Wright State University

From the Selected Works of Andrew Niklas

Summer June 8, 2017

Spectral Library Material Separability using Worldview-3 and Landsat-8 Spectral Bands

Andrew Niklas

Available at: https://works.bepress.com/andrew_niklas/3/

Spectral Library Material Separability Using WorldView-3 and Landsat-8 Spectral Bands

Andrew J. Niklas
Wright State University
3640 Colonel Glenn Hwy
Dayton, OH 45435
937-775-3333
niklas.6@wright.edu

Matthew D. Sambora
Integrity Applications Incorporated
3725 Pentagon Blvd. Suite 210
Beavercreek, OH 45431
937-320-9970
msambora@integrity-apps.com

Abstract—The WorldView-3 and Landsat-8 satellites are the most recently deployed systems in their constellations and the unique data from these sensors can positively impact environmental and military target detection applications. The research team uses spectral library data in the VNIR and SWIR spectral bands of WorldView-3 and Landsat-8 to determine the best combination of spectral bands and spectral distance measure to yield the largest spectral distance value for each target material. Spectral distance measures include Euclidean Distance, Spectral Angle Mapper, Spectral Correlation Measure, and Spectral Information Divergence. The optimal configuration results are stored in a look-up-table for implementation in an automated target detection system. The Freedman-Diaconis and Shimazaki-Shinomoto methods for optimal histogram bin width determination are applied to spectral distance measures that are cross computed for each material in the spectral library and for each sensor. The bin width determination is used to characterize material clusters based on intercluster and intracluster spectral distances. The material cluster characterization results are stored in a look-up-table for fast histogram based initialization of clustering algorithms. The research team uses the in-band spectral library data for determining endmember abundance estimates based on combinations of spectral bands, endmember combinations, spectral distance measure, and additive white Gaussian noise for both sensors. The endmember abundance estimates are optimized using Differential Evolution, Least Squares, and Linear Simplex. The numerical accuracy of the endmember abundance determination is compared across the three optimization algorithms. The completion of this foundational work increases the data exploitation potential of WorldView-3 and Landsat-8 by providing a fundamental characterization of material separability with respect to these sensors.

TABLE OF CONTENTS

1. INTRODUCTION.....	1
2. RELATED WORK.....	2
3. OPTIMAL SPECTRAL DATA CONFIGURATIONS	3
4. SPECTRAL DISTANCE CHARACTERIZATION.....	7
5. ENDMEMBER ABUNDANCE ESTIMATION.....	10
6. CONCLUSIONS AND FUTURE WORK.....	17
APPENDICES.....	18
A. SCALING OF SPECTRAL DISTANCE VALUES	18
B. ADDITIVE WHITE GAUSSIAN NOISE.....	18
ACKNOWLEDGEMENTS	18
REFERENCES.....	19
BIOGRAPHY	20

1. INTRODUCTION

This paper defines material separability as the ability of an algorithm to separate a single spectral signature from a set of spectrally unique spectral signatures. Generally, this domain of remote sensing research is called target detection [1-5]. Our current research uses the United States Geological Survey (USGS) splib06a material spectral library to create a set of unique spectral signatures. The USGS spectral library contains material spectra measured in varying bandwidths from the Ultra Violet (UV) to the Long Wavelength Infrared (LWIR) and spectral resolution as fine as 1 nanometer. This research analyzes material separability using spectral signatures as they appear to sensors onboard the WorldView-3 (WV-3) and Landsat-8 (L8) satellites [6]. The WV-3 and L-8 sensors have been selected for investigation in this research because they represent the currently operational state-of-the-art for their respective constellations.

The WV-3 and L-8 sensors have spectral bands in the Visible, Near InfraRed (VNIR) and Short Wavelength InfraRed (SWIR). Therefore, this research only considers contributions to the material spectra corresponding to the WV-3 and L-8 VNIR and SWIR spectral bands. The units of the spectral data from the USGS spectral library in the VNIR and SWIR are in units of reflectance. The in-band reflectance of a spectral signature is computed as the average reflectance of the spectral signature that is contained within the particular sensor band. In this manner, the band average reflectance of each material in the spectral library is computed for each sensor band of the WV-3 and L-8 sensors. This paper describes analysis originating with the sensor band average reflectance from the spectral library but does not contain analysis from image based exploitation of the sensor data.

There are five classes of materials that are defined in the USGS spectral library. These classes are categorized as: minerals, mixtures, coatings, manmade, and biological. There are a total of 1366 material spectra in the speclib06a spectral library. However, some of the material spectra do not have any reflectance data within the sensor bands defined by the WV-3 and L-8 sensors. This research considers only the material spectra for which there is reflectance data in each of the bands of the two sensors. As a result, 439 spectral profiles are excluded from the research and 927 spectra are included.

The WV-3 sensor has 29 sensor bands: 1 visible Panchromatic (PAN), 8 VNIR, 8 SWIR, and 12 Cloud-Aerosol-Vapor-Ice-Snow (CAVIS) bands. The research

considers only 17 WV-3 bands from the visible PAN, VNIR, and SWIR band categories. The 12 CAVIS bands are used largely for atmospheric correction of sensor image data and are therefore not considered in this research. The L-8 sensor has 11 sensor bands: 1 visible PAN, 4 VNIR, 3 SWIR, and 2 LWIR bands. The research considers only the nine L-8 bands covering the visible PAN, VNIR, and SWIR bands. The LWIR bands have been excluded from the current research because they collect data originating from thermal emission phenomenology. Figure 1 contains the relative spectral response curves for the WV-3 and L-8 bands as a reference for the sensor bands used in the research

Figure 1. WorldView-3 (top) and Landsat-8 (bottom) relative spectral response curves corresponding to the sensor bands used in the research.

The research involves using the 927 in-band material spectral reflectance signatures as the starting point to conducting three independent studies. The goal of the first study is to determine the optimal configuration of sensor bands and spectral distance measure that yields the best separability of materials based on a spectral distance value [7, 8]. The goal of the second study is to determine initial settings derived from histogram bin optimization methods for input into clustering algorithms [9]. The goal of the third study is to determine the mathematical optimization routine applied to material abundance estimation from the Linear Mixing Model (LMM) that yields the best overall accuracy performance [10]. Prior to describing in detail the three studies, it is worthwhile to explain how the research is related to Automated Target Detection (ATD).

2. RELATED WORK

The research presented in this paper is performed in anticipation of the results being directly applied in an ATD system using remotely sensed multispectral imagery. Therefore, it is useful to explain how the research is related to the problem of ATD in such applications. The fundamental test of the performance of a target detection system is the ability of the system to maximize target detections while minimizing false positive detections with high confidence [11].

One domain of multispectral image based target detection is image segmentation and classification. An active area of research in this domain is the statistical modeling of backgrounds [12, 13]. Some of the legacy approaches to background modeling involve fitting model parameters to multispectral imagery data given a mathematical function that is thought to describe the distribution. The model parameters define the Maximum Likelihood Estimates (MLE) of the data. Model fitting approaches require that the input data are segmented into clusters of a single distribution [14] before determining the MLE for a particular function.

Multispectral remotely sensed data, specifically hyperspectral data, can be very large in the spectral dimension. Although WV-3 and L-8 data do not experience spectral oversampling, the discussion of dimensionality reduction is still applicable for identification of a set of orthogonal endmembers from a larger set. The size of the data causes computational challenges that can be overcome by the discipline of dimensionality reduction [15, 16]. Dimensionality reduction typically employs Principal Component Analysis (PCA) to remove redundant spectral information from a set of measurements. In the case of a set of measurements that constitute a background, the dimensionality reduction via PCA will essentially optimize the background set to a smaller size while maintaining all of the salient spectral information necessary for accurate image segmentation and clustering.

A domain of image based target detection that presents unique challenges is subpixel target detection. In the case where a measurement is composed of multiple signatures, a mixing model similar to effective medium theory has been employed to describe the composition of the measurement. Other approaches for subpixel detection have been explored by researchers in the form of varieties of the matched filter [17].

It has been shown that the matched filter approaches for subpixel target detection are connected with the LMM [18]. The spectral variability of the background and target, as well as the nature of background interference with the target are causes for the theoretical models used for subpixel target detection not always being in agreement with ground truth control experiments. Although the models do not always replicate the truth of physical reality, the Adaptive Cosine Estimator (ACE) has been shown to be a sufficiently robust estimate of target detection for practical applications [19].

A common approach for target detection in multispectral remote sensing image data is the matched filter. The matched filter detector is essentially connected to the LMM via the abundances, which can be solved using the unconstrained least squares estimate via the normal equations to obtain the projection matrix and therefore the goodness-of-fit via the Sum of Squares Error (SSE) of the estimated abundances [20]. It has been also shown that the means and variance MLE of a Gaussian distribution are equivalent to the abundances and SSE of the abundances from the unconstrained least squares estimate.

Ultimately the likelihood ratio test is extended to the generalized likelihood ratio test (GLRT) as a function of the MLE of the background abundances and target abundance. The GLRT and ACE has been shown to be a function of the Adaptive Matched Filter (AMF). The GLRT is commonly used to determine if the target present or target absent LMM hypothesis is satisfied and therefore constitutes the target decision rule in a target detection system [21-25].

3. OPTIMAL SPECTRAL DATA CONFIGURATIONS

The spectral distance measures are metrics that can be used to determine the similarity or conversely the dissimilarity of two spectral signatures. The goal of this section of the research is to determine the optimal configuration of sensor bands and spectral distance measure that yields the best separability of materials based on a spectral distance value. Four spectral distance measures are considered in this study: Euclidean Distance (ED), Spectral Angle Mapper (SAM), Spectral Correlation Measure (SCM), and Spectral Information Divergence (SID).

The authors acknowledge that although SAM and SCM are unique spectral distance measures, SCM can hypothetically yield the same value for two signatures that are positively correlated as those that are oppositely correlated whereas SAM would yield a value that is lower for oppositely correlated signatures compared with positively correlated signatures. Furthermore, the authors recognize that a high SCM value indicates two signatures are not strongly separated and a low SCM value indicates two signatures are strongly separated. The inverse behavior is true for SAM and SID calculations and these facts have been accounted for in the research algorithms. The four spectral distance measures for the two spectral vectors $\vec{x}$ and $\vec{y}$ that contain N members are defined as

$$ED = \sqrt{\sum_{i=1}^N (x_i - y_i)^2} = \|\vec{x} - \vec{y}\|_2 \quad (1)$$

$$SAM = \cos^{-1} \left(\frac{\vec{x}^T \vec{y}}{\|\vec{x}\|_2 \|\vec{y}\|_2} \right) \quad (2)$$

$$SCM = \frac{|N \vec{x}^T \vec{y} - (\sum_{i=1}^N x_i)(\sum_{i=1}^N y_i)|}{\sqrt{|(N \vec{x}^T \vec{x} - (\sum_{i=1}^N x_i)^2)(N \vec{y}^T \vec{y} - (\sum_{i=1}^N y_i)^2)|}} \quad (3)$$

$$SID = \sum_{i=1}^N \left(\frac{x_i}{\sum_{j=1}^N x_j} \ln \left(\frac{x_i}{\sum_{j=1}^N x_j} \frac{\sum_{j=1}^N y_j}{y_i} \right) + \frac{y_i}{\sum_{j=1}^N y_j} \ln \left(\frac{y_i}{\sum_{j=1}^N y_j} \frac{\sum_{j=1}^N x_j}{x_i} \right) \right) \quad (4)$$

The $\|\cdot\|_2$ in Equation 1 and Equation 2 is the vector two-norm, and the $\ln(\cdot)$ in Equation 4 is the natural logarithm.

The procedure for computing the spectral distance values for all unique material pair combinations is as follows. The in-band material reflectance signatures are computed for the 927 materials from the spectral library from WV-3 and L-8. Each material will have 926 different materials for which the spectral distance can be computed. Computing all possible unique combinations of materials yields $927 \times 926 = 858,402$ spectral distance values. Next, the four spectral distance measures are computed for each unique material pair combination. The computations result in $4 \times 927 \times 926 = 3,433,608$ spectral distance values for a single sensor and single unique combination of sensor bands.

Selection of a sensor band configuration for each sensor enables an initial qualitative analysis of the spectral distances. The authors naively select all 17 WV-3 sensor bands and all nine L-8 bands for computing the spectral distances in the qualitative analysis. The analysis examines four graphical representations of the spectral distance data and observe any trends. The first examination is the comparison of SAM versus SID using all 17 WV-3 bands, shown in Figure 2. The comparability of the results using four different spectral distance measures is facilitated by scaling the results of the four measures as described in Appendix A: Scaling of Spectral Distance Values.

Figure 2. Graphical comparison of nearly a million SAM (horizontal axis) and SID (vertical axis) values for the WorldView-3 sensor bands.

Referring to Figure 2, the material pairs have been sorted according to increasing SAM value. The SID values follow the sorting of the material pairs by SAM value such that a single point in Figure 2 represents the x and y, or equivalently SAM and SID value, for a unique material signature pair. The graphic in Figure 2 shows that there are many material pairs for which the SAM value disproportionately increases compared to the SID value, and indicates that SAM can have advantages over SCM for material pair separability in this region.

For example, there are many material pairs following the upper bound of the plotted points for which SAM is increasing up 0.5 while SID only increases to 0.2. There are even some material pairs following the lower bound of the plotted points for which the SAM reaches 0.8 when the SID has only increased to 0.2. Recall that these observations are only valid when considering all 17 WV-3 sensor bands in the computation of the spectral distance measures. Changing the spectral band combination will conceivably change the observed patterns.

The second examination is the comparison of SCM versus SID using all nine L-8 bands, shown in Figure 3. The comparison of SCM to SID for L-8 is similar to the comparison of SAM to SID for WV-3 in that the goal is to observe qualitative patterns in the data to determine if there are regions where one measure outperforms the second measure.

Figure 3. Graphical comparison of nearly a million SCM (horizontal axis) and SID (vertical axis) values for the Landsat-8 sensor bands.

Referring to Figure 3, the material pairs are sorted according to increasing SCM value. The SID values follow the sorting of the material pairs by SCM value such that a single point in Figure 3 represents the x and y, or equivalently SCM and SID value, for a unique material signature pair. Recall that a high SCM value indicates strong correlation between signatures, whereas a low SCM value indicates the two signatures are not correlated or equivalently are highly distinguishable.

The graphic in Figure 3 shows that there are many material pairs for which both the SCM and SID values outperform one another. In general, it is observed that as SCM similarity decreases disproportionate to the SID values. The high

density of material pairs along the bottom of the graphic in Figure 3 indicates SCM to be a more advantageous measure of material separability, however, there are material pairs to the left of $SCM = 0.2$ and $SID = 0.2$ of the graphic where SID is more advantageous. These observations are only valid when considering all nine L-8 sensor bands in the computation of the spectral distance values.

The third examination is the comparison of SAM values for WV-3 versus L-8. This examination uses all 17 WV-3 bands and all nine L-8 bands for the SAM calculation. The result is shown graphically in Figure 4. The goal of this examination is to determine if one of the two sensors is preferred for distinguishing certain material pairs given the SAM distance measure.

Figure 4. Graphical comparison of nearly a million SAM values between the WorldView-3 (horizontal axis) and Landsat-8 (vertical axis) sensors.

Referring to Figure 4, the material pairs have been sorted according to increasing SAM value for the WV-3 sensor. The SAM values for L-8 follow the sorting of the material pairs by WV-3 SAM value such that a single point in Figure 4 represents the x and y, or equivalently WV-3 SAM and L-8 SAM value, for a unique material signature pair. The graphic in Figure 4 shows that there is an approximately one-to-one linear relationship between the WV-3 and L-8 SAM values. This indicates that there is no large advantage for material pair distinguishability using WV-3 or L-8 derived SAM values. Recall that these observations are only valid when considering all 17 WV-3 and all nine L-8 sensor bands in the computation of the SAM values.

The fourth examination is the comparison of SCM values for WV-3 versus L-8. This examination uses all 17 WV-3 bands and all nine L-8 bands for the SAM calculation. The result is shown graphically in Figure 5. Similar to the analysis connected with Figure 4, the goal of this examination is to determine if one of the two sensors is preferred for distinguishing certain material pairs given the SCM distance measure.

The graphic in Figure 5 shows that there is initially a one-to-one linear relationship between the WV-3 and L-8 SCM values but that the distribution of points broadens for smaller SCM values. This indicates that depending on the material

pair being considered there might be an advantage to using WV-3 over L-8 SCM and vice versa. This is in contrast to the analysis following from Figure 4 for the SAM values, where the data points do not deviate from a one-to-one slope as largely as in Figure 5.

Figure 5. Graphical comparison of nearly a million SCM values between the WorldView-3 (horizontal axis) and Landsat-8 (vertical axis) sensors.

Referring to Figure 5, the material pairs have been sorted according to increasing SCM value for the WV-3 sensor. The SCM values for L-8 follow the sorting of the material pairs by WV-3 SCM value such that a single point in Figure 5 represents the x and y, or equivalently WV-3 SCM and L-8 SCM value, for a unique material signature pair. In considering the SCM, the authors recognize that the least similar material pairs are in the bottom-left of the graphic in Figure 5 and the most similar pairs are in the upper-right of the graphic. Recall that these observations are only valid when considering all 17 WV-3 and all nine L-8 sensor bands in the computation of the SCM values.

The preceding material in Section 3 up to this point has examined four graphical representations of analysis that are possible within the scope of the research. However, the extent of analysis possible within the scope of the research extends beyond these four examinations. Therefore, the research considers how to determine the optimal configuration given the entire test domain defined by 927 materials, two sensors, and four spectral distance measures. Specifically, the optimal configuration study requires duplicating analysis similar to that which followed from Figures 2 – 4, but in a computer automated fashion and for all possible combinations of sensor bands.

The procedure for determining the optimal sensor band and spectral distance measure configuration for a single material is as follows. The optimization begins after the in-band material reflectance signatures are computed and the four spectral distance measures are computed for each unique material pair combinations. This calculation results in $4 \times 927 \times 926 = 3,433,608$ spectral distance values. This number is for a single unique combination of sensor bands for a single sensor. The number of unique combinations of a set of n elements given k elements in each combination is

given by the common mathematical grammar expression of “ n choose k ” defined in Equation 5.

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad (5)$$

And the total number of unique combinations for integer k such that $2 < k < n$ is given by Equation 6.

$$\sum_{k=2}^n \binom{n}{k} = \sum_{k=2}^n \frac{n!}{k!(n-k)!} \quad (6)$$

In the case of WV-3, $n = 17$ and there are 130,918 unique combinations of sensor bands. In the case of L-8, $n = 9$ and there are 466 unique combinations of sensor bands. Considering all band combinations for the WV-3 and L-8 sensors for a single spectral distance measure of all 927 materials therefore yields 112,380,273,036 and 400,015,332 spectral distance values, respectively. Ultimately, the optimal configuration of sensor bands and spectral distance measure to yield the largest spectral distance is determined for each material. In the case of WV-3 and L-8 with four spectral distance measures there are 523,672 and 1,864 comparisons that for a single material, respectively. The optimal configuration to maximize the separability of two materials is determined through a triage of logic. The workflow to obtain the optimal configuration is shown in Figure 6.

Figure 6. Pictured clockwise from the top-left is the workflow required to determine the optimal configuration of sensor bands and spectral distance measure for a specified pair of materials.

The result of the logic comparisons for each pair of material signatures is recorded as a key-value Look-Up-Table (LUT), which is stored on the computer in JavaScript Object Notation (JSON) format. The final LUT contains 858,402 keys, one for each unique pair of material comparisons, for each sensor. The value of a key in the LUT contains the indices to the sensor bands that optimize the spectral distance, and the spectral distance measure that yields the optimal result. The storage size of the LUT on a computer disc is approximately 230 KiloBytes (KB) for 927 materials and both sensors.

The optimal configurations which maximizes the spectral distance values of the 927 material signatures in the WV-3 sensor bands is described as follows. The WV-3 optimal configurations for all of the material signatures are defined by 48 unique configurations of sensor bands and spectral

distance measure. The results of binning the number of occurrences of each configuration across all material signature pairs is shown in Figure 7, which shows the population of the largest 32 of the 48 bins. Each of the 48 configurations contains either 16 or all 17 sensor bands and either the SCM, SID, or ED spectral distance measure. The SAM spectral distance measure does not occur in any of the 48 configurations.

An in depth look at the numerical results show that multiple spectral distance measures can have nearly the same spectral distance value for a specific sensor band combination, but the algorithm selects the optimal measure with the numerically largest distance value. The empirical manner in which the spectral measures are scaled in this research may impact this result. Referring to Figure 7, the optimal configuration with the largest bin contains 122,494 material signature pairs, and the 32nd bin contains 1,316 pairs. The four smallest of 48 configurations each have a bin count of two pairs.

Figure 7. A histogram of the number of occurrences of a unique material signature pair for a given unique optimal configuration for the WV-3 sensor bands.

The optimal configurations for maximizing the spectral distance values of the 972 material signatures in the L-8 sensor bands is described as follows. The L-8 optimal configurations for all of the material signatures are defined by 23 unique configurations of sensor bands and spectral distance measures. The results of binning the number of occurrences of each configuration across all material signature pairs is shown in Figure 8, which shows the population of the largest 16 of the 23 bins.

Each of the 23 L-8 configurations contains either eight or all nine of the L-8 sensor bands and either the SCM, SID, or ED spectral distance measure. Again, the SAM spectral distance measure does not occur in any of the 23 configurations identified by the research. Referring to Figure 8, the L-8 optimal configuration with the largest bin contains 297,464 material signature pairs and the 16th bin contains 92 pairs. The configuration with the smallest bin count contains two pairs.

Figure 8. A histogram of the number of occurrences of a unique material signature pair for a given unique optimal configuration for the L-8 sensor bands.

Comparing the results of the optimal configuration determination of WV-3 versus L-8 sensor bands it is interesting to observe that the number of unique configurations of the WV-3 results is nearly twice that of the L-8 unique configurations. What makes this an interesting observation is that nearly twice as many sensor bands are used in the WV-3 results compared with the L-8 results. The authors do not suggest a correlation, but they have also not concluded if this occurrence is a coincidence.

The conclusion of the WV-3 and L-8 comparison is that the largest spectral distance values are obtained with the most spectral information. The challenge presented by SCM being identified as one of the two preferred spectral distance measures is that SCM does not discriminate between positively and negatively correlated signatures which means logic must identify differences in signature polarity.

The benefit of the LUT is that it can be used as a reference in ATD systems to identify which algorithms will theoretically yield the largest spectral distance values for specified targets. Additionally, the LUT will provide a theoretical answer as to which sensor should be used to detect a particular target. This can be achieved using the LUT generated in this research by comparing the optimal configuration for a material signature pair and select either WV-3 or L-8 based on the largest spectral distance value.

The optimal configurations of sensor bands and spectral distance measure for a specific material are designed as input to the histogram-based distance characterization to determine the best initialization of clustering algorithms given a target material. The determination of the inter-observation distance in this manner is intended to remove the need to make assumptions about the number of background members in a population of spectral measurements. The determination of the inter-observation distance using histograms is the research discussed in the next section.

4. SPECTRAL DISTANCE CHARACTERIZATION

The spectral distance characterization study is an attempt to determine a histogram based method for the initialization of input parameters to clustering algorithms such as K-Means and Topological Anomaly Detection (TAD) because a histogram is a simple nonparametric density estimator. Typical implementations of K-Means and TAD require apriori information to be provided by the user to complete the parameter space needed by the model as defined in the clustering algorithm [26-28]. The apriori information is usually knowledge about the overall behavior of the distribution of points to be clustered.

The TAD clustering algorithm for example requires initialization of the inter-observation distance, also known as the graph resolution. The mapping function must be a similarity measure and although it is common in the TAD literature to use the Euclidean distance [29 - 31], other spectral distance measures can be used. In contrast, the K-Means algorithm is strictly formulated using the Euclidean distance.

The goal of this study is therefore to determine characterizations that macroscopically represent the overall distribution of spectral distance values for the materials, sensors, and spectral distance measures under consideration. To accomplish this goal, the research examines two techniques used for histogram bin optimization that are based on the underlying distribution of points. These two techniques are the Freedman-Diaconis (FD) and Shimazaki-Shinomoto (SS) bin width optimizations. During the discussion, the authors refer to the collection of data points within a specific bin of a histogram as a cluster of points, despite the fact that the overall distribution of points might not exhibit obvious macroscopic clustering behavior.

The FD bin width is defined as the ratio of the Inter-Quartile Range (IQR) of a distribution of points f to the number of elements N in f [32, 33]. The FD is designed to minimize the difference between the area under the distribution of points and the area under the theoretical distribution. The formula for computing the FD bin width Δ_{FD} is

$$\Delta_{FD} = 2 \frac{IQR(f)}{N^{\frac{1}{3}}} \quad (7)$$

The FD bin width formula is already in an optimized form as defined in Equation 7. In contrast to computing the FD bin width, the SS histogram optimization is an iterative formulation. The SS optimization seeks to minimize an estimate to the Mean Integrated Squared Error (MISE) between a histogram and the underlying distribution of points. The SS optimization is expected to avoid a paucity of samples in each bin, while maintaining sufficient histogram resolution such that the histogram adequately represents the underlying distribution of points.

The optimization procedure for minimizing the estimated MISE and therefore determining the optimal SS bin width Δ_{SS}

is defined by Equations 8 – 10, where N is the number of bins, k_i is the number of elements in the i^{th} bin, μ is the mean number of elements per bin, ν is the variance of the distribution of elements across the bins, and C is the objective function to be minimized [34].

$$\mu = \frac{1}{N} \sum_{i=1}^N k_i \quad (8)$$

$$\nu = \frac{1}{N} \sum_{i=1}^N (k_i - \mu)^2 \quad (9)$$

$$C(\Delta_{SS}) = \frac{2\mu - \nu}{\Delta_{SS}^2} \quad (10)$$

The optimization is performed over a range of bin numbers by averaging the objective function over a sliding window of bin widths given a constant number of bins. The optimal bin width is the bin width associated with the number of bins that minimizes the objective function.

The authors identify the optimal bin width as one of the simplest metrics for grouping one-dimensional clusters of spectral distance values because it assumes a single value for the cluster widths. Because the spectral distance values in the research are scaled to the range from zero to one, the optimal number of bins is the inverse of the optimal bin width. In research not reported in this paper, the authors investigate the effectiveness of this approach for providing iterative clustering algorithms with an initial guess of cluster diameter.

Achieving the goal of determining the optimal bin width for a given distribution of spectral distance values enables calculation of the unweighted centroid of the points within each bin. From the centroid of each bin the intra-cluster and inter-cluster distances can be computed. The intra-cluster distance of a bin is the average distance of all points within a bin to the centroid of the bin. The average intra-cluster distance of a material is the average of all of the per-bin intra-cluster distances and is identical to the inter-observation distance used in TAD. The average inter-cluster distance is the average distance between the centroids of adjacent bins. The average intra-cluster and inter-cluster distance for a material constitutes the spectral distance characterization for the material.

The description of the spectral distance characterization of a material is mathematically defined in Equations 11 – 14. The unweighted centroid of the j^{th} bin of the k^{th} material is $c_{k,j}$ defined in Equation 11, where $r_{k,j,i}$ is the i^{th} spectral distance value of the i^{th} element contained in the j^{th} bin of the k^{th} material and $T_{k,j}$ is the number of spectral distance values in the j^{th} bin of the k^{th} material. The intra-cluster distance of the j^{th} bin of the k^{th} material is $a_{k,j}$ defined by Equation 12. The average intra-cluster distance of the k^{th} material is b_k defined by Equation 13, where L_k is the number of bins for the k^{th} material. Lastly, the inter-cluster distance of the k^{th} material is d_k defined by Equation 14.

$$c_{k,j} = \frac{1}{T_{k,j}} \sum_{i=1}^{T_{k,j}} r_{k,j,i} \quad (11)$$

$$a_{k,j} = \frac{1}{T_{k,j}} \sum_{i=1}^{T_{k,j}} |c_{k,j} - r_{k,j,i}| \quad (12)$$

$$b_k = \frac{1}{L_k} \sum_{j=1}^{L_k} a_{k,j} \quad (13)$$

$$d_k = \frac{1}{L_k - 1} \sum_{j=1}^{L_k-1} |c_{k,j+1} - c_{k,j}| \quad (14)$$

There is a distribution of spectral distance values for each material signature such that each point in the distribution represents the spectral distance between each unique material pair for a single material signature. In the case of the research each distribution has 926 spectral distance values. The analysis is performed for all 927 materials in the spectral library and the result is 927 intra-cluster and inter-cluster distances; a single average intra-cluster and inter-cluster distance for each material. These calculations are computed for each of the four spectral distance measures and for each of the two sensors. The workflow for performing these computations is shown graphically in Figure 9.

Figure 9. Pictured clockwise from the top-left is the workflow required to determine the average intra-cluster inter-cluster distances for a specified material with all other materials in the spectral library.

The in-band material reflectance are the same signatures used in Section 3. The sensor band combinations represent all possible unique combinations of sensor bands for which there is at least two bands. For example, the unique combinations of a three band sensor with band names A, B, and C would be: AB, AC, BC, and ABC. The material pair construction simply builds a list of unique material pairs from the spectral library. For example, if there are three materials named A, B, and C in a library then the unique pairs are: AB, AC, and AB. The spectral distance measures are the ED, SAM, SCM and SID metrics.

The spectral distance characterization of all 927 materials are computed for the ED, SAM, SCM and SID distance metrics using both the FD and SS bin width optimization methods. The characterizations are performed for both WV-3 and L-8

and all sensor bands are used in the calculations. The results using SAM and SID for both FD and SS are shown in the following examples because they are the most representative of the results for all the spectral distance measures.

The spectral distance characterization results of materials using WV-3 and L-8 for the SAM distance metric and both FD and SS are shown in Figure 10. In the figure, the materials are identified by material index. In Figure 10, the materials are sorted based on descending values of WV-3 SAM inter-cluster distance as calculated by FD. All distances in the top and bottom graphics of Figure 10 accompany the same material sorting order in order to facilitate accurate comparison between the two graphics.

The distances in the figure have units of scaled spectral distance value. The selection of WV-3 SAM inter-cluster distance as the characterization quantity that is used to sort all of the data, as opposed to any of the other characterization quantities using either FD or SS is somewhat arbitrary. However, the advantage of sorting all of the characterization quantities with respect to a single characterization quantity is that valid comparisons can be made between the FD and SS results and also between the results for WV-3 and L-8 because all the data is dependent on the same horizontal axis.

Figure 10. Spectral distance characterizations of WorldView-3 and Landsat-8 using SAM for Freedman-Diaconis (top) and Shimazaki-Shinomoto (bottom) bin width determinations.

The top graphic in Figure 10 displays the spectral distance characterization quantities for WV-3 and L-8 using SAM and FD bin width optimization, and the bottom graphic displays the same quantities but as calculated using SAM with SS optimization. Both the top and bottom graphics of Figure 10 share the same abscissa. This means that a given material index number in the top and bottom graphics of Figure 10 refer to the same material, which enables comparison of the ordinate values of the two graphics.

Comparing the top and bottom graphics in Figure 10 demonstrates that the FD and SS average inter-cluster distances are not strongly correlated, regardless of sensor. However, the graphics in Figure 10 demonstrate that the average intra-cluster distances between the two sensors are correlated. Figure 10 also shows that the average intra-cluster distances using SS optimization for both sensors are correlated with both the WV-3 average inter-cluster distances and the WV-3 and L-8 FD optimized average intra-cluster distances. The lack of average inter-cluster distance correlation between FD and SS optimization is found by the research to be true for ED, SAM, SCM, and SID.

The comparison the two graphics in Figure 10 indicate that there is strong correlation between average inter-cluster distance using FD optimization and average intra-cluster distance using both FD and SS optimization and that strong correlations exist between average inter-cluster distance for WV-3 and L-8 and also between average intra-cluster distance for WV-3 and L-8. Ultimately, the analysis shows that all intra-cluster distances are correlated regardless of sensor or bin width optimization method, but that the inter-cluster distances are not correlated. This indicates a strong dependence of the average inter-cluster distances on bin width and therefore bin width optimization selection, whereas the average intra-cluster distances are not strongly dependent on bin width.

Due to the lack of strong correlation between the FD and SS average inter-cluster distances, the remainder of the research sorts the order of the materials separately, based on descending values of WV-3 inter-cluster distance as calculated by both FD and SS. The advantage of sorting the data in this manner is to enable comparison of sensor specific results across the characterization quantities while avoiding distortion of the spectral distance characterization by forcing SS to be sorted by the FD values or vice-versa as was done for the analysis in Figure 10.

The spectral distance characterization quantities shown in Figure 11 are computed using SS bin width determination for the case of SAM applied to WV-3 and L-8 in-band material reflectance signatures. The horizontal axis in Figure 11 is material indices as sorted by descending WV-3 SAM inter-cluster distance as calculated by SS. Although the list of materials used to generate the results in Figure 10 and Figure 11 are the same, the material indices on the horizontal axis of Figure 11 do not correspond to the material indices in Figure 10 because of the difference in sorting based on inter-cluster distance.

Figure 11. Spectral distance characterizations of WorldView-3 and Landsat-8 using SAM for Shimazaki-Shinomoto bin width determination.

The spectral distance characterizations shown in Figure 11 indicate a moderate correlation of L-8 inter-cluster distances about the descending WV-3 inter-cluster distance line. The observation is also made that the WV-3 and L-8 intra-cluster distances are strongly correlated to the inter-cluster distance quantity. In the presentation of the results, the author reiterates that the horizontal axis of Figure 11 is not ordered in the same sequence of materials as in Figure 10 and therefore comparisons between the two figures should not be made.

Although not presented in this paper, the research has verified similar qualitative trends in correlation between characterization quantities for the case of the ED and SCM metrics applied to WV-3 and L-8, and FD and SS bin width determination. Typically, the SCM results yield larger characterization quantities than the other spectral distance measures, and the ED results are similar to the SAM results. Results of the SID characterization using ED are shown in Table 1. The SID characterization quantities are smaller than the other spectral distance measures which originates from the large number of bins, and therefore small bin widths.

Figure 12. Distance characterizations of WV-3 and L-8 using SID for Freedman-Diaconis bin width optimization.

The spectral distance characterizations shown in Figure 12 indicate a strong correlation of L-8 inter-cluster distances about the descending WV-3 inter-cluster distance line. This trend indicates a stronger correlation than for the case of inter-cluster SCM using SCM from Figure 11. The observation is also made that the WV-3 and L-8 intra-cluster distances are strongly correlated to the inter-cluster distance quantity. In fact, the SID average intra-cluster distances nearly lie on top one another in Figure 12. These observations indicate that all the spectral distance characterization quantities for WV-3 and L-8 are not only correlated, but they are for all intents and purposes qualitatively identical with respect to a functional interpolation of the quantities.

Lastly, the several statistics of the characterization quantities are compared to determine how large a portion the values occupy relative to the scaled distance domain, which has a range of one. The range and median for all 16 configurations of sensor, spectral distance measure, and optimization are listed in Table 1. The notation X|Y|Z represents the X-sensor, Y-measure, and Z-optimization.

Table 1. The inter-cluster and intra-cluster distance range and median values for 16 combinations of sensor, spectral distance measure, and bin width optimization.

	Intra-cluster distance ($\times 10^{-3}$)		Inter-cluster distance ($\times 10^{-3}$)	
	Range	Median	Range	Median
L-8 SID SS	3.782	0.270	20.95	4.080
WV-3 SID SS	4.665	0.291	23.89	4.828
L-8 ED SS	7.370	4.669	31.83	23.58
WV-3 SID FD	7.944	0.973	40.18	10.18
L-8 SAM SS	8.091	4.660	36.34	22.61
WV-3 ED SS	9.123	4.369	47.19	29.26
L-8 SID FD	9.206	1.252	45.58	10.75
WV-3 SAM FD	11.16	6.850	48.49	32.02
WV-3 SAM SS	11.49	4.681	47.07	23.10
WV-3 ED FD	12.08	8.300	58.19	51.02
L-8 SAM FD	14.48	7.297	60.39	33.65
L-8 ED FD	15.59	11.29	55.40	55.27
L-8 SCM FD	25.89	20.33	101.1	82.64
WV-3 SCM FD	27.60	22.31	110.2	89.68
WV-3 SCM SS	138.7	15.17	514.0	61.76
L-8 SCM SS	142.9	10.69	508.2	43.48

The salient features observed in Table 1 are as follows. Most obviously, the intra-cluster ranges are smaller than their inter-cluster distance counterparts. Secondly, as the range of the average intra-cluster distance increases, so generally does the range of the average inter-cluster distance. Along with an increase in range of characteristic distances generally comes a corresponding increase in the median intra-cluster or inter-cluster distance value. Although not recorded in Table 1, the minima for both the intra-cluster and inter-cluster distances have much less variance than the maxima.

It appears from Table 1 that SCM yields the spectral distance characterizations with the largest characteristic distances, and

that the four largest characterizations in distance value occur for both FD and SS using SCM with both sensors. In particular, the material that has the maximum inter-cluster distance of 0.5 only has two histogram bins defining the distribution of spectral distance values. Value. Four of the five characterizations with smallest range of intra-cluster distance values are occupied by those calculated from SID. The middle region of Table 1 is occupied by spectral distance characterizations corresponding to those computed with ED and SAM.

The spectral distance characterization research enables a simplification of the spectral distance values that are cross-computed for all material combinations into an average representation of the behavior of the spectral distances for a specific material. The result of the research is a key-value LUT of spectral distance characterizations for all 927 materials recorded in JSON file format. The LUT uses approximately 90 KB of computer disc storage for each sensor, spectral distance measure, and bin width optimization method.

The spectral distance characterization research results in a coarse representation of the clustering behavior compared with more advanced clustering algorithms. The results of this research are intended to be applied strictly as initialization parameters to the problem of accurate cluster determination. The accurate determination of clusters enables confident classification of materials from which principal components representative of the classes can be used in defining linearly independent endmembers which are related to the research described in Section 5.

5. ENDMEMBER ABUNDANCE ESTIMATION

Remotely sensed multispectral image data similar to the data collected by WV-3 and L-8 sensors are comprised of an array of pixels, with each pixel representing a photodetector on a sensor chip assembly. As the sensor images the ground, the photodetector will often record energy arriving from multiple materials within the photodetector ground footprint. In this case the pixel corresponding to the photodetector experiences subpixel mixing and such a pixel is referred to as a mixed pixel. Because the studies in this paper focus strictly on spectral library data, the mixed pixel will be referred to as a composite material. The set of materials that contribute to the composite material signature are referred to as the component materials.

An additional term called the endmembers comprise a set of signatures to be evaluated in the LMM. The endmember set can contain both component and non-component materials. Although the endmember set can contain signatures of materials not present in the target composite material, it must contain the signatures of the composite material. This study assumes that the set of all component materials of a composite material is always included in the endmember set. The size of the endmember set is always greater than or equal to the set of component materials.

The goal of the endmember abundance estimation study is to determine endmember abundances and from the abundance estimates to build an accurate reconstruction of the composite material signature using mathematical optimization algorithms. The study will examine specific configurations of endmembers, endmember abundances, Additive White Gaussian Noise (AWGN), and optimization algorithms using material signatures in all 17 of the WV-3 sensor bands. Although the study comprises a handful of examples with specific materials to deduce patterns, the research is ongoing to test the entire design space, including extension to L-8 sensor bands. The workflow for testing the complete design space is shown in Figure 13.

Figure 13. Pictured from the top-left to bottom-right is the workflow to study the complete design space of abundance estimation and signature reconstruction.

The mathematical optimization routines examined in this study are Linear Least Squares (LS), Differential Evolution (DE), and Linear Simplex (SX). The LS optimization solves exactly determined and overdetermined linear systems of equations [35]. The DE optimization is a nonlinear stochastic optimization and as such it has the advantage that the objective function can be either linear or nonlinear [36, 37]. The disadvantages of DE are that the solutions can be nondeterministic, the number of iterations to converge are not bounded, and the solution is not guaranteed to be the global solution. The SX optimization is a linear programming technique of operations research that searches for the optimal feasible solution to a system of linear constraint equations given a linear objective function [38]. The advantage of SX over LS is that it can typically solve underdetermined as well as overdetermined systems of equations. The advantage of SX over DE optimization is that if an optimal feasible solution exists, SX optimization will find the solution in a predictable maximum number of steps.

The LS optimization programmed for the research uses forward and backward substitution to solve an exactly determined system of linear equations. The LS optimization is not used in the research to solve underdetermined systems because typically no solution exists or there are infinitely many solutions, although a LS minimum norm solution might exist for such underdetermined systems. If the system of equations is overdetermined then the research uses the

Singular Value Decomposition (SVD) approach to solving the normal equations of the LS problem because it is guaranteed stable and works for rank deficient linear systems. The direct solution of the normal equations is strictly avoided when solving overdetermined LS problems because they can be strongly ill-conditioned and also result in a singular normal equation matrix.

In order to facilitate classification of the composite material $\vec{z}$ into its component materials, the research uses the LMM which is a function of additive noise $\vec{w}$, the m^{th} endmember material abundance a_m , and the m^{th} endmember material signature $\vec{s}_m$. The total number of endmember material signatures is M . The LMM is then defined as

$$\vec{z} = \vec{w} + \sum_{m=1}^M (a_m \vec{s}_m) \quad (15)$$

The LMM requires that $M \geq N$, where M is the number of endmembers and N is the number of sensor bands. If M is less than N , the system is underdetermined and a unique solution does not exist. The requirement therefore guarantees that the LMM will have unique solutions.

If there are P component materials in the composite material and $Q = M - P > 0$ endmember materials not in the set of component materials, then the LMM is expanded using a_p , $\vec{s}_p$, a_q , and $\vec{s}_q$ to represent the component material abundances and signatures, and non-component abundances and signatures, respectively. The equation for the expanded form is provided in Equation 16. In this manner, the endmembers are shown as one set of component material signatures and one set of non-component material signatures.

$$\vec{z} = \vec{w} + \sum_{p=1}^P (a_p \vec{s}_p) + \sum_{q=P+1}^M (a_q \vec{s}_q) \quad (16)$$

The variables that are optimized in all three of the optimization routines are the endmember abundances. The formula in Equation 15 and Equation 16 equally define a linear system of equations that can be solved by the optimization routines. The objective function C_{DE} in Equation 17 that is used in the DE optimization for this research is the square of the nonlinear Euclidean distance two-norm equation derived from Equation 15, with number of sensor bands N , as

$$C_{DE} = \sum_{i=1}^N \left(z_i - w_i - \left(\sum_{m=1}^M (a_m \vec{s}_m) \right)_i \right)^2 \quad (17)$$

The objective function C_{SX} for the SX optimization is defined by Equation 18. Unlike the LS and DE optimizations, which as presented in this paper are unconstrained formulations, the simplex algorithm in linear programming is a constrained optimization. The corresponding SX constraints are defined in Equations 19 – 20.

$$C_{SX} = \sum_{i=1}^N \left(z_i - w_i - \left(\sum_{m=1}^M (a_m \vec{s}_m) \right)_i \right) + \text{Max} \left(\vec{w} + \left(\sum_{m=1}^M (a_m \vec{s}_m) \right) \right) \quad (18)$$

$$0 \leq a_m \leq 1 \quad (19)$$

$$\sum_{m=1}^M a_m = 1 \quad (20)$$

The presence of the $\text{Max}(\cdot)$ function in Equation 18 is to guarantee that the objective function never goes negative, which also guarantees that the function has a minimum.

The remainder of this section is organized primarily into two subsections. Both subsections use all 17 bands of WV-3 and do not use the L-8 sensor bands. The first subsection contains examples that examine abundance estimation for the case of a 10% sand, 20% seawater, and 70% old black asphalt (10-20-70) composite material, called composite 1, for various optimizations and AWGN levels. The second subsection contains examples of a 33.33% sand, 33.33% seawater, and 33.33% old black asphalt (33-33-33) composite material for various AWGN levels and purposefully selected endmembers using SX optimization exclusively. The equal mixture composite is called composite 2. A table of the AWGN levels used in the examples throughout the remainder of this section can be found in Table 3 of Appendix B: Additive White Gaussian Noise.

The first example in the study examines the abundance estimation and signature reconstruction of a 3-component composite material using LS, DE, and SX optimization. The example compares the accuracy performance for the cases of three endmembers and eleven endmembers. In the case of three endmembers all the endmembers are in the composite material, whereas the eleven endmember case has eight endmembers which are not in the composite material. These eight endmembers are included in an attempt to challenge the optimization routines.

The eight endmembers used to challenge the optimization are: light grey road concrete, red paving brick, new black tar paper, cotton bond paper, tan asphalt shingle, light grey cinder block, coated steel girder, and light grey painted Aluminum. These material spectra were selected subjectively based on a familiarity of the authors with the materials. The composite material is the 10-20-70 mixture in both the three and eleven endmember cases. The component materials and the composite material (Composite 1) signatures are shown graphically in Figure 14. The motivation for the Composite 2 signature in Figure 14 will be addressed later in this section but is not germane to the current 10-20-70 mixture example.

Figure 14. Spectral reflectance signatures of Sand, Seawater, Asphalt, and two composite signatures. Composites 1 and 2 have abundance percentages of {10, 20, 70} and {33, 33, 33}, respectively.

The results of the abundance estimation are shown in Figure 15 for the case of 2x AWGN. The research also performed the test using 0x AWGN, that is without any noise, and all optimizations exactly estimated the correct endmember abundances for the cases of three and eleven endmembers. The AWGN signal is never explicitly provided to the optimization routines. In Figure 15, LS-3 is LS with three endmembers, LS-11 is LS with eleven endmembers, SX-3 is SX with three endmembers, DE-3 is DE with three endmembers, SX-11 is SX with eleven endmembers, and DE-11 is DE with eleven endmembers. As can be seen from the figure, LS-11 yields the least accurate abundance estimates.

Figure 15. Abundance estimation results for Least Squares, Linear Simplex, and Differential Evolution using 3 and 11 endmembers in the case of 2x AWGN.

The analysis of the results in Figure 15 indicate that all three optimizations qualitatively performed with acceptable accuracy in the case of three endmembers. The LS optimizations yielded the least accurate abundances for both the three and eleven endmember cases. Both DE and SX performed good in terms of abundance accuracy. Incidentally, the LS computational runtime was the fastest of the three optimizations while in contrast, the DE runtime was the longest.

The optimized abundances for the 10-20-70 mixture of sand-seawater-asphalt are used to generate reconstructed composite signatures in the presence of 2x AWGN. The

reconstructed composite signatures are shown in Figure 16. In the figure, the horizontal axis are the WV-3 sensor bands, starting with the PAN band, and indexed by increasing wavelength beginning with the first visible band (V1) and concluding with the last SWIR band (S8).

Figure 16. Reconstructed composite signature using Least Squares, Linear Simplex, and Differential Evolution for 3 and 11 endmembers in the case of 2x AWGN.

The reconstructed composite signatures in Figure 16 qualitatively indicate that DE-11 has the most accurate composite signature reconstruction of all the optimizations. The results also indicate that DE-3, SX-3, and LS-3 have better signature reconstruction accuracy than DE-11, SX-11, and LS-11, respectively. This result is intuitive because abundance estimation with three endmembers which are the component materials is a much less challenging optimization than the case with eleven endmembers, eight of which are non-component endmembers.

The SX optimizations yield least accurate signature reconstructions as can be seen in Figure 16. The most surprising observation after comparing Figure 15 and Figure 16 is that even though LS optimization results in the least accurate abundance estimates, the LS signature reconstructions are qualitatively equal in accuracy to the DE reconstructions. This indicates that LS optimization is selecting energy from the eight non-component endmembers in a sufficiently allocated manner that the reconstructions are accurate although the component abundances are significantly incorrect. To quantitatively determine the best performing optimization routines, the authors refer to Figure 17.

The graphic in Figure 17 shows the signature reconstruction error with respect to the true noise-free signature for the LS-3, LS-11, SX-3, SX-11, DE-3, and DE-11 reconstructions in Figure 16, in the presence of 2x AWGN. The results of the error analysis quantitatively show that LS-3, DE-3, and DE-11 are the most accurate signature reconstructions. However, the DE-3 and DE-11 reconstructions originate from the correct allocation of material abundance, whereas the LS-3 reconstruction originates from an incorrect allocation of material abundance. The next closest to the noise-free signature is LS-11, which obfuscates the inaccuracy with LS

abundance estimates. These results are reinforced by the qualitative analysis shown in Figures 16 – 17.

Figure 17. Errors of the reconstructed WV-3 3-composite signature using Least Squares, Linear Simplex, and Differential Evolution for 3 and 11 endmembers in the case of 2x-AWGN.

The graphic in Figure 17 also shows for the Composite 1 material, the SX reconstruction error has a larger variance in the SWIR sensor bands. Additionally, and unexpectedly, the SX-3 reconstruction error is larger than that of the SX-11 reconstruction with respect to the SWIR sensor bands. The SX reconstruction error in the SWIR bands is also discernable in the Graphic in Figure 16. The range in least accurate reconstructed reflectance based on sensor band for all optimizations shown in Figure 17 is between 6 – 11% of the true signal value.

The graphic in Figure 18 shows the abundance error of LS-3, LS-11, SX-3, SX-11, DE-3, and DE-11 increasing levels of AWGN. The noise levels used in this example can be found in Table 3 of Appendix B: Additive White Gaussian Noise under the Composite 1 column. The results of the error analysis show that LS yields the least accurate abundance estimates of the three optimization routines. The maximum error in Figure 18 for LS-11 is off the chart at approximately 12 error units. Overall, SX and DE yield accurate abundance estimates indicated by their correspondingly low errors. The error calculated in Figure 18 is the two-norm error.

Figure 18. Errors of the reconstructed WV-3 3-composite signature using LS, SX, and DE for 3 and 11 endmembers.

Generally, the trend of increasing error as a function of AWGN level is seen in Figure 18. This trend is true for all three optimization routines in this example. Not surprisingly, the 3-endmember case is less susceptible to noise than the 11-endmember case. The slight dip in error shown in Figure 18 at 10x AWGN is explained by the particular structure of the noise signal; although the 10x noise standard deviation is ten times that of the 1x noise case, the signals are also fundamentally different in shape due to the random nature of their generation.

The overall result of analyzing a 10-20-70 composite material for LS, SX, and DE optimization, three and eleven endmembers, and 2x-AWGN indicates that although LS yields accurate signature reconstructions, SX and DE are the preferred optimizations because they yield the most accurate abundance estimates and they can accommodate overdetermined and underdetermined systems of equations while still yielding acceptable levels of accuracy. Even in the case of square linear systems which are investigated next, the LS yields abundance estimates which are sufficiently incorrect in the presence of noise that the authors do not report the result for LS optimization in the remainder of the section.

The second subsection of examples follows for the remainder of the section on endmember abundance estimation. In these remaining examples, the research examines the case of the 33-33-33 composite material, with equal parts of sand, seawater, and old black asphalt contributing to the spectral reflectance signature. These examples will use all 17 WV-3 sensor bands and will also use 17 total endmembers. This means that there are 14 endmembers available for confusing the optimization routine. The authors consider 14 confusion endmembers of a total of 17 endmembers to be one of the more challenging endmember allocations possible for WV-3 abundance estimation. The SX and DE optimizations are selected for these examples based on the results of the analysis of the 10-20-70 composite material abundance estimation.

Whereas the eight endmembers from the 10-20-70 composite material abundance estimation example are selected somewhat arbitrarily, each of the 14 excess endmembers in the 33-33-33 composite material example will be selected with purpose. The reflectance signature of this composite material (Composite 2) is shown in Figure 14. Two cases will be examined for this example. The first case will select the 14 endmembers to be those materials in the spectral library that have the largest SAM values compared with the component materials, and the second case will select the 14 endmembers corresponding to the smallest SAM values. It is expected that the case of endmembers with high SAM values will not challenge the abundance optimization as much as the case of low SAM values.

First, the SAM values are computed for sand, seawater, and old black asphalt with all of the other 924 material reflectance signatures in the USGS spectral library; the component materials are not compared with one another. Next, five, five,

and four unique materials are selected with the largest SAM values for sand, seawater, and old black asphalt, respectively. The set of five, five, and four materials become the 14-endmember set. This set of high SAM valued endmembers represents the materials that are the least similar from the three component materials that comprise the composite material. Some of the 14 endmembers will have varying degrees of similarity amongst themselves, however they are generally dissimilar but not strictly linearly independent to their respective component signatures.

After selecting the 14 endmembers, the AWGN noise identified in Table 3 in Appendix B under the Composite 2 column is used to test the accuracy of the SX and DE abundance optimizations to correctly estimate the endmember abundances. The AWGN for the example is added to the LMM as multiples of 0, 1, 2, 4, 6, 8, and 10. The abundance estimation results are shown graphically in Figure 19.

Figure 19. Abundance estimation results for Linear Simplex (top) and Differential Evolution (bottom) using 17 endmembers and integer multiples of AWGN. The endmembers are materials with high SAM values.

The abundance estimation using SX optimization of a 3-composite material for seven levels of AWGN and 14 endmembers is shown in Figure 19. The horizontal axis of the figure is labeled by endmember name. The category All Others in Figure 19 indicates the summation of all abundances for a particular noise level of the non-component endmembers. The materials in All Others comprise materials with a high SAM value, equivalently dissimilar and high distinguishability, with respect to the component materials

and for which the optimization incorrectly allocate contributions in the abundance estimation.

The observation is made from the graphics in Figure 19 that as the noise levels increase, so does the inaccuracy of the abundance estimation for both SE and DE. Most profoundly, the figure shows that the DE optimization yields more accurate abundance estimates than SX for increasing noise levels. This indicates that DE optimization is not as adversely affected by noise compared with SX optimization. It is also observed from Figure 19 that both the SX and DE optimizations accurately estimate the abundance allocations in the presence of 14 confusion endmembers and low levels of AWGN.

Referring to the top graphic of Figure 19, the SX optimization yields monotonically increasing non-component abundance allocations with increasing noise up to 8x AWGN. In contrast, a jump in non-component abundance allocation is observed for DE optimization at 6x AWGN. It is also observed that the Asphalt category of Figure 19 experiences the strongest fluctuations in abundance allocation as the noise increases, followed by seawater. The errors in the reconstructed composite signatures from the abundance estimations depicted in Figure 19 are shown graphically in Figure 20.

Figure 20. Errors of the reconstructed WV-3 3-composite signature using Linear Simplex (top) and Differential Evolution (bottom) abundance estimation for the case of 17 high SAM value endmembers and increasing AWGN amplitude.

The reconstruction error for integer multiples of AWGN relative to the noise-free signature shown in Figure 20 quantitatively demonstrates that the 8x and 10x AWGN yields the least accurate signature reconstruction. More generally, the reconstruction error shows that the error tends to increase as the AWGN level is increased. Most notable in Figure 20 is the lower range of reconstruction error for DE compared with SX optimization. The DE optimization has a global range of approximately 19×10^{-3} whereas SX has an error range of approximately 47×10^{-3} . The observation that the SX derived signature reconstruction error is 2.5 times greater than the corresponding DE signature reconstruction error, combined with the results of Figure 19, demonstrate the advantage of using DE optimization over SX optimization in the presence of noise. The authors mention that the abundances for all endmembers are used in the signature reconstruction, not only the abundances associated with the component materials because in a real application the component materials are not known a priori.

The final example of the subsection is similar to the previous 33-33-33 composite signature example, except that this example will examine the accuracy performance of the SX and DE optimization in the case of 17 endmembers with low SAM values. In this example there are still 17 endmembers, and the composite materials is composed of equal parts sand, seawater, and old black asphalt. The additive noise is again included as multiples of 0, 1, 2, 4, 6, 8, and 10 to the AWGN defined in Table 3 from Appendix B. The sensor bands remain all 17 WV-3 bands.

As before, the SAM values are computed for sand, seawater, and old black asphalt with all of the other 924 material reflectance signatures in the USGS spectral library; the component materials are not compared with one another. Next, five, five, and four unique materials are selected with the smallest SAM values for sand, seawater, and old black asphalt, respectively. The set of five, five, and four materials become the 14-endmember set. This set of low SAM value endmembers represent the materials that are the most similar to the three component materials that comprise the composite material. Some of the 14 endmembers might have varying degrees of similarity amongst themselves, but they are generally similar to their respective component signatures. The authors have constructed this example in an attempt to maximally challenge the accuracy of the abundance optimization result.

After selecting the most similar 14 endmembers with respect to the component signatures, the AWGN noise identified in Appendix B is used to test the ability of the SX and DE abundance optimization to correctly estimate the endmember abundances in the presence of noise. The AWGN for the example is added to the LMM as standard deviation multiples of 0, 1, 2, 4, 6, 8, and 10. The noise signals used in the highly similar endmember example are identical to the noise signals in the case of dissimilar endmembers. The abundance estimation result for the highly similar endmembers are shown graphically in Figure 21.

Figure 21. Abundance estimation results for LS (top) and DE (bottom) using low SAM value endmembers and integer multiples of AWGN.

The abundance estimation using SX optimization of the 33-33-33 3-composite material for seven levels of AWGN and 14 of the highly similar endmembers is shown in Figure 21. The category All Others in Figure 19 indicates the summation of all abundances for a particular noise level of the non-component endmembers. The horizontal axis of the figure is labeled by endmember material name. The category All Others in the figure comprise materials with a low SAM value with respect to the component materials and for which the optimizations incorrectly allocate contributions in the abundance estimation.

Comparing Figure 19 with Figure 21, it is deduced that the abundance estimation is more challenged by the low SAM value endmembers than by the high SAM endmembers, regardless of SX or DE optimization. In the case of low SAM value endmembers, the abundance estimation incorrectly allocates an increasing amount abundance to the non-component material endmembers as the noise levels increase. In contrast, the high SAM value endmember abundance estimation only changes abundance allocation within the set of component materials and does not appreciably allocate abundance to non-component material endmembers, even as the noise increases to 10x AWGN. The errors in the reconstructed composite reflectance signatures from the abundance estimations depicted in Figure 21 are shown graphically in Figure 22. Although not shown in the figure, the LS solution does not yield accurate abundance or reconstruction results when a non-zero noise contribution is considered.

Figure 22. Errors of the reconstructed WV-3 3-composite signature using Linear Simplex (top) and Differential Evolution (bottom) abundance estimation for the case of 17 low SAM value endmembers and increasing AWGN amplitude.

The reconstruction error for varying multiples of AWGN relative to the noise-free signature are shown in Figure 22 demonstrates that the 10x AWGN yields the least accurate signature reconstruction, which agrees with the abundance analysis for the same example. Also in agreement with the analysis of Figure 20, the reconstruction error shows that the error increases as the AWGN amplitude is increased. The authors mention that the abundances for all 17 endmembers are used in the signature reconstruction, not only the abundances associated with the component materials.

It is observed by comparing Figure 20 with Figure 22 that the range of reconstruction error for high SAM value endmembers is greater than the range of error for the low SAM value endmembers. Specifically, the DE optimization has a global error range of approximately 17×10^{-3} whereas SX has an error range of approximately 37×10^{-3} , which equates to an SX error twice as large as the DE error. Therefore, the range of reconstruction error for dissimilar endmembers is shown to be greater than that of high similarity endmembers for both DE and SX optimization. Although the example of dissimilar endmembers outperforms that of high similarity endmembers in terms of abundance estimation accuracy, the accuracy of the reconstructed signature in the highly distinguishable case can actually be less than the accuracy of the least distinguishable case.

A summary of reconstruction two-norm error in reflectance units using SX and DE optimization is provided in Table 2. The two-norm error using DE optimization reconstruction increases from low to high similarity endmembers as expected, except for 4x and 8x AWGN. Surprisingly, the two-norm error using SX optimization decreases from low to high similarity endmembers for all noise levels. This observation is counterintuitive but it has been verified for correctness by multiple redundant computations.

Table 2. Composite signature reconstruction error versus AWGN in units of reflectance using endmembers with high and low SAM values relative to component signatures.

	Error ($\times 10^{-3}$), High SAM		Error ($\times 10^{-3}$), Low SAM	
	SX	DE	SX	DE
0x Noise	1.01	1.01	1.01	1.01
1x Noise	8.60	3.75	8.52	4.66
2x Noise	17.5	7.85	16.3	7.91
4x Noise	41.4	7.26	32.2	5.97
6x Noise	62.5	14.8	48.1	16.7
8x Noise	97.3	25.7	93.5	23.8
10x Noise	84.4	13.8	78.6	14.1
Global	21.5	4.95	19.5	4.87

In conclusion to this section, the research shows that for the cases studied, LS optimization provides the least accurate abundance estimation in the presence of noise and that SX and DE optimizations are acceptably accurate for both square and non-square systems of equations in the presence of noise. Ultimately, DE optimization outperformed SX optimization for both abundance estimation and signature reconstruction in the cases examined by the research. The Global row in Table 2 quantitatively shows that the global two-norm reconstruction error of DE optimization for all noise levels is four times less than the corresponding error of SX optimization in cases of low and high endmember similarity. The research also shows that increasing the AWGN contribution to the LMM also increases the error in abundance estimation regardless of optimization routine.

Although dissimilar endmembers relative to the component materials yield the most accurate abundance estimation, the most similar endmembers yield the most accurate signature reconstruction. This statement seems to be a contradiction, but the results are contained in the analysis within the limited scope of the research results and are observed in Figures 19 – 22. Further proof is shown in the last row of Table 2, the Global row, which quantitatively demonstrates that the similar endmembers have an average two-norm error over all noise levels that is less than that of the dissimilar endmembers. This global result holds true for both SX and DE optimization as represented in the last row of Table 2.

A more confident statement regarding the performance of the optimizations algorithms and signature reconstruction is

answered with an expanded design of experiments as outlined in Figure 13. The complete design space will compare abundance estimation and signature reconstruction accuracy for all possible unique combinations of sensor bands, endmembers, abundance combinations and permutations, noise, for all three optimization routines and for both WV-3 and L-8 sensor bands. Without conducting an extended design of experiments, the research is highly confident of two conclusions: introduction of increasing AWGN levels decreases the accuracy of the abundance estimation and signature reconstruction, and excluding endmembers that are highly similar from the component material signatures yield the most accurate abundance estimates.

6. CONCLUSIONS AND FUTURE WORK

This paper provides results from a large design of experiments implementation that are being integrated with an ATD system. The optimal spectral data configurations result of Section 3 provides a LUT for retrieval of optimal processing configurations that improve the ability to detect the presence of target materials from a set of background materials. The optimal configurations derived in Section 3 are intended to be used in generating the histogram-based initializations depending on the target material and sensor.

The spectral distance characterization of Section 4 introduces a methodology for calculating histogram-based initialization values for clustering algorithms used in automated segmentation and classification systems. An outcome of the research in Section 4 is a LUT of intra-cluster and inter-cluster distances for material clusters that are loaded into automated segmentation and classification systems for initialization of parameters. The clusters generated by advanced clustering algorithms initialized with the histogram-based distance characterizations are intended to be used in determining classes of background materials for use in the LMM for subpixel mixing determination.

The endmember abundance estimation of Section 5 provides several case studies of the effects of noise and endmember population on the accuracy of abundance estimation and composite signature reconstruction. The confident outcomes of Section 5 are that, increases in noise negatively impact abundance estimation accuracy, and endmember selection strongly impacts abundance estimation. Additionally, the SX and DE optimizations have been demonstrated to outperform LS in terms of abundance accuracy, and they are both robust to the dimensionality of the LMM. The DE optimization, based on the cases examined in the research, yields the most accurate abundance estimation and signature reconstruction of the three optimization algorithms investigated for this research.

The next step for future work regarding the optimal configuration of sensor bands and spectral distance measure for material separation from a background set is the application of the results to real WV-3 and L-8 remotely sensed data. This data is already acquired and testing has

begun. The next steps for the spectral distance characterization work is the application of the results to real WV-3 and L-8 remote sensing data. The authors have obtained trusted computer code for numerous clustering algorithms and the cluster distances are being integrated into the algorithms and applied to the multispectral image data.

Future work will be performed to determine the effect of using the optimal configurations in the histogram-based initializations for the clustering algorithms. The work will also assess the accuracy of TAD using all four spectral distance measures identified in this research, which will extend TAD beyond only using ED as the mapping function similarity measure. The researchers anticipate the investigation of additional clustering algorithms for the segmentation and classification of image data. Additional research will also investigate the use of the SID-SAM spectral measure for enhanced signature discrimination [39-40].

The next steps for the endmember abundance estimation research is to extend the design of experiments to include all possible test cases. This work has already begun and the outcome is expected to enable the researchers to determine a confidence metric for predicting the accuracy performance of solutions to various configurations of the LMM system. Linear independence is related to the SAM distance, and similarity of endmembers causes abundance estimation challenges. Therefore, future research will investigate identifying linearly independent principal components from classes of materials as endmembers to reduce the effects of endmember similarity. Lastly, the future work will more thoroughly experiment with the input settings of the DE algorithm to determine the most robust configuration that also maintains reasonable computational runtimes, and accuracy of endmember abundance estimation.

APPENDICES

A. SCALING OF SPECTRAL DISTANCE VALUES

The comparability of the results using spectral distance measures is facilitated by scaling the results of the four measures such that the smallest possible value is zero and the largest possible value is one. The measures considered are ED, SAM, SCM, and SID. Scaling the ED, SAM, and SCM in this manner is straightforward because the extrema are known before any calculations are performed. The SID however requires additional consideration for scaling.

The dependence of SID on the probability distributions defined by the relative entropy of self-information causes an inability to theoretically determine the exact range of the possible SID values. This is addressed by computing SID for all unique material signature pairs of a unique sensor band combination and scaling the SID values so that the range of values is unity. This technique is considered valid for sufficiently large numbers of material signatures of sufficiently varying spectral composition.

The SCM is naturally scaled with a range of unity. The theoretical range of ED and SAM is $\sqrt{N}$ and $\pi/2$, respectively, where N is the number of sensor bands. However, theoretically scaling the SID is complicated because the values computed depend on the relational complexity of the ratios of the two signatures as defined in Equation 4. Numerical testing using randomly generated reflectance and differential evolution optimization of Equation 4 for up to 17 sensor bands indicates that the range of SID values essentially increases with an increasing number of bands and an approximate maximum range of at least 40.5 is achievable. Ultimately a theoretical limit on the range of SID values has not been derived and therefore all spectral measures are scaled relative to their respective largest spectral distance value which effectively determines the relative discriminatory power of each measure with respect to a reference signature pair and enables comparison across all the spectral measures [41-42]. Although each spectral measure is likely to not have the same signature pair as the reference pair, the authors are confident the comparisons are valid.

B. ADDITIVE WHITE GAUSSIAN NOISE

The AWGN referred to in Section 5 is generated by a Gaussian random sample using integer multiples of a pre-defined standard deviation. The starting standard deviations are given in Table 3 in the 1x noise row for composite 1 and composite 2 materials used in Section 5. The standard deviation is multiplied by an integer factor for each noise level and the AWGN is generated.

For example, a 4x noise signal is generated for composite 1 by multiplying a standard deviation of 1.219×10^{-3} by a factor of four to obtain a larger standard deviation of 4.875×10^{-3} . The larger standard deviation is used as input into the Gaussian function to generate the 4x noise AWGN signal. The mean of the Gaussian function used to generate the noise is set to zero. The discrete noise signal is repeatedly generated until the standard deviation of the discrete noise signal is within a precision of 10^{-6} of the theoretical standard deviation, at which point the noise signal is accepted.

Table 3. The standard deviations used to generate AWGN for the two composite materials in Section 5.

	Standard Deviation ($\times 10^{-3}$)	
	Composite 1	Composite 2
1x Noise	1.21871	1.65740
2x Noise	2.43741	3.31479
4x Noise	4.87483	6.62958
6x Noise	7.31224	9.94437
8x Noise	9.74965	13.2592
10x Noise	12.1871	16.5740

ACKNOWLEDGEMENTS

The authors would like to thank Wright State University Interdisciplinary Applied Science and Mathematics program and Integrity Applications Incorporated for providing funding, research facilities, and personnel collaboration.

REFERENCES

- [1] N. Keshava, J. Kerekes, D. Manolakis, G. Shaw, "An Algorithm Taxonomy for Hyperspectral Unmixing," *Algo. and Tech. for Multi., Hyper. and Ultraspectral Imagery VI* 4049 (2000): 42-63.
- [2] D. Manolakis, R. Lockwood, T. Cooley, J. Jacobson, "Is There a Best Hyperspectral Detection Algorithm?," *Algo. and Tech. for Multi., Hyper. and Ultraspectral Imagery XV* 7334 (2009).
- [3] D. Manolakis, G. Shaw, "Detection Algorithms for Hyperspectral Imaging Applications," *IEEE Signal Processing Magazine* 19.1 (2002): 29-43.
- [4] D. Malanokis, "Taxonomy of Detection Algorithms for Hyperspectral Imaging Applications," *Optical Engineering* 44.6 (2005): 066403.
- [5] D. Manolakis, E. Truslow, M. Pieper, T. Cooley, M. Brueggeman, "Detection Algorithms in Hyperspectral Imaging Systems," *IEEE Signal Processing Magazine* 31.1 (2014): 24-33.
- [6] F. A. Kruse, S. L. Perry, "Mineral Mapping using Simulated Worldview-3 Short-Wave-Infrared Imagery," *Remote Sensing* 5.6 (2013): 2688-703.
- [7] D. Manolakis, "Detection Algorithms for Hyperspectral Imaging Applications," *Lincoln Laboratory Journal* Volume 14, Number 1 (2003).
- [8] J. A. Richards, X. Jia, "Remote Sensing Digital Analysis: An Introduction," 4th Ed., Springer-Verlag Berlin Heidelberg (2009).
- [9] C. Gingles, M. Emre Celebi, "Histogram-Based Method for Effective Initialization of the K-Means Clustering Algorithm," *Proceedings of the Seventh International Florida Artificial Intelligence Research Society Conference, Association for the Advancement of Artificial Intelligence* (2014).
- [10] D. Manolakis, C. Siracusa, G. Shaw, "Hyperspectral Subpixel Target Detection using the Linear Mixing Model," *IEEE Transactions on Geoscience and Remote Sensing* 39.7 (2001): 1392-409.
- [11] M. L. Pieper, D. Manolakis, E. Truslow, T. Cooley, M. Brueggeman, "False Alarm Mitigation Techniques for Hyperspectral Target Detection," *Algo. and Tech. for Multi., Hyper. and Ultraspectral Imagery XIX* 8743 (2013): 874304.
- [12] D. Manolakis, M. Rossacci, J. Cipar, R. Lockwood, T. Cooley, J. Jacobson, "Statistical Characterization of Natural Hyperspectral Backgrounds using t-Elliptically Contoured Distributions," *Algo. and Tech. for Multi., Hyper. and Ultraspectral Imagery XI* 5806 (2005): 56-65.
- [13] D. Manolakis, M. Rossacci, D. Zhang, J. Cipar, R. Lockwood, T. Cooley, J. Jacobson, "Statistical Characterization of Hyperspectral Background Clutter in the Reflective Spectral Region," *Applied Optics* 47.28 (2008): F96-F106.
- [14] D. B. Marden, D. Manolakis, "Modeling Hyperspectral Imaging Data," *Algo. and Tech. for Multi., Hyper. and Ultraspectral Imagery IX* 5093 (2003): 253-62.
- [15] D. Manolakis, D. Marden, "Dimensionality Reduction of Hyperspectral Imaging Data using Local Principal Components Transforms," *Algo. and Tech. for Multi., Hyper. and Ultraspectral Imagery X* 5425 (2004): 393-401.
- [16] M. Rossacci, D. Manolakis, J. Cipar, R. Lockwood, T. Cooley, J. Jacobson, "Effects of Dimensionality Reduction on the Statistical Distribution of Hyperspectral Backgrounds," *Imaging Spectrometry XI* 6302 (2006): H3020.
- [17] R. S. DiPietro, D. Manolakis, R. Lockwood, T. Cooley, J. Jacobson, "Performance Evaluation of Hyperspectral Detection Algorithms for Sub-Pixel Objects," *Algo. and Tech. for Multi., Hyper. and Ultraspectral Imagery XVI* 7695 (2010): 76951W.
- [18] D. Manolakis, R. Lockwood, T. Cooley, J. Jacobson, "Hyperspectral Detection Algorithms: Use Covariances or Subspaces?," *Imaging Spectrometry XIV* 7457 (2009).
- [19] D. Manolakis, M. Pieper, E. Truslow, T. Cooley, M. Brueggeman, S. Lipson, "The Remarkable Success of Adaptive Cosine Estimator in Hyperspectral Target Detection," *Algo. and Tech. for Multi., Hyper. and Ultraspectral Imagery XIX* 8743 (2013): 874302.
- [20] D. Manolakis, G. Shaw, N. Keshava, "Comparative Analysis of Hyperspectral Adaptive Matched Filter Detectors," *Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VI* 4049 (2000): 2-17.
- [21] R. S. DiPietro, D. G. Manolakis, R. B. Lockwood, T. Cooley, J. Jacobson, "Hyperspectral Matched Filter with False-Alarm Mitigation," *Optical Engineering* 51.1 (2012): 016202.
- [22] D. Manolakis, G. Shaw, "Directionally Constrained or Constrained Energy Minimization Adaptive Matched Filter," *Imaging Spectrometry VII* 4480 (2002): 57-64.
- [23] D. Manolakis, C. Siracusa, D. Marden, G. Shaw, "Hyperspectral Adaptive Matched Filter Detectors: Practical Performance Comparison," *Algo. and Tech. for Multi., Hyper. and Ultraspectral Imagery VII* 4381 (2001): 18-33.

- [24] D. B. Marden, D. Manolakis, "Constrained Detectors for Hyperspectral Remote Sensing Applications: Theory versus Practice," *Imaging Spectrometry VIII* 4816 (2002).
- [25] C. Rozell, D. Manolakis, "Matched Filter Performance for Unequal Target and Background Covariance Matrices," *Algo. and Tech. for Multi., Hyper. and Ultraspectral Imagery X* 5425 (2004): 109-17.
- [26] B. Basener, E. J. Lentilucci, D. W. Messinger, "Anomaly Detection using Topology," *Algo. and Tech. for Multi., Hyper. and Ultraspectral Imagery XIII* 6565 (2007): J5650-1.
- [27] T. J. Doster, D. S. Ross, D. W. Messinger, W. F. Basener, "Anomaly Clustering in Hyperspectral Images," *Algo. and Tech. for Multi., Hyper. and Ultraspectral Imagery XV* 7334 (2009): 73341P.
- [28] M.G. Gartley, D. W. Messinger, "Topological Anomaly Detection Performance with Multispectral Polarimetric Imagery," *Algo. and Tech. for Multi., Hyper. and Ultraspectral Imagery XV* 7334 (2009): 73341O.
- [29] B. Basener, D. W. Messinger, "Enhanced Detection and Visualization of Anomalies in Spectral Imagery," *Algo. and Tech. for Multi., Hyper. and Ultraspectral Imagery XV* 7334 (2009): 73341Q.
- [30] J. A. Albano, D. W. Messinger, "Euclidean Commute Time Distance Embedding and its Application to Spectral Anomaly Detection," *Algo. and Tech. for Multi., Hyper. and Ultraspectral Imagery XVIII* 8390 (2012): 83902G.
- [31] L. P. Dorado Munoz, D. W. Messinger, A. K. Ziemann, "Target Detection using the Background Model from the Topological Anomaly Detection Algorithm," *Algo. and Tech. for Multi., Hyper. and Ultraspectral Imagery XIX* 8743 (2013): 87430M.
- [32] D. Freedman, P. Diaconis. "On the Histogram as a Density Estimator - L2 Theory," *Zeitschrift Fur Wahrscheinlichkeitstheorie Und Verwandte Gebiete* 57.4 (1981): 453-76.
- [33] R. Hyndman, "The problem with Sturges' rule for constructing histograms," 5 July 1995.
- [34] H. Shimazaki, S. Shinomoto, "A Method for Selecting the Bin Size of a Time Histogram," *Neural computation* 19.6 (2007): 1503-27.
- [35] B. N. Datta, "Numerical Linear Algebra and Applications," Second Edition, Society for Industrial and Applied Mathematics (2010).
- [36] A. K. Qin, V. L. Huang, P. N. Suganthan, "Differential Evolution Algorithm with Strategy Adaption for Global Numerical Optimization," *IEEE Transactions on Evolutionary Computation* Vol. 13 No. 2 (April 2009): 398-417.
- [37] R. Storn, K. Price, "Differential Evolution - A Simple and Efficient Heuristic for Global Optimization Over Continuous Spaces," *Journal of Global Optimization* 11.4 (1997): 341-59.
- [38] R. Bronson, G. Naadimuthu, "Theory and Problems of Operations Research," Second Edition, McGraw-Hill (1997).
- [39] Y. Du, et al., "New Hyperspectral Discrimination Measure for Spectral Characterization," *Optical Engineering* Vol. 43 Issue. 8 (August 2004): 1777-1786.
- [40] M.N. Kumar, et al., "A New Hybrid Spectral Similarity Measure for Discrimination of Vigna Species," *International Journal of Remote Sensing*, 32(14), (2011) 4041-4053.
- [41] H. Chauhan, B. K. Mohan, "Effectiveness of Spectral Similarity Measures to Develop Precise Crop Spectra for Hyperspectral Data Analysis," *ISPRS Annals of Photog., Rem. Sens. And Spat. Info. Science*, Volume II-8 (2014) 83-90.
- [42] C.I. Chang, "An Information-Theoretic Approach to Spectral Variability, Similarity, and Discrimination for Hyperspectral Image Analysis," *IEEE Transactions on Information Theory*, Vol. 46, No. 5, (August 2000) 1927-1932.

BIOGRAPHY

Andrew J. Niklas received a B.S. in Physics from the University of Cincinnati in 2007, a M.S. in Physics from Wright State University in 2012, and is currently a Ph.D. candidate at Wright State University. His research interests include signal and image processing of spectral and temporal remote sensing systems, computational electrodynamics, and operations research.

Dr. Matthew D. Sambora is a retired Colonel of the United States Air Force and is a Principal Systems Engineer at Integrity Applications Incorporated. He earned a B.S. degree in Electrical and Computer Engineering from Clarkson University in 1988, a M.S. degree in Electrical Engineering from Southern Illinois University at Edwardsville in 1996, and a Ph.D. in Computer Engineering from the Air Force Institute of Technology in 2008. His assignments have been in both leadership and technical areas and have encompassed a variety of engineering, research and development, and program management jobs.