committee Membership:

Dr. M. R. Hawks
Chair

Dr. K. C. Gross
Member

Dr. M. A. Marciniak
Member

LtCol J. D. Clark, PhD
Member

AFIT-ENP-MS-15-M-88

Abstract

A lack of objective techniques to assess the level of coating degradation creates a significant cost to the Air Force in both time and money. In this research infrared spectra were examined as possible methods of determining aircraft coating degradation. Single Value Decomposition(SVD) then Linear Discriminant Analysis(LDA) or multiple class linear discriminant analysis(MDA), were applied to measured IR spectra. When measuring infrared emittance spectra only 52% classification accuracy was achieved. Raman spectroscopy has higher classification accuracy at 70.4% when using the same SVD-LDA algorithm. However the best performing method was infrared reflectance with classification accuracy 94-100% depending experimental factors.

For the Diffuse Reflectance Infrared Fourier Transform System(DRFITS) data, a fingerprint region with higher classification accuracy was identified as 865.6 - 1238.7 cm^{-1} reducing the classification error by half. Feature selection, ReliefF and forward selection, were applied to determine possible filter locations for a multi-spectral measurement. When simulating the optimal filters and commercially available filters an accuracy of 95% and 94% were achieved using a mere five filters.

Acknowledgements

I would like to offer my gratitude to the following individuals without whom this project would not have been possible.

My advisor Dr. Michael Hawks for his continued guidance throughout the project specifically on some of the most challenging aspects of the project.

Major Timothy Zens for assisting me with getting started project and working through the academic planning and bureaucracy regarding the project.

Captain Hans Korth for providing me with a walk through of his classification algorithm techniques as well as information on some of the classification techniques he did not have time to develop and for the DRIFTS data for sample sets A and B.

Dr. Kevin Gross for providing code and algorithms regarding the Bomem 154 FTS and Telops Hyperspectral Camera and putting me in direct contact with many of the other individuals helpful to the completion of the project.

Dr. Michael Marciniak for providing assistance regarding the radiometry aspect of using the Bomem 154 FTS and by extension the Telops Hyperspectral Camera.

Mr. Jake Martin for walking through the Telops Hyperspectral Camera and assisting me in troubleshooting data collection and data analysis with the Telops data.

Dr. Adam Cooney and AFRL/RXCA for sponsoring the project, providing sample sets, data and the DRIFTS equipment.

Mr. Joe Costantino for providing unrestricted access to sample set C which were the primary experiment samples throughout most of the project.

Kody A. Wilson

Table of Contents

Page

Abstract iv

Acknowledgements v

List of Figures viii

List of Tables xi

I. Introduction 1

1.1 Motivation 1
1.2 Previous Work 1
1.3 Objective 2

II. Theory 5

2.1 Aircraft Coating and Polyurethane Degradation 5
2.2 Fourier Transform Infrared Spectroscopy 5
2.3 Hyperspectral Imaging 9
2.4 Blackbody Radiators and Emittance 9
2.5 Specular and Diffuse Reflectance 10
2.6 Radiometric Correction and Calibration 11
2.7 Raman Spectroscopy 13
2.8 Classification 14
Overtraining 15
Linear Discriminant Analysis 18
Decision Trees 19
K-Fold Cross Validation 20
2.9 Feature Selection 21
Forward Selection 22
Relief 23
ReliefF 24
2.10 Feature Extraction: Single Value Decomposition 25

III. Experimental Methods 30

3.1 Sample Preparation 30
3.2 Diffuse Reflectance Infrared Fourier Transform Spectroscopy(DRIFTS) 31
3.3 Surface Optics Corporation SOC-100 Hemispherical Directional Reflectometer 33
3.4 Stand-off Measurement Bomen FTS Reflectance 35
3.5 Telops Hyperspectral Reflectance Imaging 39

Page

3.6 Stand-off Measurement Bomen FTS Emittance 41
3.7 Raman 42
3.8 Classification Technique 43

IV. Results and Discussion 47

4.1 DRIFTS 47
4.2 Surface Optics Corporation SOC-100 61
4.3 Bomen Fourier Transform Spectrometer Reflectance 64
4.4 Hyperspectral Imaging Telops 70
4.5 Instrument Comparison 77
4.6 Bomen Fourier Transform Spectrometer Emittance 79
4.7 Raman Spectroscopy 83
4.8 Feature Selection-DRIFTS 88

V. Conclusions 93

5.1 Achievements 93
5.2 Future Work 94
References 97

List of Figures

Figure Page

1. Fourier Transform Example 6
2. Fourier Transform Spectra Example 8
3. Raman Scattering Illustration 14
4. Overtraining Accuracy Prediction 16
5. Two Dimensional Overtraining Example 17
6. LDA example 19
7. Decision Tree Example 20
8. SVD Example Spectra Page 26
9. SVD Coefficients Example 27
10. SVD Example Spectra with Noise 28
11. SVD Coefficients Example with Noise 29
12. DRIFTS Spectra for Data set B 32
13. SOC-100 Diagram 34
14. Bomen Experimental Setup 36
15. Telops Experimental Setup 40
16. Bomem Emittance Experimental Setup 42
17. Average Reflectance Spectrum for Data Set B 43
18. Eigen Spectra for Sample Set B 44
19. SVD Constructed Signals 45
20. SVD-MDA for Data Set B 46
21. SVD coefficients and discriminant lines for sample set A 49
22. SVD, discriminant lines and PDF's for sample set A 51

Figure Page

23. Reflectance Spectra for Sample Set A 52

24. Absolute Reflectance for Sample Set C 53

25. Normalized Reflectance for Sample Set C 54

26. Normalized SVD coefficient space for Sample Set C 55

27. Normalized SVD coefficient space for Sample Set A 56

28. Set B Accuracy with Respect to SNR 58

29. Multiple Layer Reflectance Example 59

30. Test Layers for Source of Reflectance Spectra 60

31. DRIFTS Top Coat Examination 61

32. SOC-100 Reflectance vs Angle Test 62

33. SOC-100 Average Reflectance Spectra by Sample 63

34. SOC-100 Classification Accuracy 64

35. Bomen Calibration Test 66

36. Bomen Diffuse Reflectance Spectra for Sample Set C 67

37. SVD Coefficients for Bomen Diffuse Reflectance Spectra 68

38. Bomen Specular Reflectance Spectra for Sample Set C 69

39. Telops Reflectance Image 71

40. Telops Average Spectra for Each Sample 72

41. Telops PCA False Color Image 73

42. Telops PCA Feature Space 74

43. Telops Classification Accuracy 75

44. Telops Classification Pixel Map 76

45. Telops 3rd Principle Component Pixel Map 77

46. Comparison of Infrared Reflectance Instruments 78

Figure Page

47. Bomen Radiance Spectra for Heated Samples 80

48. Temperature Estimates for Heated Samples 81

49. Bomem Emittance Spectra for Weathered Samples 82

50. Raman Variations with Respect to Input Power 85

51. Raman Variation with Focus Depth and Integration Time 86

52. Raman Accuracy vs the Number of SVD Coefficients Retained and Input Power 87

53. Forward Selection Classification Accuracy Vs Number of Retained Features 88

54. ReliefF Classification Accuracy Vs Number of Features 89

55. DRIFTS Selected Features 90

List of Tables

Table Page

1. All Samples Generated for Degradation Classification 30

2. Filter Center Wavenumbers and Bandwidths 91

NON-DESTRUCTIVE TECHNIQUES FOR CLASSIFYING AIRCRAFT COATING DEGRADATION

I. Introduction

1.1 Motivation

The current operational method of non-destructive measurement of aircraft coating degradation is visual inspection done by maintenance personnel. This visual inspection technique is subjective, time consuming and provides little early indication of degradation before the aircraft requires depot maintenance. The immediate and unexpected removal of the aircraft from flight status and addition to depot maintenance causes significant problems for scheduling both maintenance and flight operations. Air Force aircraft unavailability due to degradation maintenance generates 2.1 million hours of non-availability or 16 days for every aircraft every year [1]. There are scientifically reliable methods of determining degradation levels but the methods destroy the sample and thus are inapplicable for aircraft maintenance inspection. A non-destructive measurement technique to determine aircraft coating degradation levels is required. This non-destructive technique will need to be highly mobile and require little training so it may be easily and quickly incorporated into regular maintenance checklists.

1.2 Previous Work

The most recent work on developing a non-destructive technique was done by Captain Hans Korth at Air Force Institute of Technology. In the method Captain Korth

developed, aircraft paint samples were measured using the commercially available Agilent 4100 Exoscan Diffuse Reflectance Infrared Fourier Transform Spectroscopy system (DRIFTS) [2]. The DRIFTS system is handheld Fourier Transform Infrared Spectroscopy (FTIR) device. While less accurate than a normal bench-top device, and measures reflectance spectra rather than transmission spectra, this device is small enough to be carried by maintenance personnel. The classification technique developed by Captain Korth applied Single Value Decomposition (SVD) to the finger print region of the DRIFTS spectra and then classifies using Linear Discriminant Analysis (LDA) or multiple class linear discriminant analysis(MDA) to the first two or three SVD coefficients. Captain Korth's method was able to accurately classify the degradation levels of four artificially degraded paint samples for 98% of the measured spectra. Unfortunately when this same classification method was applied to a different sample set half the treated sample measurements were misclassified therefore more research is required.

1.3 Objective

Motivated and informed by the work Captain Korth has done, we can continue to build toward an employable technique for measuring coating degradation levels. From previous results it is expected that long-wave infrared reflectance measurement will perform the best in determining coating degradation and so multiple instruments will be used to validate this. Additional reflectance spectra measurements will be made with a variety of instruments including the Surface Optics Corporation SOC-100 Hemispherical Directional Reflectometer and Bomem Fourier Transform Spectrometer (FTS). The potential for hyperspectral imaging to determine a degradation image will be tested with the Telops Long-wave Hyperspectral Imaging. In addition the potential for emission and Raman spectral measurements for determining degradation

classification will be tested using the Bomem 154 FTS and Raman spectroscopy data provided to us by ChromoLogic LLC. The numerous measurement techniques will provide insight into the best measurement technique for determining aircraft coating degradation.

Modifications to the classification techniques will be tested along with determining the best measurement techniques for classifying aircraft degradation. Captain Korth identified several spectral fingerprint regions and selected the region which generated the highest classification accuracy with his technqiue. A more exhaustive search of the possible fingerprint regions will be conducted. In addition, different classification algorithms such as quadratic discriminant analysis (QDA) and decision trees (DT), will be tested to determine if more complex classification algorithms may be more appropriate for degradation classification in the SVD feature space.

To this point the measurement techniques, feature extraction and classification algorithms have been completed in a manner that provide limited knowledge regarding which spectral features actually contribute degradation information. These techniques require hyperspectral measurements of large portions of the infrared spectrum. Knowledge of which spectral features change could support future research in identifying the chemical bonds which degrade, which could improve future coating development. Further, this knowledge could contribute to a classification technique which only requires a mutli-spectral measurement instead of a hyperspectral measurement. The multi-spectral instrumentation is generally significantly less expensive than the hyperspectral equivalent. To this end feature selection will be performed on some of the measured spectra to determine which spectral features contribute information required for degradation and thus inform us on which molecular bonds may be degrading and which regions of the spectrum must be measured. The feature selection algorithms will include forward selection, Relief, and ReliefF. The feature selection

algorithms will be paired to LDA or MDA to determine the classification accuracy for the number of retained features. Actual measurement using multi-spectral instrumentation will be outside the scope of this study. The spectral features identified by feature selection and large bandwidth spectral data will then be used to simulate multi-spectral measurements for degradation classification.

II. Theory

2.1 Aircraft Coating and Polyurethane Degradation

The primary ingredient in the aircraft coatings of interest is polyurethane which has many degradation methods, such as exposure to visible and ultraviolet light, water, oxygen, and high temperatures. [3] [4]. It is anticipated aircraft coatings will degrade faster in warm humid climates than aircraft coatings in cold dry climates. Using this insight, the aging process can be artificially accelerated by baking a sample in an autoclave at a higher than normal temperature and humidity [5]. The accelerated aging allows these samples to be generated within days rather than months and with known levels of degradation. However it should be noted that the faster a sample is aged the less it represents a slowly aged sample [6]. In addition, degradation may be less uniform sample around the edges where the sample is cut and moisture is able to get between the coating and the substrate.

2.2 Fourier Transform Infrared Spectroscopy

Spectroscopy involves measuring the intensity of incoming radiation for a number of wavelength bands. Traditionally instruments will accomplish this by spatially separating the incoming radiation using a prism or diffraction grating onto an array of detectors or a single moving detector. The physical construction of the grating, prism and the relative size and location of the detector elements is related to the bandwidth each detector element observes.

Fourier Transform Infrared Spectroscopy measures the spectral response but does not separate the incoming radiation along a spatial dimension. Fourier Transform Infrared Spectroscopy instead uses a Michelson Interferometer to measure the interference pattern resulting from splitting the incoming radiation beam into two separate

paths and then varying one of the path lengths [8]. When the two beams recombine the electric fields will combine destructively or constructively depending on the optical path length difference. Since the input radiation is not separated along a spatial dimension for detection, only a single unmoving detector is required. Then Fourier Transforms can be used to determine the spectra of the signals. Two signals of two different wavelengths are shown are shown in Figure 1 with the combined signal.

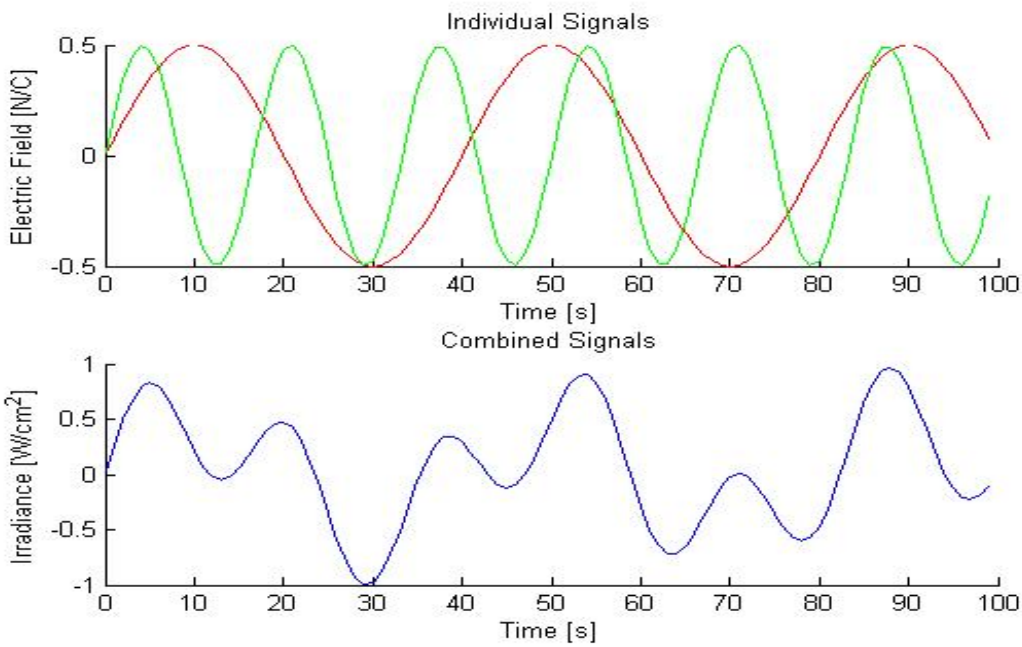

Figure 1. Top: The electric field for each individual signal vs time. Bottom: The measured irradiance for the combined signals shown on top. The two signals alternate between destructive and constructive interference in a repeating pattern.

The interference pattern shown in Figure 1 can then be transformed from a time space to a frequency domain. Therefore the Fourier transform is used, given by

$$E(f) = \int_{-\infty}^{\infty} E(t) e^{ift} \, \mathrm{d}t. \tag{1}$$

Here, $E(x)$ is the measured irradiance $E(t)$ is the transformed irradiance, t is time,

and f is frequency.

A true Fourier Transform requires a continuous sampling of irradiance and observation for infinite path length difference. Therefore the discrete Fourier transform

$$E_n(f) = \frac{1}{N} \sum_{j=1}^{N} E_j(s_j) e^{ift_j} \tag{2}$$

is used where N is the number of samples, $E_j(s)$ is the measured irradiance at time t_j and $E_n(f)$ is the transformed irradiance at frequency f. Applying the discrete Fourier Transform to the interferogram in Figure 1 results in the transformed spectrum shown in Figure 2. The interferogram is known to be the result of combining two spatial frequencies at 50 and 120 cm^{-1}. The transformed spectra displays signals near these two wavenumbers but each contains some bandwidth. This is due to the finite sampling distance. If the interferogram was sampled for infinite distance then the transformed spectrum would appear as two delta functions at 50 and 120 cm^{-1}.

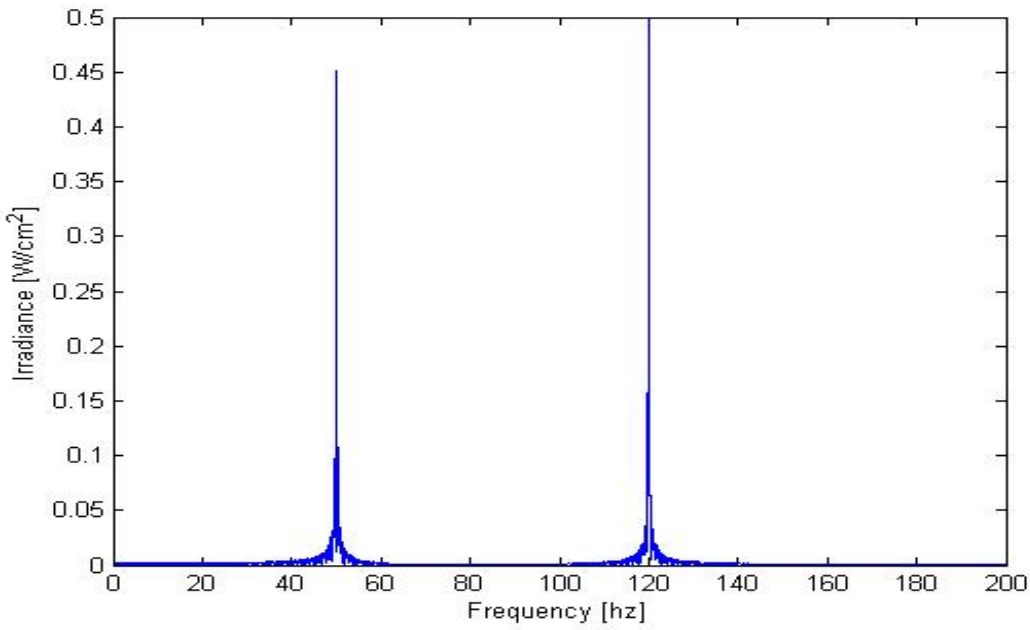

Figure 2. The Fourier Transform spectra for the pair of combined signals shown in figure 1. The measured spectra is known to be the combination of two signals at 50 and 120 cm^{-1}

When using the discrete Fourier Transform, the spectral resolution will increase as the maximum optical path length difference increases. The maximum measurable frequency is proportional to the distance between sample measurements. A problem known as aliasing can occur if a measurement is under sampled where measurements of higher frequency spectra may appear as lower frequency in the calculated spectra [9]. Depending on the instrument build, Fourier Transform Spectrometers(FTS) allow the experimenter some control over the spectral resolution and wavenumber range that prism and grating spectrometers do not. One disadvantage of a FTS over prism and grating spectrometers is that the interferogram requires time to collect. Therefore, a FTS may be inappropriate for scenes that are rapidly changing.

2.3 Hyperspectral Imaging

A commonly used imaging technique is a visual multi-spectral image. A visual multi-spectral image is generated by combining the measured intensity in multiple spectral bands. A hyperspectral image measures intensity for a large number of electromagnetic bands. This is normally accomplished by sending the signal through a vertical slit then separating the data onto an array of detectors. Spectral information being separated along the horizontal direction. The vertical slit then scans across the image to collect a three-dimensional data cube of two spatial dimensions and one spectral dimension. With this measurement method a rapidly changing scene will effect the image in the horizontal spatial dimension.

In a Fourier Transform hyperspectral imaging system such as the Telops an interferogram is produced on each pixel of the image plane. With this measurement method a rapidly changing scene will effect the image in the spectral dimension. In the case of the long-wave Telops hyperspectral camera, the measured voltages are calibrated using two onboard blackbodies.

2.4 Blackbody Radiators and Emittance

A blackbody is an ideal model where all incident radiation is absorbed. The blackbody model relates the emitted radiance of an object to its temperature

$$L_{BB}(\lambda, T) = \frac{2hc^2}{\lambda^5} \frac{1}{e^{hc/\lambda kT} - 1} \tag{3}$$

where L is the emitted radiance, λ is the wavelength, T is the temperature of the blackbody in Kelvin, h is Planck's constant, c is the speed of light, and k is Boltzman's constant [10]. It is also common to use blackbody radiance with wavenumber rather

than wavelength,

$$L_{BB}(\tilde{\nu}, T) = 8\pi hc\tilde{\nu}^3 \frac{1}{e^{hc\tilde{\nu}/kT} - 1} \tag{4}$$

where $\tilde{\nu}$ is frequency in wavenumbers and the remaining variables are the same as those for Equation 3.

Unfortunately, not every object emits as a blackbody. For those objects that do not, a wavelength dependent emittance $\epsilon(\lambda)$ term is added to blackbody model

$$L(\lambda, T) = \epsilon(\lambda) L_{BB}(\lambda, T). \tag{5}$$

In the event that $\epsilon(\lambda) = 1$ we have returned to the ideal blackbody model. For objects such that $\epsilon(\lambda) < 1$ but constant for all wavelengths the object is referred to as a graybody. The remainder of the objects are modelled with a combination of the blackbody and the emittance. The emittance of each object is dependent upon the material properties of the object itself. By measuring the emittance of an object information can be attained regarding the material composition of an unknown object.

2.5 Specular and Diffuse Reflectance

Ideal reflection models assume all radiation reflects off surfaces diffusely or specularly. Diffuse reflectance is the fractional quantity of incident radiation which is reflected to all non-specular angles. This is attributed to rough surfaces where incoming light is scattered in various directions. In an ideal diffuse reflector, the reflected radiation is evenly spread across all possible observations angles. This is also referred to as a Lambertian reflector. The specular reflectance is the fractional quantity of incident radiation which is reflected at the specular angle. The specular angle is the angle predicted by the law of reflection. A perfect specular reflector would be an ideal

mirror and is associated with no surface roughness in the reflector.

The most complex reflectance model is bidirectional reflectance distribution functions (BRDF) which describes how much radiation is reflected for each solid angle and each incident angle. An intermediate model between ideal reflectors and BRDF is to assume all reflectance is a combination of diffuse and specular reflectance. Separating these two types of reflectance provides a more accurate model than assuming the reflecting surface is an ideal specular or diffuse reflector. In addition the difficulty in measuring a BRDF and relative simplicity of measuring a sample's specular and diffuse reflectance provides sufficient motivation for using the intermediate reflectance model. The Surface Optics Corporation SOC-100 Hemispherical Directional Reflectometer instrument measures the specular and diffuse reflectance.

2.6 Radiometric Correction and Calibration

Photon detectors do not measure radiances directly. Each incident photon will raise the temperature for a thermal detector or generate an electron via the photoelectric effect for a photon detector [11]. The method which relates the temperature, voltage, or current to the incident radiation is referred to as calibration. Instruments such as the Agilent 4100 Exoscan Diffuse Reflectance Infrared Fourier Transform Spectroscopy system, or the Surface Optics Corporation-100 HDR Hemispherical Directional Reflectometer measure reflectance rather than radiance. These instruments often include a gold reference mirror, which has near 100% reflectance in the infrared. A steady illumination source is reflected off the gold mirror and the voltage signal is recorded representing 100% reflectance. Sample reflectance is calculated by comparing the measured signal reflected from the sample to the signal reflected from the gold mirror. These instruments are internally calibrated and therefore, report only the calibrated spectra to the experimenter.

For other instruments such as the Bomem 154-Fourier Transform Spectrometer (FTS) or the Telops Hyperspectral Camera, the radiometric correction is completed by the experimenter in a software suite such as Matlab. The radiometric correction method applied to the Bomem 154-FTS and Telops Hyperspectral Camera uses a two-point calibration, which assumes that the voltage signal of the detectors follows the response

$$V(\lambda) = G(\lambda) * L(\lambda) + O(\lambda) \tag{6}$$

where V is the voltage, G is the gain, and O represents an offset [12]. Two blackbodies of different temperature provide known radiances, assuming each blackbody covers the entire field of view of the detector. For the Telops, the two blackbodies are part of the instrument. For the Bomem 154-FTS, separate external blackbodies are used. Then two instances of Equation 6, where V and L are known can then be solved for the detector gain and offset then substituted back into Equation 6. Solving for the input radiance yields

$$L(\lambda) = \frac{[V(\lambda) - V_1(\lambda)] * [L_1(\lambda, T_1) - L_2(\lambda, T_2)]}{V_1(\lambda) - V_2(\lambda)} + L_1(\lambda, T_1) \tag{7}$$

where L_1 and L_2 are the theoretical blackbody radiances at temperatures T_1 and T_2 with V_1 and V_2 the corresponding measured voltage responses when measuring the blackbodies [13].

The calibration will continue to be accurate as long as the measurement conditions continue to match the conditions during calibration. Atmospheric absorption can change throughout the experiment, most notably due to changes in concentrations of CO_2, H_2O, or dust particulate in the air. Gases such as CO_2, H_2O have high absorptivity such that minute changes in atmospheric absorption bands, even from

people breathing in the lab, can degrade the accuracy of a calibration. Changes in path length such as the Telops onboard calibration may reduce calibration accuracy. In addition if measuring outside, changing temperatures, wind speed, and solar illumination angle can also affect calibration accuracy. Lastly the instrument and detector should reach stable temperatures before calibration. Calibrating an instrument at a cold temperature and then measuring as the instrument changes temperature may result in bad data as well. Due to the numerous environmental and experimental problems that can alter a calibrated measurement, calibrations should be taken often.

2.7 Raman Spectroscopy

The dominant scattering process for photons is Rayleigh scattering where an incoming photon is absorbed and re-emitted in a random direction. In Rayleigh scattering the scattered photon has the same energy as the incoming photon. In Raman scattering the scattered photon can gain or lose energy associated with the vibrational energy levels in the molecule [14]. This change in energy is referred to as the Raman shift. Stokes scattering is when a photon of lesser energy and a phonon are generated from an incident photon. Anti-Stokes Scattering is when outgoing photon is the energy of the incident photon and phonon. Stokes and Anti-Stokes scattering are shown in Figure 3.

Raman Spectroscopy takes advantage of the Raman Scattering effect to measure the vibrational energy levels of the molecule. A laser is shined onto the sample. Then the high intensity Rayleigh scattering beam is ignored. The remaining scattered photon energies are measured and a Raman Shift is calculated. The energy change of the photons can then be associated with the vibrational energy levels of the sample.

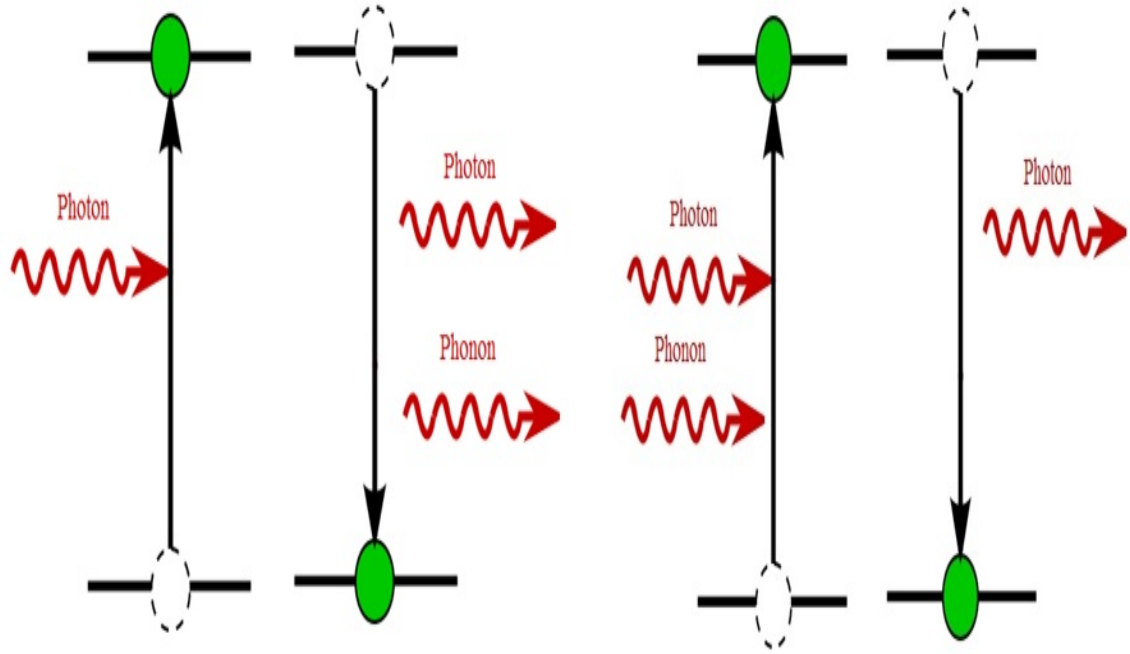

Figure 3. Left: Demonstrates Stokes Scattering where a portion of the incoming photon energy is left behind in the form of a phonon resulting in an scattered photon with decreased energy. Right: Demonstrates Anti-Stokes Scattering where the scattered photon has increased energy due to gaining energy from a phonon.

2.8 Classification

Classification is the method of determining which class or group a given set of observations belong to based upon a set of measured characteristics referred to as dimensions or features. In this project the measured dimensions are the individual measured spectral channels for the various instruments. There are two types of machine learning algorithms; unsupervised and supervised. In unsupervised machine learning the true class of the observations are unknown and the observations are separated using statistical analysis. In supervised machine learning there are two sets of data. The first is the training data where the actual classes of each observation is known. The training data is then used to generate the classification rules based

on statistical analysis of the class observations. Common or classic classification algorithms estimate a probability density function based on the variation of each data class and then the classification rules are developed based on the estimated probability density functions for the various classes. A second set of data referred to as the validation data is then used to estimate how well the classification rules will generalize to a larger data set. Validation is an important step to estimate if the classification rules developed on the training data, have been overtrained.

Overtraining.

When classifying data, overtraining or overfitting is when the training predicted classification error decreases with respect to classification algorithm complexity while the validation classification error increases with respect to classification algorithm complexity as shown in Figure 4 [18]. The classification rules themselves can increase in complexity in a variety of ways. Moving from a linear to quadratic discriminant function in a classification algorithm or increasing the number of used dimensions when generating classification rules will increase the classification complexity. When using feature extraction, such as with Single Value Decomposition, complexity can be increased by increasing the number of retained coefficients. Classification complexity will be increased using each of these methods throughout this project.

Overtraining is particulary problematic when data is limited. If the trained classification rules are not validated with an additional data set then determining if overtraining has occurred can be especially difficult. This can be observed in Figure 4 where the prediction error for the training data continues to steadily decrease with no apparent indication of when optimal training has occurred. However if the training data is split in half to generate a validation data set, then the estimation of the class probability density functions and resulting classification rules will be inherently less

accurate. Thus the easiest way to prevent overtraining is by acquiring more data for an accurate estimation of the probability density function, generating classification rules and testing the classification rules on the validation data. Unfortunately additional data may be impossible or expensive to collect, in these cases the techniques outlined in Section 2.8 will be used to estimate the probability of overtraining.

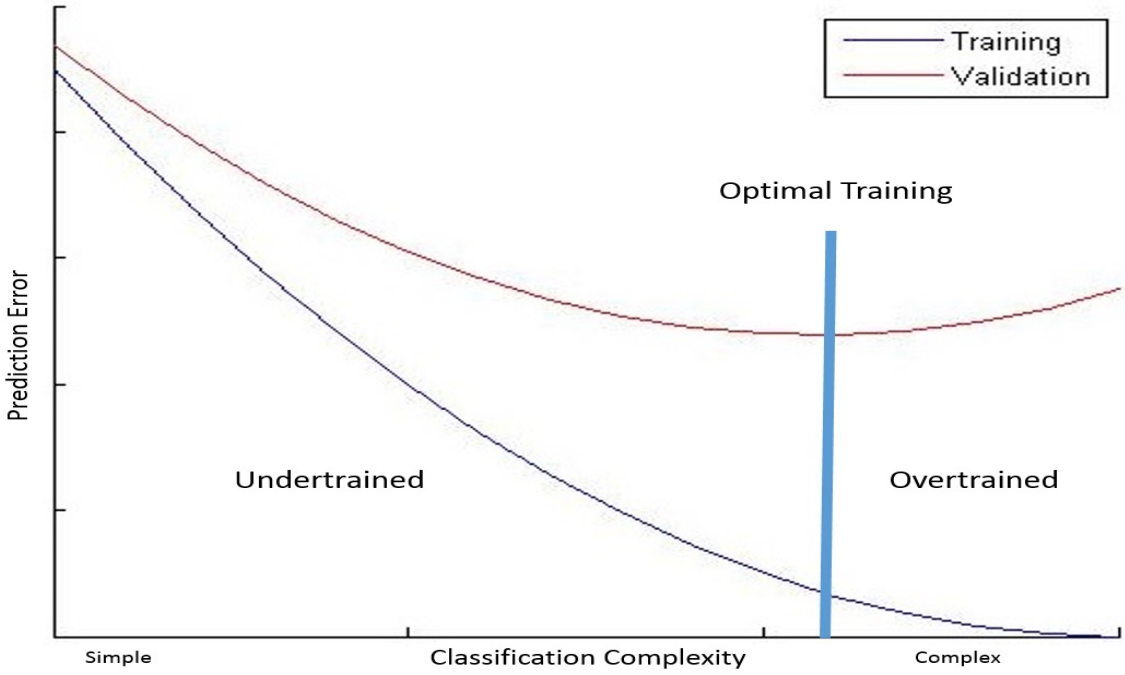

Figure 4. As the complexity of the classification algorithm increases, the prediction error of the training data normally decreases. However when testing on validation data, the accuracy will increase until overtraining is present. If only training data is available, predicting when the classification complexity has been overtrained can be difficult.

A two dimensional example will be assessed to help better understand how overtraining occurs. The truth probability density functions for two example classes and a small set of observations used for training data are shown in Figure 5. The blue shaded area representing a truth probability density function associated with that class' sample observations displayed as blue triangles. The red shaded area represent-

ing a truth probability density function associated with the class sample observations displayed as red triangles.

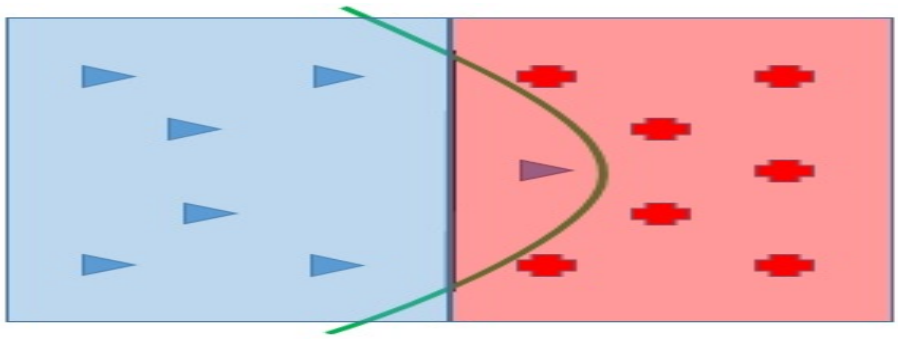

Figure 5. Two example class's probability functions, blue and red shaded areas, with a set of example training data, blue triangles and red crosses. The more complex quadratic discriminant curve classifies better on training data but will classify with reduced accuracy on a larger data set with the probability density functions shown.

When using a linear (simple) classification rule, the blue triangles and the red crosses can be classified correctly with the exception of one blue triangle which appears within the red cross probability density function due to noise or outlier effects. Due to this single misclassification, a more complicated non-linear decision boundary can be developed as shown with the green curve. Using the non-linear classification curve, the two classes can be perfectly classified for the training data. Unfortunately by adjusting the classification rule to a single outlier to increase training classification accuracy, the classification will be less accurate when applied to additional data. This additional misclassification is represented by the blue area to the right of the green curve and the red area to the left of the curve. Overtraining is most prevalent when small amounts of noisy training data are available. As more training data becomes available, a better estimate of the probability density function can be attained and by extension better classification rules.

Linear Discriminant Analysis.

Linear Discriminant Analysis (LDA) is a supervised classification method. In a two dimensional sense, a line is placed in the data plane onto which the data can be projected. This line is chosen to give the best discrimination possible between the various classes as shown in Figure 6. The line of separation can be represented as a single value when projected onto the dimension calculated for discriminant analysis [15]. When the original data is represented in more than two dimensions, a hyperplane is projected through the data hyper cube and used to discriminate the classes [16]. Despite being difficult to visualize, the process is mathematically identical regardless of the number of dimensions. When there are more than two classes it is referred to as multiple class linear discriminant analysis (MDA) and instead of a single discriminant function there will be a discriminant function for each class. More complex discriminant functions can be developed using the same reasoning such as quadratic discriminant analysis (QDA), however the more complicated the discriminant function the greater the chance of overtraining.

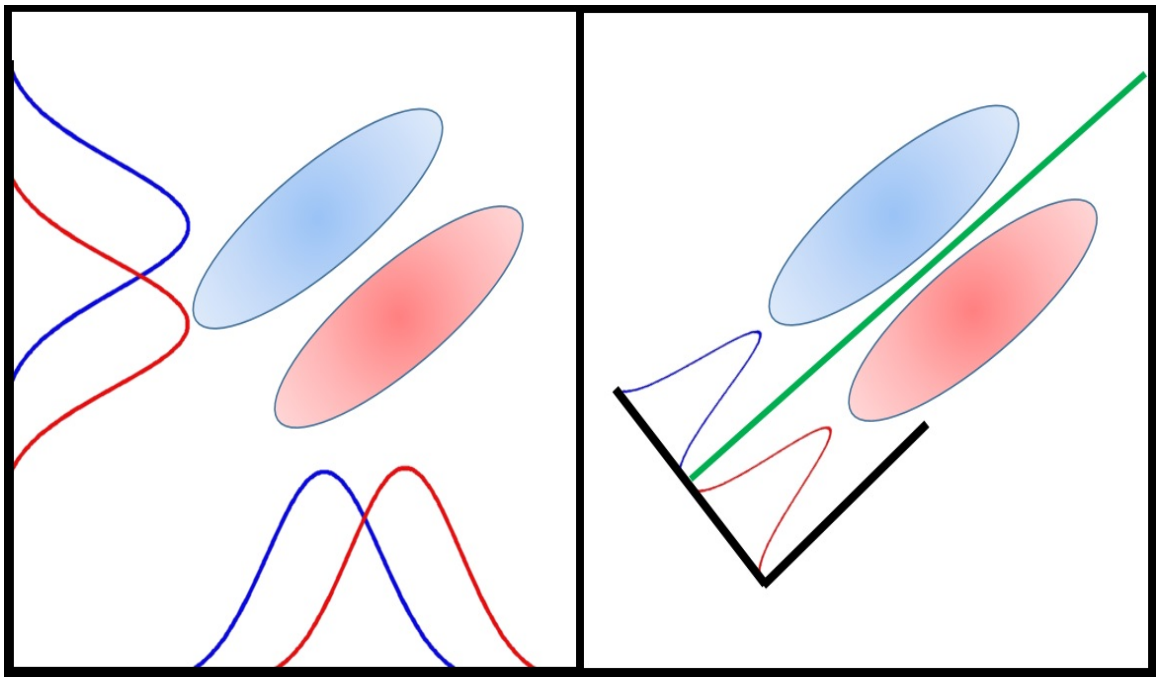

Figure 6. The probability density functions for two classes are shown in red and blue in this two dimensional feature space. The data representing the two classes can be projected onto a single dimension that may not be one of the original dimensions(left). This dimension is chosen to maximize the separation between the two classes. The discriminant function in the original feature space forms a line(right).

Decision Trees.

Decision Trees represent another supervised classification algorithm. Decision trees classify sample instances by repeatedly dividing the feature space into smaller and smaller subsections or branches, each subsection representing a different classification. Each division is chosen for maximum additional classification accuracy on each side of the additional division. The number of sections is increased until a pre-defined branch size, number of sections, or accuracy is achieved [17]. After the full decision tree is created, the decision tree is pruned by removing sections which do not increase accuracy above a pre-defined threshold or do not contain a large number of sample observations within the branch. Sufficient pruning increases the generalization

of the classifier by removing subsections that may have been over-fitted to outliers or noise in the data. The exact shape of the decision tree branches do not need to be pre-chosen. Each decision tree branch may take on a shape most appropriate for that class' estimated probability density function as shown in Figure 7. Since the decision tree branch shapes can adapt to each data set, decision trees have the capability to change their complexity based on characteristics in the data rather than be predefined by the user. This makes decision trees versatile and they generally report higher accuracy on training data than other classification algorithms such as LDA. However this versatility can pose problems due to easily overtraining the decision tree.

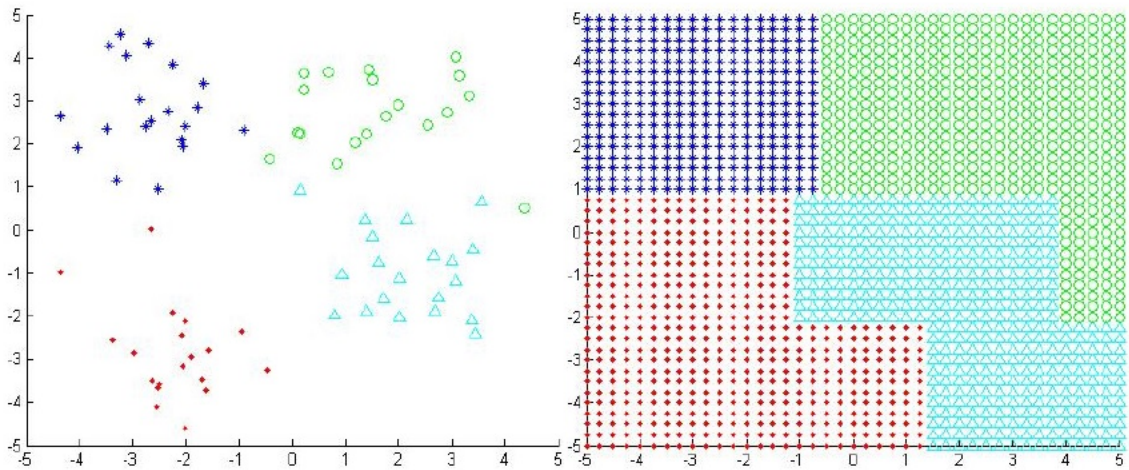

Figure 7. Left: Example training data for a decision tree including 4 classes with 20 observations per class. Right: The decision tree algorithm divides the feature space subsections or branches where each branch represents a different class. Complex shapes can be generated in the feature space that would not be possible when using LDA.

K-Fold Cross Validation.

Cross validation is used to estimate how well the classification rules, trained on a relatively small data set, will generalize to a new data set. In K-fold cross validation, the training data is divided randomly into k equal subsets or 'folds'. Classification rules are generated using $k-1$ of the folds and then validated or tested using the

remaining fold. This process is then repeated k times; normally around 10 [19]. Depending how the classifier is trained, the classifier accuracy can continue to improve while the k-fold cross-validation accuracy remains the same. This is a useful indication that the classification algorithm has been overtrained because it requires only the training data set and not an additional validation data set to determine when the classification algorithms have been overtrained.

While k-fold cross-validation is a useful indicator of overtraining, the random method of choosing the data associated with each fold produces variance in the reported k-fold classification accuracy across multiple runs. When using k-fold it is assumed that the variation in the training data set is representative of future experimental sets. Like most classification algorithm problems, additional data can improve the expected result and decrease the variance in reported k-fold classification accuracy across multiple runs[20]. As more data is attained, the predicted classification accuracy and k-fold cross-validated accuracy will converge. Cross validation provides a conservative estimate of the expected field performance of a given classification algorithm rather than relying only on the predicted classification accuracy, while still allowing a larger training data set to generate better probability density function estimates and better classification rules.

2.9 Feature Selection

Feature selection refers to identifying the data dimensions which contain classification information or information that separates the data into distinct classes. Classifying with the properly selected dimensions will generate higher classification accuracy with less data. Ideally it would be possible to test every combination of dimensions to determine the highest performing combinations. This is referred to as an exhaustive search. For example the DRIFTS data associated with this project

contains 1700 spectral bins resulting in more than 10^{300} possible combinations of spectral features. Such large numbers of possible combinations prevent exhaustively searching. Most feature selection algorithms rely on non-exhaustive search methods where the number of tested combinations is significantly reduced. When utilizing a non-exhaustive search method, it is possible to miss the optimal combination of dimensions. For this reason non-exhaustive feature selection algorithms have been developed to minimize the chance of missing the optimal feature combinations..

Forward Selection.

Forward selection is a non-exhaustive feature selection method which, when applied to spectral data relevant in this project, identifies the spectral dimensions containing information pertinent to classifying degradation levels. In forward selection, the classification algorithm is applied to each dimension/spectral feature separately. For this research Linear Discriminant Analysis will be used as the classification algorithm. Then the spectral feature which generates the highest classification accuracy is retained for future use. The next dimension is identified by combining the previously retained dimensions with an additional test dimension and applying the classification algorithm of choice to the combination. Then the test dimension which, when combined with the previously retained dimensions, generates the greatest increase in classification accuracy is added to the list of retained dimensions. This process is repeated until a predefined stopping criterion, such as accuracy or number of dimensions, is achieved [16]. The selected features are tied to the classification algorithm such that the forward selected features for one classification algorithm may not be the same selected features when applying a different classification algorithm. In addition this feature selection method requires application of the classification method once for each non-retained spectral dimension for every selected feature identified.

Depending on the amount of data, number of spectral dimensions and complexity of classification method, even this non-exhaustive feature selection technique may not be appropriate for the amount of processing power available.

Relief.

When processing power is an issue Relief provides a feature selection method which identifies spectral features important to classification without numerous applications of the classification method. Relief achieves this by calculating a weight to each dimension and ranking each dimension by its associated weight[22]. The dimension weight is calculated by choosing a number m of random sample instances and applying the weight formula,

$$W_{n+1}(X) = W_n(X) - diff(X, R, H) + diff(X, R, M), \tag{8}$$

to each feature dimension and each random sample instance. Here X is one of the dimensions, R is the randomly chosen element, H is the nearest hit (member of the same class), and M is the nearest miss (member of the a different class). The difference function shown is

$$diff(X, R, H) = \frac{dist(R, H)}{m * [Max(X) - Min(X)]}, \tag{9}$$

where $dist(R, H)$ is the one-dimensional Euclidean distance in dimension X. $Max(X)$ and $Min(X)$ are the maximum and minimum measured values in the dimension X. The weight of each dimension is initialized to zero and is updated for each random sample instance.

The weight will increase if the distance to the nearest member of the same class is small and the distance to the nearest member of a different class is large. This

should result in dimensions where there is a large variation between the classes and small variation within the classes being weighted the highest. Then a user specified number of dimensions with the largest weights are included for classification. If a certain accuracy is required the classification algorithm is applied to the highest weighted dimension and then additional high weighted dimensions are included until the predefined accuracy is achieved. Since the testing of an additional dimension requires only one application of the classification method, identifying spectral features with Relief can be significantly faster than selecting features with forward selection for data with large number of dimensions.

ReliefF.

ReliefF is an improvement but extremely similar to its predecessor Relief. Depending on which software suite is being used Relief can be identified when in actuality the ReliefF algorithm is applied. ReliefF's selected features will be affected by the data distribution in the same manner as Relief [23]. The random selection of sample instances in Relief provides the possibility that multiple applications of the Relief algorithm may identify different features. One method to address this is to increase the number of randomly chosen sample instances. Another method which is used in the ReliefF algorithm is to choose a user defined number of nearest hits and nearest misses, usually around 10, when calculating the dimension weights. The contribution of multiple near hits and near misses are averaged,

$$W_{n+1}(X) = W_n(X) - \sum_{i=1}^{k} \frac{diff(X, R, H) + diff(X, R, M)}{k}, \tag{10}$$

where k is the user defined number of nearest hits and misses, and the $diff()$ function is the same $diff()$ function shown in Equation 9. The use of multiple nearest neighbors mitigates the effect of noise and outliers that can adversely affect the out-

come of the Relief algorithm. In addition, increasing the number of nearest hits and misses used to calculate the dimension weights requires less processing time for the same performance when compared to just increasing the number of random sample instances in the Relief algorithm.

2.10 Feature Extraction: Single Value Decomposition

Single Value Decomposition(SVD) accomplishes data reduction by projecting the data into a different feature space[21]. The dimensions of this feature space are ordered according to the amount of variation they represent within the data. This allows the dimensions which represent negligible variation within the feature space to be removed without sacrificing information relevant to classification. Ideally the first few SVD dimensions will capture all the variation within the data and the remaining dimensions will contain only noise information. This reduces the data to be stored and transmitted while decreasing subsequent processing time. SVD's data reduction is particulary helpful when dealing with large data sets such as a hyper-spectral data cube. In addition the new feature space may allow for more accurate classification than in the original feature space.

SVD accomplishes this data reduction by identifying the eigenvectors and relative amount of each eigenvector which best represents the data[2]. First the matrix representing the data is decomposed into three different matrices,

$$\mathbf{D{=}UEV{*}}, \tag{11}$$

where D is the mean subtracted matrix representing the data (mean of each dimension subtracted from each observation), U is a m x m unitary matrix ($U * U = I$), E is a m x n diagonal matrix, and V^* is a n x n unitary matrix. For this work, m is the

number of sample observations and n is the number of spectral channels. The relative amounts of each eigenvector in the data are represented by the SVD coefficients,

$$\mathbf{C}=\mathbf{DV}. \tag{12}$$

Then the largest coefficients can be retained while the remaining coefficients are discarded. The number of coefficients retained is user or application specific normally either a predefined number of coefficients or the number of coefficients required to represent a user defined amount of variation in the data. Previous research in this topic used 2-3 SVD coefficients, the possibility of using more than 3 coefficients for degradation classification will be explored as well as determining the number of SVD coefficients required to attain a given classification accuracy.

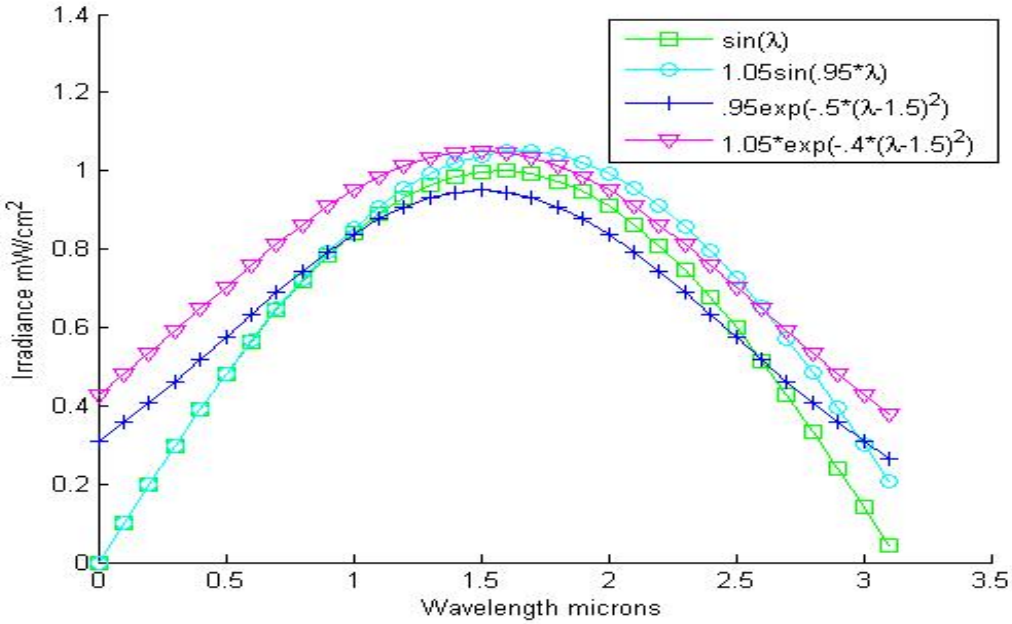

Figure 8. Four visually similar but mathematically distinct potential spectral responses. These test spectra are difficult to distinguish and classify in original feature space.

As an example four visually similar functions, representative of possible spectral

measurements, are shown in Figure 8. These functions are difficult to distinguish both with the naked eye and many existing classification algorithms. However after applying SVD to these four spectra the data can be represented by the two most important SVD coefficients as shown in Figure 9. Any number of coefficients can be retained, but retaining two coefficients is more practical for visualization purposes. As shown, the SVD coefficients representative of the four possible spectra are now very distinct from each other. By representing the data in this new feature space, classification can be made more accurate and faster than classification in the original feature space, as well as easily visualized.

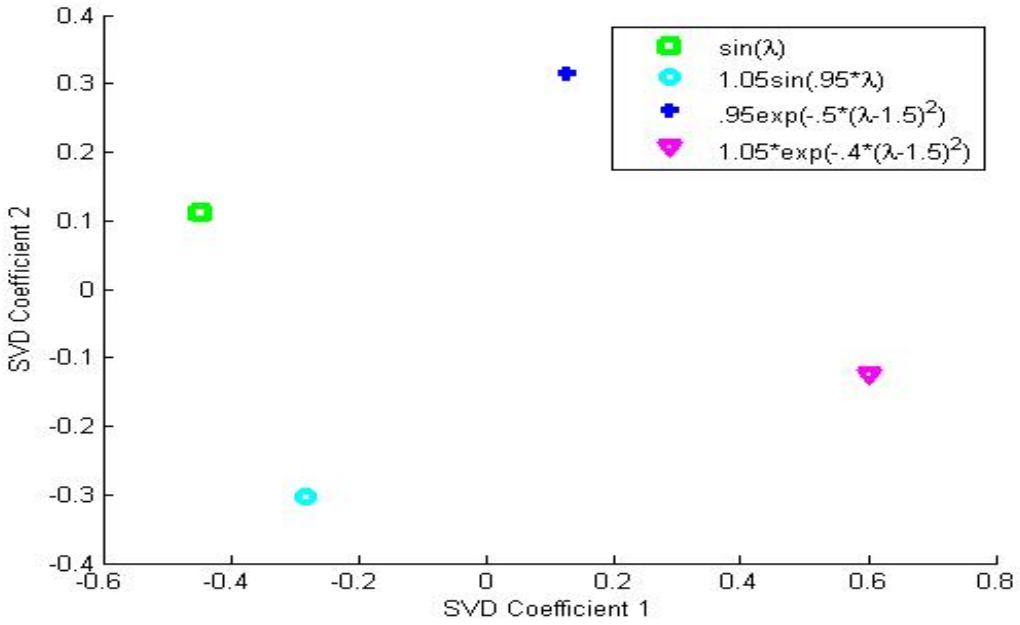

Figure 9. The two most significant SVD coefficients for the four possible spectral responses shown in Figure 8. The four test spectra are now easily classified in the new feature space.

In addition if there is random noise in the incoming signal the four example spectra become even more difficult to distinguish as shown in Figure 10. Random Gaussian noise was added to 10 instances of sample signals to generate 10 signals with signal

to noise ratio of 100. As can been seen, many of the methods that could discriminate between the four perfect signals shown in Figure 8 may be of little use for noisy signals.

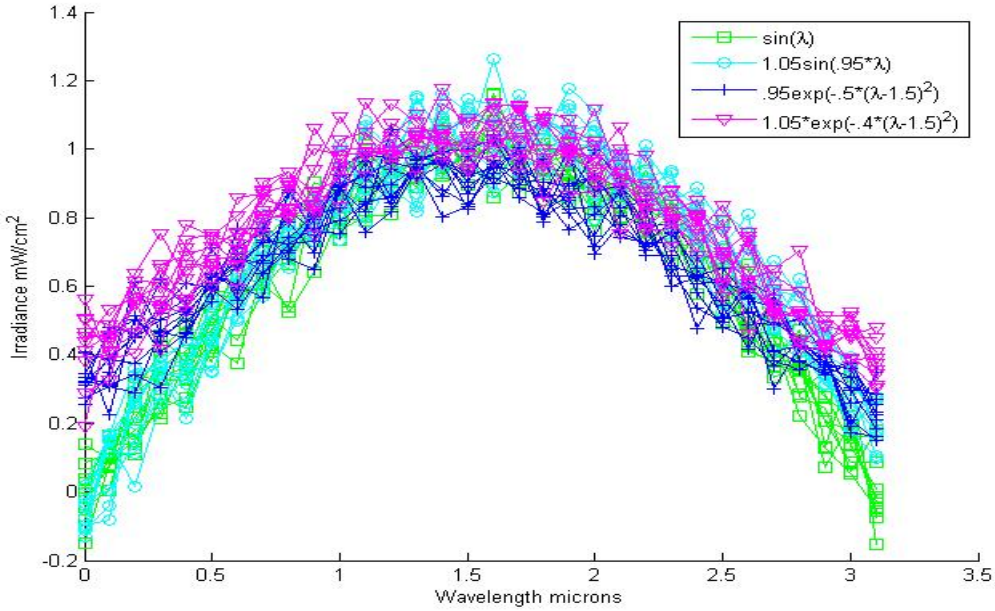

Figure 10. White Gaussian noise added to multiple instances of the original four example spectra. The added noise limits the effectiveness of many techniques used to classify the four example spectra in the original feature space.

Fortunately the majority of the signal variation caused by noise is accounted for in the later SVD coefficients leaving the first two coefficients still largely representative of the actual signal. When examining the signal in the SVD coefficient space, the four signals remain easily distinguishable as shown in Figure 11 despite the addition of significant noise.

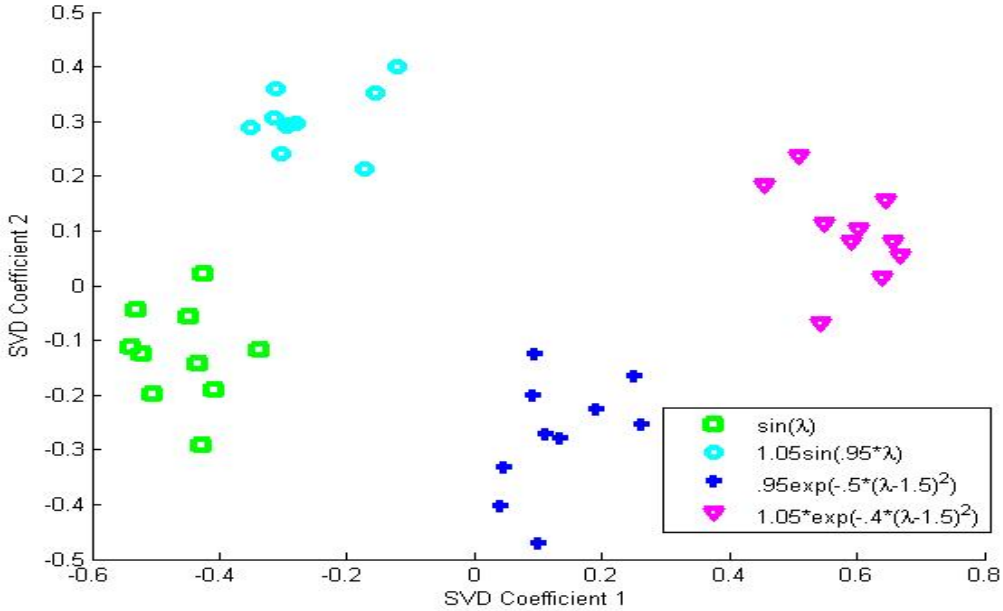

Figure 11. The addition of noise does effect the location of the first two SVD coefficients in the modified feature space. However the SVD coefficients for the noisy example spectra can still be easily classified with basic techniques in the SVD coefficient feature space.

III. Experimental Methods

3.1 Sample Preparation

Four sample sets were generated for determining the best method of degradation classification. Each sample set was generated to be representative of coatings used on Air Force aircraft. A list of the samples for each sample set is shown in Table 1.

Table 1. All Samples Generated for Degradation Classification

Sample Set A	**Sample Set B**	**Sample Set C**	**Sample Set D**
Autoclave 95 C	Autoclave 101 C	Weathered	Autoclave 101C
2 Each	1 Each	1 Each	3 Each
REC/REC	***TC/REC***	***AF Dark Paint***	***Deft TC***
0 hrs, 144hrs	0,24,48,96 hrs	0,2100,2600 hrs	0,24,48,96 hrs
		AF Light Paint	***Grey TC***
		0, 3000 hrs	0,24,48,96 hrs
			Black REC (BREC)
			0,24,48,96 hrs
			DEFT TC, BREC
			0,12,24,48,72,96 hrs
			Grey TC,BREC
			0,12,24,48,72,96 hrs

Sample set A included four samples of multiple combined rain erosion coats (REC). Two of the samples were left in pristine condition while the remaining two samples were cured in an autoclave at 95 °C with 130% relative humidity for 144 hours. This sample set was used to determine if the classification rules generated for the training set could generalize too samples not in the training set. This was validated by classifying the remaining pristine and degraded samples.

Sample set B included four samples of top coat (TC) and REC combinations with smaller variations in the amount of degradation. One sample was kept pristine, while the remaining samples were cured in an autoclave at 101 °C with 130% relative humidity for 24, 48, and 96 hours. Sample set B was used to determine if classification

could still be achieved with samples which had smaller differences in degradation levels.

Sample set C was generated using an artificial weathering process where the samples were exposed to cycles of 90 minutes simulated solar radiation and 30 minutes of simulated precipitation. One sample of low reflectance Air Force paint was left pristine while two additional samples were artificially weathered for 2100 and 2600 hours. An additional two samples of a high reflectance Air Force Paint were also generated as part of this sample set. This pair of samples degraded significantly slower due to the higher reflectance of these samples. One was left pristine while the second was artificially weathered for 3000 hours.

Sample set D is the most extensive sample set and includes 81 samples of different variations of deft topcoat, grey topcoat, black rain erosion coat on aluminum and silicon substrates. Then a subset of each combination was cured in an autoclave at 101 °C for 0, 12, 24, 48, 72, and 96 hours. Unfortunately, due to delays outside of our control, only the pristine samples from sample set D arrived in time to be included in this report.

3.2 Diffuse Reflectance Infrared Fourier Transform Spectroscopy(DRIFTS)

Every sample set was measured using an Agilent 4100 Exoscan Diffuse Reflectance Infrared Fourier Transform Spectroscopy System (DRIFTS). Sample set A was measured by Capt Hans Korth, in the two days following sample preparation, with 36 measurements per sample. For sample set A, the measurements were taken with the sample on a table and the Agilent facing down on the sample. This unfortunately left marks on the samples. Later measurements included procedures to prevent the Agilent from pressing down on the sample.

For sample sets B and C, the Aiglent was laid on its side with the samples held

vertically to the detector. Sample set B was measured by Capt Hans Korth over a six week period with 75 measurements per sample. The average spectra for each degradation class are shown in Figure 12. Sample set C was measured by First Lieutenant Kody Wilson with 10 measurements per sample. For sample set C, half the measurements were made with the detector laying on its right side. The detector laid on its left side for the second half of the measurements. For sample set C, the detector was re-calibrated between every set of five sample observations. Five spots on each rectangular sample, the four corners and the center of the sample, were measured in a repeating pattern. For each observation the sample was normal to the detector normal axis and the sample filled the full field of view of the detector.

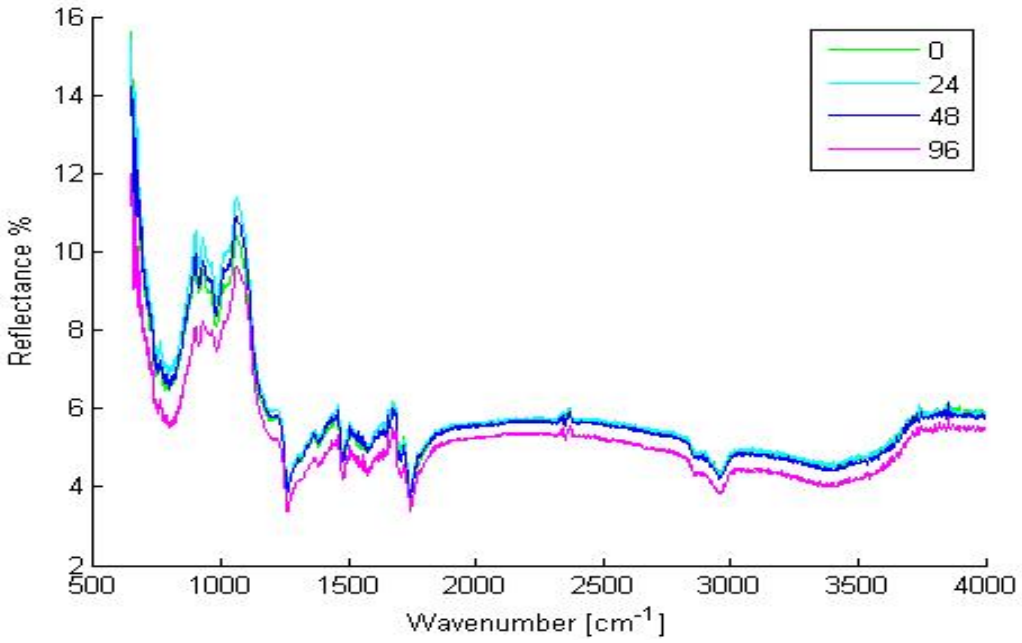

Figure 12. Average DRIFTS reflectance spectrum for each degradation class in sample set B. The highly treated sample appears clearly distinct but the remaining three are difficult to distinguish

For some of the pristine samples in set D, the detector was mounted in a stand with the detector facing upward. Sample Set D was measured by First Lieutenant

Kody Wilson and Mr. Billy Kelley. A flat sample mount fit over the detector with each sample laying horizontally over the detector. When measured by Kody Wilson a repeating pattern of five measurement locations on each circular sample formed an X. When measured by Mr. Billey Kelley, the center of each sample was measured once and and then all sample measurements were repeated for four iterations. It should be noted that when measuring sample set D, the sample was not perfectly horizontal across the detector when measuring every location. This is due to a slight protrusion of the detector above the surface of the sample mount. When measuring near the edge of the sample, a small angle forms when part of the sample rests on the sample mount and another part rests on the protrusion of the detector. When measuring near the center of the sample, all resting parts of the sample rest on the detector protrusion and appear to lie perfectly horizontal to the naked eye. One sample was measured at varying radii from the center to determine if increasing this angle would effect measurement. No apparent correlation between measured spectra and measurement radii was immediately noticeable.

3.3 Surface Optics Corporation SOC-100 Hemispherical Directional Reflectometer

The Surface Optics Corporation SOC-100 Hemispherical Directional Reflectometer illuminates the sample using a gold hemi-ellipsoid mirror with the blackbody source at one focus and a horizontal sample at the second focus. The emitted light from the blackbody reflects off the hemi-ellipsoid mirror and hemispherically illuminates the sample. Then a mirror reflects light into an Nicolet FTIR to take spectral measurements at multiple angles with respect to the vertical as shown in Figure 13. This measures hemispherically illuminated directional reflectance (HDR) which is radiometrically equivalent to directionally illuminated hemispherical reflectance (DHR).

The diffuse illuminated directional reflectance (DDR) is measured by placing a black stopper at the specular angle so the sample is only diffusely illuminated. The specular directional reflectance (SDR) is then calculated by subtracting the DDR from the HDR [24]. In addition, different polarizations states can be measured using a rotatable and removable polarizer in the detector.

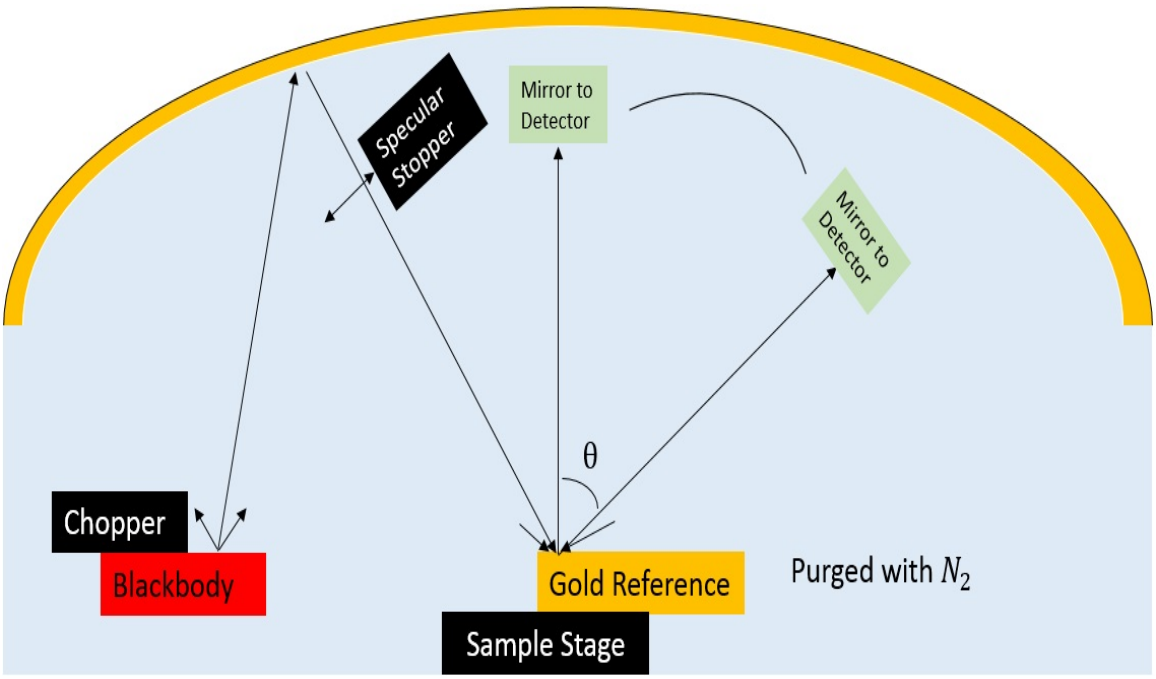

Figure 13. Side view of the SOC-100. The blackbody is placed at one focus. The gold reference and second sample are sequentially moved into position at the second focus. Detector rotates about the sample depending on desired measurement angle.

The SOC-100 is automatically calibrated with a gold reference mirror. The gold reference mirror is assumed to have a reflectance of 100%. The sample and reference mirror are measured with a black chopper preventing the blackbody from illuminating the sample and reference. This provides an estimate of self-emission and path radiance for the gold reference and the sample. The reflectance is then calculated by

$$R(\lambda) = \frac{L(\lambda)_{SampleIlluminated} - L(\lambda)_{SampleNot-Illuminated}}{L(\lambda)_{ReferernceIlluminated} - L(\lambda)_{ReferenceNot-Illuminated}}, \tag{13}$$

where $L(\lambda)_{Sample-Illuminated}$ is the radiance measured when the sample is illuminated by the blackbody, $L(\lambda)_{SampleNot-Illuminated}$ is the radiance measured when the sample is not illuminated, $L(\lambda)_{Reference-Illuminated}$ is the radiance measured when the gold reference mirror is illuminated and $L(\lambda)_{ReferenceNot-Illuminated}$ is the radiance measured when the gold reference mirror is not illuminated. The detector is not calibrated to measure radiance, but the digital units are assumed to be linear with the incident radiances. Each reflectance measurements is the average of eight sample and gold reference scans. Each spectrum is the average of four reflectance measurements. Further, the test chamber is sealed and purged with gaseous nitrogen to remove absorption from water and CO_2. The HDR, DDR, and SDR for each sample from sample set C were measured at seven angles, vertical, horizontal and unpolarized states.

3.4 Stand-off Measurement Bomen FTS Reflectance

The method outlined in section 2.6 used to calibrate the Bomem MR-154 Series Fourier Transform Spectrometer (FTS) was tested by measuring one blackbody at three different temperatures; 50 °C, 60 °C, and 70 °C. Taking care that each blackbody source covered the entire field of view of the detector. Using the 50 °C and 70 °C temperatures as calibration sources, the calibrated spectra for 60 °C could be compared to a theoretical prediction for a blackbody of that temperature. This provided an estimate of the accuracy of the two-point calibration used across the measured spectra.

Following the testing of the calibration scheme, the Bomem MR-154 FTS was used to acquire the reflectance spectra for the three dark paint samples in sample set C. The experimental setup for acquiring the reflectance spectra of the three samples is shown in Figure 14. Unlike previous measurements where the detector was required to be in contact with the sample or even enclose sample, the Bomem MR-154 FTS

allowed for sample measurement with a standoff distance between the detector and the sample. Measurements were taken of two known blackbody references 140 cm from the detector, one 'hot' blackbody at 150 °C and one 'cold' blackbody at 50 °C used to calibrate the instrument. The blackbodies were measured just prior to measuring the samples' reflectance spectra and immediately after measuring the samples. This provided an estimate of possible atmospheric changes and changes in instrument response during the several repeat measurements of a single sample, which required approximately 20-30 minutes. Each measurement is the average of 256 interferograms. Care was taken so that the reference blackbody and the source blackbodies had the same path length to the detector to prevent overestimating or underestimating the amount of atmospheric absorption in the measured spectra.

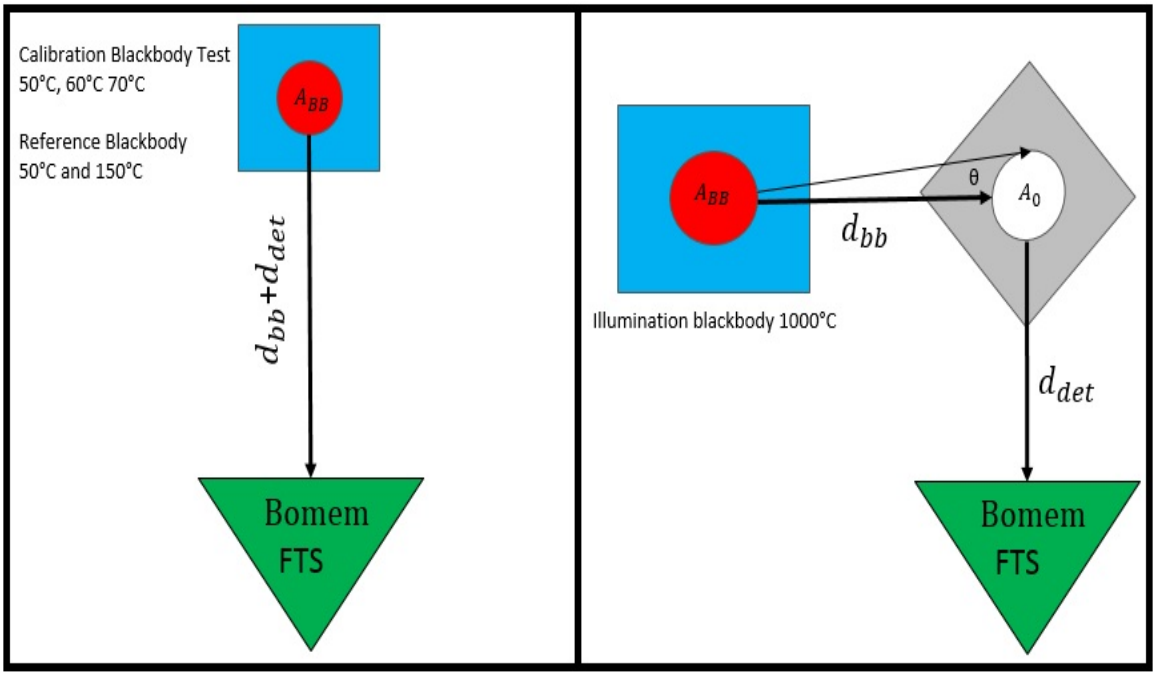

Figure 14. (Left)The experimental positions of blackbodies for calibration test and experiment calibration. (Right) The experimental position for illumination blackbody. For this experiment the path length from each blackbody source to the detector were equal to ensure accurate atmospheric correction.

A 1000 °C source blackbody was placed approximately 40 cm from the samples. Three sample illumination angles were measured with a rotating stage: 30°, 45°, and 60° with respect to the sample surface normal. The detector was placed at three measurement angles; 60°, 45° and 30° with respect to the sample surface normal. The detector used a telescope to focus on the sample approximately 1m away. The illuminated measurements include reflected radiation from the source, reflected background radiation, self-emission and path radiance and is represented by

$$L_{Sample} = (L_{ReflectedSource} + L_{ReflectedBack} + L_{Emission})\tau_{d_{tel}}(\lambda) + L_{PathRadiance}. \quad (14)$$

To remove radiation not reflected from the blackbody source, each sample was measured without source illumination to determine the amount of reflected background radiation, self-emission and path radiance and is represented by

$$L_{Back} = (L_{ReflectedBack} + L_{Emission})\tau_{d_{tel}}(\lambda) + L_{PathRadiance}. \quad (15)$$

Subtracting the measured background radiance from the measured sample radiance, the reflected radiance can be determined. The reflectance at a given angle of measurement can be represented by a Bi-Directional Reflectance Distribution Function (BRDF) $f(\theta_i, \phi_i, \theta_r, \phi_r, \lambda)$. In this experiment, the blackbody source, sample and detector were in plane with each other making $\phi_i = 0$ and $\phi_r = \pi$. Knowing this and assuming the blackbody is a Lambertion emitter, the measured radiance can be related to the source blackbody temperature, sample BRDF and experimental dimensions. Then reflectance at the specular angle is assumed to be represented by the specular reflectance ρ_S while the non-specular measurements are assumed to be

represented by the diffuse reflectance ρ_d. These changes yield.

$$L_{Sample} - L_{Back} = \rho_d(\lambda) L_{BB}(\lambda, T_S) \tau_{d_{bb}}(\lambda) \tau_{d_{tel}}(\lambda) \frac{A_{BB} cos(\theta)}{\pi d_{BB}^2} \tag{16}$$

for the diffuse reflectance, where the experimental dimensions are shown in Figure 14 and $\tau_{d_{bb}}(\lambda)$ and $\tau_{d_{tel}}(\lambda)$ are the atmospheric transmittance for each path. Similarly for the specular case

$$L_{Sample} - L_{Back} = \rho_S(\lambda) L_{BB}(\lambda, T_S) \tau_{d_{bb}}(\lambda) \tau_{d_{tel}}(\lambda). \tag{17}$$

Area, field of view and length are constant with respect to wavelength. Assuming the experiment is not changed between measurements the a normalized reflectance factor can be calculated which is equal to reflectance times a constant. For the specular case in Equation 17 it is assumed that the room temperature radiance significantly smaller than the blackbody radiance and can be ignored. Then solving for the specular reflectance factor yields

$$\rho_S C_n' = f(45°, 0, 45°, \pi, \lambda) C_n'' = \frac{L_{Sample} - L_{Back}}{L_{BB}(\lambda, T_S)}, \tag{18}$$

where the C_n are constants that account for the constant geometric factors. Similarly the diffuse reflectance factor,

$$\rho_d \frac{C_n}{\pi} = f(60°, 0, 30°, \pi, \lambda) C_n = f(30°, 0, 60°, \pi, \lambda) C_n = \frac{L_{Sample} - L_{Back}}{L_{BB}(\lambda, T_S)}. \tag{19}$$

The normalized specular and diffuse reflectance factors are independent of experimental setup and more advantageous for possible fielded use where the exact dimensions of the experiment may change often or be uncontrollable.

3.5 Telops Hyperspectral Reflectance Imaging

The DRIFTS and Bomem 154 FTS are restricted to measuring a single point per measurement. It would be more advantageous to image degradation across the aircraft using an imaging technique. Hyperspectral imaging provides the necessary spectral information for degradation classification and allowed for imaging the large portions of the aircraft during one measurement.

As will be discussed below, results of the previous experiments showed infrared reflectance measurement resulted in higher classification accuracy. For these measurements a 6-inch wide-area blackbody at 250°C or 300°C was used to illuminate the dark Air Force paint samples. A specular reflectance measurement was taken with the wide area blackbody at roughly 15° to the sample surface normal and the telops at the opposing 15° to the sample surface normal. A diffuse reflectance measurement was taken with the wide area blackbody at roughly 45° to the sample surface normal and the Telops at 0° to the sample surface normal. The experimental setups for the diffuse and specular reflectance measurements are shown in Figure 15, where each sample was measured in each numbered sample location.

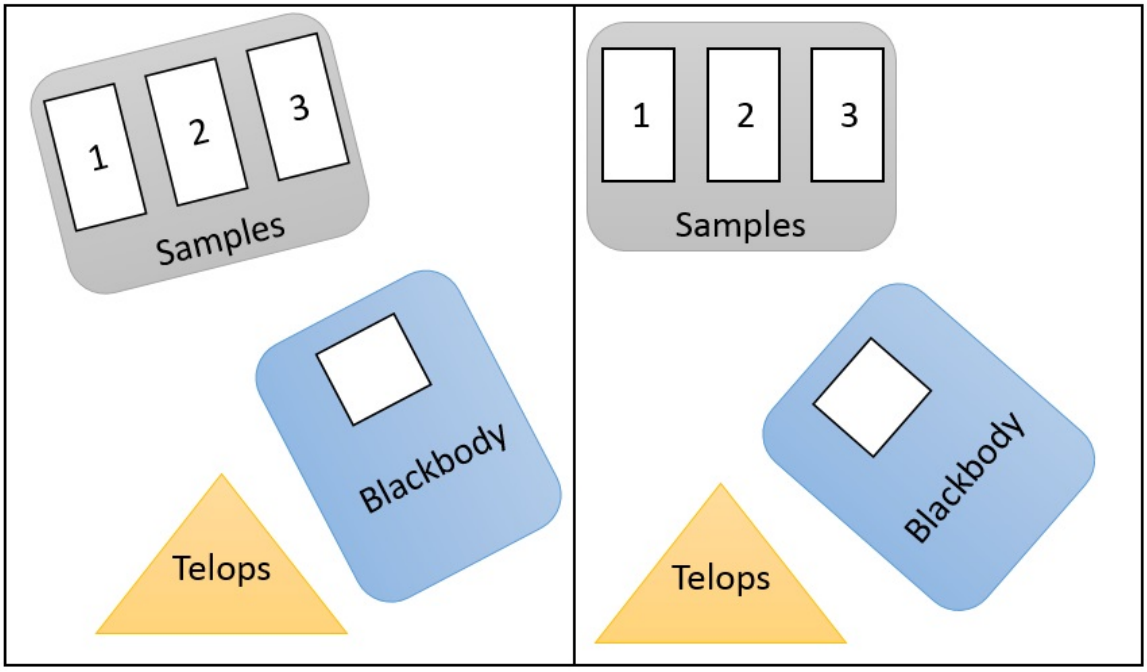

Figure 15. Left: The samples and the blackbody are oriented so the Telops measures the specularly reflected blackbody illumination. Right: The samples and blackbody are oriented so the Telops measures the diffusely reflected blackbody illumination.

Each Telops observation was calibrated using two onboard blackbodies at 25°C and 30 − 45°C. Two measurements are required to calculate a normalized reflectance data cube. One Telops observation measured the sample's self emittance and reflected room radiance. A second Telops observation measured the sample's self emittance, reflected room radiance and the reflected wide area blackbody radiance. The normalized reflectance factor data cube is then calculated based on a pixel by pixel application of

$$\rho_S C_n = \frac{L_{Sample} - L_{Back}}{L_{BB}(\lambda, T_S)\tau(\lambda)} C_n. \tag{20}$$

where $\rho_S C_n$ is the normalized reflectance factor, L_{Sample} is the measured radiance when the sample is illuminated, L_{Back} is the measured radiance when the blackbody

is not illuminating the sample, $L_{BB}(\lambda, T_S)$ is the theoretical blackbody radiance, C_n is the normalization constant and $\tau(\lambda)$ is the atmospheric transmission which will be assumed to be one. Additional assumptions are each pixel of both data cubes represents the same point on the samples, the sample temperatures have not changed, and the atmospheric conditions have not changed appreciably. For this experiment, the samples and illumination blackbody were stationary on an optical table with the Telops on a tripod in a lab. The temperature near the samples only varied by ±1°C throughout the experiment; thus the assumptions can be warranted.

3.6 Stand-off Measurement Bomen FTS Emittance

The samples in sample set C have an average reflectance of 0.18 leading to low reflected radiance and requiring high temperature blackbodies for the measured signal to be significantly above the background radiation. According to Kirchhoff's law of thermal radiation, a bad reflector should also be a good emitter [25]. These samples are expected to have an average emittance of 0.82. With the low reflectance and high emittance, the self-emission of a heated sample is expected to dominate the observed signal in a room temperature environment.

For this experiment, the three degraded samples of dark Air Force paints were set 1m away from the detector and heated using a calibrated hot plate to 30°C, 40°C and 50°C. Throughout the experiment, the hotplate displayed the set temperature ±1°C. The Bomem MR-154 FTS was calibrated applying the same two-point calibration used previously and outlined in Section 2.6 with a 30°C cold blackbody reference and a 120°C hot blackbody reference. The blackbody references were measured once prior to measuring the samples and then again following the measurement of the samples. The sample and hotplate rested at a 60° angle above the horizontal as shown in Figure 16.

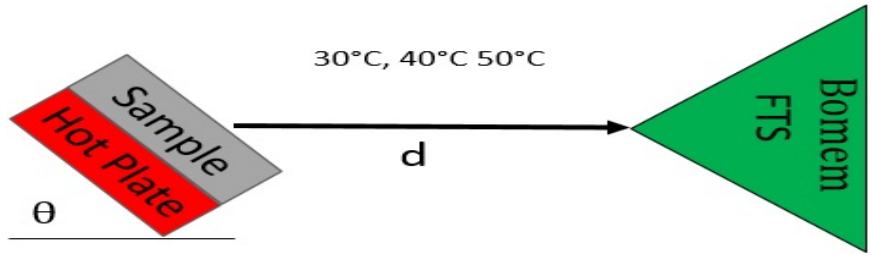

Figure 16. Horizontal view of the Bomem emittance measurement.

A laser pointer was aimed at the sample from the detector to ensure a ceiling tile and not a fluorescent light was at the specular angle. The samples and the blackbodies all used the same detector field of view and both covered the entire field of view of the detector. It was assumed that with the low reflectance and high emittance the sample's self-emission dominated the detected signal. With this assumption the emittance can be calculated using Equation 5 where T is the sample's known temperature or a best fit solution to the measured radiance.

3.7 Raman

All Raman data was collected by the contractor ChromoLogic LLC and provided to the Air Force Institute of Technology, Department of Engineering Physics (AFIT/ENP) by Air Force Research Laboratories (AFRL). ChoromoLogic measured the four samples in set B using 11 different incident laser power levels. Each sample was measured with 3, 4 or 10 spots across the surface of the sample. The incident laser was focused at different depths, varying from 0-40 mils (1/1000 inch), to access different layers in the sample. Unfortunately the focusing of the laser created a very high localized irradiance which burned small spots on the sample surfaces. ChromoLogic smoothed the data and did a background subtraction before providing the data to AFIT.

3.8 Classification Technique

This section will walk through one example using the SVD of Data set B along with multiple-class LDA. Each data set was put through this same process with very few differences which will be discussed each time the data processing differs from this sections classification technique.

Following data collection the spectra from each measurement technique are then put through nearly the same classification technique. The average spectrum for each degradation class for data set B are shown in Figure 17. The primary difference in classification techniques at this point are the number of retained SVD coefficients when applying linear discriminant analysis or multi-class linear discriminant analysis.

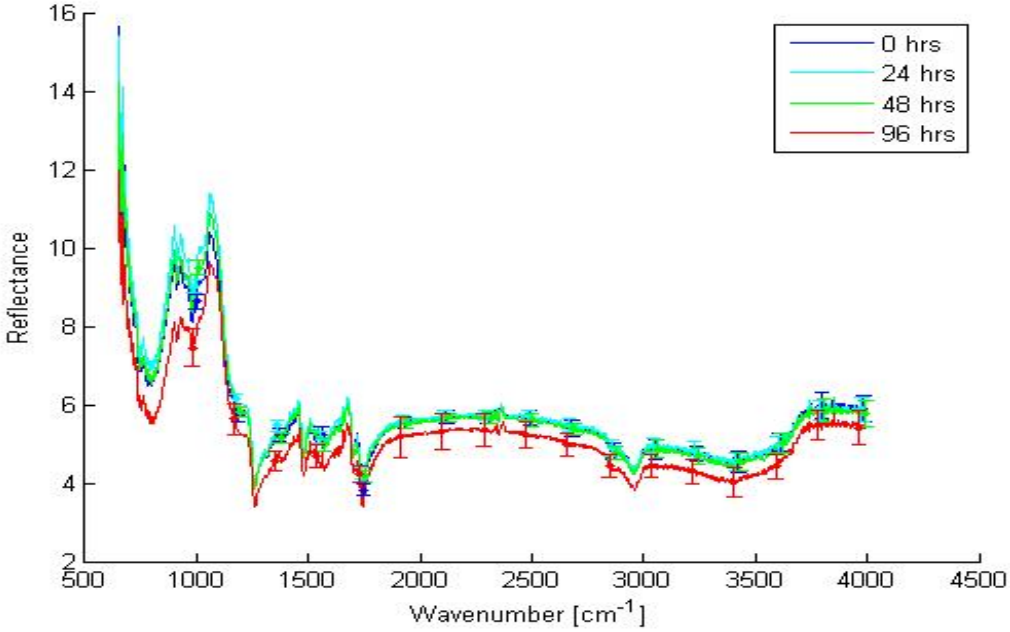

Figure 17. The average reflectance for each degradation class in data set B. The four degradation classes are difficult to distinguish based on the reflectance spectra in this domain. The standard deviation across 75 DRIFTS measurements for each degradation class are shown as error bars.

After applying SVD, the measured spectra are represented by multiple eigen-

spectra. The first four eigenspectra, for data set B are shown in Figure 18. These eigenspectra represent the variation in the data and may not necessarily represent the spectra.

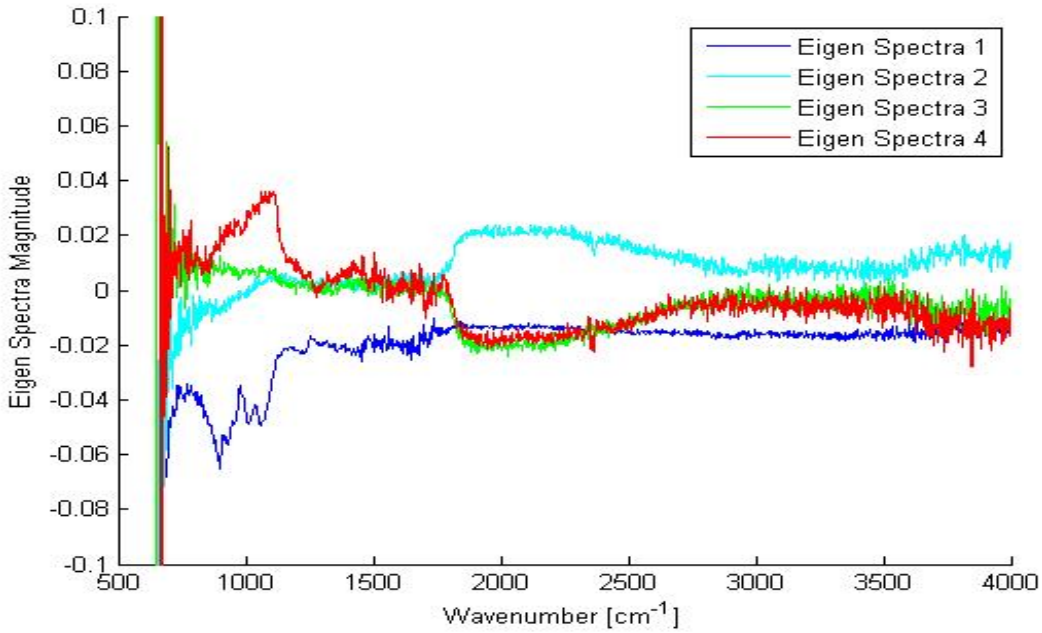

Figure 18. The first four SVD eigenspectra associated with sample set B.

The average constructed signal for each degradation class based on the first two eigenspectra and SVD coefficients for sample set B are shown in Figure 19. The constructed signals are easier to distinguish than the original signals.

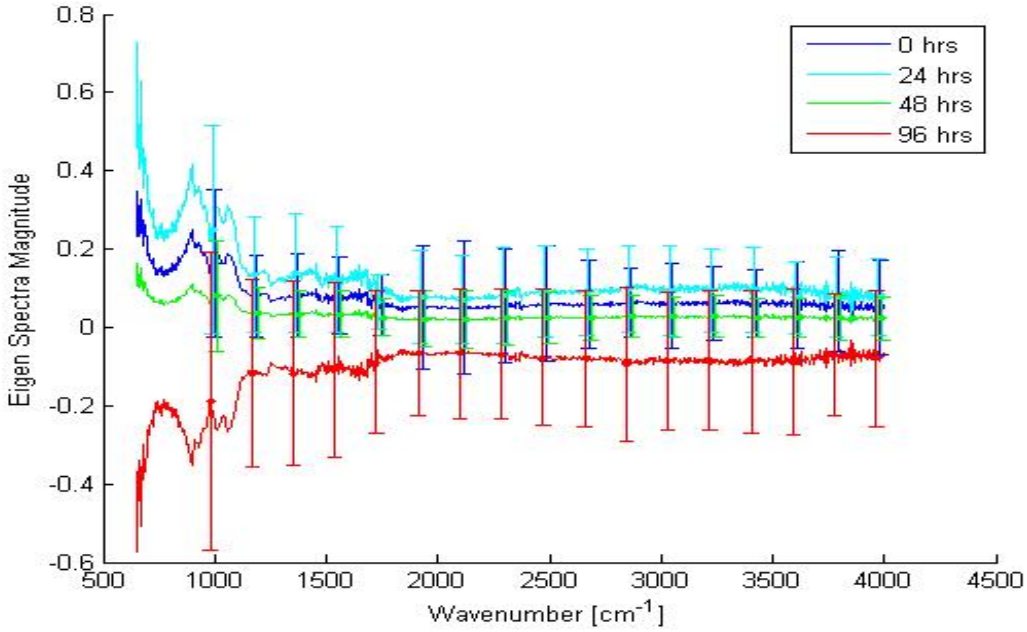

Figure 19. The sum of the first two eigenspectra times the respective SVD coefficients. The four average spectra are more distinguishable but the two-SVD coefficient constructed signals are not representative of the original spectra. The standard deviation across 75 DRIFTS measurements for each degradation class are shown as error bars.

Once the spectra are represented by the SVD coefficients in the SVD coefficient feature space, linear discriminant analysis or multi-class linear discriminant analysis can be applied with high accuracy as shown in Figure 20.

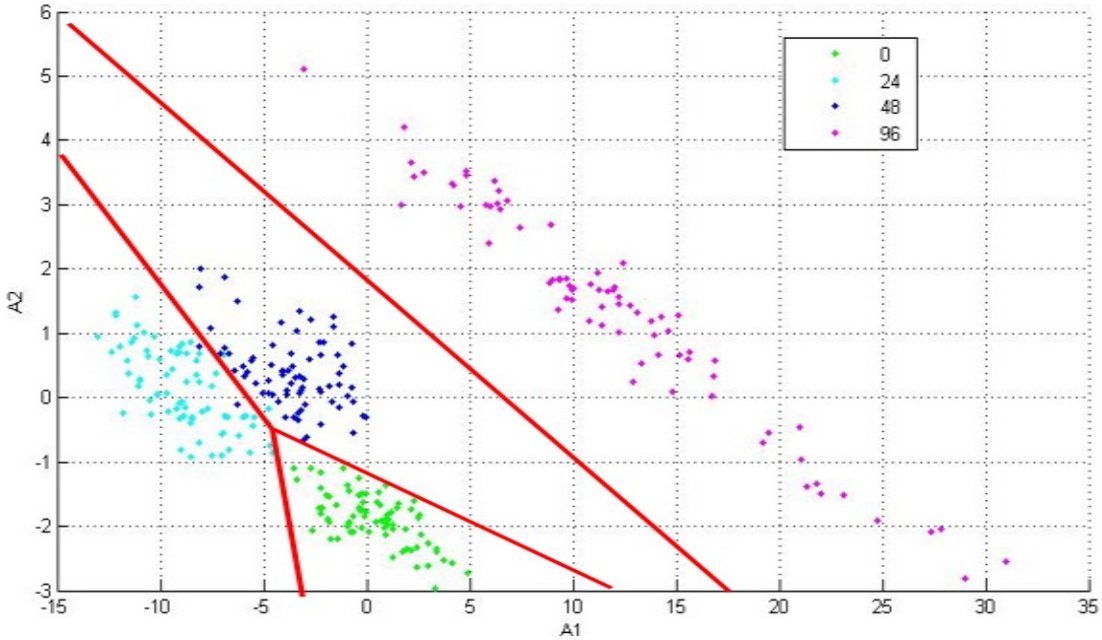

Figure 20. The multi-class linear discriminant boundaries for data set B. In the SVD coefficient domain the four degradation classes are classifiable with high accuracy.

IV. Results and Discussion

First the most recent research will be examined for possible improvements with DRIFTS data acquisition and subsequent processing on sample sets A and B. What is learned from this was applied to data acquisition for sample sets C and D. With specific focus on possible systematic errors contributing to problems with classification. Then the results of similar infrared reflectance measurements such as the SOC-100, Bomem-154 FTS, and Telops hyperspectral will be examined and how well various infrared measurement techniques compare. Then the results of two different measurement techniques, Raman spectroscopy and infrared emittance, will be examined. This will provide multiple independent estimates of the normalized reflectance factor. In addition infrared reflectance's classification accuracy will be compared to classification accuracy of other measurement techniques. Lastly the results of feature selection will be examined to determine if classification using multi-spectral measurement techniques has potential.

4.1 DRIFTS

The fingerprint region Capt Korth identified was selected by determining the region with the highest classification accuracy from ten test regions selected based on literature. Capt Korth determined that the fingerprint region with the highest accuracy was 850.7-1220 cm^{-1} with a resulting error of 3.00% when applied to sample set B. Lack of automation in applying the SVD-LDA technique limited the number of spectral regions that could be compared. The classification technique was later automated by Lt Wilson using Matlab and tested on the existing data for sample sets A and B to validate the automation matched the results generated by Capt Korth. With an automated technique, a more exhaustive search including 16,200

test fingerprint regions was completed to determine the best fingerprint region with a spectral resolution of 18.6 cm^{-1}. Using this test method, the fingerprint region with the highest accuracy was 865.6-1238.7 cm^{-1} and resulted in an error of 1.67% when applied to data set B. This slight modification to the fingerprint region for this top coat and rain erosion coat combination reduced the classification error rate by nearly one half.

Once the optimal fingerprint region was identified it was used with the SVD-LDA classification technique on half the data from sample set A and validated on the other half of sample set A. The validation accuracy of this technique was lower than expected. The linear discriminant line determined from training on the SVD coefficients for the A1 untreated sample and the A2 treated sample is shown as the black line in Figure 21. The training data predicted a mere 1.39% error but the error rate when validated on the A3 untreated and A4 treated samples is 36.1%. The method Capt Korth used to reduce this error was to adjust the cost of misclassification between the treated and untreated samples by increasing the weight misclassifying the treated sample as untreated to 100 times more important than the weight of misclassifying an untreated sample as treated. This will shift the LDA classification line to the left and is shown as a red line in Figure 21. This will reduce the validation error to 11.1%; however, the reasoning used to adjust the weight of misclassification in this instance is not easily applied to other data sets, such as data set B where six relative misclassification weights would have to be determined. In addition it requires at least some knowledge of the expected SVD coefficients locations to identify which misclassification weights should be increased. A number of other technique modifications can be applied which do not require significant knowledge of the expected data distribution.

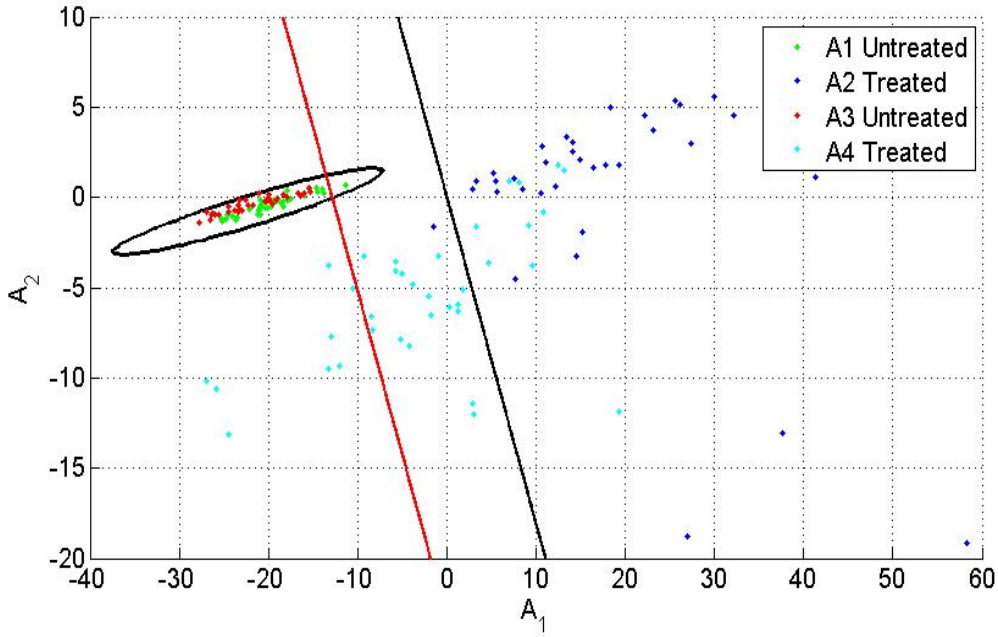

Figure 21. The first two SVD coefficients for the two treated and two untreated samples in sample set A. The linear discriminant line is shown in black, showing significant classification error. The weight-adjusted linear discriminant function is identified by red and still shows some classification error. The quadratic discriminant function results in 100% classification accuracy.

The first technique is modifying how the training data is chosen. Collecting additional data would allow more accurate estimation of the treated and untreated class' probability density functions, and thus more accurate discriminant lines and higher classification accuracy. In this instance, if the discriminant functions were trained using all the sample data SVD, coefficients 100% prediction accuracy can be attained. This accuracy can also be attained by training on half the data from each sample and validating on the remaining half or by training on the A3 untreated and A4 treated data sets and validating on the A1 untreated and A2 treated data sets. In the event this increased training is unacceptable due to the possibility of errors appearing in future training data sets, the classification technique can be modified as opposed to

increasing the training data.

The simplest modification to this classification technique is to use quadratic discriminant analysis (QDA) instead of linear discriminant analysis. In quadratic discriminant analysis, a quadratic function is used to discriminant the various classes instead of a linear function. The quadratic discrimination function when training on A1 and A2 data sets is shown in Figure 21 centered tightly around the untreated data and results in 100% classification accuracy. In addition this technique modification still yields 98% accuracy when applied to data set B. It is expected that a more complex classification rule would result in improved classification accuracy but the reason for such significant improvement becomes apparent when looking at the estimated probability density functions for the treated and untreated classes shown in Figure 22. Both linear discriminant lines divide the feature space such that significant misclassification error of the treated class is expected. A quadratic discriminant line is a more appropriate division for these two classes based upon the estimated probability density functions.

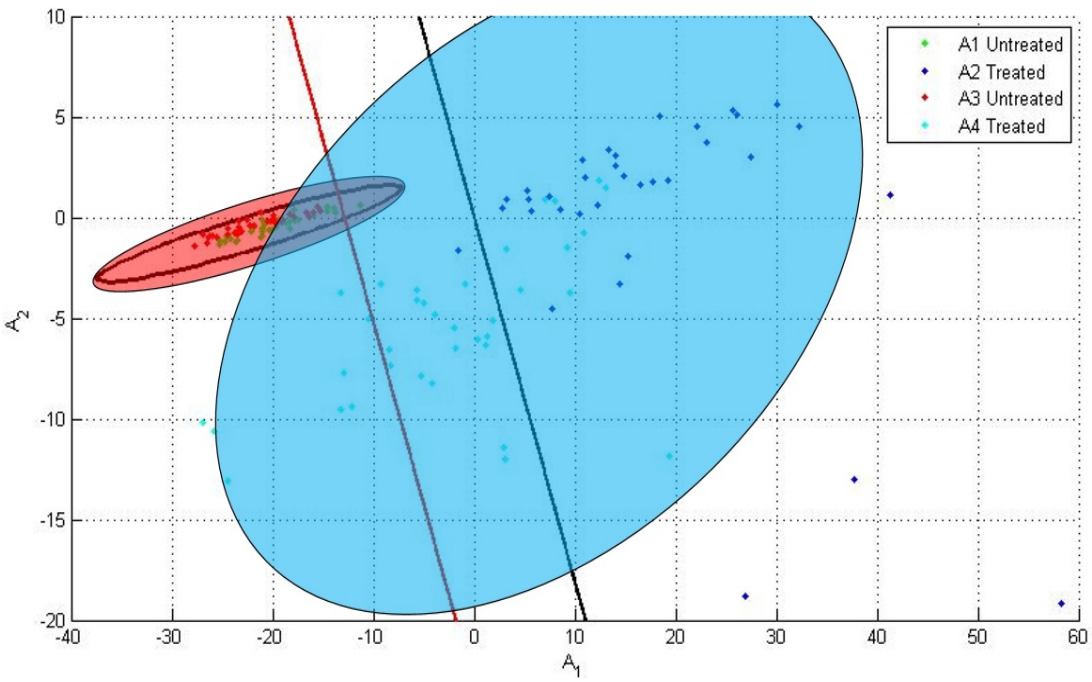

Figure 22. The first two SVD coefficients are are shown for the four samples in set A. The discriminant lines for LDA, cost adjusted LDA and QDA are shown with the estimated probability density functions for the treated(red) and untreated(blue) classes. In this instance it appears LDA is not an appropriate classification method.

Throughout this experiment, the A2 Treated and A4 Treated sets have been regarded as part of the same class. However it can be seen that the variation within the treated class can be largely attributed to variation between the A2 treated sample and the A4 treated sample. When examining the two samples in the SVD coefficient space, the two treated samples appear nearly distinct from one another. When examining the reflectance spectra for the two treated samples each spectrum is barely within the error bars of the other samples' reflectance spectrum as shown in Figure 23. The two treated average reflectance spectra appear very similar in shape however there is a magnitude difference of about 2% across most of the two treated reflectance spectra. From our knowledge of the autoclaving process which is representative of the fielded aircraft coating degradation, these two samples should be indistinguish-

able [26]. The remaining possibility is the two treated reflectance spectra are different due to the way the two samples were measured. This possibility will be explored with sample set C where an experiment can be designed to test this possibility.

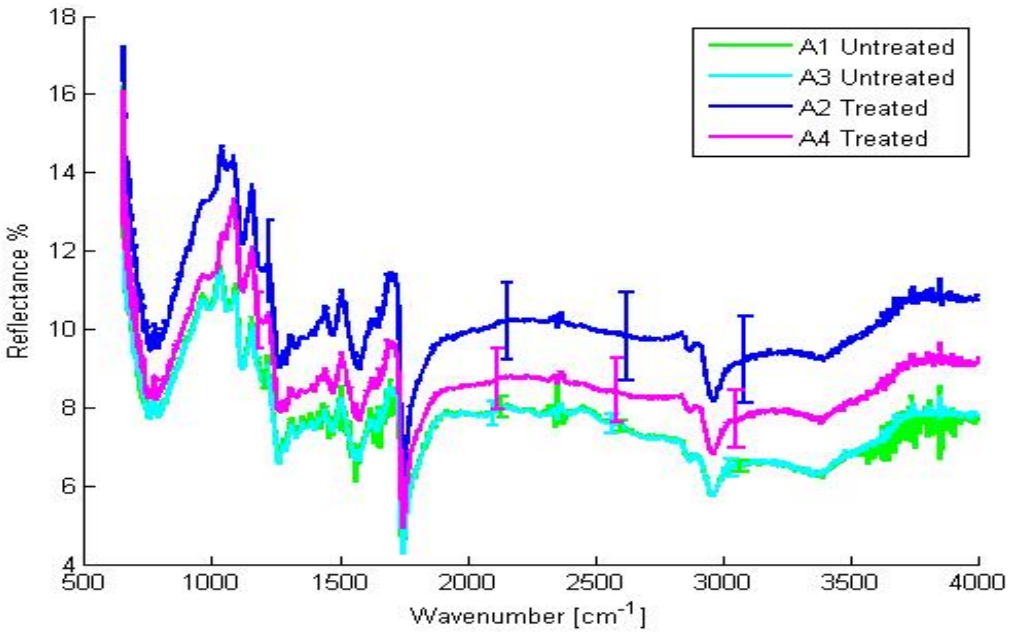

Figure 23. The average reflectance spectra of each sample in sample set A with the standard deviations. The two untreated samples are very similar with small standard deviations. The two treated samples are similar but contain a uniform magnitude difference between the two along with large uncertainties.

In sample set C, half of the measurements were taken with the DRIFTS on its right side following a 30-minute warm-up period. After taking approximately 25 measurements, five measurements per sample, the DRIFTS was rotated to its left side then used to take another 25 measurements on the same samples. The averages for each set of five measurements per sample and for instrument position are shown in Figure 24. The primary spectrum of interest is that for the pristine sample measured with the instrument on its left side. Its reflectance is vastly different than that for the pristine sample measured with the instrument on its right side.

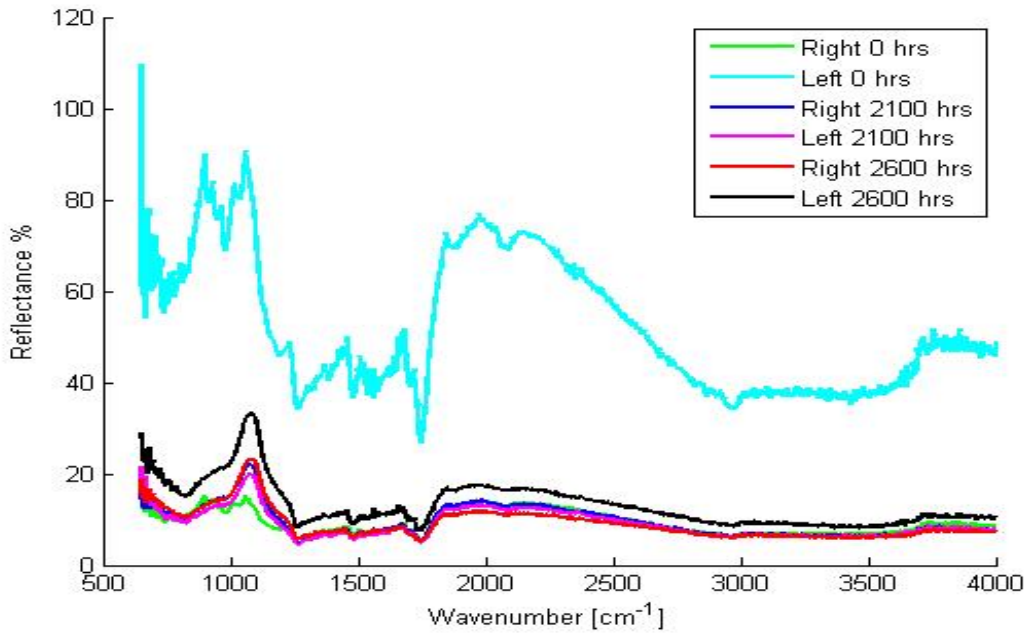

Figure 24. Average reflectance spectra of the 3 samples in sample set C for two different instrument uses. The pristine sample was the first to be measured after rotating the instrument and the average pristine spectra shows significant magnitude variation when compared to previous pristine measurements.

Interestingly the two degraded samples have similar measurements regardless of instrument positions. It is believed this is because measurements were paused after the pristine sample was measured with the instrument on its left side due to a laboratory visitor. During this time the instrument underwent an additional unscheduled warm-up period. It is possible that rotating the instrument generates mechanical oscillations or a changing thermal environment internal to the instrument which must dampen out or return to a stable equilibrium before absolute reflectance measurements can be made again. It is not conclusive but it is possible that the large variation between the two treated samples in sample set A could be attributed to this and similar effects caused by moving the instrument. Ideally the DRIFTS data on these samples would be retaken with the procedures used for later sample sets. Unfortunately sample set

A was no longer available for this research.

Fortunately this instrument movement effect appears to be approximately spectrally uniform. This means that normalizing the data, according to the largest peak, may be all that is necessary to mitigate the instrument movement effect. The normalized reflectance measurements for sample set C are shown in Figure 25. The spectra are now separable by degradation class.

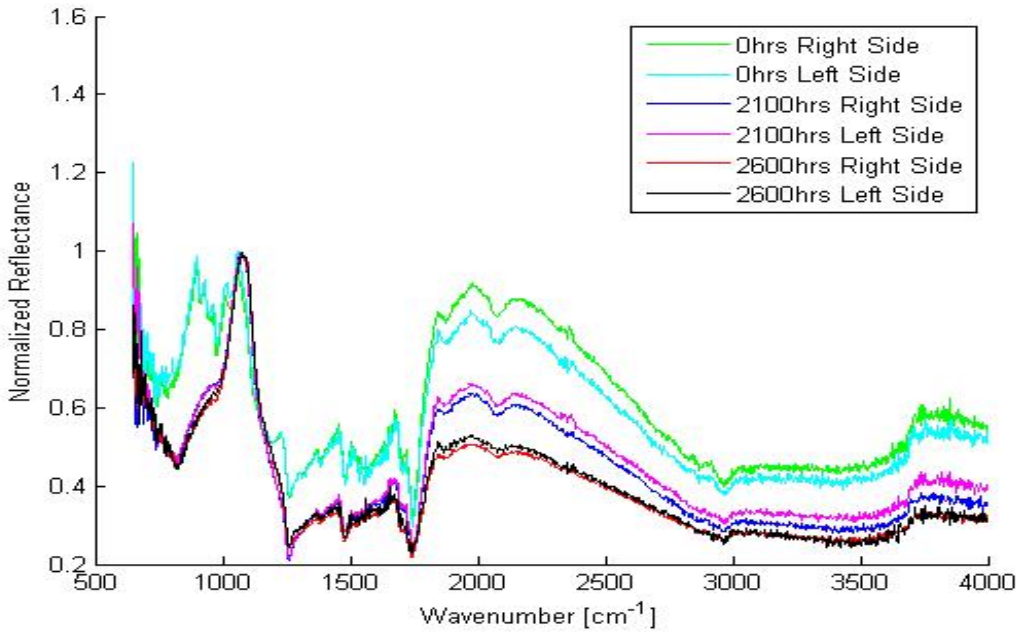

Figure 25. The average normalized reflectance values for sample set C. As shown the sample spectra for each of the three degradation classes can now be more easily distinguished.

Following the normalization of each individual spectrum, the original SVD-LDA algorithm can be applied. For more accurate normalization, care must be taken to not normalize to the edge of the instrument response around 650-700 cm^{-1}. Noise in this region can cause significantly different normalized reflectance spectra sometimes resulting from a impossible measured reflectance of greater than 100%.

When applying the SVD-LDA algorithm with two SVD coefficients, the samples

from sample set C are classified with 90% accuracy. Each of the misclassifications is between the 2100 hr and 2600 hr degradation classes. When examining the samples in the SVD coefficient feature space, shown in Figure 26, it becomes apparent that the misclassification is the result of a single outlier of the 2600 hr sample. It is expected that accuracy would improve with additional data so the effect of individual outliers are reduced.

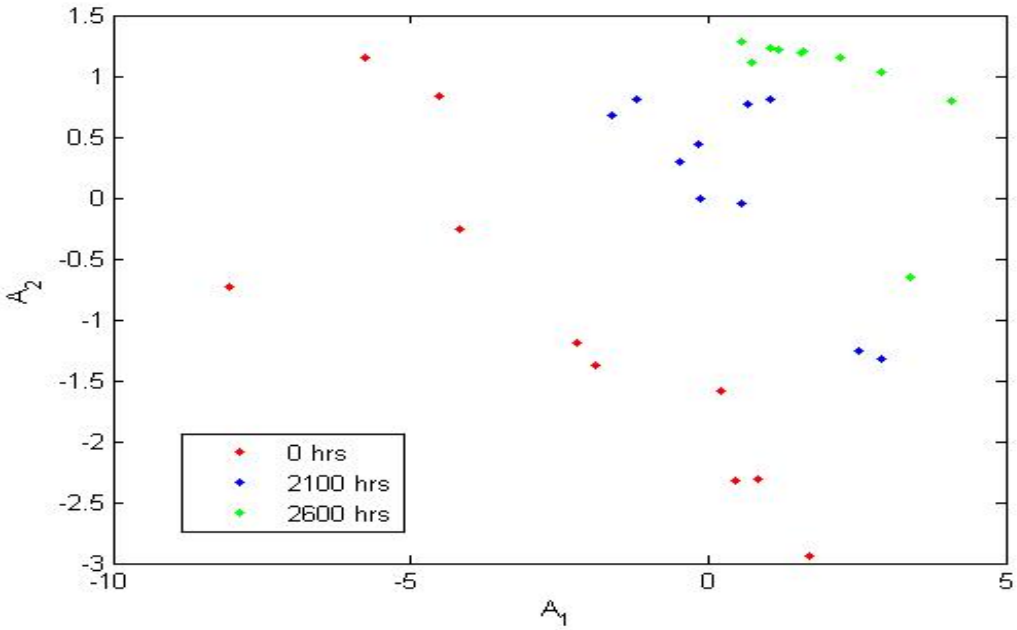

Figure 26. SVD coefficient space for sample set C. The three degradation classes are well separated with the exception of the single outlier in the 2600hr degradation class. This one outlier shifts the classification line such that three sample measurements are misclassified.

Taking the information learned about the instrument from sample set C, the same normalization process can be applied to the data from sample set A. The normalization process was disregarded in previous literature because the normalization reduced classification accuracy for sample set B from 98.3% to 93.3% for two SVD coefficients. However the effect of instrument movement and position on the data was ignored at

that time. By taking the instrument movement effect into account, mitigating with reflectance spectra normalization, degradation classification can be made significantly more accurate for sample set A without resorting to a more complex classification algorithm. Figure shows 27 four samples from sample set A can be classified with 100% accuracy using the normalized spectra SVD coefficient space with only two coefficients. There still appears to be some differences between the two treated samples but the probability density functions for the two treated samples have significantly more overlap than shown previously in the unnormalized spectrum's SVD coefficient space in Figure 22. Due to this, the classification rules generated by training on one pair of treated/untreated samples are more accurate when validated on the remaining pair of treated/untreated samples than previously possible without normalization.

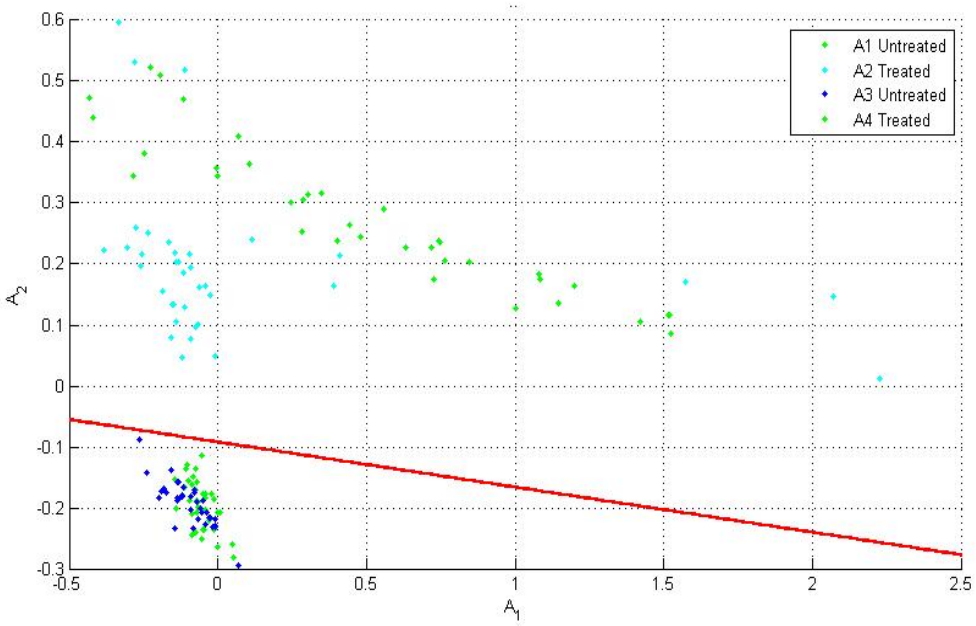

Figure 27. The SVD-coefficient space for sample set A after each spectrum was normalized. The two treated samples now have significantly more overlap in SVD-coefficient space than previously without the normalization. The resulting discriminant line trained on one pair of samples is then more accurate when validated on the remaining pair of samples.

Another factor of interest is how detector noise may affect classification accuracy. The DRIFTS is a relatively low-noise instrument in comparison to the Bomem or Telops hyperspectral camera. It is anticipated that increased noise in the spectral signature will decrease classification accuracy despite the mitigating effects of the SVD-LDA technique. To examine what signal-to-noise ratio (SNR) may be necessary for a specified classification accuracy, artificial zero-mean Gaussian noise was added to each spectrum in sample set B with an SNR varying from 10^0 to 10^4. Then the resulting noise-degraded data was put through SVD and the resulting SVD coefficients were classified using LDA and Decision Trees (DT) and then the LDA result was cross-validated using K-Fold analysis. The results of each algorithm for varying levels of noise are shown in Figure 28. Treating the average spectrum for each degradation class as a pristine spectra the SNR of the DRIFTS data is estimated to be 2875.

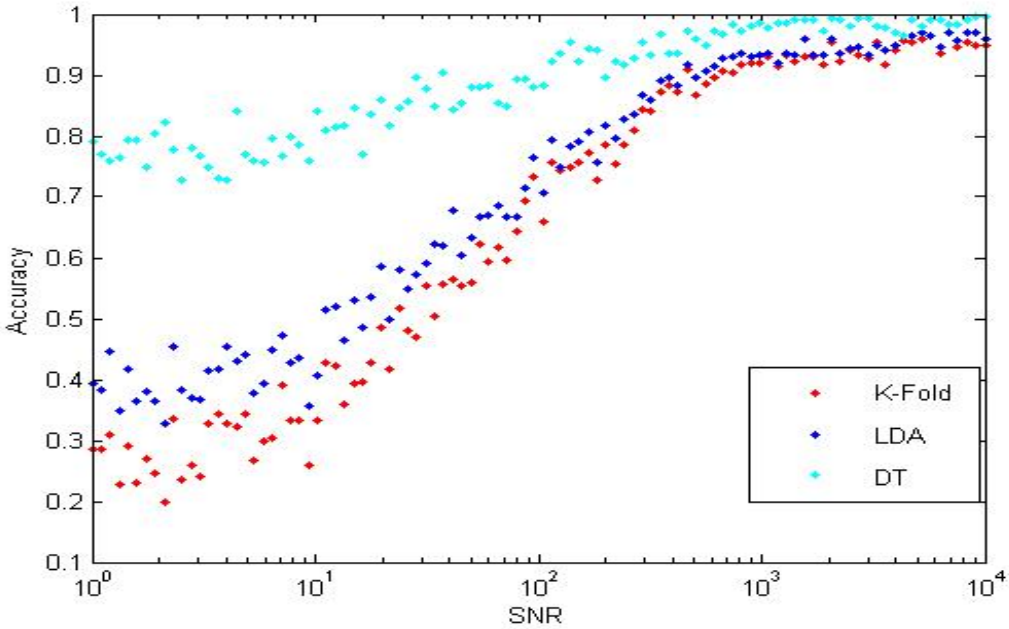

Figure 28. Accuracy for Decision Trees, LDA and K-Fold cross validation for LDA. Decision Trees are heavily overtrained as seen by the high reported accuracy even at near zero signal. High classification accuracy requires fairly high signal to noise ratio's while using the current SVD-LDA technique. The K-fold estimate of overtraining for LDA is within acceptable limits as well.

Decisions Trees were considered as a possible substitute classification algorithm over LDA but the decision tree classification algorithm was determined to be too susceptible to overtraining as shown in Figure 28 where decision trees still report around 70% accuracy with virtually no signal. LDA has reduced potential for overtraining due to the less complex classification rules. In addition the LDA overtraining can be predicted using K-Fold cross validation as shown in Figure 28 where LDA still reports nearly 40% accuracy without signal while K-Fold reports around 25% or random accuracy without signal. Additional variations on LDA, such as QDA, diagLinear, diagQuadratic, pseudoLinear and pseduoQuadratic discriminant analysis were identified as possible substitutes. However, it was determined that LDA's simplicity and

reduced risk of overtraining made it the best suited for this degradation classification problem.

An additional topic of interest is how transmission through and reflection from the various layers affect the reflectance spectrum. Figure 29 shows two possible mechanisms for the reflectance. On the left side the top coat either absorbs or reflects all the incident radiation and the lower layers do not contribute to the reflectance spectrum. If this were the case, all degradation information obtained with a reflectance measurement would only pertain to the degradation of the top coat and no information is attained about the lower layers. On the right side the top coat still absorbs and reflects, thus contributing to the reflectance spectrum, but it also transmits some of the light which is then reflected at the TC/REC boundary, the REC layer reflects, absorbs and transmits, and finally another reflection happens at the REC/substrate boundary. While this mechanism is much more complicated, information is attained about all three layers.

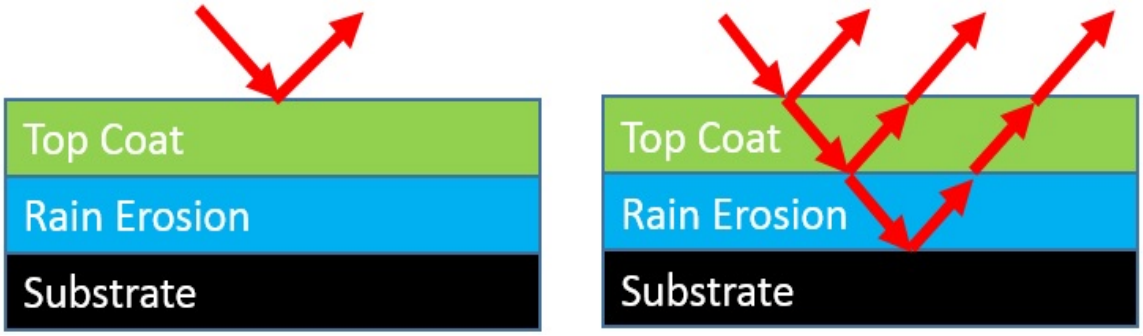

Figure 29. There is the possibility that the measured reflectance is due to only the material prosperities of the top coat as shown on the left. If the layers are not opaque then the top coat, rain erosion coat and substrate all contribute to the measured reflectance as shown on the right.

To test if the reflectance spectrum is only the result of reflections from the top coat, multiple combinations of the various layers were tested. Three layer combinations

were tested; the original TC/REC combination on the substrate, the top coat on the substrate and the REC on the substrate. This set up is shown in Figure 30. In addition, two types of top coats and RECs were tested as it is possible that one top coat is responsible for the entire reflectance spectra while another top coat allows light to transmit to lower layers. Testing the effect of the substrate was also planned. Unfortunately the test samples with a different substrate were unavailable. If the reflectance spectrum is only dependent on the top coat, then the first and second samples should have nearly identical reflectance spectra.

Figure 30. If the top coat is the only layer responsible for the reflectance spectra then the first and second samples should have identical reflectance spectra. If the first and second spectra are distinct the third sample will provide information on the reflectance spectra of the middle layer and later assist in determining of the substrate contributes to the reflectance spectra.

The reflectance spectra for TC1 on silicon and TC1/REC1 combination on silicon substrate display minor differences primarily around 1800-3000 cm^{-1} as shown in Figure 31. Meanwhile the reflectance spectra for TC2 and TC2/REC2 combination both on silicon substrates showed significant differences around numerous sections of the reflectance spectra. Additional measurements of these combinations on an aluminum substrate are planned upon arrival.

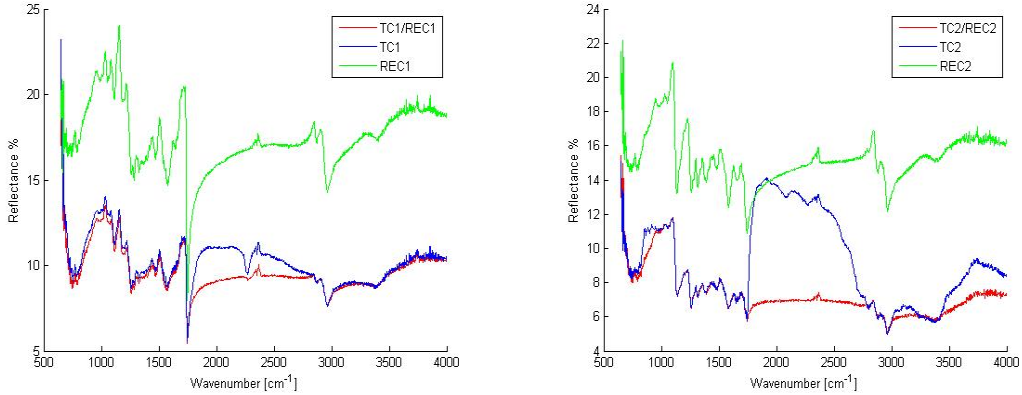

Figure 31. DRIFTS reflectance measurement for comparing two TC/REC combinations on silicon. With two TC's and two REC's alone on silicon. The reflectance is shown to be dependent upon the lower layers in addition to the TC layer.

4.2 Surface Optics Corporation SOC-100

The first part of the SOC-100 experiment was to determine the difference between measuring the diffuse and specular reflectance of the weathered samples at different angles. As expected, the average reflectance over all measured wavelengths, increases with angle for diffuse and specular reflectance but at the same rate for each degradation class as shown in Figure 32. This suggests that normalizing each spectrum may be sufficient to compare sample measurements taken at different angles. This is particularly useful when taking future measurements of samples that may have multiple angles throughout the measurement such as during a hyperspectral image of an aircraft.

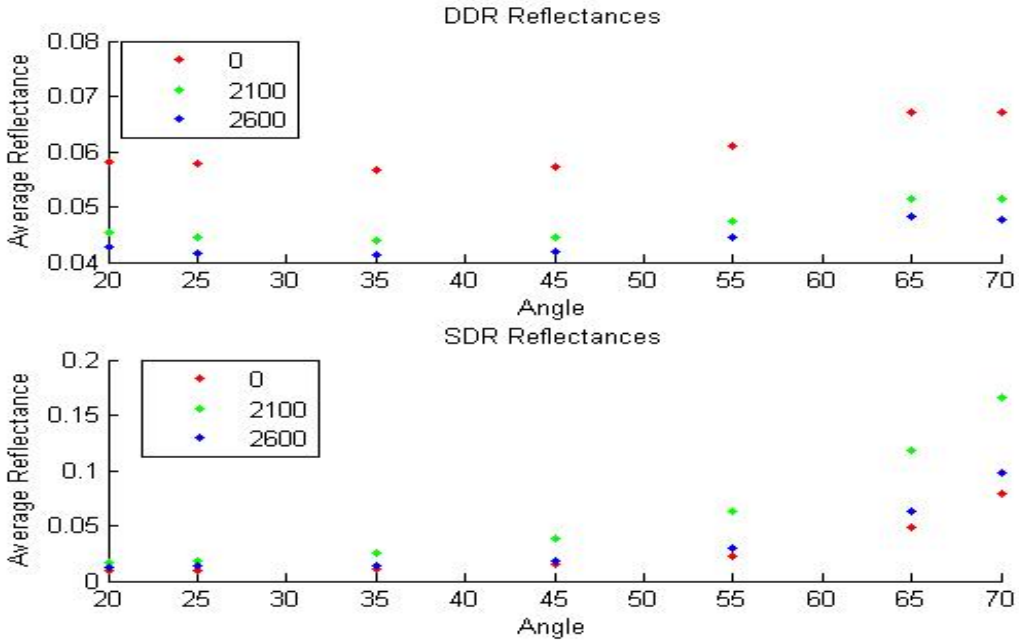

Figure 32. Average diffuse and specular reflectance between 800 **and** $3500cm^{-1}$ **for the three artificially weathered samples. The relative magnitudes of the average reflectance remain roughly the same as the observation angle changes.**

Unfortunately simply normalizing each reflectance spectrum did not remove all variation due to angle. Fortunately a large portion of this variation appears to come from the overall magnitude of the reflectance while still maintaining similar spectra. This is believed to be due to problems inherent in normalizing on noisy data. This happens when normalizing to the the largest peak inside the lower noise region of 800-3000 cm^{-1}. Normalizing outside this region is possible but there is significant noise below 800 cm^{-1} and above 3000 cm^{-1}. Therefore including the spectra outside this region in normalization can result in normalizing to physically impossible reflectance values. Some detector noise is still present. If noise is present around the large peak an apparent shift in the magnitude of the normalized reflectance is observed. The issue now being that each degradation spectrum contains significant variation within each degradation class. For accurate degradation classification, small variation within

class and large variation between classes is desired. As shown in Figure 33, it appears the spectra are distinct, but each degradation sample spectrum is within the standard deviation of the other samples' reflectance spectra as well. Despite the magnitude change and associated variation, the reflectance spectra still retain the same shape and thus the SVD-LDA technique is still expected to provide high accuracy since SVD is adept when dealing with overall magnitude shifts.

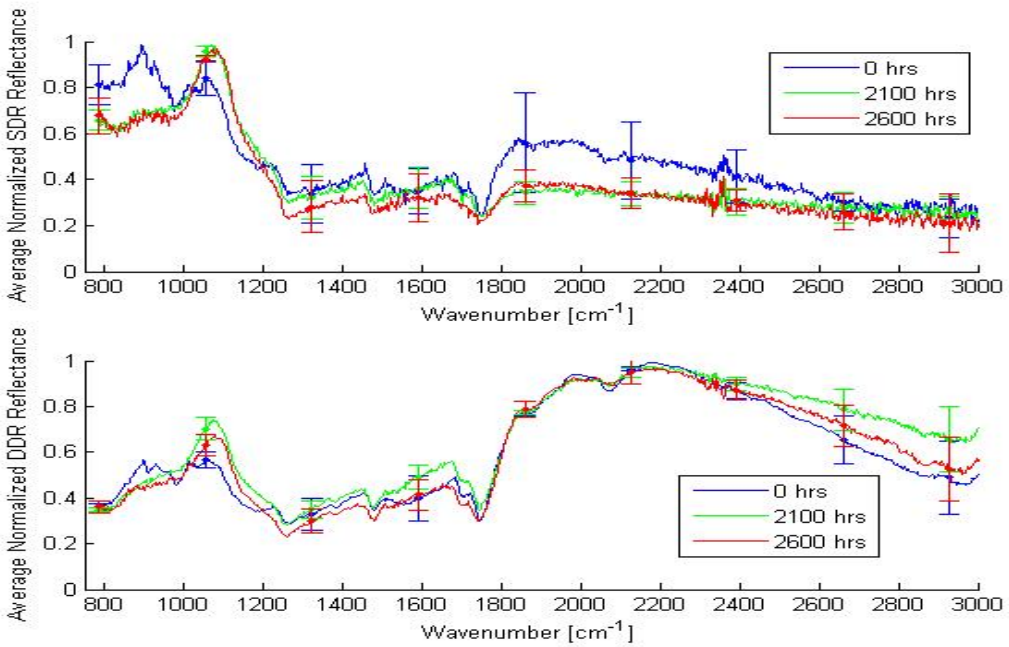

Figure 33. The average normalized spectrum across all measured angles for each degradation class after each spectrum was normalized for both specular and diffuse reflectance.

The combined single value decomposition and linear discriminant analysis technique was applied to the normalized DDR, SDR, and hemispherical directional reflectance (HDR) independently. The resulting accuracy for each reflectance measurement is shown in Figure 34 where 100% accuracy was attained using DDR and HDR. Unfortunately SDR never attained 100% and becomes noticeably overtrained with a larger number of retained SVD coefficients. This implies that for these samples, dif-

fuse reflectance contains more information relating to degradation classification than the specular reflectance. Therefore in future measurements, it is expected that a diffuse reflectance measurement will give a higher accuracy than a specular reflectance measurement.

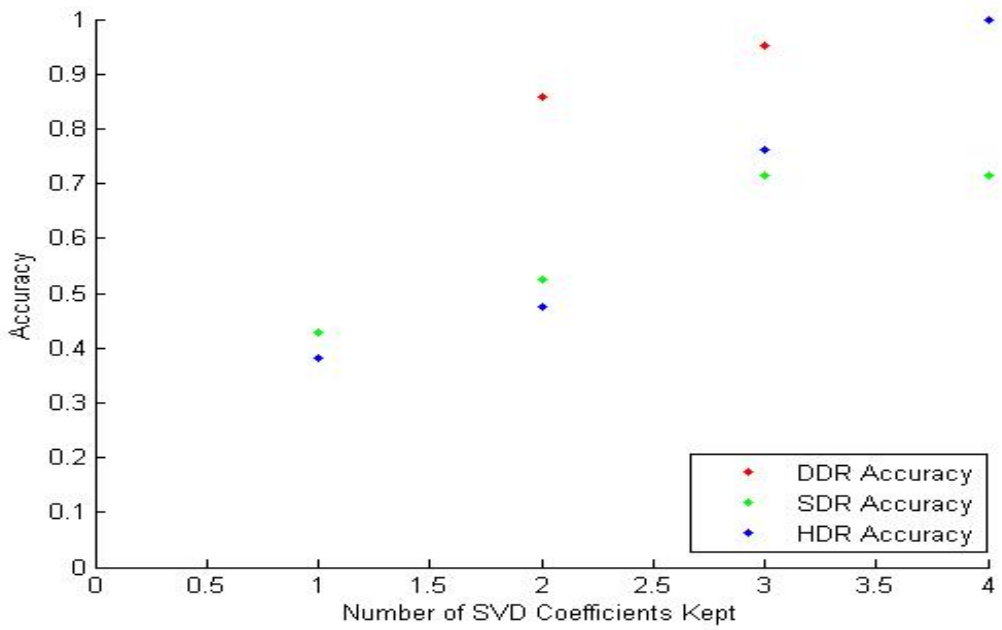

Figure 34. Classification accuracy for the normalized measurements from the SOC-100 for diffuse, specular and hemispherical reflectance. Diffuse and hemispherical reflectance appear to out perform the specular reflectance for maximum classification accuracy.

4.3 Bomen Fourier Transform Spectrometer Reflectance

The DRIFTS and SOC-100 reflectance measurements require being in contact or enclosing the sample. A stand-off measurement may be more useful for maintenance procedures. The Bomem Fourier Transform Spectrometer (FTS) provides such a measurement technique. The first step of the Bomem experiment was to determine if the calibration technique outlined in section 2.4 yielded reliable results for blackbodies

at 50 °C, 60 °C, and 70 °C. When this experiment was originally attempted the coldest blackbody was measured and then the warmer blackbodies placed in front of it. The result was a path length change of approximately 25 cm. This led to an issue where the amount of atmospheric absorption was overestimated for the closer blackbodies and underestimated for the blackbodies at a distance. This was later corrected and all sample measurements were taken with the blackbody to detector path length remaining the same for all measurements.

The calibrated spectra, with blackbody positions carefully monitored, and theoretical blackbody spectra for the 60 °C blackbody are shown in Figure 35. As shown the calibration data is remarkably close to the theoretically expected radiance spectra with a normalized mean squared error(MSE) of 0.276. Further, the vast majority of that error is found in the spectral regions 1380-1800 cm^{-1} and 2340-2380 cm^{-1} which are associated with water and CO_2 absorption respectively. Removing these regions, the normalized MSE is reduced to 0.015. For the Bomem data, it is expected that data taken in these atmospheric absorption regions will be of less use in classifying degradation levels than spectral regions where atmospheric absorption is lower and varies less with time.

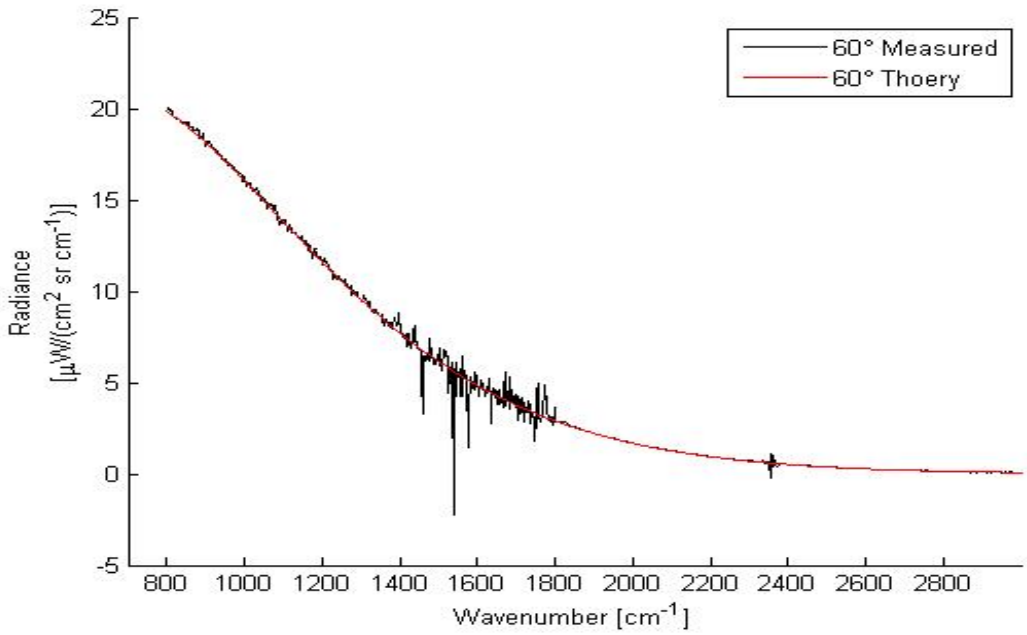

Figure 35. The calibration test for the Bomem 154 FTS. The theoretical expectation and measured data for a $60\,^{\circ}\mathrm{C}$ **blackbody are nearly identical outside the water and** $\mathbf{CO_2}$ **absorption regions around** $1380 - 1800$ **and** $2340 - 2380\ cm^{-1}$ **respectively.**

The reflected radiance of a $1000\,^{\circ}\mathrm{C}$ blackbody was then measured using the Bomem after it reflected off each of the three degraded samples. The Bomem calibrated reflected radiance data was converted to normalized reflectance factor using Equations 18 and 19. The normalization constant was removed by normalizing the spectra to a band around 2950 cm^{-1}. This band was chosen because it was uniform across the three samples and outside of all atmospheric absorption bands identified in the calibration test. The Bomem data included a normalized diffuse reflectance measurement and a specular reflectance measurements shown in Figures 36 and 38. Due to the large difference between the specular and diffuse reflectance, the classification on the diffuse reflectance measurements must be done separately from the classification on the specular measurements.

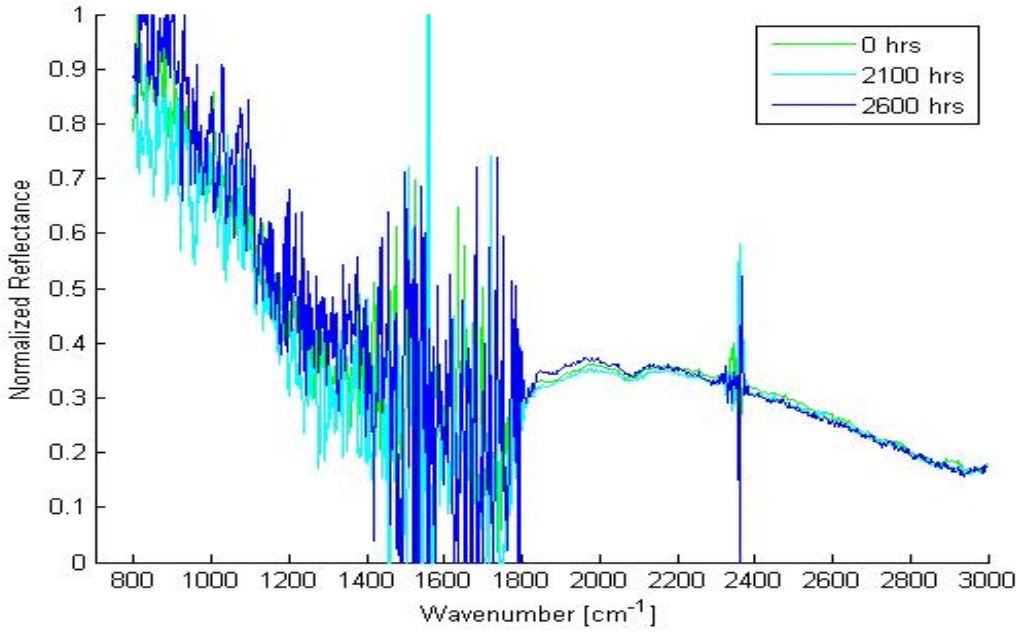

Figure 36. The average Bomem measured diffuse reflectance for each degraded sample in sample set C. High noise and small variation between the signals makes classification difficult.

Unfortunately when attempting a diffuse reflectance spectral measurement with the Bomem, the majority of incident radiation is reflected in directions other than the detector. The resulting reflectance signal-to-noise ratio is relatively low as shown in Figure 36 including high noise in regions outside of the atmospheric absorption bands.

In spite of the significant noise in the spectra, the SVD-LDA technique still shows reasonable results because the noise is captured in higher dimensions, while the first few SVD coefficients are mostly the signal. The SVD coefficients still display remarkable separation as shown in Figure 37. Similar to the SVD example in Section 2.10 classification in the SVD coefficient space can result in higher accuracy than classifying in the original feature space. The SVD-LDA classification algorithm resulted in 97% accuracy. This places the Bomem stand-off measurement accuracy within

similar range of that with the DRIFTS reflectance measurements.

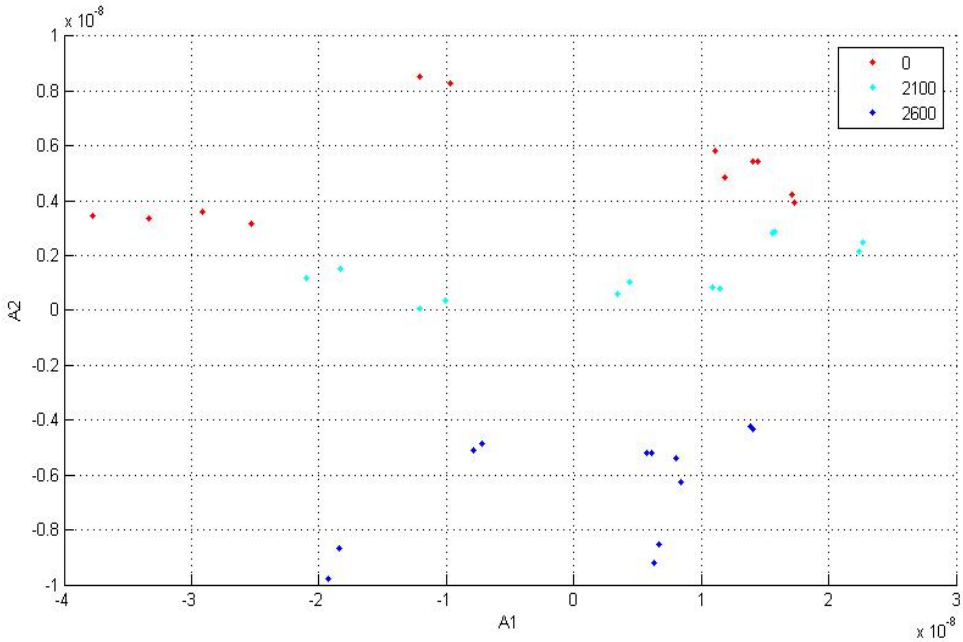

Figure 37. The first two SVD coefficients for the three degraded samples in sample set C. The SVD coefficient feature space provides relatively high classification accuracy when using LDA.

While there was a large amount of noise in the diffuse reflectance data, analysis on the normalized specular reflection factor demonstrated a much higher signal-to-noise ratio as shown in Figure 38. Unlike the diffuse reflectance measurement where the majority of reflected radiation was not reflected toward the detector, placing the detector at the specular angle allowed for a much higher signal and signal-to-noise ratio. As expected the water and H_2O absorption regions did not provide useful spectral information. However outside these two absorption regions, the sample spectra are clearly distinct enough for accurate classification as shown in Figure 38.

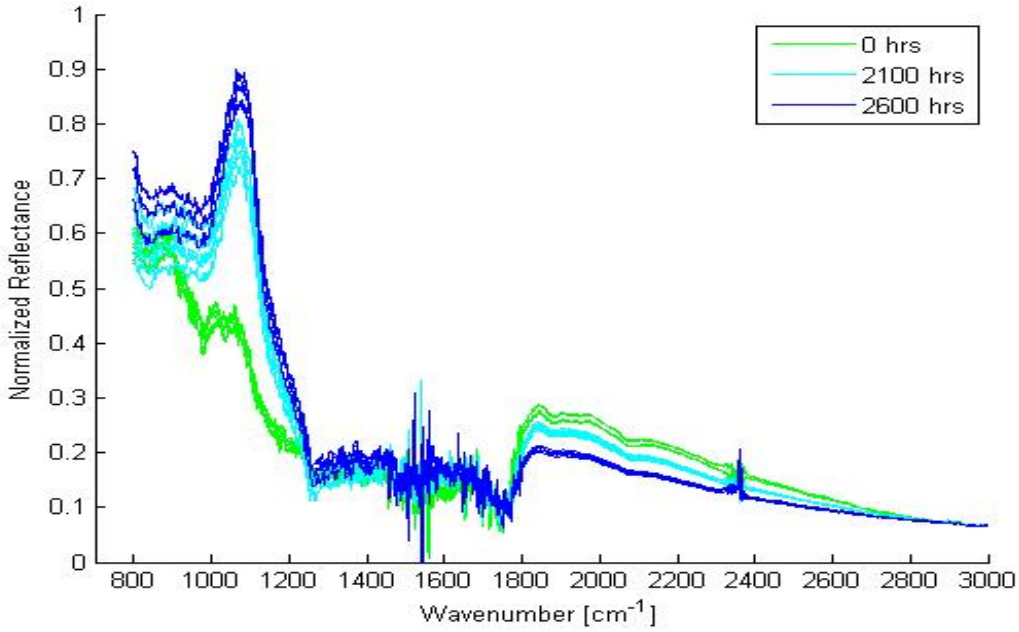

Figure 38. The normalized specular reflectance for the artificially weathered sample set C. There is good separation between the data for the three different sample degradation classes.

Linear discriminant analysis was then applied directly to two spectral bands centered around 1300 and 2000 cm^{-1}. LDA and K-Fold analysis both predicted 100% accuracy for the three artificially weathered samples. This suggests for these samples and measurement technique, the more complicated SVD-LDA algorithm may be unnecessary and that a simpler classification technique may be used. In addition a full spectral measurement may not be necessary as a multi-spectral measurement could be made once the appropriate spectral bands for classification are identified. The comparison of the various instruments' measured reflectance spectra will be discussed in detail in Section 4.5.

4.4 Hyperspectral Imaging Telops

There are two primary advantages using a hyperspectral camera has over previous measurement techniques. The first was the spatial information gained by taking an image of the sample. One hypothesis was blisters and wrinkling may be visible using hyperspectral imaging before they became visibly apparent. In addition, spatial non-uniform degradation is expected to increase the longer the sample has been aged[29]. The second advantage comes from quickly collecting a large number of spectral measurements uniformly spaced across the sample. All previous measurement techniques were averaged across large areas of the samples while location and spacing across the samples were varied by hand.

Unfortunately the reflected radiance at each pixel across the samples is dependent upon the incident radiance and thus sample location relative to the illuminating blackbody. Since the incident radiation can vary largely across multiple sample surfaces, the reflectance for each pixel was normalized to the largest peak between 900-1250 cm^{-1}. All reflected radiance values are scaled by a constant as outlined in Section 3.5. Normalizing the reflectance spectra yields normalized reflectance factors for each pixel and allows for a direct comparison of all pixels across the image to determine if blisters, wrinkling or non-uniform degradation across the sample are present. The average normalized reflectance for the field of view of the detector is shown in Figure 39. For these sample measurements, the three samples appear roughly uniform across the sample surface with no noticeable spatial features apparent in the sample surfaces to suggest blisters, wrinkling or non-uniform degradation can be detected using longwave infrared normalized reflectance factors.

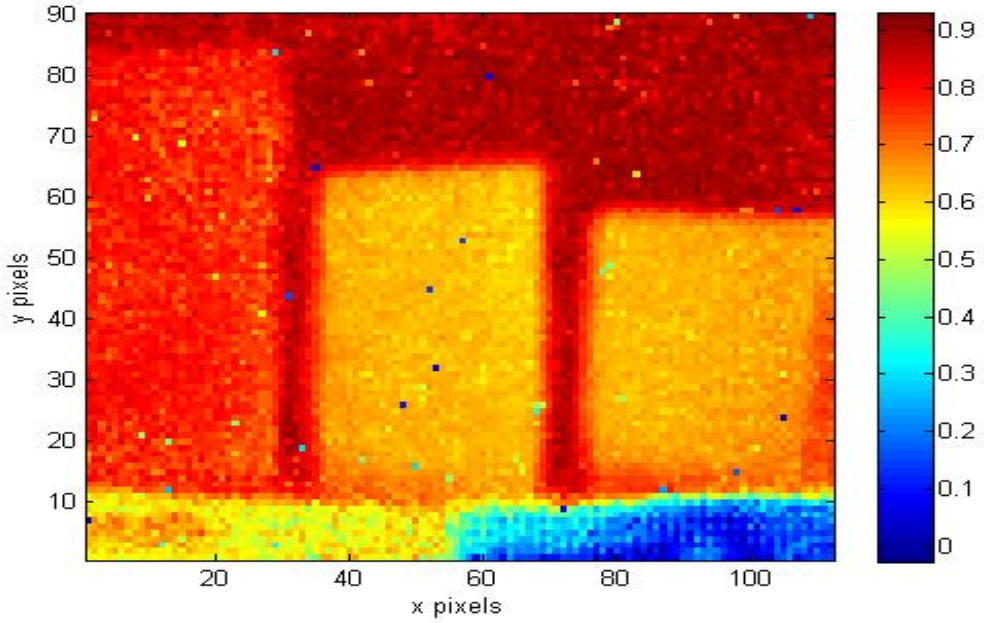

Figure 39. The average normalized reflectance for the three samples and nearby surroundings. The pristine sample on the left, 2100 hours of degradation in the center and 2600 hours of degradation on the right.

While many more measurements are possible in the same amount of time using a hyperspectral camera, each individual pixel's spectrum is relatively noisy in comparison to previous measurement techniques. This is expected since each pixel's field of view and detector size is smaller than with previous techniques. However the spectra across multiple pixels can be averaged if noise is an issue. The spectra for each sample's surface were averaged individually as shown in Figure 40 with very little noise apparent. In addition it is expected that noise will have a limited effect on the SVD-LDA classification technique as outlined in Section 2.10. The measured spectra for the three samples are similar to spectra measured in previous techniques.

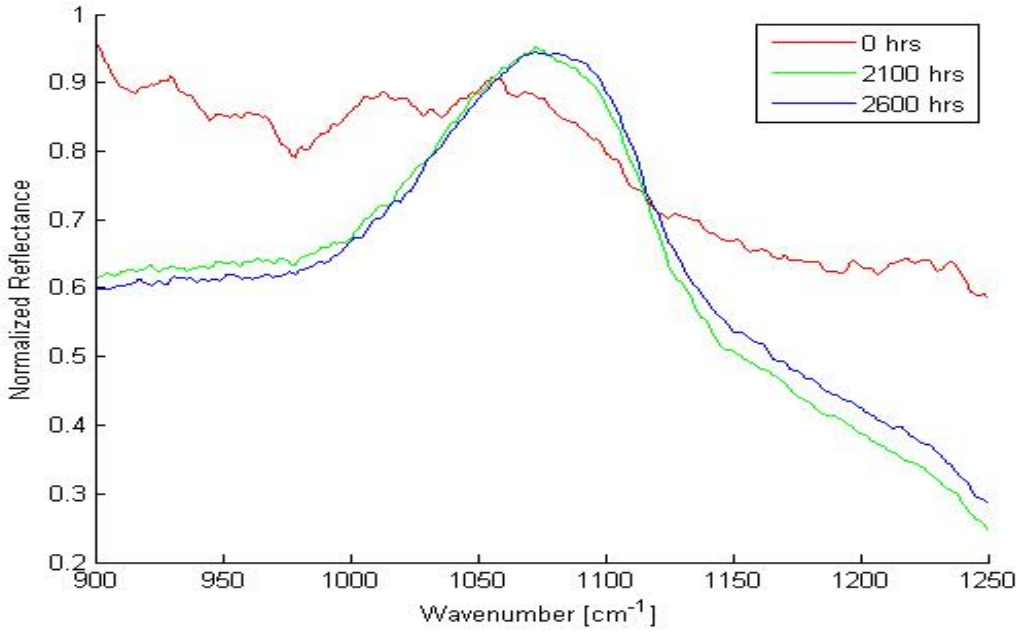

Figure 40. The spectra averaged across the surface area of each sample. As expected the pristine sample is clearly distinguishable from the two degraded samples.

Next, SVD was applied to just the pixels associated with the three samples in an attempt to highlight possible blisters and non-uniform degradation across the samples. A false color image of the first three principle components which represent 91% of the variation in the data is shown in Figure 41. Unfortunately SVD did not appear to highlight any non-uniformities across the sample surfaces. It did however highlight the differences between the samples.

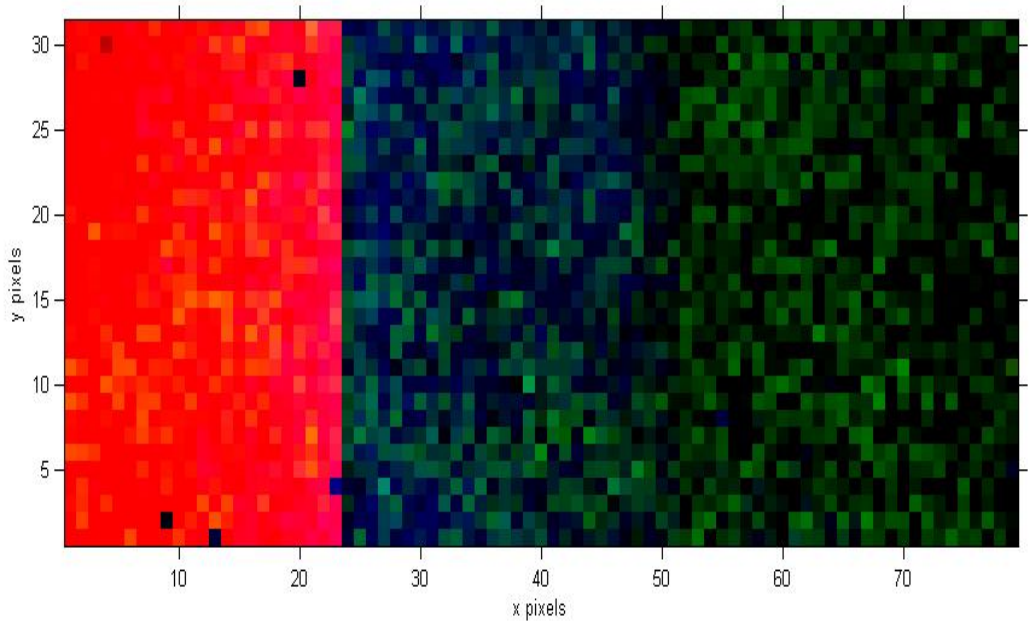

Figure 41. A false color image where the first three principle components represent the RGB components of the image. The pristine sample is clearly distinguishable. The two degraded samples can be classified but with less accuracy than the pristine sample.

When examining each pixel independently in the principle component space, the first and third principle components are the most useful for degradation classification. The first principle component is responsible for determining the pristine sample from the others and the third principle component determines the difference between the two degraded samples as shown in Figure 42.

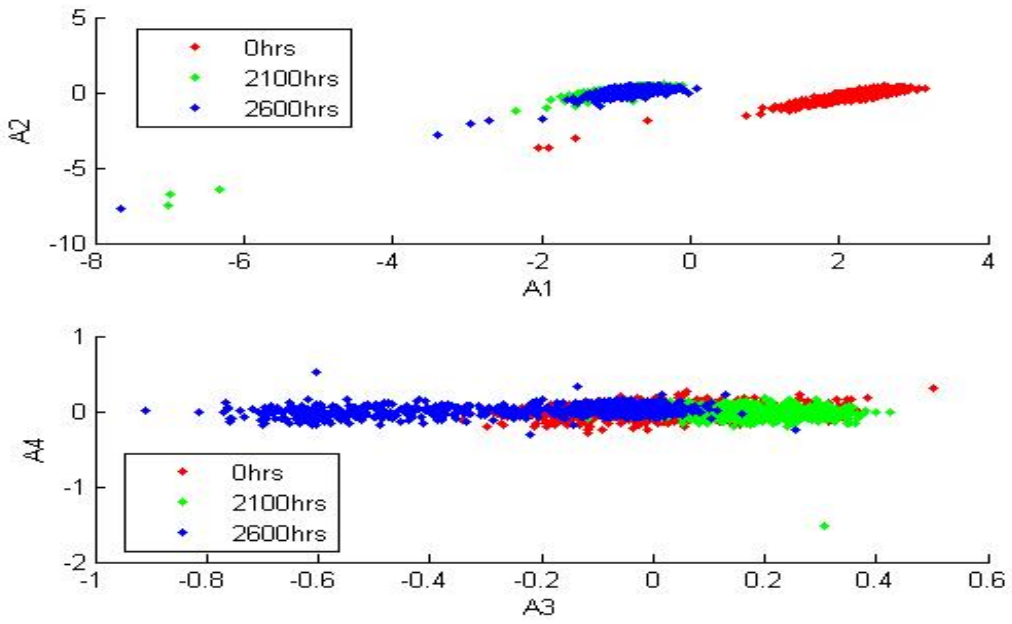

Figure 42. When examined in the principle component space it quickly becomes apparent than the first and third principle components will be the most useful for degradation classification.

Accuracy increases as the number of principle components kept for LDA classification increases as shown in Figure 43. Classification accuracies were tested with more than three coefficients; however, there was only marginal increases in accuracy beyond retaining three principle components for LDA classification. The classification accuracy with three principle components retained is similar to the classification accuracies reported by the other techniques. In addition, because of the large number of sample observations, the effect of noise and outliers on the classification rules is reduced. This results in the K-fold cross validation and the LDA predicted classification accuracy to return similar results.

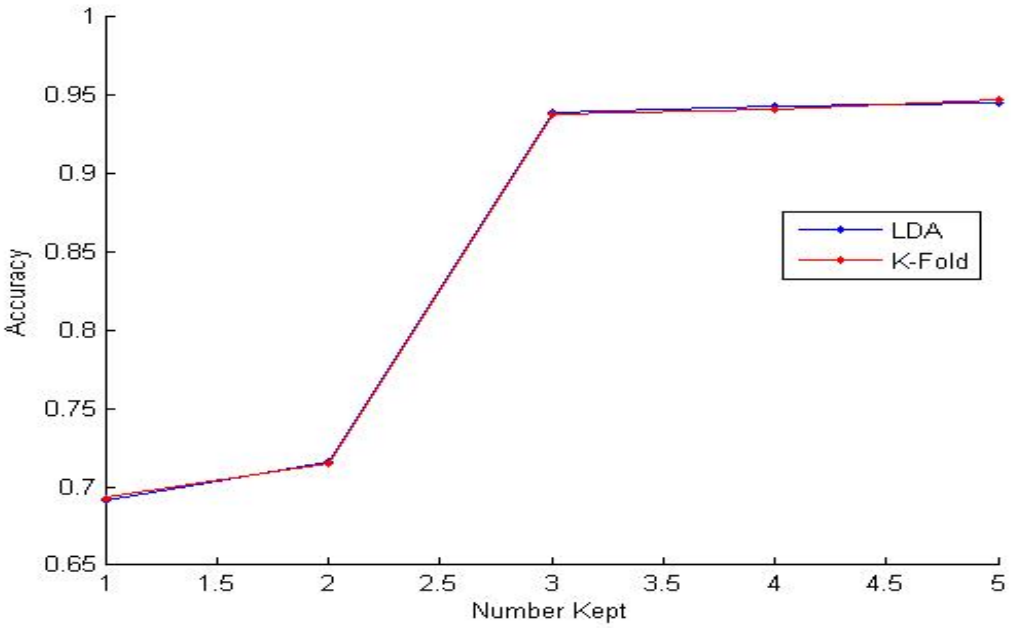

Figure 43. The accuracy for classifying all sample pixels. Nearly all degradation classification information is present within the first three principle components. In addition the close match between K-Fold cross validation and the LDA predicted accuracy suggest that these samples were measured a sufficient number of times such that noise and outliers do not effect the classification rules.

For uniformly degraded samples, classification across the samples is expected to be distributed randomly across the surface of each sample. While examining the location of the misclassifications, it becomes apparent that the majority of misclassifications occur near the boundary between the 2100 hour and 2600 hour samples as shown in Figure 44. The prevailing hypothesis is that the center of the three samples is illuminated with high intensity and measured mostly at the specular angle. As you move from the center of the field of view to the edges the incident radiance is decreased. This effect should be removed when moving to the normalized radiance factor but it is possible that classification errors could be attributed to remnants of such an effect. In addition to the change in incident radiance, the reflectance becomes

increasingly diffuse as you move from outwards from the center of the samples due to the Telops being set up at the specular angle for the center of the sample. This means diffuse and specular reflectance may be compared, reducing the overall classification accuracy.

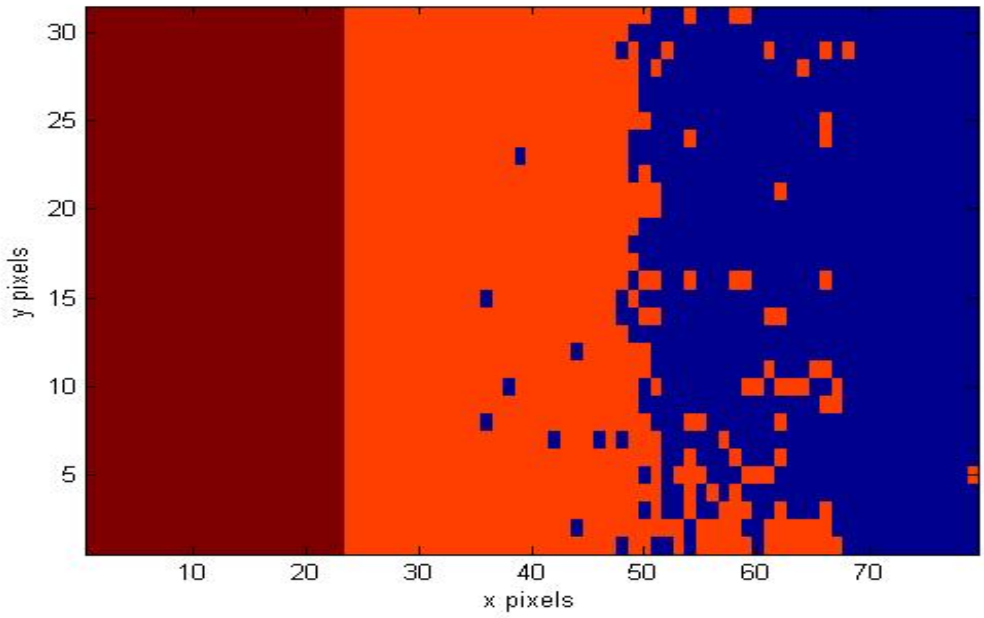

Figure 44. A pixel map of which image pixels classify as which degradation class. The location of misclassification between the middle and right samples suggests that non-uniform illumination across the three samples may be responsible for some of the misclassification.

When examining the principle components it becomes clear that either non-uniform illumination or a change from specular to diffuse reflectance, has an effect on the measured normalized reflectance factor as shown in the third principle component in Figure 45. Measuring all three samples at a non-specular angle may be a solution. Unfortunately attempts measure non-specular reflected radiance yielded insufficient reflected radiance for accurate classification. Therefore with the current experimental setup, the solution is to only measure samples that have the same surface area

as the illumination blackbody. For the current experiments this would result in only measurements made on the sample when it was in the center and near uniformly illuminated being used for classification.

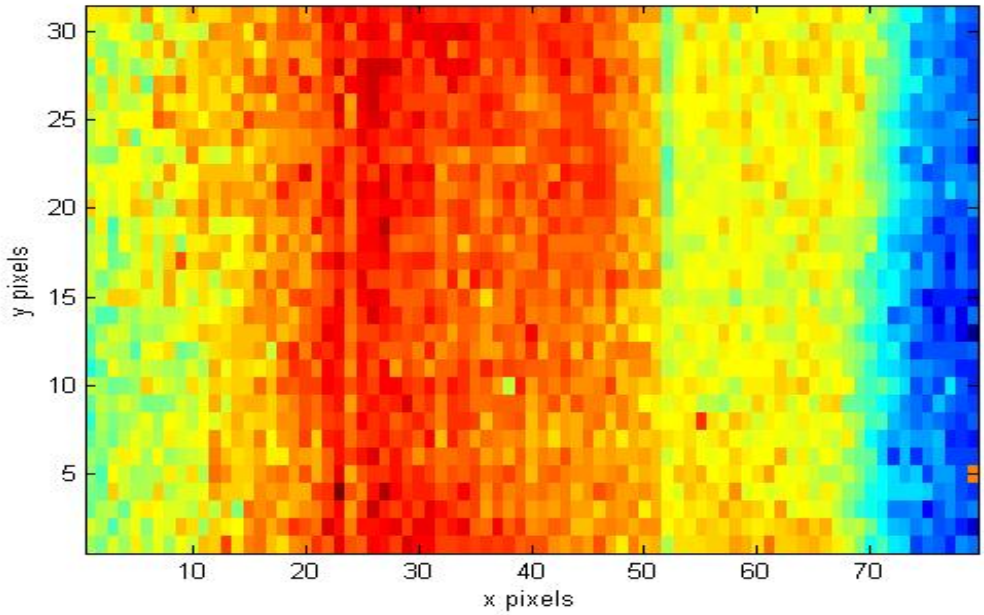

Figure 45. A pixel map of the magnitude of the third principle component. With the blackbody illumination centered on the middle sample it appears the third principle component may contain large variation associated with sample position rather than degradation level.

4.5 Instrument Comparison

At this point, there may be some concern about the validity of comparing results from spectra measured with different instruments. Due to sample and instrument availability, many of the samples have not been measured by every instrument. This makes it difficult to attribute differences in measured reflectance to change in samples or change in instrument. Thus far sample set C has been measured using all instruments mentioned except for Raman spectroscopy. A comparison of the measured

reflectance for each instrument is shown in Figure 45. Analysis on some of the instrument techniques focused on a normalized reflectance measurement over an absolute reflectance measurement so all absolute reflectance measurements were normalized for comparison.

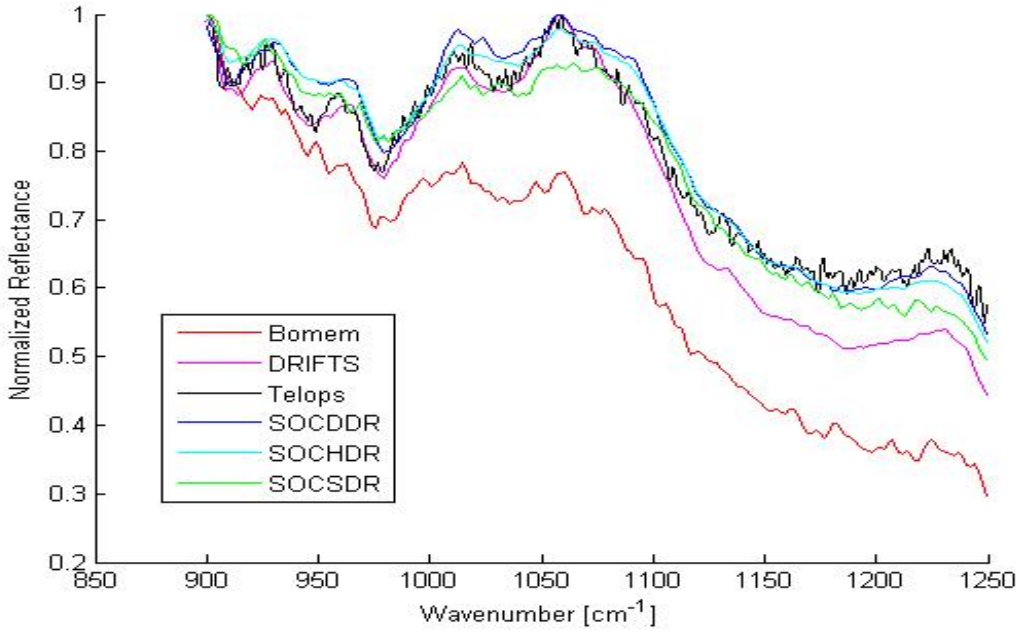

Figure 46. Comparison of the various instruments used to measure the longwave infrared reflectance of the degraded samples. All measurements shown are of the pristine sample of sample set C. The Bomem outlier may have been affected by untracked sample heating during the measurement.

The majority of the measured reflectance spectra agree well with the exception of the measured Bomem reflectance spectra. This may be due to sample heating during the measurement due to prolonged exposure to a 1000° C a mere 40 cm away. First the sample's self-emission, reflected background radiance and path radiance were measured. During these measurements, the sample was at room temperature with the blackbody aperture closed. Then the sample's reflectance was measured with the blackbody aperture open. These reflectance measurements were taken several

times at different angles. During this time, the samples heated due to proximity to the blackbody. The samples became noticeably warm to touch but the sample temperatures were not measured. Because of this error, the subtracted self-emission does not match the actual self-emission of the now heated sample. Further this change in the sample's self-emission is not spectrally uniform and thus persists after normalization. Lastly the experiment could not be repeated due to lack of instrument availability once the measurement error was identified and priority shifting to hyperspectral measurements.

4.6 Bomen Fourier Transform Spectrometer Emittance

The low reflectance samples required high temperature blackbody illumination to generate a measurable reflected radiance signal. Emittance measurements do not require 1000 °C blackbody illumination. Measuring and classifying emittance is desirable over classifying with reflectance if the two measurement techniques have similar accuracies.

To test if an emittance measurement can be used to classify degradation, the emitted radiances from the three degraded samples from sample set C were measured at 30 °C, 40 °C, and 50 °C all above room temperature. The expectation was the higher the samples' temperatures, the more the samples' self-emission would dominate the measured radiance and thus generate more accurate classification. The average measured radiances for each sample at each temperature are shown in Figure 47. Each set of curves line up well with the expected curve for a blackbody at that temperature. The water and CO_2 absorption regions are again expected to provide little spectral information regarding the samples due to large amounts of noise in these regions related to changes in the atmosphere and not changes in the samples. In addition the low signal seen at higher wavenumbers is also expected to be problematic for

classification due small variation in measured radiance resulting in large variation in the calculated reflectance.

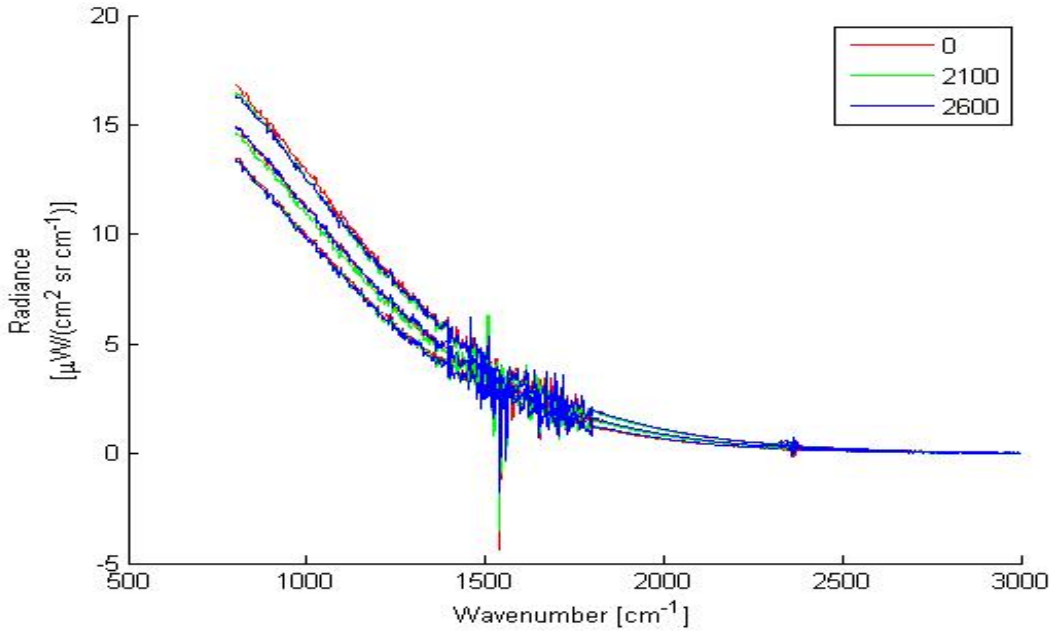

Figure 47. The emitted radiance of the three artificially weathered samples at three different temperatures. Each the spectra group well with other measurements at the same temperature. Unfortunately the water and CO_2 absorption regions provide little spectral information about the samples. In addition the signal approaches zero at higher wavenumbers.

Throughout the experiment the hot plate reported temperature variations of ±1°C. The blackbody spectrum for each sample degradation level and temperature are shown in Figure 47 were each fitted to the emittance and blackbody model outlined in Section 2.4. The fitted temperatures are shown in Figure 48; as expected, the average standard deviation of the fitted temperatures was 1.04°C. In addition there appeared to be no relationship between the average variation and the degradation level of the samples. Every sample fitted temperature underestimated the hot-plate reported sample temperature. It is also possible that the temperature underestimate

may be the result of the emittance being less than one across the spectrum. The magnitude of the underestimate increases as the sample temperature increases. It is believed that when the sample reaches thermal equilibrium the sample surface temperature, the temperature of the emitting surface, is lower than the reported hotplate surface temperature. This effect is more pronounced as the difference between room temperature and hot plate surface temperature increases. This should not affect the classification algorithm as long as the temperature is estimated for the sample surface rather than assumed to be equal to the reported hot-plate temperature.

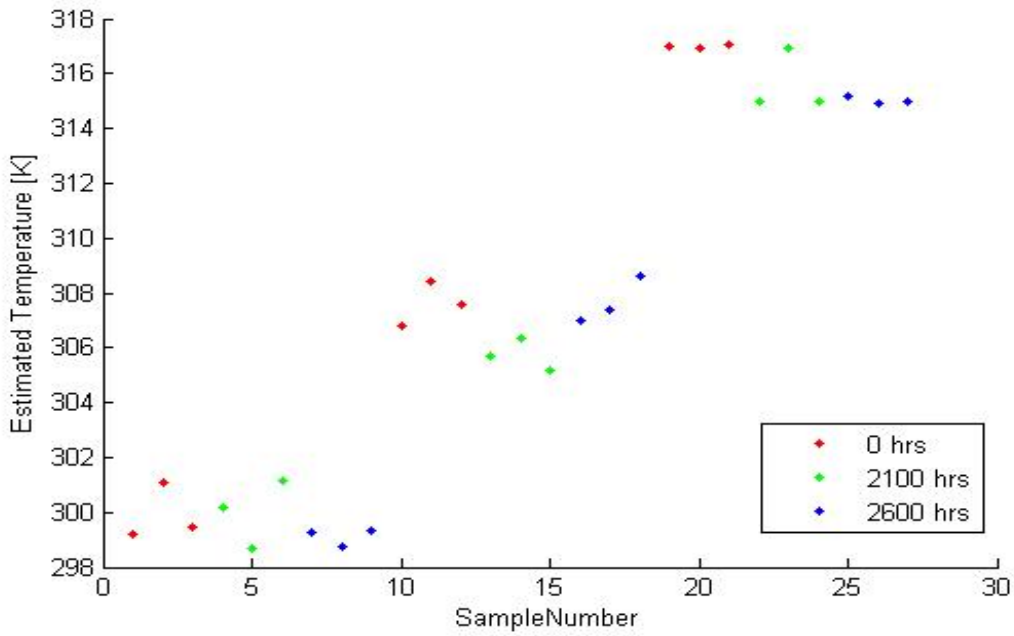

Figure 48. Sample temperature estimates using measured radiance spectra. The sample radiances appear to come from a samples of lower temperatures than reported by the calibrated hotplate. Indicating a temperature difference between the sample surface and the hot plate surface.

The estimated emittance of the samples resulted in low signal-to-noise ratio spectra. The average fitted emittance spectra for each sample and fitted temperature are shown in Figure 49. The water and CO_2 absorption regions were removed due

extreme noise and physically impossible emittances. This was expected due to calibrations problems with these regions. In addition the extremely low radiance signal measured for high wavenumbers resulted in high noise in the high wavenumber region for the fitted emittance. As such the spectra beyond the CO_2 region was removed as well.

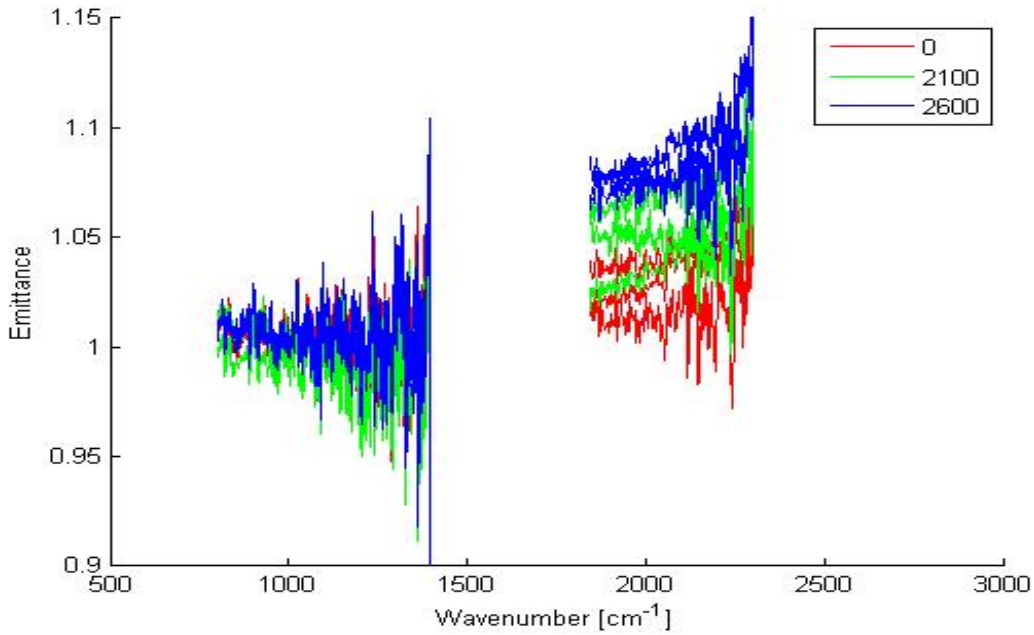

Figure 49. Bomem emittance measurements after removing the absorption and low signal regions. The measurement has a relatively low signal to noise ratio and the spectra for the three different degradation classes are not well separated.

It appears there may be sufficient spectral information for classification between the water and CO_2 absorption regions where there is some separation between the degradation classes. Two bands centered at 1300 and 2000 cm^{-1} were averaged and then linear discriminant analysis applied to each. Classification yielded 48% accuracy which is above random classification of 33% so sample degradation information is present but not sufficient for high classification accuracy. Single Value Decomposition (SVD) combined with LDA proved to be effective in the DRIFTS data and in the noisy

Bomem reflectance data. Thus SVD-LDA was applied to each region separately shown in Figure 49. Unfortunately this technique yielded only slightly better performance in the first region, 800-1456 cm^{-1}, classifying with 52% accuracy, while the second region, 1842-2301 cm^{-1}, was 49% accurate. The same technique was then used with the entire emittance spectrum with only 49% accuracy.

Following the failed attempts with LDA and SVD-LDA techniques, feature selection was applied. The ReliefF feature selection algorithm was utilized to identify any specific spectral bands that were useful for classification. Unfortunately ReliefF algorithm only identified spectral features associated with noise. This determined that while it may be easier experimentally to set up an emittance measurement, the necessary spectral information for degradation classification may not be present or easy to measure. A reflectance measurement may be more difficult and time consuming to set up, requiring a high temperature blackbody to be near the sample, but such a measurement does provide the necessary spectral information for classification with similar accuracy to applying the SVD-LDA algorithm to the DRIFTS data.

4.7 Raman Spectroscopy

The first processing done on the Raman data collected on sample set B was completed by the ChromoLogic LLC. A fifth-order polynomial was fit to each measured Raman spectrum and then the polynomial fit was subtracted from each spectrum. Then each spectrum was smoothed using a Savitzky-Golay filter [27]. This process removed the increased background signal and increased the apparent signal-to-noise ratio. Both signal to noise and background signal become increasing problematic for high energy Raman Spectroscopy where temperature changes in the sample may be present [28]. In this experiment, the laser was not high powered but it was focused, generating a high irradiance and possible temperature changes on small spots across

the sample that may not be present in an unfocused system with similar laser powers.

Unfortunately very few measurements were made with identical instrument and experiment parameters. The large variation within the experiment parameters also generated large variation within each degradation class' measured Raman spectra. This large within-class variation made classification difficult due to the comparatively small variation between the degradation classes. For this reason, the collected Raman data was separated by instrument parameters which caused significant variation and by degradation class.

Varying the input power of the laser causes the largest variation due to varying instrument parameters as can be seen in Figure 50. In addition to an overall change in peak height for each input power, the relative peak heights change, and in some cases, certain peaks cannot be distinguished from neighboring peaks. Unfortunately when applying the SVD-LDA algorithm on Raman with multiple laser powers, only 70.4% accuracy could be attained. This accuracy required retaining 25 SVD coefficients. Further it not apparent that this reported accuracy is not due to overtraining.

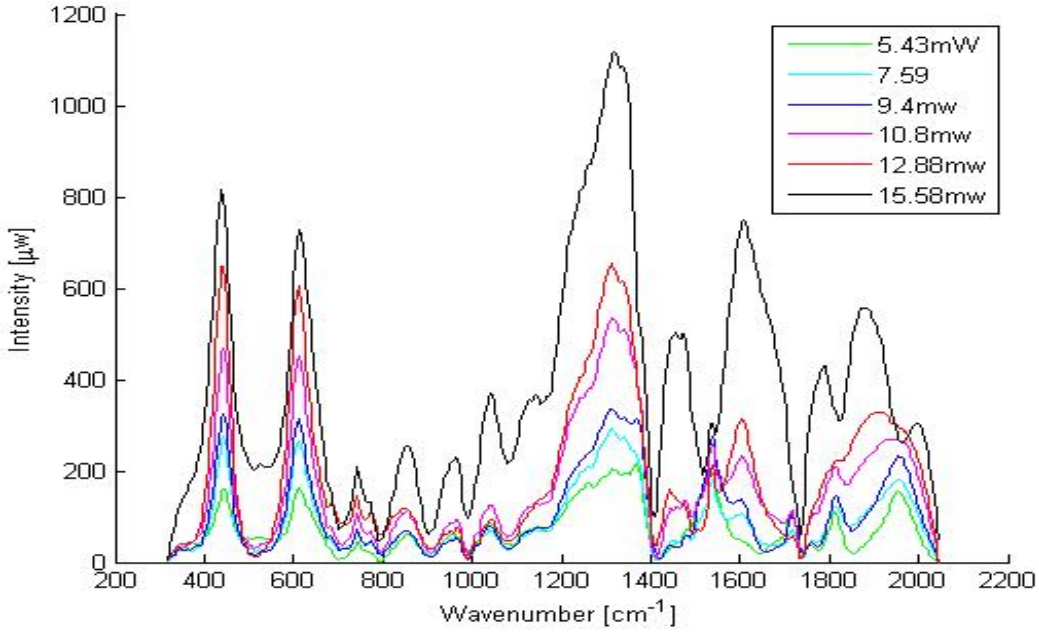

Figure 50. The Raman spectra for varying powers of incident laser is highly dependent on incident power. Classification training and feature selection must be done separately for each laser power.

Variation caused by changing the focus depth and integration time can also be noticed as shown in Figure 51. However unlike the variation caused by changing the input power, the variation caused by focus depth and integration time is primarily a change in peak height as shown in Figure 51. The location of each peak remains noticeable and the relative height of each peak remains the same. Ignoring these two parameter changes is necessary to generate enough observations from the available data for statistical significance for feature extraction and classification.

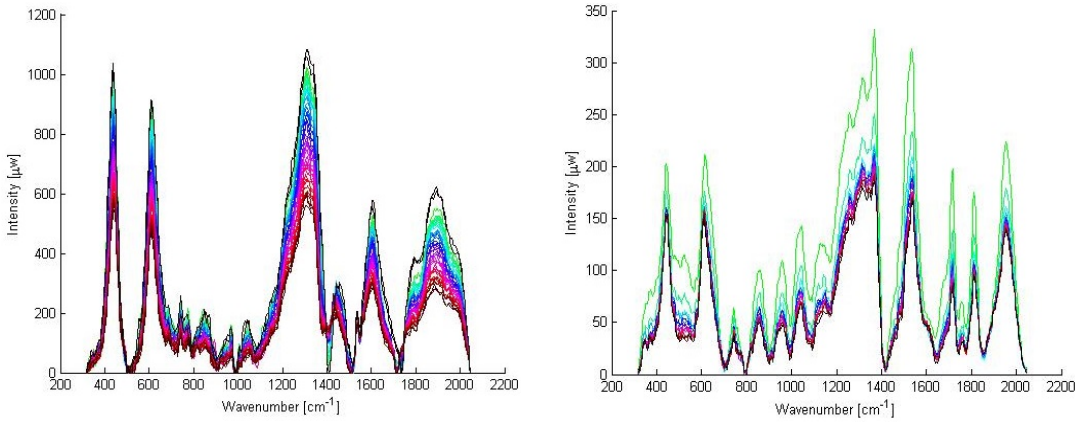

Figure 51. The Raman spectra for varying focus depths shown on left and changing time profile shown on right. The variation due to these changing instrument parameters is noticeable but observations taken at different focus depths and time profile's will be treated as part of a the same class

The four degradation levels were classified via the combined SVD-LDA algorithm for each input power separately. The results for each power level are shown in Figure 52. It appears that for the same number of SVD coefficients and equal classification complexity, certain lower laser powers resulted in higher classification accuracy. When applying the SVD-LDA technique to DRIFTS, only two or three coefficients were required to attain high accuracy. When classifying using the SVD coefficients of the Raman data, it is necessary to retain additional SVD coefficients to attain similar accuracy to the DRIFTS data. This is because for DRIFTS the first three coefficients represent 99% of the variation in the data, while in Raman the first 16 coefficients represent 90% of the variation in the data and significantly more are required to represent 99% of the variation in the data. Thus it is expected that more coefficients are required to produce similar accuracy.

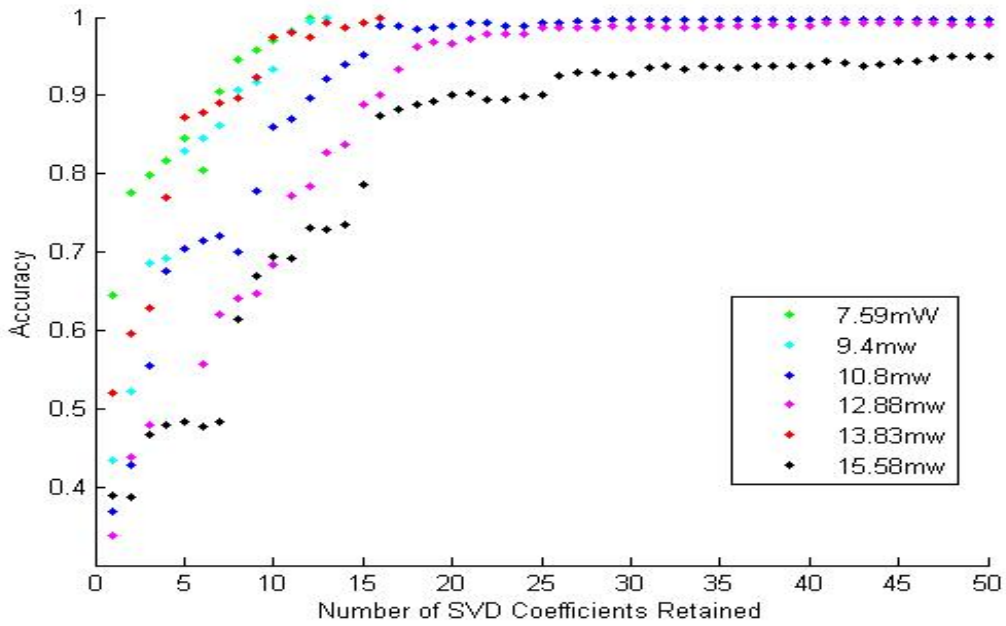

Figure 52. SVD-LDA classification accuracy with Raman data improves as additional SVD coefficients are retained for LDA. High accuracy is attained as more variation within the data is represented by more SVD coefficients. In addition lower laser input power appears to provide higher accuracy.

Raman spectroscopy can be used to classify the degradation levels of aircraft. However variation in input power affects the accuracy of the classification significantly such that it would be necessary for a Raman measurement to use the same input power across all training and validation measurements. The focus depth of the Raman measurement can cause additional, possibly non-uniform, variation throughout data collection. The engineering of a portable Raman system which can maintain focus depth accuracy on the order of millimeters is necessary and may prove difficult. Infrared reflectance spectroscopy is less susceptible to problems with changing input power due to the ease in which the infrared reflectance measurements can calibrate out the input power variation. Thus it can be concluded that infrared reflectance measurements show more promise for degradation classification than Raman spec-

troscopy.

4.8 Feature Selection-DRIFTS

The SVD-LDA classification technique useful in determining degradation but no information is provided about which parts of the spectrum are important for classification, and by extension, any information about the chemistry of the degradation process. By applying feature selection and then MDA, we can determine the vibrational energies associated with the degradation process. Forward selection requires very few of the 1700 possible features for high accuracy as shown in Figure 53.

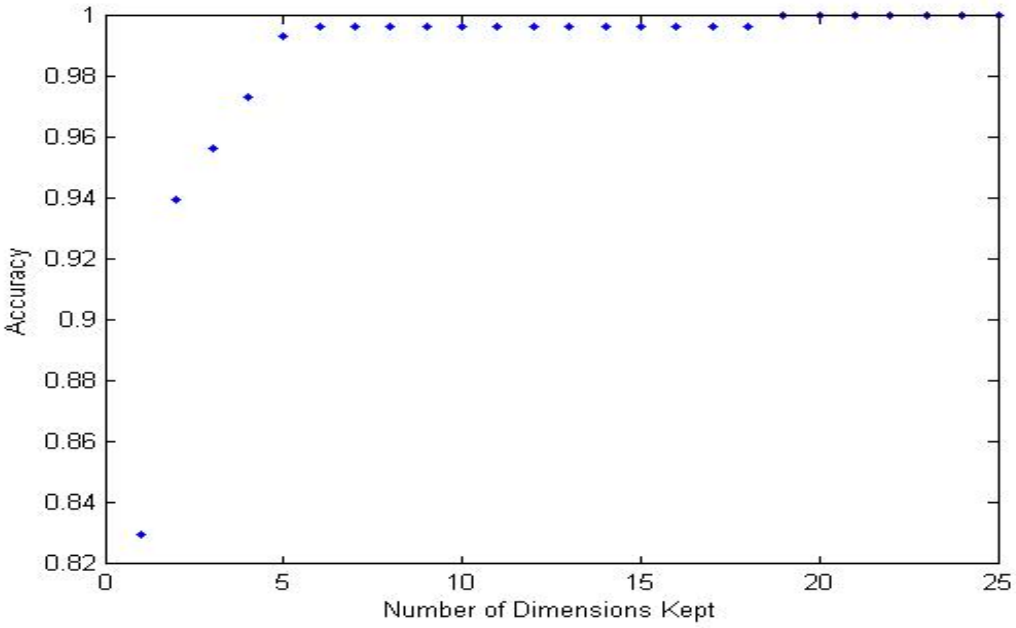

Figure 53. Forward selected features' MDA accuracy increases rapidly for the first few retained dimensions. Then additional dimensions offer diminishing returns until 100% accuracy is attained.

ReliefF weights each dimension independently of the other dimensions. Neighboring spectral channels are likely to be weighted similarly. When retaining additional dimensions each dimension added may be weighted highly but not contain degradation

information different than that already retained. The result being more dimensions must be retained for similar accuracy to that of Forward Selection where dimensions which contain new degradation information are more likely to be retained.

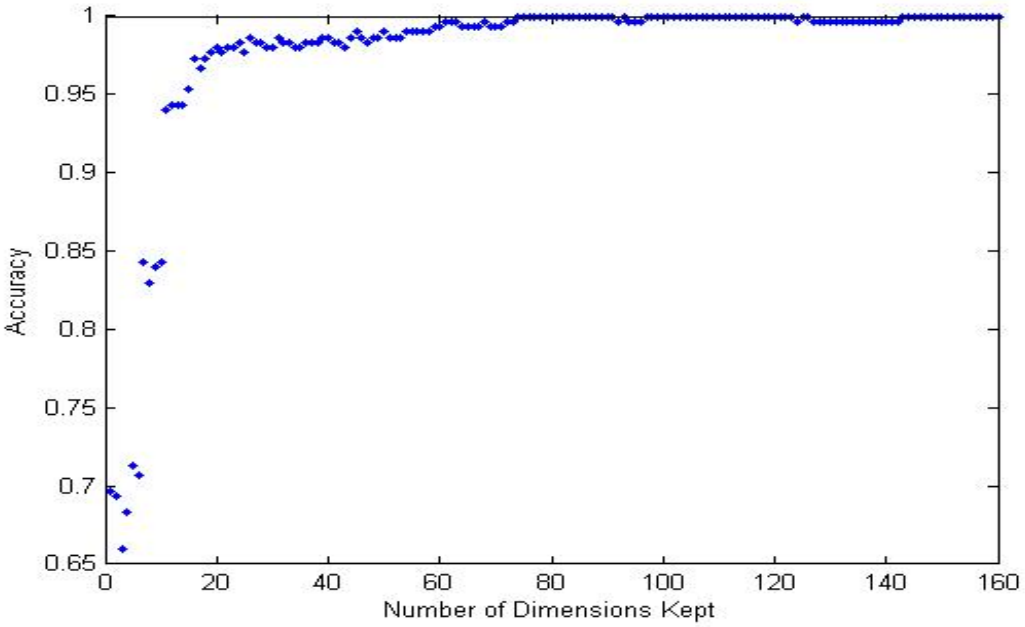

Figure 54. ReliefF selected features's MDA accuracy increases much slower than for Forward Selection. Similar to Forward Selection added dimensions eventually add dimensioning returns until 100% accuracy is attained.

The first 16 ReliefF features and the first six forward selection features are shown in Figure 55 for the DRIFTS measurements of sample set B. These number of features were chosen because they are the number of features required to classify with 99% accuracy.

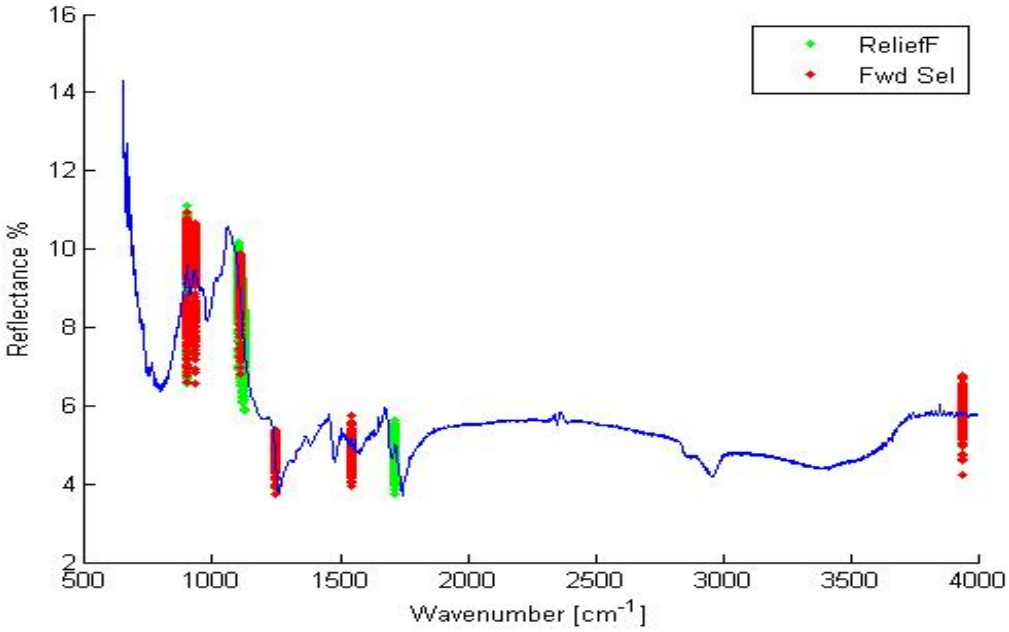

Figure 55. The first six forward selected features and 16 ReliefF features. Many of the selected features are appear within a single band of selected features.

Although there are is significant number of features, many of the features fall within the same band, or set of neighboring spectral channels. This is expected since not all spectral channels are independent, but are treated as independent for the feature selection algorithms. In reliefF, two neighboring spectral channels with similar degradation information will each be ranked similarly for selection. During classification, the inclusion of an additional neighboring spectral channel is not expected to greatly increase classification accuracy because neighboring channels tend to be highly correlated. Forward selection only uses classification accuracy to determine features and thus is less likely to select neighboring features which lack additional information compared to the earlier selected features.

With the information provided by feature selection, four spectral bands were chosen as potential filter locations for a simulated multi-spectral measurement to determine degradation classification. In addition, a fifth 'normalization' band was chosen

Table 2. Filter Center Wavenumbers and Bandwidths

Custom Filter		Comerically Available	
Center Wavenumber $[cm^{-1}]$	Bandwidth $[cm^{-1}]$	Center Wavenumber $[cm^{-1}]$	Bandwidth $[cm^{-1}]$
2229.3 (Normal Filter)	37.310	2218.1(Normal Filter)	97.008
1546.5	3.7310	1554.0	74.622
1244.3	3.7310	1268.6	63.429
1106.3	18.655	1100.7	37.311
901.06	70.890	945.83	33.580

at a location where the spectra were visually indistinguishable for each degradation class. The normalization band is used to scale the other measured spectral bands. This scaled or normalized feature space is then independent of many factors that could affect an absolute reflectance measurement such as distance from target, integration time, and spectrally uniform calibration errors. This simplifies instrument calibration. If the measured spectrum is accurate within a multiplicative constant, this measurement technique will work. This reduces the training time and possibility for mistakes in the potential fielded multi-spectral measurement technique.

After the bands were selected, the bandwidths were determined by simulating different filter bandwidths for each spectral band. The simulated bandwidths ranged from 2-200 cm^{-1}. The spectral bands and bandwidth combination which generated the highest classification accuracy were then selected as the optimal multi-spectral filter locations. The optimal bands and bandwidths are shown in Table 2. The simulated multi-spectral diffuse reflectance measurement was then used for classification with LDA resulting in 95% accuracy.

This approach unfortunately requires using expensive custom made filters. Using commercially available filters with similar central wavenumbers and bandwidths could reduce experimental and fielded costs. With this is mind, commercially available filters were chosen from Spectrogon's public list of long-wave bandpass filters [30]. A gaussian was used to approximate the filter response function for each of the filters

listed in Table 2. The resulting accuracy for this limited filter set was 94%, a mere 1% lower than the custom filters despite the majority of the filters having larger-than-optimal bandwidths. These results support the hypothesis that a multi-spectral imager can be used for degradation classification.

V. Conclusions

5.1 Achievements

This research investigated whether infrared reflectance, infrared emittance and Raman spectral measurements can be used to classify aircraft coating degradation. Each measurement method was able to classify aircraft coating degradation with greater than random accuracy. Applying the SVD-LDA algorithm to infrared emittance spectral measurements classified the worst, achieving only 52% classification accuracy. The SVD-LDA algorithm applied to Raman spectroscopy measurements, classified better with a classification accuracy of 70.4%. The SVD-LDA algorithm applied to multiple infrared reflectance measurements classified the best achieving 100%, 99.83%, and 94.4% classification accuracy when using the Bomem FTS, DRIFTS, and Telops respectively. While the Telops hyperspectral images had lower success rates than other possible instruments measuring infrared reflectance, it was able to image an entire sample at once while the other instruments would require numerous measurements to cover the entire sample. These results suggest that infrared reflectance still outperforms other measurement techniques for classifying aircraft coating degradation.

The SVD-LDA classification technique pioneered by Captain Hans Korth was automated for use with all measurement techniques in this experiment. With the automatic implementation of the SVD-LDA classification technique, a more exhaustive search for the spectral region with the highest classification accuracy was conducted for sample set B. The new fingerprint region was identified as 865.6-1238.7 cm^{-1}, which is near the previously identified fingerprint region of 850.7-1220 cm^{-1}. This region performed better than the previously identified fingerprint region and reduced the classification error from 3.00% to 1.67%, an error reduction of 45% for data set

B. A possible measurement error was identified that explained much of the classification error associated with DRIFTS measurements for sample sets A and C. A slight algorithm change, to include reflectance spectra normalization prior to applying the SVD-LDA classification technique, significantly increased classification accuracy from 63.9% to 100% for sample set A. With only slight modifications to the SVD-LDA classification technique and examined fingerprint region region, large classification accuracy increases were observed. Lastly when feature selection was applied to the DRIFTS data, it became apparent only a few spectral bands are required to achieve similar accuracy to using the entire measured spectral range. The five selected spectral bands were identified as probable filter locations for a multi-spectral measurement. Using these five filter locations, a multi-spectral measurement was simulated using the DRIFTS data. Utilizing only five filters produced a classification accuracy of 95%, compared to the classification accuracy of 99% when using the full spectra. Following the success of the simulated optimal filters, commercially available filters near these bands were simulated, yielding a minor drop in classification accuracy to 94%.

5.2 Future Work

Subsequent work regarding this research would likely begin with similar measurement and classification to be repeated on the entirety of sample set D, specifically, the DRIFTS and Telops infrared reflectance measurements which yielded the highest classification accuracy and show the most promise for testing future instruments that could be used in aircraft maintenance; both systems are likely still available with AFIT and AFRL. Unfortunately all of the samples in sample set D were not available for measurement during this research. Should the classification technique no longer work for this new data set, re-examining the fingerprint region search may yield a different fingerprint region for the different coating combinations in that data set.

In addition, modifications to the feature extraction, features selection or classification algorithms could be done. More complicated classification algorithms discussed throughout the work but not tested thoroughly would likely increase classification accuracy. It should also be anticipated that when classifying sample set D, the classification accuracy will probably decrease because the sample ages are more closely spaced than in previous sets. Therefore a direct comparison of classification accuracy between the sample sets may not be an appropriate method of comparison to determine if the SVD-LDA classification technique works for sample set D another method of comparison or metric of success may be necessary.

The potential of multi-spectral measurements to provide a cheaper measurement and classification technique than hyper-spectral imaging should be investigated. Probable filter locations were determined using feature selection on the DRIFTS data for sample set B. The Bomem and Telops data suggests that features in the CO_2 and H_2O may not be of much use due to rapidly changing atmospheric absorption in these regions. A multi-spectral measurement can be artificially generated using the Telops hyperspectral data, but only a few of the selected features are within the Telops spectral range. It is expected that multi-spectral imaging will provide lower classification accuracy than hyperspectral imaging; however, this reduction in classification accuracy may be acceptable for the reduced price and required training for equipment.

A very small number of available algorithms were chosen for this project. There are numerous additional algorithms that can be used for the feature selection, feature extraction and classification algorithms. An additional feature selection algorithm which may be of use is the Fast Correlation Based Filter (FCBF). FCBF is useful for removing selected features which are highly correlated to other selected features [31]. It is believed that some of the selected spectral features are highly correlated

with one another. FCBF may identify additional spectral regions which are not highly correlated with previously selected ones. The simplest change to the feature extraction method is applying a decomposition other than SVD when doing principle component analysis.

References

1. Forman, David A. and others *The Annual Cost of Corrosion for Air Force Aircraft and Missile Equipment.* LMI Govnerment Consulting, Report MEC81T2; 2009. `http://corrdefense.nace.org/corrdefense_summer_2009/images/AFCostofCorrosion.pdf.`

2. Korth, Hans G. *Investigation of the Potential for FTIR as a NonDestructive Inspection Technique for Aircraft Coating Degradation.* MS thesis, AFIT/ENP/14-M-19 School of Engineering and Management, Air Force Institute of Technology, Wright-Patterson AFB OH, March 2014. (ADA598862).

3. Yang, Xiong F. and others, "Degradation of Low Glosss Polyurethane Aircraft Coating Under UV and Prohesion Alternating Exposures," *Polymer Degradation and Stability.* 80:51-58. (Feb 2003).

4. Perera, Dan Y. "Physical Ageing of Organic Coatings," *Progress Organic Coatings.* 47:61-76. (July 2003).

5. Johson, B. W. and R. Mcintyre, "Analysis of Test Methods for UV Durability Polymer Coatings," *Proceedings of the 20th International Conference in Organic Coatings Science and Technology.* 27:95-106. (April 1996).

6. Yang, Xiong F. and others. "Blistering and Degradation of Polyurethane Coatings Under Differenct Accelerated Weathering Tests," *Polymer Degradation and Stability.* 77:103-109. (2002).

7. Yilgor, Iskender and others, "FTIR Investigation of the Influence of Disocyanate Symmetry on the Morphology Development in Model Segmented Polyurethanes," *Polymer.* 47:4105-4114. Istanbul: Guildf, (May 2006).

8. Griffiths, Peter R. and James A. DeHaseth. *Fourier Transform Infrared Spectrometry.* (2nd Edition). New Jersey: Wiley-Interscience, 2007.

9. Zill, Dennis G., and Michael R. Cullen. *Advanced Engineering Mathematics.* (3rd Edition). Boston: PWS-KENT, 760-75. 1992.

10. Dereniak, Eustace L. and Glenn D. Boreman. *Infrared Detectors and Systems.* New York: Wiley, 1996.

11. Gasiorowicz, Stephen. *Quantum Physics.* (3rd Edition). Hoboken, NJ: Wiley, 2003.

12. Eismann, Michael T. *Hyperspectral Remote Sensing. Bellingham.* Washington: SPIE, 2012

13. Revercomb, Henry E. and others, "Radiometric Calibration of IR Fourier Transform Spectrometers: Solution to a Problem with the High-Resolution Interferometer Sounder." *Applied Optics.* 27:3210-3218. (1988).

14. Leng, Yang. *A Materials Characterization: Introduction to Microscopic and Spectroscopic Methods.* Singapore: J. Wiley, 2008.

15. Hastie, Trevor and others. *The Elements of Statistical Learning Data Mining, Inference, and Prediction.* (2nd Edition). New York: Springer, 106-113. 2009.

16. Alpaydin, Ethem. *Introduction to Machine Learning.* (2nd Edition). Massachusetts: MIT, 2010.

17. Cha, Sung-Hyuk, and Charles Tappert. "A Genetic Algorithm for Constructing Compact Binary Decision Trees," *Journal of Pattern Recognition Research.* 4:1-13. (2009).

18. Braga-Neto, Ulisses. M. and Edward R. Dougherty. "Is Cross-validation Valid for Small-sample Microarray Classification?"" *Bioinformatics.* 20:374-380. (2004).

19. Bengio, Yoshua, and Yves Grandvalet. "No Unbiased Estimator of the Variance of K-Fold Cross-Validation," *Journal of Machine Learning Research.* 5:1089-1105 (2004).

20. Rodriguez, Juan D. and others. "Sensitivity Analysis of K-Fold Cross Validation in Prediction Error Estimation," *IEEE Transactions On Pattern Analysis and Machine Intelligence.* 32:569-575. (January 2010).

21. Baker, Kirk. "Singular Value Decomposition Tutorial," 2013. 11 March 2015 `http://www.ling.ohio-state.edu/~kbaker/pubs/Singular_Value_Decomposition_Tutorial.pdf`.

22. Kira, Ken J. and Larry A. Rendell. "A Practical Approach to Feature Selection," *Proceedings of the Ninth International Workshop on Machine Learning.* Spain: Univsity of Granada, 249-256. (July 1992).

23. Sun, Yijun, and Jian Li. "Iterative RELIEF for Feature Weighting," *Proceedings of the 23rd International Conference on Machine Learning.* Florida: Interdisciplinary Center for Biotechnology, 913-920. 2006.

24. *SOC-100 User's Manual.* San Diego: Surface Optics Corportation, 2009.

25. Verdeyen, Joseph Thomas. *Laser Electronics.* (3rd Edition). New Jersey: Prentice-Hall, 234-238. 1981.

26. Cooney, Adam. Air Force Research Labratories Material Directorate, Wright-Patterson Air Force Base OH. Personal Correspondence. 7 May 2015.

27. Luo, Jianwen and others. "Properties of SavitzkyGolay Digital Differentiators," *Digital Signal Processing.* 15:122-136. (2005).

28. West, Yvonne D. " "Study of Sample Heating Effects Arising During Laser Raman Spectroscopy,"" *The Internet Journal of Vibrational Spectroscopy.* (2004).

29. Yang, Xiong F. and others. "Blistering and Degradation of Polyurethane Coatings under Different Accelerated Weathering Tests," *Polymer Degradation and Stability.* 77:103-109. (2002).

30. Spectrogon. *Narrow Bandpass Filters.* 22 September 2014. `http://www.spectrogon.com/product-services/optical-filters/spectrogon-us/narrow-bandpass-filters.`

31. Yu, Lei and Huan Liu. "Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filer Solution," *Proceedings of the Twentieth International Conference on Machine Learming.* Arizona: Department of Computer Science and Engineering Arizona State Univeristy 2003

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704–0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202–4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD–MM–YYYY)*
30–03–2015

2. REPORT TYPE
Master's Thesis

3. DATES COVERED *(From — To)*
Sept 2014 — Mar 2015

4. TITLE AND SUBTITLE
Non-destructive Techniques for Classifying Aircraft Coating Degradation

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Wilson, Kody A, 1LT

5d. PROJECT NUMBER
15P133

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Institute of Technology
Department of Engineering Physics (AFIT/ENP)
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT-ENP-MS-15-M-88

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research laboratories, Material Directorate
Dr. Adam Cooney
11th Street WPAFB OH 45433-7765
DSN 785-6573, COMM 937-255-6573
Email: adam.cooney@us.af.mil

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A:
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES
The views expressed in this document are those of the author and do not reflect the official policy or position of the United States Air Force, the United States Department of Defense or the United States Government. This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

14. ABSTRACT
In this research non-destructive techniques were examined as possible methods of determining aircraft coating degradation. Single Value Decomposition(SVD)-Linear Discriminant Analysis(LDA) algorithms were applied to measured spectra. When applied to infrared emittance spectra only 52% classification accuracy was achieved. When applied to Raman spectroscopy a higher classification accuracy of 70.4% is attained when using the same SVD-LDA algorithm. However the best performing measurement was using infrared reflectance classification accuracies were 100%, 99.83% and 94.4% when using the Bomem FTS, DRIFTS and Telops respectively for one of the sample sets.

For DRIFTS data a more accurate fingerprint region was identified 865.6 - 1238.7 cm^{-1} decreasing classification error by 50%. Feature selection was applied to determine filter locations for multi-spectral measurements. Simulating the optimal and commercially available filters accuracies of 95% and 94% were achieved using 5 filters. Infrared reflectance produces high classification accuracy when using the DRIFTS, Bomem FTS, Telops and a multi-spectral imager.

15. SUBJECT TERMS
Aircraft Coatings, Degradation, Infrared Reflectance, DRIFTS, Telops Hyperspectral

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dr. M. R. Hawks, AFIT/ENP
U	U	U	UU	112	19b. TELEPHONE NUMBER *(include area code)* (937) 785-3636, x4828; Miachael.Hawks@afit.edu

Standard Form 298 (Rev. 8–98)
Prescribed by ANSI Std. Z39.18