Hyperspectral Band Selection with the N-dimensional Spectral Solid Angle and
Its Utilization in the Discrimination of Spectrally Similar Targets

by

Yaqian Long

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Earth and Atmospheric Sciences
University of Alberta

Abstract

The discrimination of earth surface materials using hyperspectral sensing can be facilitated by selecting a subset of spectral bands that focuses on essential features. Materials in detailed classes such as rock types and tree species often present great spectral similarity, producing challenges for band selection. A method named the N-dimensional Solid Spectral Angle (NSSA) was proposed to select the most dissimilar spectral regions amongst targets for their maximum spectral separation, however, the use and performance of this method in practical application needed to be explored.

In chapter 2, the NSSA method was applied to two real datasets of geologic relevance to establish guidelines for the selection of parameters that will allow non-expert users to exploit this method. This study demonstrated that the NSSA method is a robust tool for feature identification, since bands selected from the two datasets not only captured absorption feature position, and shape and depth, but also showed improved class separation.

In chapter 3, the NSSA method was applied in a hierarchical manner to address the inter- and intra-class variability among materials. Two datasets were analyzed, including airborne image endmembers for geological mapping and leaf spectra for tree species discrimination. Bands were separately selected from different hierarchies of those categorized materials using the NSSA and combined into a single band set. The agreement between bands selected by the hierarchical strategy and by experts suggested that the hierarchical band selection using the NSSA method is both practical and effective in addressing the spectral variability.

In chapter 4, an ensemble of multiple band selection methods encompassing random forest, minimum redundancy maximum relevance, and the NSSA was used to select and characterize longwave infrared features of leaves for the discrimination of tree species which display great

spectral similarity. The selected features could be related to leaf constitutional compounds such as cellulose and oleanolic acid. Meanwhile, the band selection improved the classification using a regularized logistic regression method by 3%. These results can be useful to future image mapping of tree species at large scales. The ensemble strategy was recommended for the band analysis of vegetation.

Chapter 5 proposed a strategy that simultaneously employed band selection and endmember selection by incorporating the NSSA into the Spatial Spectral Endmember Selection (SSEE) method in order to select bands that enhance the spectral contrast of endmembers and hence improve the estimation of fractional abundances from hyperspectral images. The detailed methodology was described and an airborne image that was acquired for the mapping of mafic and ultramafic rocks was used to evaluate the proposed method. The results showed that the integration of NSSA and SSEE automates band selection in spectral mixture analysis and reduces the efforts in field investigations for feature identification.

The results of this thesis demonstrated that the NSSA method, whether used in a hierarchical manner or integrated with other methods, was robust in the analysis of spectral libraries collected from field samples or from hyperspectral imagery collected from laboratory or airborne imaging systems. Its effectiveness also spans the visible near-infrared, shortwave infrared to the thermal infrared range of the data. The band selection results were evaluated by both classification performance and the physical meaning of spectral features, which balanced the need for high accuracies in statistical learning algorithms and application significance highlighted by remote sensing experts. The proposed method, guidelines, and experimental designs provided in this thesis contribute in identifying meaningful features from data encompassing a small number

of labeled samples for the discrimination of spectrally similar material in a variety of fields including geology, ecology, urban and agriculture.

Acknowledgments

First and foremost, I would like to express the greatest appreciation to my supervisor Dr. Benoit Rivard for his patient guidance, insightful critique, and constant support. He is not only a wonderful educator who leads me to be joyfully involved in the research and study but also a role model that has long-lasting influences on my attitude toward work and life. His charisma for science drives me to become passionate about research, and he always makes helpful comments to my research. He shows responsibility to his students, in every detail, which teaches me to be professional in work. He is also an open, kind and righteous person, creating a relaxing environment for my Ph.D. studying.

I gratefully acknowledge Dr. Derek Rogge for his great help with my research. He made constructive suggestions and assisted the composition of my research papers. I am grateful to Dr. Jilu Feng for his guidance and help. He introduced me to my Ph.D. study and the research area, which is significant to my entire career. I also express my appreciation to my committee members: Dr. Arturo Sánchez-Azofeifa, Dr. Russell Greiner, Dr. Christopher Herd. They offered useful comments during my preparation for the candidacy exam and the later research.

This thesis would not have been completed without financial support from the China Scholarship Council (CSC) and the research grant provided by my supervisor Dr. Rivard.

Special gratitude is given to Dr. Rob Desjardins and Dr. Heather Graves for their instructions on academic writing.

I would also like to express my thankfulness to the fellows in our EOSL laboratory. Iman Entezari, Dominica Harrison, Minghua Tian are particularly acknowledged for their professional assistance in my research. Meimei Chong as being the lab manager is appreciated for her support. Thanks to my labmates for their friendship, encouragements and help: Branko Hilje Rodriguez,

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Background and research motivation

Hyperspectral remote sensing, also known as imaging spectroscopy, is an advanced technology that collects spectral information across the electromagnetic spectrum spanning the visible, near-infrared to the thermal-infrared range for the detection or identification of earth surface materials (Borengasser et al., 2007; Van der Meer and De Jong, 2011). Hyperspectral sensors onboard ground, airborne or spaceborne platforms collect the radiation intensity (usually reflectance values) as a function of wavelength (a spectrum) for a point observation or imagery. A hyperspectral image can be regarded as a three-dimensional data cube consisting of adjacent pixels with a spatial dimension (x, y) and a third dimension (λ) that describes the spectral radiance. A spectrum is thus captured for each pixel of the image. The wavelength interval, or the spectral channel that stores the radiant information is referred to as a 'band' or 'bandwidth'. Each layer of the data cube is called a band image. Hyperspectral sensors acquire hundreds or thousands of contiguous and narrow bands, typically smaller than 20nm (e.g., AVIRIS, 224 bands, 0.4-2.5 μm). In contrast, multispectral sensors have wider bands with only a small number of non-contiguous bands (e.g., Landsat 7 ETM+: 7 bands, 0.4-2.5 μm) (Landgrebe, 2002). Narrower bands enable the detection of finer spectral features and thus more detailed material discrimination. Specific classes such as different rock types can be distinguished by hyperspectral sensing, as opposed to multispectral sensing that can only classify broad classes (Van der Meer and De Jong, 2011; Pu, 2017).

A variety of statistical methods exist for the classification of hyperspectral data and the discrimination of targets (Plaza et al., 2009). Conventional tools such as the Spectral Angle Mapper

(SAM), maximum likelihood and clustering analysis have been commonly used, and they are easily accessible in remote sensing software (e.g., ENVI). With advances in computational capacity in hardware and software for statistical learning, machine learning techniques offer many effective tools to analyze hyperspectral data (Chang, 2000; 2013). For example, Support Vector Machines, Random Forest, neural networks and logistic regression have been widely used to classify geologic, ecological, urban and water materials (Camps-Valls et al., 2013; Fassnacht, 2016; Van der Meer, 2011). Compared to conventional methods, machine learning classifiers show advantages in processing large amounts of spectral data. However, the performance of these methods is sensitive to the number of labeled samples. Another technique for classifying hyperspectral imagery is Spectral Mixture Analysis (SMA) (Keshava and Mustard, 2002). A linear SMA model treats each pixel of the imagery as a mixture of distinct components, hence, the mixed spectrum for that pixel can be represented by a linear combination of spectrally pure signatures (endmembers). The process to extract the distinct endmembers from the imagery is referred to as 'endmember extraction'. Once the endmember set is defined, one can retrieve the proportion of each endmember present in the pixel, namely fractions or abundances, and this process is referred to as 'spectral unmixing'. The accuracy of SMA is largely impacted by the endmember variability in materials of interests. The endmember set must therefore capture not only the broad spectral classes in the scene (e.g., water, soil, rock, vegetation) but also more subtle differences amongst spectrally similar materials of a given class (e.g., within the rock class). The inter-class variability, referring to the similarity among endmembers of a given class of materials, introduces challenges to accurate discrimination of materials (Somers et al., 2011). To address the variability, effort has been put into the use of spatial information in endmember extraction and the use of multiple endmembers for each class in an iterative unmixing process (Bateson et al., 2000; Plaza et al.,

2002; Rogge et al., 2006). The Spatial-Spectral Endmember Extraction (SSEE) and Iterative Spectral Mixture Analysis (ISMA) are two representative approaches proposed to address the issue of intra-class spectral variability and subtle spectral differences amongst classes (Rogge et al., 2006; 2007). The two approaches have been applied together to enhance the discrimination of spectrally similar lithologies in the context of regional geological mapping and mineral exploration (Rogge et al. 2014). Despite these innovations, the process of defining a final endmember set as input into the unmixing process generally still requires the aid of expert knowledge.

Focusing on the most important spectral features can facilitate the discrimination of targets of interest since it avoids the influence of other spectral bands or regions where similarity and noise negatively impact spectral discrimination in data of high dimensionality. The performance of classifiers can decrease beyond a certain dimensionality, a phenomenon known as the "curse of dimensionality" (Chang, 2013). A common approach for reducing the high dimensionality is to select a subset of spectral bands from the hyperspectral data. The band selection improves the accuracy of classification models by reducing the strong correlation between contiguous bands, and more importantly, by resolving the problem of the disparity in the number of training samples (typically scarce) in contrast to the numerous variables (number of spectral bands). The dimensionality can also be reduced using feature extraction approaches, such as principal component analysis that conducts a linear transformation of the spectral bands. Feature or band selection approaches can retain the original information in the data, allowing for a meaningful interpretation of spectral features (e.g., peaks or troughs) and a better understanding of the drivers (feature origin) for the target separation (Van der Meer, 2004). For example, leaf pigments, water content are associated with wavelength regions in the visible-near infrared and shortwave infrared

(Asner, 1998). Diagnostic bands for capturing related absorption features are often selected for the discrimination of vegetation types (Fassnacht et al., 2016).

Band selection methods generally fall into two groups, supervised or unsupervised, based on the availability of prior knowledge of labeled classes or large training samples (Bajcsy and Groves, 2004). Supervised methods such as Random Forest (RF) and Support Vector Machine (SVM) models (Chan and Paelinckx, 2008; Abdel-Rahman et al., 2013; Feilhauer et al., 2015) are prevalent in a wide range of applications. For hyperspectral analysis, the band importance is quantified, and bands are selected based upon the prediction accuracy of classification or regression (Tuia et al., 2015) models. Unsupervised methods such as cluster analysis and Non-negative Matrix Factorization (NMF) are preferred when prior knowledge is lacking or insufficient (Jia and Qian, 2009; Jia et al. 2016). These methods often use distance measurements in spectral data, also known as "spectral similarity" (Chang, 2000; Keshava, 2004; Du and Yang, 2008; Li et al., 2014). For example, the method of Variable-Number Variable-Band selection (VNVBS) has been proposed for feature selection from a single spectrum based on the similarity between this spectrum and a reference spectrum (Harsanyi and Chang, 1994; Wang and Chang, 2007). Another example is the Band Add-On (BAO) (Keshava, 2004) method that attempts to maximize the separability of targets by using the "Spectral Angle"(SA), the most common measure of similarity in hyperspectral analysis. Both methods, however, are limited to the discrimination of two classes. The need for unsupervised methods that are designed for the discrimination of multiple classes arises.

Few band selection methods have been proposed specifically for discriminating spectrally similar targets. The spectral contrast of targets in many applications is high, making the separation (e.g., vegetation vs. soils) of targets using classification tools a comparatively easy task. However,

in several applications such as rock type or plant species discrimation, the targets display great similarity in their material make-up (e.g., mineral abundance), which results in low spectral contrast amongst targets of interest. The difficulty in differentiating low-contrast targets stimulates the need for robust methods for feature identification. A method named the N-dimensional Solid Spectral Angle (NSSA) was recently proposed by Tian et al. (2016) to select bands for the improved discrimination of objects with similar spectral properties. By selecting bands characterized by the largest solid spectral angles (e.g., NSSA values), one can capture the most important spectral information to distinguish a set of spectra. NSSA is an unsupervised method that overcomes the limitations of the large number of training samples and the numbers of classes (e.g., two) that affects many classification methods. Prior to my research, the method was assessed on simulated data, and no information was available on the selection of key methodological parameters. This thesis aims to fill this gap while conducting a test with datasets acquired from natural targets spanning two disciplines. Because the NSSA can be calculated from a suite of spectra or endmembers, it is also well suited for use within spectral mixture analysis and therefore to analyze point data and imagery. Thus this application was pursued as part of my research.

A majority of band selection methods, especially those driven solely by classification accuracy, can be problematic in that they overlook the physical meaning of spectral features. Spectral bands are treated as variables rather than important information revealing physical or chemical properties of targets. In practice, remote sensing experts rely on spectroscopy to select meaningful bands (Clark, 1999) and to identify spectral absorptions or peaks associated with specific chemical bonds in targets. Throughout my research, I tried to examine the results of band selection in the context of the known spectroscopy of targets investigated to provide an applied meaning to the results.

Another issue that I identified in my early research is that novel methods are often evaluated using "standard" publicly available hyperspectral datasets typically those of Cuprite, Pavia and Indian Pines (Bioucas-Dias et al., 2013; Williams et al., 2017). However the users of these datasets are typically not familiar with the intricacies of the ground data that was assembled or with the characteristics of the scene, and they are unaware of flaws in the data or how the data can be misused. Throughout the development of my research, I aimed to design band selection methods that capture realistic features for target separation and test the feasibility of the methods in well-designed case studies for data that are well understood.

1.2 Thesis Objectives

This research aims to investigate the use of the N-dimensional Solid Spectral Angle (NSSA) for band selection in hyperspectral data, either from spectral libraries or imagery; and its integration with spectral mixture analysis for improved mapping of ground targets. The specific objectives are:

1) to establish guidelines for the selection of the two parameters that will enable the use of the NSSA band selection method for practical applications; and to evaluate the performance of the method in detecting subtle features from spectral data of geologic relevance (Chapter 2);

2) to identify key spectral features for geological and ecological applications using Hierarchical Band Selection with the NSSA Method (Chapter 3);

3) to investigate the use of band selection methods including the NSSA method for the discrimination of spectrally similar targets, here, diverse tree species of a tropical dry forest using thermal infrared hyperspectral data (Chapter 4);

4) to explore the use of the NSSA band selection method to improve the mapping of spectrally similar rock types by integrating this method with an endmember extraction method as part of spectral unmixing. In doing so, I aim to develop a new methodology that incorporated the NSSA band selection within the SSEE process (defined as "NSSA-SSEE) and examine the impact of band selection on endmember selection and spectral unmixing. This would imply the per pixel band selection in imagery (Chapter 5).

1.3 Thesis Outline

This thesis compiles three research papers and an IEEE conference paper that are either published, submitted, or ready for submission. The thesis starts with an introduction chapter followed by four chapters that respectively address the four objectives of the research. Chapter 6 provides conclusions.

Chapter 2 explores the use of the NSSA method for band selection from point spectral data of natural targets, which is a continuing of the methodological development by Tian et al., 2016. The paper has been published in 2019 as: Long, Y., Rivard, B., Rogge, D., & Tian, M., "Hyperspectral band selection using the N-dimensional Spectral Solid Angle method for the improved discrimination of spectrally similar targets" in the International Journal of Applied Earth Observation and Geoinformation, 79, 35-47. The two datasets used for analysis were provided by Dr. Jilu Feng and Dr. Derek Rogge, as part of previous research conducted in the Earth Observation Science Laboratory (EOSL) at the University of Alberta. The code for the NSSA algorithm was modified in the ENVI/IDL software based on the original version written by Dr. Minghua Tian, who is in the author list. Dr. Rogge made constructive comments on the manuscript, and Dr. Benoit Rivard was the supervisory author and helped the composition and editing of the manuscript.

Chapter 3 proposes a hierarchical strategy for band selection based on the NSSA method and applies the strategy in two geological and vegetation case studies, through which the paper demonstrates how band selection can be used for analyzing targets that show spectral variability. This study was presented as an oral presentation at the 2018 IEEE International Geoscience and Remote Sensing Symposium held in Valencia, Spain in July, 2018. The publication appears as Long, Y., & Rivard, B. (2018). Hierarchical Band Selection Using the N-Dimensional Solid Spectral Angle Method to Address Inter-and Intra-Class Spectral Variability. IEEE International Geoscience and Remote Sensing Symposium (pp. 8377-8380). The data for the geologic and vegetational applications are from prior studies by Dr. Rogge and Dominica Harrison. The paper was edited by Dr. Rivard as the supervisory author.

Chapter 4 applies multiple band selection methods to identify spectral features in Longwave Infrared (LWIR) spectra of leaves of tree species in the tropical dry forest of Costa Rica. The NSSA band selection method was used as an independent band selector and also combined with other methods to form an ensemble band selection that is then used for species classification. The resulting manuscript has been edited and prepared for publication as a research paper. Dominica Harrison provided the thermal infrared spectral data for the analysis in this paper. Dr. Rivard was involved with structuring and editing the paper with further edits required by Dr. Sanchez-Azofeifa.

Chapter 5 proposes a novel band selection method that incorporates the NSSA method in the spatial selection of spectral endmembers for a better distinction of similar targets. The resulting manuscript has been submitted to the International Journal of Applied Earth Observation and Geoinformation in March 2019. Dr. Rogge contributed to the design of the methodology and editing of the manuscript following edits by Dr. Rivard.

References

Abdel-Rahman, E. M., Ahmed, F. B., and Ismail, R., 2013. Random forest regression and spectral band selection for estimating sugarcane leaf nitrogen concentration using EO-1 Hyperion hyperspectral data. International Journal of Remote Sensing 34(2): 712-728.

Asner, G. P., 1998. Biophysical and biochemical sources of variability in canopy reflectance. Remote Sensing of Environment 64(3): 234-253.

Bajcsy, P., and Groves, P., 2004. Methodology for hyperspectral band selection. Photogrammetric Engineering & Remote Sensing 70(7) :793-802.

Bateson, C. A., Asner, G. P., and Wessman, C. A., 2000. Endmember bundles: A new approach to incorporating endmember variability into spectral mixture analysis. IEEE Transactions on Geoscience and Remote Sensing 38(2): 1083-1094.

Bioucas-Dias, J. M., Plaza, A., Camps-Valls, G., Scheunders, P., Nasrabadi, N., and Chanussot, J., 2013. Hyperspectral remote sensing data analysis and future challenges. IEEE Geoscience and Remote Sensing Magazine 1(2): 6-36.

Borengasser, M., Hungate, W. S., and Watkins, R., 2007. Hyperspectral remote sensing: principles and applications. CRC press.

Camps-Valls, G., Tuia, D., Bruzzone, L. and Benediktsson, J.A., 2013. Advances in hyperspectral image classification: Earth monitoring with statistical learning methods. IEEE signal processing magazine, 31(1): 45-54.

Chan, J. C. W., and Paelinckx, D., 2008. Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. Remote Sensing of Environment 112(6): 2999-3011.

Chang, C-I., 2000. An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis. IEEE Transactions on Information Theory 46(5): 1927-1932.

Chang, C-I., 2013. Hyperspectral data processing: algorithm design and analysis. Hoboken, NJ: John Wiley & Sons.

Clark, R. N., 1999. Spectroscopy of rocks and minerals, and principles of spectroscopy. Manual of Remote Sensing 3(3-58): 2-2.

Du, Q., and Yang, H., 2008. Similarity-based unsupervised band selection for hyperspectral image analysis. IEEE Geoscience and Remote Sensing Letters 5(4): 564-568.

Fassnacht, F.E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L.T., Straub, C. and Ghosh, A., 2016. Review of studies on tree species classification from remotely sensed data. Remote Sensing of Environment 186: 64-87.

Feilhauer, H., Asner, G. P., and Martin, R. E., 2015. Multi-method ensemble selection of spectral bands related to leaf biochemistry. Remote Sensing of Environment 164: 57-65.

Goetz, A. F., 2009. Three decades of hyperspectral remote sensing of the Earth: A personal view. Remote Sensing of Environment 113: S5-S16.

Harsanyi, J. C., and Chang C-I., 1994. Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach. IEEE Transactions on Geoscience and Remote Sensing 32(4): 779-785.

Jia, S., and Qian, Y., 2009. Constrained nonnegative matrix factorization for hyperspectral unmixing. IEEE Transactions on Geoscience and Remote Sensing 47(1): 161-173.

Jia, S., Tang, G., Zhu, J., and Li, Q., 2016. A novel ranking-based clustering approach for hyperspectral band selection. IEEE Transactions on Geoscience and Remote Sensing 54(1): 88-102.

Keshava, N., 2004. Distance metrics and band selection in hyperspectral processing with applications to material identification and spectral libraries. IEEE Transactions on Geoscience and Remote Sensing 42(7): 1552-1565.

Keshava, N., and Mustard, J. F., 2002. Spectral unmixing. IEEE signal processing magazine 19(1): 44-57.

Landgrebe, D., 2002. Hyperspectral image data analysis. IEEE Signal Processing Magazine 19(1): 17-28.

Li, S., Qiu, J., Yang, X., Liu, H., Wan, D., and Zhu, Y., 2014. A novel approach to hyperspectral band selection based on spectral shape similarity analysis and fast branch and bound search. Engineering Applications of Artificial Intelligence 27: 241-250.

Long, Y. Q, Rivard, B., and Rogge, D., 2019. Hyperspectral band selection using the N-dimensional Spectral Solid Angle method for the improved discrimination of spectrally similar targets. International Journal of Applied Earth Observation and Geoinformation 79: 35-47.

Plaza, A., Benediktsson, J.A., Boardman, J.W., Brazile, J., Bruzzone, L., Camps-Valls, G., Chanussot, J., Fauvel, M., Gamba, P., Gualtieri, A. and Marconcini, M., 2009. Recent advances in techniques for hyperspectral image processing. Remote Sensing of Environment 113: S110-S122.

Pu, R., 2017. Hyperspectral remote sensing: fundamentals and practices. CRC Press.

Rogge, D. M., Rivard, B., Zhang, J., and Feng, J., 2006. Iterative spectral unmixing for optimizing per-pixel endmember sets. IEEE Transactions on Geoscience and Remote Sensing 44(12): 3725-3736.

Rogge, D. M., Rivard, B., Zhang, J., Sanchez, A., Harris, J., and Feng, J., 2007. Integration of spatial-spectral information for the improved extraction of endmembers. Remote Sensing of Environment 110: 287-303.

Rogge, D., Rivard, B., Segl, K., Grant, B., and Feng, J., 2014. Mapping of NiCu–PGE ore hosting ultramafic rocks using airborne and simulated EnMAP hyperspectral imagery, Nunavik, Canada. Remote Sensing of Environment 152: 302-317.

Somers, B., Asner, G. P., Tits, L., and Coppin, P., 2011. Endmember variability in spectral mixture analysis: A review. Remote Sensing of Environment 115(7): 1603-1616.

Tian, M., Feng, J., Rivard, B., and Zhao, C., 2016. A method to compute the n-dimensional solid spectral angle between vectors and its use for band selection in hyperspectral data. International Journal of Applied Earth Observation and Geoinformation 50: 141-149.

Tuia, D., Flamary, R., and Courty, N., 2015. Multiclass feature learning for hyperspectral image classification: Sparse and hierarchical solutions. ISPRS Journal of Photogrammetry and Remote Sensing 105: 272-285.

Van Der Meer, F., 2004. Analysis of spectral absorption features in hyperspectral imagery. International Journal of Applied Earth Observation and Geoinformation 5(1): 55-68.

Van der Meer, F. D., and De Jong, S. M. (Eds.), 2011. Imaging spectrometry: basic principles and prospective applications (Vol. 4). Springer Science & Business Media.

Wang, S., and Chang, C-I., 2007. Variable-number variable-band selection for feature characterization in hyperspectral signatures. IEEE Transactions on Geoscience and Remote Sensing 45(9): 2979-2992.

Williams, M., Parody, R., Fafard, A., Kerekes, J., and van Aardt, J., 2017. Validation of abundance map reference data for spectral unmixing. Remote Sensing 9(5): 473.

Chapter 2 Hyperspectral Band selection Using the N-dimensional Spectral Solid Angle Method for the Improved Discrimination of Spectrally Similar Targets

2.1 Introduction

Hyperspectral remote sensing has been used in a wide array of applications over the last thirty years (Goetz, 2009; Bioucas-Dias et al., 2013). Among those applications, the mapping of ground   targets (e.g., rock, vegetation, soil, etc.) in remote and inaccessible areas greatly benefits from the availability of high spectral resolution data. Hyperspectral data consist of measurements acquired for a large number of contiguous spectral bands. As the number of spectral bands increases, which is typically commensurate with an increase in spectral resolution, the abundant information facilitates more accurate discrimination and identification of targets. However, some information may be redundant and not critical to the discrimination, specifically the strong correlation between contiguous bands can negatively impact data analysis algorithms (Peng et al., 2005; Van der Meer and Jia, 2012). The effectiveness of these algorithms for class separation can decrease once it reaches a certain number of bands, a phenomenon known as the 'curse of dimensionality' (Chang, 2013). The high data dimensionality (e.g., hundreds to thousands of bands) also results in a computational burden that can hinder hyperspectral analysis. These negative effects can be mitigated by dimensionality reduction techniques. Of these techniques, the process of band selection (i.e., feature selection), aims to identify a subset of original bands from the data, focusing on the most essential spectral features (e.g., peaks or troughs) to distinguish the targets of interest (Bajcsy & Groves, 2004; Keshava, 2004). Accordingly, band selection alleviates the negative effects on both accuracy and computational burden.

Most published methods of band selection perform effectively to discriminate objects that are spectrally distinct, but they may perform poorly when faced with objects displaying spectral similarity. For example, geological applications often involve similar rock types characterized by almost identical spectra (Rogge et al., 2007, 2014). Discrimination of plant species can also be impacted by the similarity in their reflectance spectra (Rock et al., 2016). Therefore, Tian et al. (2016) proposed the N-dimensional Solid Spectral Angle (NSSA) as an extension of the Spectral Angle (SA) to the n-dimensional space (Figure 2.1a), to select bands and improve the discrimination of objects with similar spectral properties. The earth observation remote sensing community is well versed with the use of the spectral angle mapper (SAM) method that enables the measurement of the spectral angle between two vectors. One of the most extensive use of this tool is to classify satellite or airborne multi or hyperspectral imagery using an input spectral library. The NSSA is an extrapolation of the SAM to more than 2 vectors and the resulting band selection method making use of the N-dimensional Solid Spectral Angle is intended to improve discrimination of spectrally similar targets. However Tian et al. (2016) conducted a limited assessment of the NSSA band selection method using synthetic data that left an important knowledge gap as to the applicability of the method for the investigation of natural targets and more importantly no guidelines were established for the parameter selection in the NSSA band selection method. The later would be a deterrent to the exploratory use of the method by the wider user community. Therefore, the primary objective of this paper is to examine the use of the NSSA method for different spectral libraries from natural targets, here of geological relevance. As well we aim to establish guidelines for the selection of the two parameters that will enable the use of the band selection method for practical applications including whether the two parameters can be assigned fixed values for a given application. As an outcome of this study we provide access to

the code for the method on a public forum (Github) and together with the guidelines resulting from this study, other researcher will be able to expand the scope of its application. Here we illustrate its use to analyze spectral libraries, comparing spectra with low spectral contrast that has applications for mapping using airborne and satellite hyperspectral data though other uses are possible that are the subject of future publications.

To achieve these objectives, this paper is organized as follows. The geological sample suite and spectral data collected are described in section 2.2. The method to examine the selection of the two parameters and for the evaluation of bands selected by the NSSA and VNVBS methods is explained in section 2.3. Results are presented in section 2.4 followed by a discussion. We also compare the performance of the NSSA band selection method that of its competitor, the Variable-Number Variable-Band selection method (VNVBS) (Harsanyi and Chang, 1994; Wang and Chang, 2007) though this method is limited to the selection of bands for the discrimination of two classes.

2.2 NSSA method

2.2.1 NSSA definition

While the spectral angle (SA) describes the similarity between two spectra (either collected from point spectrometer or hyperspectral image) by calculating the angle between them, the NSSA measures the solid angle or the similarity among n (more than two) spectra (Figure 2.1a). Treating spectra as vectors, the solid angle in n-dimensional space can be calculated (detailed equation seen in the Appendix and Tian et al., 2016), but the number of spectra defining the angle must be equal to the number of bands over which they are compared. For example, the NSSA among seven

spectra can only be calculated for a dimensionality of seven bands. This is a paramount characteristic of the NSSA method.

2.2.2 Use of the NSSA for band selection

Following the description of the computation of the NSSA, Tian et al. (2016) described an approach for band selection based on the characteristics of the NSSA for varying band sets. By selecting a subset of spectral bands characterized by the largest solid spectral angles (e.g., NSSA values), one can capture the most important spectral information to distinguish a set of spectra. For this method, the NSSA is computed by moving a sliding window with varying band intervals (denoted by "k") over the available spectral range (Figure 2.1). The band interval "k" is a user-defined parameter and can be adjusted in order for the sliding window to cover spectral features of varying widths; with larger k values covering wider features. For the example shown in Figure 2.1, seven spectra are used for the computation of the NSSA. When the band interval is zero (k=0), the window first encompasses reflectance values from the first seven contiguous bands (band1-band7). Note that the computation of the NSSA requires that the number of bands is equal to the number of spectra. This data matrix is used for the computation of the NSSA for the middle band (band 4) of the window. Then the window slides to the data subset encompassing band2-band8 for the determination of the NSSA value for band 5 and so forth, ultimately producing a continuous profile of NSSA values (red line in Figure 2.1) from which bands can be selected based on a given threshold value (dashed red line on Figure 2.1). If k=1, then the starting window is enlarged; and the window encompasses band 1, 3, 5, 7, 9, 11, and 13 to account for a one band interval while retaining a filter that encompasses 7 bands for 7 spectra. Bands with lower NSSA values (regions

lacking distinguishing features) are deemed less important because they reflect greater similarity amongst spectra within the specific subset spectral regions. Bands with NSSA values above the user-defined parameter threshold are retained. Thus, the threshold determines the number of bands selected. The two parameters, k and threshold, must be determined for NSSA band selection.

2.3 Data

This study makes use of hyperspectral data collected from two geological sample suites. The first sample suite consists of powdered clay minerals and the second consists of mafic and ultramafic rock samples. The spectral data from these two sample suites represent respectively a simple and more complicated case of band selection with the NSSA method. The powdered clay mineral samples are pure, simplifying their spectral discrimination in contrast to the rock samples, which in this case, encompass minerals juxtaposed to lichen coatings resulting in greater spectral variability. In addition, both sample suites were measured with spectrometers encompassing different spectral ranges and with a different number of bands. These differences allow for testing of the use and performance of the NSSA method.

2.3.1 Sample suites

2.3.1.1 Clay minerals

A suite of seven powdered clay minerals was obtained from the Clay Mineral Society repository (Figure 2.2a). These include clays with little or no structural water which for this study are Ca-Montmorillonite and Na-Montmorillonite [$(Na,Ca)_{0.33}(Al,Mg)_2(Si_4O_{10})(OH)_2 \cdot nH_2O$]. Also

included is a synthetic hectorite [$Na_{0.3}(Mg,Li)_3Si_4O_{10}(OH)_2$] which, as seen from spectra, is not water free. The remaining four clays are water-bearing and include palygorskite [$(Mg,Al)_2Si_4O_{10}(OH)\cdot4(H_2O)$], kaolinite [$Al_2Si_2O_5(OH)_4$] high and low defect, and a synthetic Mica-Montmorillonite [K partially substitutes for Na or Ca]. These clay samples underwent mineral characterization in a series of baseline studies by the Clay Minerals Society (Costanzo and Guggenheim, 2001; Costanzo, 2001). The samples show a high degree of purity enabling comparison of their spectra without the potential complications introduced by mineral mixtures. However, the use of powders as opposed to solid samples results in greater spectral contrast and thus should facilitate the distinction of spectra based on a reduced band set. Finally, clay minerals were selected because they represent a key mineral class in mining exploration and as such their spectra and spectral discrimination will be familiar to remote sensing geologists (Speta et al., 2016).

2.3.1.2 Mafic and ultramafic rock samples

The second sample suite consists of unaltered naturally exposed weathered surfaces of mafic and ultramafic rocks as described by Rogge et al. (2014). The 77 samples encompass six rock types namely dunite (6), peridotite (25), olivine-pyroxenite (9), pyroxenite (11), gabbro (16) and basalt (10). The first four are referred to as ultramafic rock types while the last two are mafic rock types. The sample surfaces display variable and typically extensive lichen cover as seen in Figure 2.2b. As reported by Rogge et al. (2014) using X-ray diffraction analysis, these weathered and metamorphosed rocks are comprised of similar mineralogy and differ in the relative abundance of three dominant minerals namely antigorite [$(Mg,Fe)_3(Si_2O_5)(OH)_4$], actinolite [$(Ca_2(Mg,Fe)_5Si_8O_{22})(OH)_2$], and clinochlore [$(Mg_6(Mg_4Al_2)[Si_6Al_2O_{22}](OH)_{16}$] with plagioclase also present in gabbro and basalt (Lesher, 2007). Dunite and peridotite have the highest abundance

of antigorite, olivine-pyroxenite and pyroxenite have a mixture of antigorite, actinolite, and clinochlore, and basalt and gabbro have actinolite and clinochlore. The presence of lichen on the weathered rock surfaces complicates the discrimination of rock types from spectra and as such represents a more challenging case study of band selection.

2.3.2 Acquisition of spectral data

The seven powdered clay samples were scanned at the University of Alberta using a SisuROCK imaging spectrometer (Specim Ltd., www.specim.fi). As discussed in the next section, clays show discernible spectral variations in the Short-wave Infrared (SWIR) region, so radiance measurements were collected for 256 bands in the wavelength range of 0.970-2.530 μm. The first eleven bands were omitted owing to lower responsivity, leaving 245 bands extending from 0.997-2.530 μm. A reflectance spectrum for every pixel in the image was then obtained by normalizing the radiance obtained from the sample to that of a Spectralon™ panel (~100% reflectance). The seven clay samples were imaged at a spatial resolution of 1 mm per pixel. For data analysis, a first dataset referred herein as ClaySamples consisted of the image data encompassed by a region of interest (ROI) over the most uniform part of each clay powdered surface. The ROI's on average captured 460 pixels per sample (min 163, max 668). From these data, a second dataset referred herein as ClayAverages, consisted of the average spectrum captured by each ROI resulting in seven clay spectra.

Spectra for the seventy-seven mafic and ultramafic rocks were collected as part of a prior study, with one spectrum reported per sample (Rogge et al. 2014). The spectra were measured using an Analytical Spectral Devices (ASD) spectrometer that has 2151 bands ranging from 0.350 to 2.500 μm and an average band sampling interval of 2 nm. Thus these spectral data display a

considerably higher spectral dimensionality (number of bands) and resolution than that obtained

for the clay samples with implications, as seen below, on the selection of k values in the band

selection process. For each rock sample, spectral measurements were initially taken at 4–11

random points (2 cm diameter), as illustrated by the numbered paper on the photo of sample

surfaces seen in Figure 2.2b. The spectral measurements collected for each sample were then

averaged to obtain a representative spectrum per sample. Note that these spectra encompass

contributions from weathered rock and lichen. For data analysis, a first dataset, referred herein as

RockSamples, consisted of the average spectrum for each of seventy-seven samples. From these

data, a second dataset, referred herein as RockAverages, consisted of the average spectrum per

rock type resulting in six spectra. The spectral range of 0.426-2.426 µm (2001 bands) was retained

for analysis to minimize regions of poorer system responsivity.

2.3.3 Spectral characteristics of samples

To assist in the evaluation of the band selection results, diagnostic spectral features observed for

each sample suite are summarized below. These features were labeled on the basis of features

documented in the literature and serve as a reference band set.

2.3.3.1 Spectral features of clay minerals

The spectral characteristics of clay minerals are described in the literature (Hunt, 1977;

Clark et al., 1990; Bishop et al., 2008). The spectra of the seven clays (ClayAverages) fall into two

broad categories based on diagnostic absorption features (Figure 2.3). The absorption feature

centered near 1.4 µm is present in all the clay minerals of this study, but its depth, position, and to

some extent, its asymmetry varies across the spectral suite. The first category includes kaolinite

high and low defect, minerals with little to no structural water and their spectra display a weak

absorption near 1.9 um. Kaolinites present a distinct triplet absorption feature centered at 1.4 μm and a doublet at 2.162 and 2.205 μm. The spectra of both kaolinites in this study are distinguishable on the basis of weak differences in reflectance in certain wavelength regions and in the depth of the strongest absorptions at 1.4, 1.9 and 2.386 um. Minerals in the second category (hectorite, palygorskite, and three montmorillonites) are water-bearing and display a strong absorption feature at 1.9 μm. These minerals can be discerned on the basis of the shape and position of the water absorptions near 1.4 and 1.9 μm and of a singlet cation-hydroxyl feature near 2.200 or 2.311 μm. Hectorite has no obvious features around 2.200 μm but has an absorption band at 2.311 μm. Palygorskite and the Na and Ca montmorillonites can be distinguished on the basis of the detailed location of a singlet absorption near 2.200 μm. These two montmorillonites have absorptions located respectively at 2.180 um to 2.205 μm and the feature of palygorskite is closer to 2.218 μm. The mica-montmorillonite is a synthetic sample that displays two absorptions near 2.100 and 2.180 μm. Thus the band selection process is expected to reveal bands that capture the presence of key absorptions and their subtle positional differences across minerals.

2.3.3.2 Spectral features of mafic-ultramafic rocks

Figure 2.4 displays the average spectra of the six rock units (RockAverages) that include the spectral contributions of lichen and weathered rock (e.g., minerals). Their broad overall spectral similarity is attributed to the extensive coverage of lichen. As explained in Bechtel et al. (2002), spectra of rock encrusting lichen can vary significantly in the visible and near-infrared (0.4–0.8 μm) due to light-absorbing pigments that explain their black, grey, white, yellow, green, and orange color (Figure 2.4a). The reflectance of lichens beyond the VNIR (>1 μm) increases reaching a maximum near 1.860 μm with a small and relatively narrow absorption feature at 1.730 μm attributable to cellulose. Pronounced absorptions near 1.4 and 1.9 μm are due to water. Lichen

spectra are very similar in the SWIR. For pure lichen surfaces, two broad absorption features also caused by cellulose are seen near 2.1 and 2.3 μm, but their detail is modified by the presence of three key minerals for the samples of this study. As shown by Rogge et al. (2014), spectra of powdered antigorite, actinolite and clinochlore taken from the USGS spectral library (http://speclab.cr.usgs.gov/spectral- lib.html) present diagnostic absorption features at [2.285, 2.325], [2.245, 2.315, 2.386] and 2.345 μm respectively, which are caused by various overtone and combination tones of the OH stretching (Clark et al. 1990; Clark, 1999). This information translates into key observations for the discrimination of the six rock units. As seen from the detailed view of spectra in the shortwave infrared (Figure 2.4b), the decreasing relative abundance of antigorite versus actinolite leads to a subtle shift in a prominent absorption from 2.323 to 2.315 μm and the progressive increase of an absorption at 2.386 μm progressing from dunite, peridotite, olivine-pyroxenite, and pyroxenite. In contrast, basalt and gabbro contain more clinochlore, resulting in a relatively flat and broad absorption spanning 2.315 to 2.345 μm.

The observed absorptions for lichens and minerals for these rocks led Rogge et al. (2014) to manually select bands for the lithologic analysis of airborne data acquired with the Specim AisaDUAL system. A relevant subset of these bands is listed in Table 2.1 and serves to evaluate the NSSA band selection given the successful mapping results obtained and validated by Rogge et al. (2014) using field evidence. Seven bands were selected from 0.45-0.75 um with an approximate interval of 0.05 um because lichen species and some non-geological materials are more readily discriminated in this region. Note that some of the bands listed in Table 2.1 were selected from AISA airborne imagery with 10nm spectral resolution, a resolution considerably lower than that of the ASD data of this study and thus disparities in the fine detail of band selection results (e.g., band location) can be expected.

2.4 Methods

This section describes three methodological aspects of this study. Working with the hyperspectral datasets obtained from the clay minerals and rock samples, we first describe adjustments to the k parameter to examine impacts on NSSA profiles and consequently band selection. Next, we determine an approach to select the k and threshold parameters to define a band set for each dataset. Afterward, the selected bands are evaluated in regards to the agreement with known features and the impact of band selection on class separability. Lastly, the bands selected with NSSA are compared to that obtained from the VNVBS method.

2.4.1 Parameter adjustment

For this aspect of the study, we made use of the average spectra per clay sample (7 spectra of ClayAverages) and the average spectra per rock type (6 spectra, RockAverages).

2.4.1.1 Adjustment of k

As stated in section 2.2.2, k defines the width of the spectral region over which the NSSA is measured and thus relates to the width of the spectral features seen in spectra that influence the measurement of the NSSA. The maximum width of the sliding window cannot exceed the entire available spectral range. The maximum k value cannot exceed (l-n)/n, where l and n denote the number of bands and spectra in a given dataset. Based on this equation, the maximum value of k for ClayAverages (245 bands, 6 spectra) and RockAverages (2001 bands, 7 spectra) is 39 ((245-6)/6) and 284 ((2001-7)/7) respectively. When possible, one can further constrain the k values to

investigate based on a priori knowledge of realistic feature widths encountered in the spectra of targets of interest.

For clay minerals, the broadest features expected occupy less than one-quarter of the full wavelength range available so less than 60 bands. Therefore, the maximum value for k could be further reduced to 9 (bands 1, 11, 21, 31, 41, 51, 61). In the same manner, the maximum k for the rock type can be reduced to 63 on the basis that absorptions do not exceed 500 bands in width. Based on the above constraints, continuous values of k (0-9) were used for the clay dataset. However when the number of bands is large, as is the case in the rock dataset (2001 bands), the use of continuous k values can impose an overwhelming computational burden. Thus for this dataset, we examine a subset of profiles (k=0,1,3,7,15,31,63) by doubling the window size starting with k=0 (7 bands wide). For each dataset, a continuous profile of NSSA values as a function of wavelength was produced for each k value. It thus follows, as will be shown in section 2.5 and discussed in section 2.6.2, that the spectral dimensionality, namely the number of bands, which also typically relates to spectral resolution (smaller bandwidths), can require the selection of a subset of k values for the band selection process.

2.4.1.2 Adjustment of the threshold

As stated in section 2.2.2, a threshold can be applied to a given continuous profile of NSSA values to retain bands with greater NSSA values and thus of higher importance in target discrimination. Early research revealed that NSSA values decrease remarkably with increasing k as results will show below. This implies that a single threshold cannot be applied to continuous profiles of NSSA values obtained for different k. To define how a threshold could be selected for a range of k values, we examined the graphical representation of ranked NSSA values obtained for

each k value used in section 2.4.1.1. These graphs can then be compared and, if a general pattern can be observed for various k values, a threshold can be defined from the derivative analysis of the ranking curve. This strategy was explored to estimate a threshold for each k value.

As will be seen below, the first result establishes the selection of a subset of k values and their corresponding threshold for the clay and rock datasets. From this process, bands are selected for each k value, and these are then combined in a single band set that can then be evaluated for the given dataset.

2.4.2 Evaluation of selected bands

Bands selected for ClayAverages and RockAverages were first evaluated by examining the agreement between spectral regions encompassed by the selected bands and that of spectral features documented in the literature. For clay minerals, these features are described in section 2.3.3 and for the rock types these features were documented by Rogge et al. (2014) and are listed in Table 2.1. Then the bands selected for ClayAverages and RockAverages were respectively used for ClaySamples (clay image data, 245 bands and an average of 460 pixels per mineral) and RockSamples (average spectrum of each rock sample, 2001 bands and 77 samples) to examine class separation. Specifically, for each case, we compared the class separation with and without band selection. In doing so, we assess if distinctive and informative bands are selected (Khoder et al., 2015). Principle Component Analysis (PCA) was chosen for the analysis of ClaySamples as this dataset has highly correlated variables. Because the clay samples are compositionally homogeneous (e.g. pure), results of classification accuracy are close to 100% accuracy with or without band selection and do not offer valuable insights. Such insights are however obtained

when the data distribution is shown as scatterplots of the loadings of selected principal components. PCA was not used for the RockSamples because it requires that the number of samples be considerably larger than the number of bands. Thus, Linear Discriminant Analysis (LDA) was used to visualize class separation by plotting loadings of selected discriminants that are the linear combinations of original features (Hastie et al., 2008).

2.4.3 Method comparison

Bands selection for ClayAverages using NSSA were compared to that selected by the Variable Number Variable Bands Selection (VNVBS) method. This method has been proposed for feature selection from a single spectrum based on the similarity between this spectrum and a reference spectrum. It measures an Orthogonal Subspace Projection (OSP) - Band Prioritization Criterion (BPC) by applying a matrix transformation on the two spectra until all undesired features are removed and the spectra have been decomposed into two orthogonal components. Bands with high priority scores derived from this OSP-BPC are retained after being compared against the scores of the reference spectrum (Harsanyi and Chang, 1994; Wang and Chang, 2007). This method, like NSSA, can be applied to a spectral library; but it requires a reference spectrum against which all other spectra are compared (Wang and Chang, 20007). In this study, the reference spectrum is the average spectrum of the seven clay mineral spectra, an approach used in past studies with the VNVBS method (Sun et al. 2014). After selecting the reference spectrum, two sets of results are obtained for each mineral that include a band set and a profile of scores as a function of wavelength measured by the OSP-Band Prioritization Criterion (BPC). The profile can be interpreted as the band-importance for the given mineral spectrum.

2.5 Results

2.5.1 Parameter adjustment

Continuous profiles of NSSA values as a function of wavelength were produced for selected k values using ClayAverages (Figure 2.5) and RockAverages (Figure 2.6). These profiles reveal wavelength regions of high NSSA values where the spectra of interest are best discerned. Because k defines the width of the sliding window for the NSSA computation, the profiles can highlight regions, and thus features or groups of features, of greater width as the k value increases as seen near 2.2 μm in Figure 2.5. With increasing k values there is a general increase in the NSSA values, for example, values in the order of 10^{-15} at k=0 and 10^{-12} at k=3 for clay minerals (Figure 2.5). On this figure, one can also observe the progressive loss of some regions of higher NSSA values with increasing k value, as seen near 1.4 μm. Referring to Figure 2.3 this region encompasses a relatively narrow wavelength interval where displacement of the hydroxyl or water absorption feature is observed across several clay minerals (from 1.395-1.414 μm). This key interval is thus captured at low k values despite being part of a water feature that is relatively broad. For clay minerals, another region of high NSSA value is observed near 2.2 μm for k=0 and 1. This region gets broader extending to about 2.4 μm for the highest k values. However, within this range, two prominent excursions of NSSA values near 2.2 and 2.3 μm are seen in profiles for k values > 2. On that basis, and to retain the regions at 1.4 and 1.9 μm observed at low k values, band selection for ClayAverages was performed using NSSA profiles at k=0, 1 and 3. The profile for k=2 is omitted because it offers largely redundant information to that of k=1. In doing so, the primary regions of high NSSA values seen across all k values for ClayAverages are encompassed. Band selection can then proceed from this selection of NSSA profiles. To define how a threshold of

NSSA values could be selected for band selection, Figure 2.5 also shows the graphical representation of ranked NSSA values obtained for each of the selected k values. From these graphs, a general pattern is observed characterized by an asymptotic decrease in NSSA value as a function of the number of bands. The presence of a graphical pattern enables a method of band selection. Here the number of bands was selected setting a threshold (black arrows on Figure 2.5b) that is defined as the maximum value of the second derivative of the ranking curve (e.g., slope of 0). From this process bands retained at each of the selected k values are combined in a single band repository.

For RockAverages, the six NSSA profiles shown on Figure 2.6 span a wide range of k values from 0 to 63. Profiles at k=3,7,15 appear distinct from those at k=31,63. For the first group, regions of high NSSA values are located around 0.5, 1.4 and 2.3 μm, but for the second group the regions near 1.4 and 2.3 μm are not apparent and the one near 0.5 μm shifts to longer wavelengths and is at its broadest for k=63. As was done for ClayAverages, these observations guided the selection of profiles that capture the range of dominant regions of high NSSA values. Thus, as shown in Figure 2.6b, profiles at k values of 3, 15, and 63 were used to obtain a graphical representation of ranked NSSA values and conduct band selection.

2.5.2 Evaluation of selected bands

On the basis of the observations described in the prior section, a total of 44 and 265 bands were selected from clay minerals and rock types. The location of these bands is shown on Figures 7 and 8 along with the position of known relevant mineral features described in section 2.5.2.1 (see features in Table 2.1 for the rock types) as part of background knowledge. These figures also show the spectra of the clay minerals and rock types respectively to facilitate the evaluation of the

selected bands. Table 2.2 lists the 44 bands of clay minerals and the wavelength regions encompassed by the more numerous bands (265) selected for rock types that have an almost tenfold greater spectral dimensionality.

2.5.2.1 Agreement with known features

As expected for clay minerals, features near 1.4 and 1.9 μm are selected (Figure 2.7). The absorption feature centered near 1.4 μm is present in all the clay minerals of this study but its depth, position, and to some extent its asymmetry, varies across the spectral suite. The relatively narrow wavelength range occupied by the selected bands (1.383-1.415 μm) with respect to the full width of the feature and the position of the selected bands near the center of the absorption indicates that spectral differences across the clay minerals near 1.4 um are primarily capturing the variable position of the absorption center (1.395-1.414 μm). In contrast, the feature depth at 1.9 μm is highly variable in the clay minerals of this study and these minerals display variability in the shape of this absorption feature. Consequently, a larger number of bands (1.886-1.949 μm) are selected to account for these differences and for the observed range of absorption position centers (1.904-1.917 μm). The largest number of selected bands is in the SWIR ranging from 2.149 to 2.287 μm (Figure 2.7b) and, with the exception of hectorite, encompassing the full width of the most pronounced absorption features of these clays. The near contiguous set of SWIR bands accounts for the need to capture differences in the number, position, depth, and asymmetry of absorptions that occur in the same spectral region. The absorption for hectorite near 2.3 μm is distinct and consequently few bands were selected to capture this feature. The bands selected for clay minerals encompass spectral regions that are consistent with the mineral relevant spectroscopy knowledge used to discriminate these minerals.

For the rock types seen in Figure 2.8, the bands selected encompass three primary regions that essentially include the bands listed in Table 2.1 and selected by Rogge et al. (2014) on the basis of spectral knowledge of non-geological materials, lichens, and minerals relevant to the rocks investigated. In the VNIR, this study selected bands in three regions (0.436-0.493, 0.586-0.677, 0.700-0.713 μm) consistent with variability in light-absorbing pigments that explain their black, grey, white, yellow, green, and orange color. Rogge et al. (2014) selected seven bands uniformly distributed from 0.45-0.75 μm, and these were primarily selected to distinguish non-geological materials, not the rock types. It is thus not surprising that the bands selected in this study within this wavelength interval encompass more defined spectral regions. Next one notes the selection of a few bands near 1.4 μm on the edge of an absorption feature of variable depth across rock types. Importantly the NSSA method effectively selects a near contiguous suite of bands in the SWIR encompassing four of the five wavelength regions selected by Rogge et al. (2014) to account for key absorptions features of the primary rock-forming minerals, namely antigorite, actinolite, and clinochlore. The band at 2.245 μm, which is missing in the selection, sits immediately outside of the lower extent of the region selected by the NSSA method (starting at 2.256 μm).

2.5.2.2 Impact of band selection on class separability

The bands selected above for ClayAverages and RockAverages were respectively used for ClaySamples (average of 460 pixels per clay mineral image) and RockSamples (77 rock samples) to compare class separation with and without band selection. From the PCA conducted on ClaySamples and the LDA conducted on RockSamples, Figures 2.9 and 2.10 show the distribution of these respective data as scatterplots of the loadings of selected components or discriminants.

For clay minerals, two combinations of the first three principal components are shown as these provided the best class separation. The most significant improvement resulting from band

selection is the ability to discern three classes namely kaolinite L, kaolinite H, and Ca-montmorillonite. In particular, the first two classes show extensive overlap prior to band selection (Figure 2.9a, 2.9b). This should come as no surprise as seen by the similarity in these spectra in Figure 2.3. Band selection also resulted in a decrease in intra-class variability particularly in Figure 2.9d that provides complete class separation for all data points. The reduced intraclass variability is particularly apparent for Ca-montmorillonite.

For rock types, Figure 2.10 shows plots of the first three linear discriminants. The mafic rocks (gabbro and basalt) are readily discernable from the four ultramafic rocks along the LD1 direction, with or without band selection. Amongst the ultramafic rocks, dunite can be discerned from all other rocks using a particular combination of linear discriminants without (LD 1 and 2) or with (LD 2 and 3) band selection. Olivine pyroxenite is only distinguished from all other rock types using band selection (LD 2 and 3). Peridotite (Figure 2.10 cross symbol) is consistently mixed with pyroxenite though several pyroxenite samples of this study are better distinguished from those of peridotite using band selection (LD 2 and 3). Note that these ultramafic rock types have similar mineralogy and differ primarily in their abundance of the serpentine mineral with peridotite having a relatively lower and higher abundance than dunite and pyroxenite respectively. Thus, the observed overlap of peridotite and pyroxenite on Figure 2.10 is consistent with the mineralogical continuum displayed by these rocks. However, band selection improves the discrimination among dunite, olivine-pyroxenite and pyroxenite, as seen from the three dotted circles in Figure 2.10a and 2.10b as well as the distinction of olivine pyroxenite from all other rocks, and to some extent that of pyroxenite as well. Generally, the band selection enhances discrimination among similar ultramafic rocks.

32

2.5.3 Comparison with bands selected using the VNVBS method

According to the rules in VNVBS, the reference spectrum has to provide information from the original spectra (Wang and Chang, 2007). Through several experiments using VNVBS, the average spectrum of the seven clay mineral spectra was selected as the suitable reference spectrum. Other reference spectra were tested, for example, a random combination including some minerals (kaolinite L*0.2 + hectorite*0.3) or including all spectra ([random weights]*[spectra of seven minerals]). When the spectrum of any mineral was discarded, subtle changes in the calculation processes appeared, while, using all spectra and only adjusting their weights produced little difference. However, in all trials, no bands were selected. To ensure that there were no issues with the functionality of the method, tests were conducted with targets presenting a high spectral contrast, for example, a comparison of soil and water spectra. In such a case the result showed that the method could effectively select a subset of bands. Though no bands were selected for the datasets of this study, the scores calculated based on the OSP-BPC can be used for a comparison of VNVBS and NSSA results.

For the VNVBS method, a pair of scores from two orthogonal components provides key information on a spectrum (Pr1) of interest with respect to the reference spectrum (Pr2). The score for the spectrum of interest (Pr1) can be used to prioritize bands (e.g., highest score for highest importance) (Wang and Chang, 2007). Figure 2.11 plots the band importance represented by the VNVBS band prioritization score for each mineral. For the two kaolinites, the water feature centered near 1.9 μm has high importance values (Figure 2.11). Because no bands are selected, a specific number of bands or the width of captured features could not be known. The same limitation applies to the known absorption features near 2.2 μm seen as high values in the profiles for Palygorskite, and Ca-and Na-Montmorillonite. A wide region from 1.0-1.4 μm is highlighted

in the profiles of hectorite and Mica-Montmorillonite; however little variation was observed from the spectrum of these minerals (Figure 2.3) suggesting that this region has low significance. Also, an absorption of hectorite near 2.3 μm, that distinguishes it from the other clay minerals, is not highlighted. Unlike the NSSA results, the VNVBS results are not as highly consistent with known distinguishing mineral spectral features and do not provide adequate information for the selection of bands that are key to the discrimination of these clay minerals.

2.5.4 Computing time

The applicability of the NSSA band selection method for analysis of spectral data with high dimensionality (e.g., hyperspectral data) is dependent on the computing efficiency of the program used to compute the NSSA. As part of this study, we implemented improvements in ENVI/IDL on the code for the calculation of the NSSA made available by Tian et al. (2016). Here we present observations on the performance of the calculation. Figure 2.12 reveals that the computing time is dependent on the data size and increases with the numbers of spectra (n) and the number of bands (l). For our datasets, l is much larger than n; thus the computing time is primarily impacted by the number of bands, i.e., the spectral dimensionality. The maximum computing time for our data is approximately10 minutes when the number of bands is in the thousands, and the number of spectra is around 20. Most airborne and spaceborne hyperspectral data do not exceed two to three hundred bands thus it is entirely feasible to apply the use of the NSSA to the analysis of such imagery. The current version of the NSSA program was run on a computer equipped with an Intel(R) Core (TM) i5-4590 @3.30GHz CPU, with 8 GB of installed memory (RAM) and a 64-bit Operating System.

2.6 Discussion

2.6.1 Merits of the NSSA band selection method

There are several valuable band selection methods reported in the literature, but few can be applied to datasets encompassing a relatively small number of spectra and to select bands that enable the discrimination of spectrally similar materials. The NSSA band selection method can do both and is thus potentially useful in a range of hyperspectral applications and related analysis scenarios. The primary aim of this study was to establish guidelines for the selection of key parameters using this band selection method and thus focus was given to one use of the method, namely the analysis of spectral libraries, with dataset examples pertinent to a geologic audience and materials that are challenging to discriminate because of their spectral similarity.

Our findings indicate that bands selected from the two datasets are in good agreement with known features. Of significance is that the selected bands encompass a range of distinguishing and often subtle spectral characteristics that include absorption feature position (2.25-2.35 μm for rocks and 1.383-1.415 μm for clays), and shape (asymmetry) and depth (1.9 μm for clay). These results indicate that the NSSA method is sensitive to the key discriminating spectral characteristics of these materials, the same that are recognized by experts, and thus can be used to assist experts in identifying key features when using a limited number of spectra. Bands selected for the two datasets resulted in the improved separation of targets as seen in reduced intra-class variability (e.g., Ca-montmorillonite) and generally better discrimination of specific classes (e.g., ultramafic rocks). Kaolinite L and H, materials that present very similar spectra, could readily be distinguished following band selection.

The assessment of the current computational burden of the method has relevance for its future use in a wider range of data analysis scenarios. Prior to this investigation, several hours were consumed to process a data set of fewer than 10 spectra and 200 bands for the NSSA calculation whereas 200 seconds are now required due to the improved efficiency of the codes resulting from this study. One can now consider using this method in image analysis where data sizes exponentially increase.

2.6.2 Guidelines for the use of the NSSA band selection method

A key outcome of this study is that it is now feasible to provide general recommendations for the selection of the k and threshold parameters that define the use of the NSSA method.

The first guideline consists in constraining the maximum k value based on the spectral dimensionality of the widest significant spectral feature (typically an absorption feature) expected from the materials under study. Therefore, the maximum window size should cover the widest feature. For example, if the broadest feature seen in a group of 8 spectra encompasses a width of 90 bands, the maximum k value under consideration would be 10 (calculated as 90/(8+1)). Furthermore, if the multiple continuous k values impose a computational burden, we recommend examining k at intervals by doubling the window size (e.g., k=0,1,3,7,15…). With the resulting permissible k values defined and using a zero threshold, a profile of the NSSA value as a function of wavelength for each k value can be produced. These profiles will reveal wavelength regions highlighted through the range of k values.

The second guideline applies to the selection of k values based on these NSSA profiles. This step aims to capture the primary wavelength regions of high NSSA values that can be compiled across all k values and exclude, when possible, k values that contain redundant features.

In the last guideline, the threshold parameter should be defined for each k selected, and the value should be estimated from a graph of the NSSA value as a function of the number of bands as illustrated in Figure 2.5b. The methodology for this estimation was described in section 2.4.1.2 and results shown in section 2.5.1.

2.6.3 Additional applications of the NSSA band selection method

The use of the NSSA band selection method on the geological datasets of this study resulted in the selection of band sets consistent with known spectral features. Future work can thus expand to fields outside of geology. The use of vegetation spectroscopy in ecologic applications is a prime target as it typically involves materials with high spectral similarity (e.g., optical region) and/or low contrast (e.g., longwave infrared). Materials in urban settings or soils can also present very similar spectra, and studies attempting to discriminate these materials based on the use of spectral libraries could also benefit from this band selection method as should any study involving spectrally similar targets.

Another application concerns the analysis of airborne and satellite hyperspectral imagery and the use and extraction of spectral endmembers that can vary as a function of time (e.g., phenological cycle) and sensor spatial and spectral resolution (Somers et al., 2011). Because the selection of bands using the NSSA method can help to increase class separation and minimize intra-class variability, there is potential in integrating this method with the endmember selection

process as part of spectral Mixture analysis (SMA). Regional mapping or monitoring investigations can often encompass imagery covering hundreds of square kilometers, and it is a challenging and time-consuming process to assemble spectral endmembers prior to unmixing. With the upcoming launch of hyperspectral satellite missions such as PRISMA (PRecursore IperSpettrale della Missione Applicativa) and the Environmental Mapping and Analysis Program (EnMAP) and the Hyperspectral Infrared Imager (HyspIRI), the NSSA method may prove valuable to assemble and analyze such datasets.

2.7 Conclusions

The NSSA band selection method can detect subtle differences among spectrally similar targets and select bands for their discrimination. The method can be used for the analysis of a limited number of spectra and is thus well suited to the analysis of spectral libraries either resulting from the collection of field spectra or the extraction of endmembers from imagery. When applied to such data, the results indicate that spectral regions with highest contrast in spectral shape (depicted by the largest NSSA values) are captured. For the geological datasets investigated, the spectral regions captured are consistent with documented features of mineral spectra. Thus, the NSSA band selection method is a valuable tool to assist experts in feature identification and spectral analysis.

This study was also able to define guidelines to constrain the selection of the two key parameters behind the method. The first guideline consists in constraining the maximum k value based on the spectral dimensionality of the widest significant spectral feature (typically an absorption feature) expected from the materials under study. Therefore, the maximum window size should cover the widest feature. Furthermore, if the multiple continuous k values impose a

computational burden, we recommend examining k at intervals by doubling the window size (e.g., k=0,1,3,7,15…). The second guideline is to use a profile of the NSSA value as a function of wavelength for each permissible k value to capture the primary wavelength regions of high NSSA values. These can be compiled while excluding, when possible, k values that contain redundant features. In the last guideline, the threshold parameter for each k is estimated from a graph of the NSSA value as a function of the number of bands as illustrated in Figure 2.5b. The guidelines on the parameter definition allow non-expert users to select a subset of bands while capturing both narrow and broad discriminating features. The NSSA method is a powerful tool for band selection that should be of value to any study involving spectrally similar targets.

Acknowledgments

Table 2.1 Bands used to evaluate the NSSA band selection for the rock datasets.

Spectral range	Band location (μm)	Source
VNIR	0.454, 0.500, 0.546, 0.603, 0.651, 0.698, 0.746	Table 3, Rogge et al. (2014) from AISA imagery
SWIR	2.245, 2.285, 2.315, 2.345, 2.386	Figure 7, Rogge et al. (2014) from USGS spectral library

Table 2.2 Bands or wavelength regions obtained from the NSSA band selection method.

Dataset	Bands or wavelength regions (µm) of selected bands
Clay minerals	1.383, 1.389, 1.396, 1.402, 1.408, 1.415, 1.886, 1.892, 1.899, 1.905, 1.911, 1.917, 1.924, 1.930, 1.936, 1.943, 1.949, 2.124, 2.130, 2.137, 2.149, 2.155, 2.162, 2.168, 2.174, 2.180, 2.187, 2.193, 2.199, 2.206, 2.212, 2.218, 2.224, 2.231, 2.237, 2.243, 2.249, 2.256, 2.262, 2.274, 2.281, 2.287, 2.306, 2.355
Rock types	0.436-0.493, 0.586-0.677, 0.700-0.713, 1.391-1.404, 2.256-2.386, 2.405-2.411
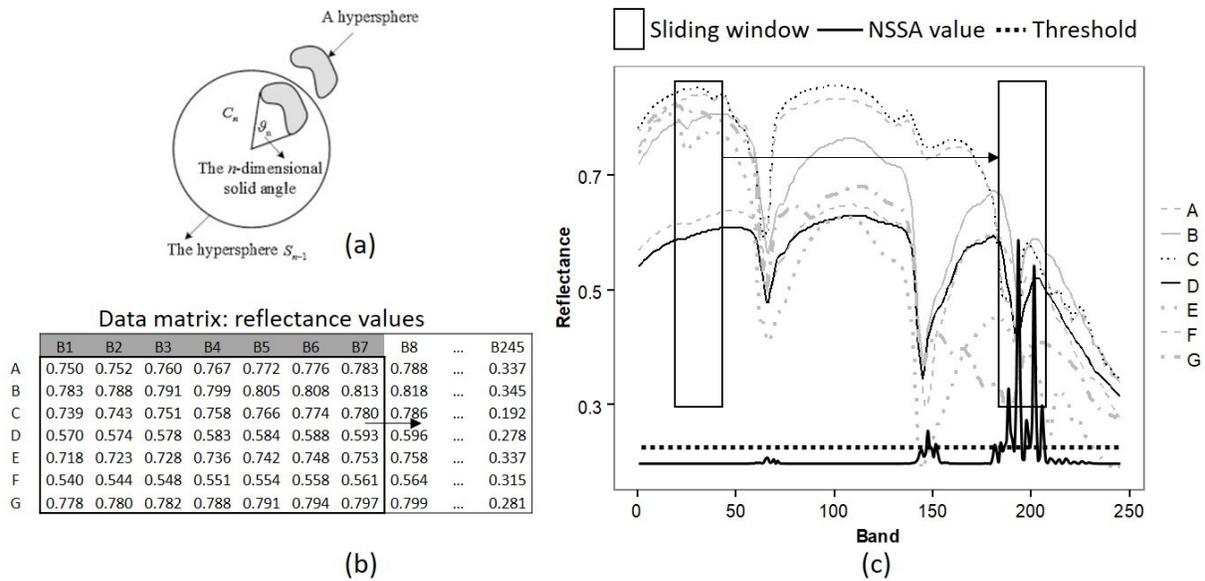
Figure 2.1 Schematic representation of the NSSA (a) and its use for band selection (b) and (c). The example provided is for 7 spectra (A-G). The sliding window, calculated NSSA profile, and threshold are also shown.
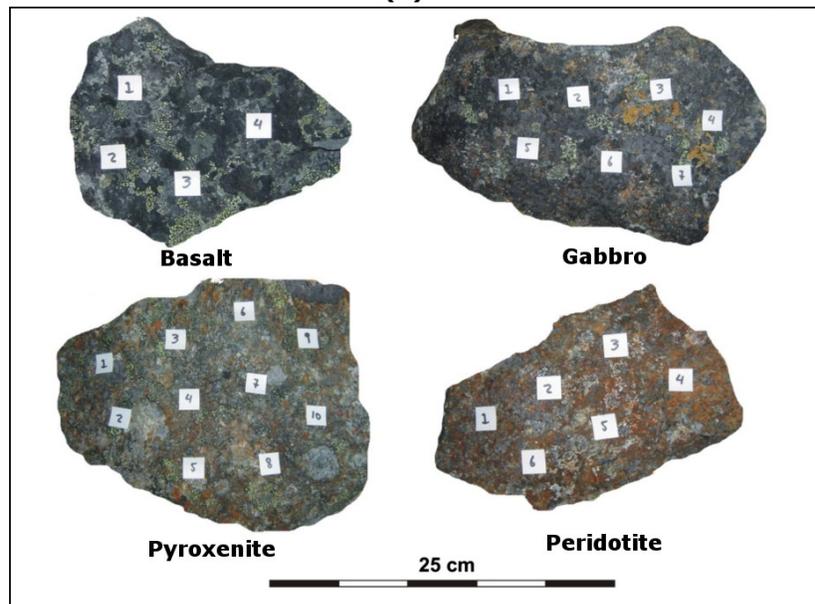
Figure 2.2 Photographs of samples. Powder clay minerals (a), and mafic and ultramafic weathered and partially lichen covered rock samples (b). The numbered white paper squares in (b) indicate the location where point spectra were collected.
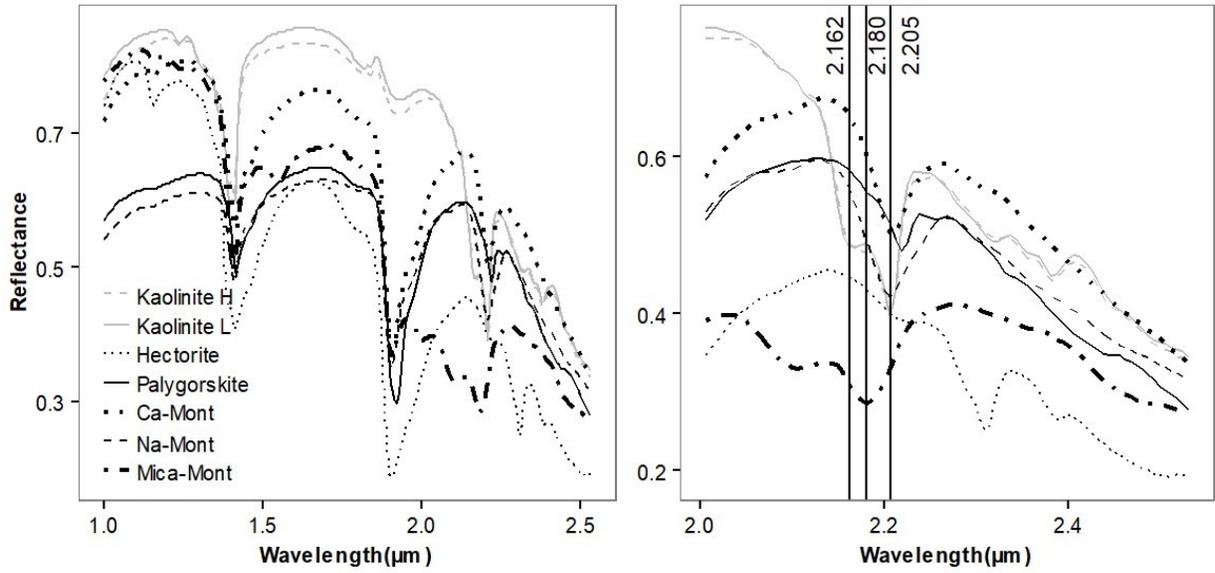
Figure 2.3 Averaged spectra of clay mineral samples (ClayAverages) (a), and enlargement of features observed between 2.00-2.53 μm (b).
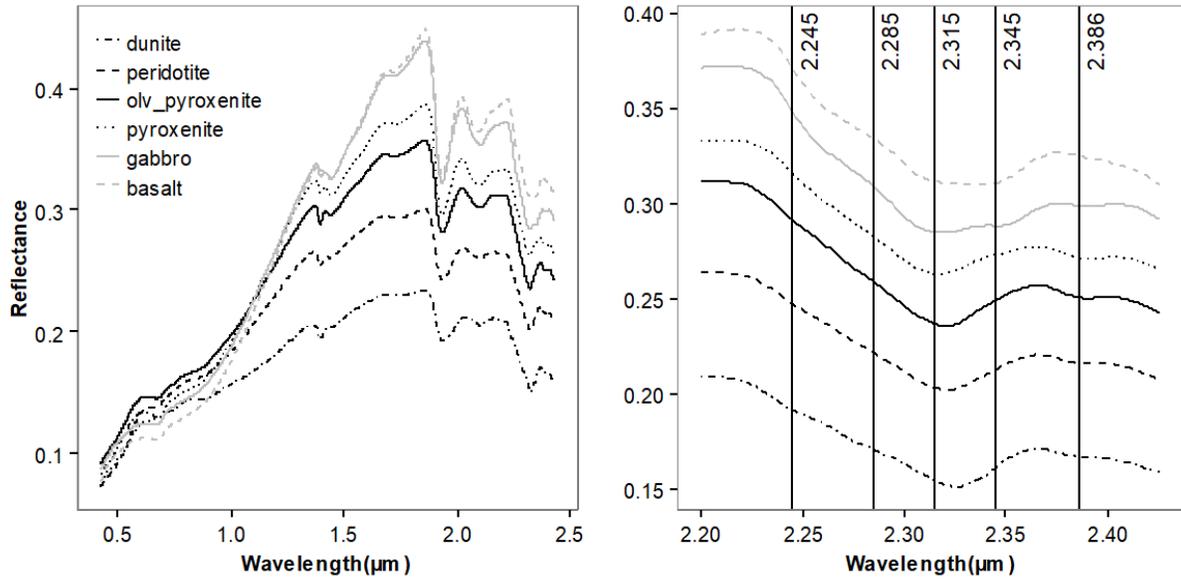
Figure 2.4 Averaged spectra for each rock type (RockAverages) (a) and enlargement of mineral features observed between 2.0-2.5 μm (b).
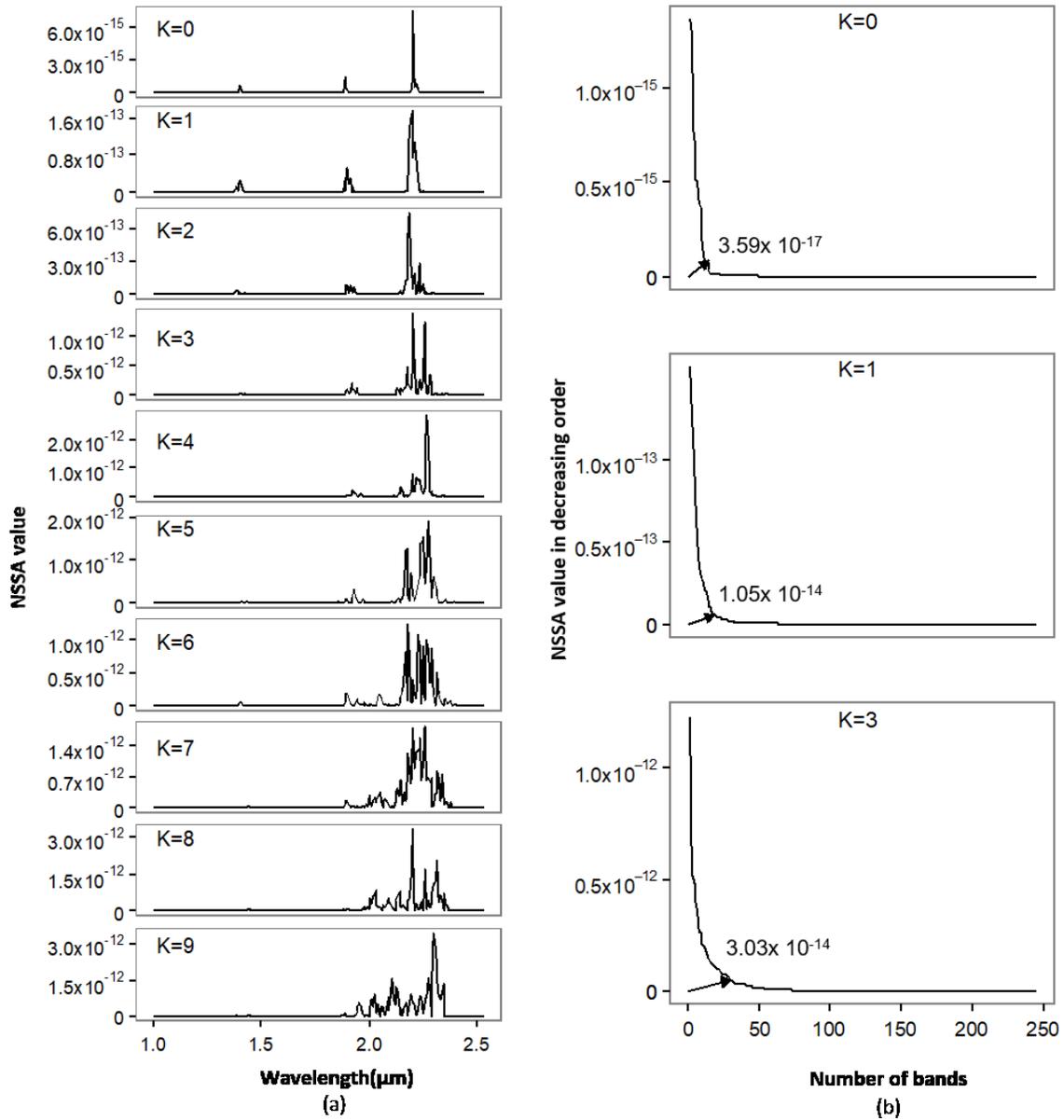
Figure 2.5 Parameter adjustment for clay minerals. (a) plot of the NSSA value as a function of wavelength for k=0-9, and (b) plot of ranked NSSA value as a function of the number of bands for k=0,1,3. The black arrow on each of these three plots defines the number of bands selected.
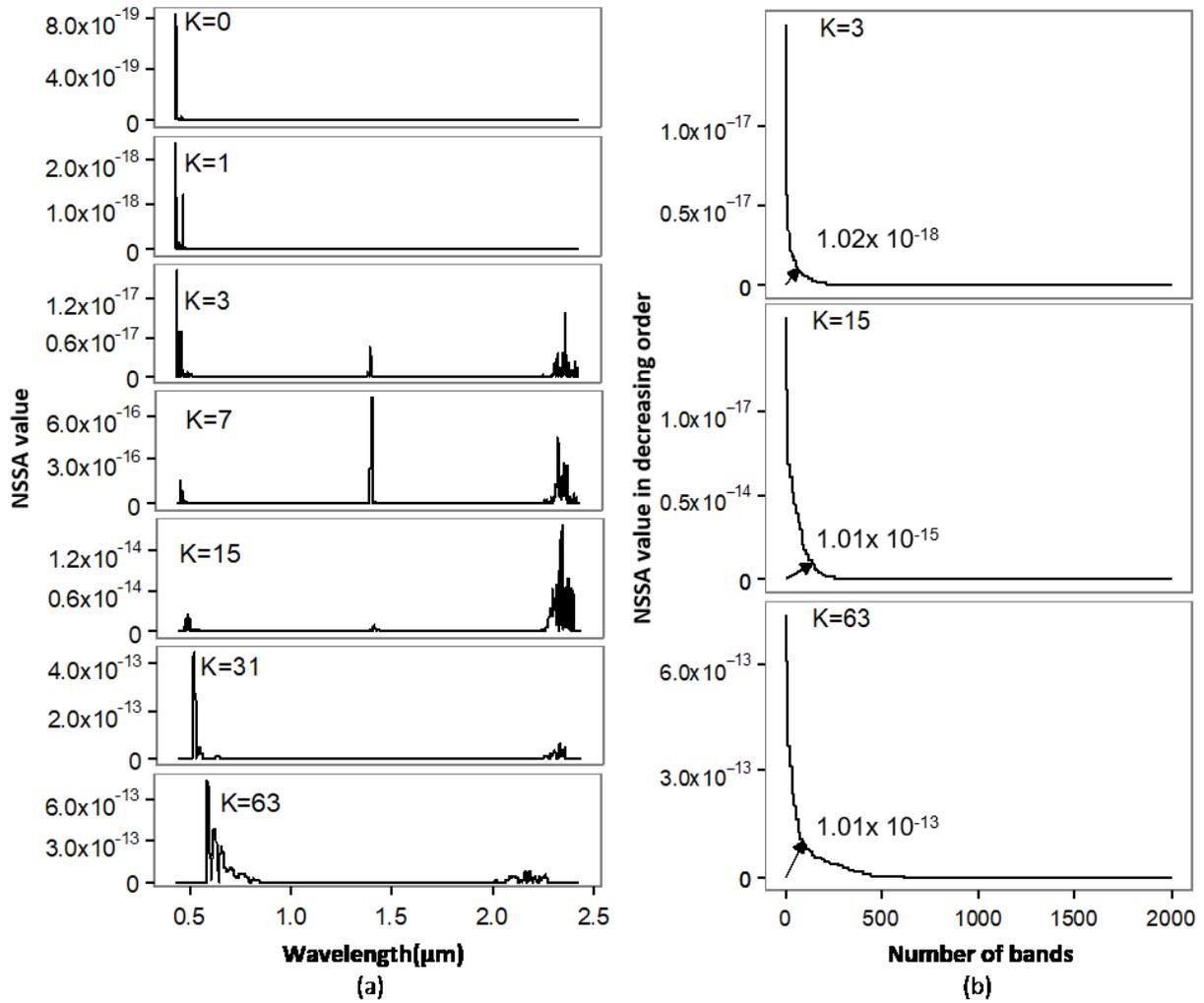
Figure 2.6 Parameter adjustment for rock types. (a) plot of the NSSA value as a function of wavelength for k=0,3,7,15,31,63 and (b) plot of ranked NSSA value as a function of the number of bands for k=3,15,63. The black arrow on each of these three plots defines the threshold value (i.e., the number of bands selected).
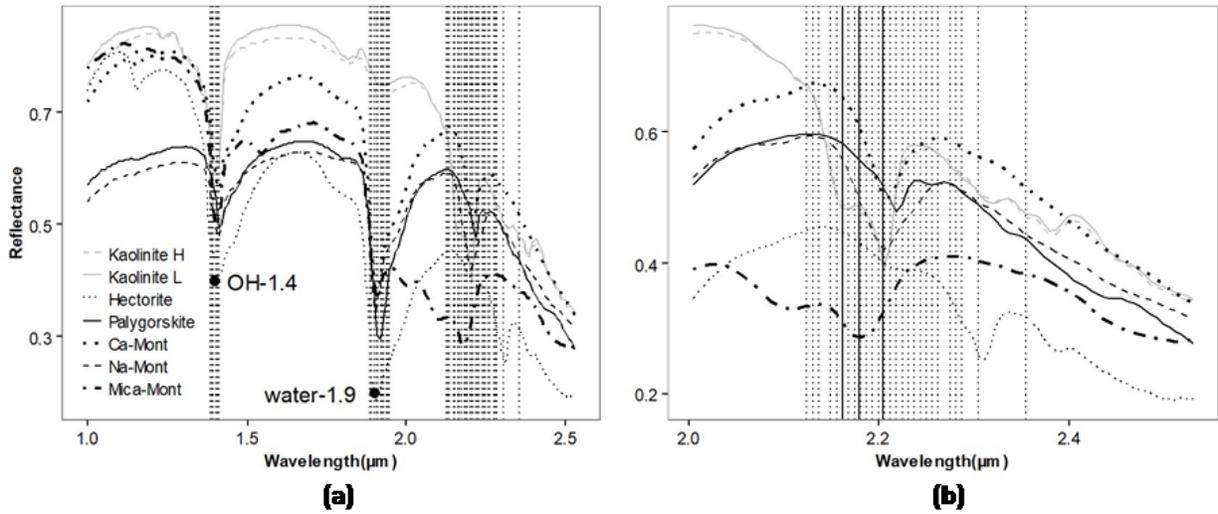
Figure 2.7 Average spectrum of each clay mineral (ClayAverages) and bands selected with the NSSA method shown as vertical dashed lines. (a) full wavelength (0.97-2.50 μm), and (b) the 2.00-2.50 μm region. Water and cation-hydroxyl features summarized in section 2.2.3.1 are labeled by dots in (a) and by vertical solid lines in (b).
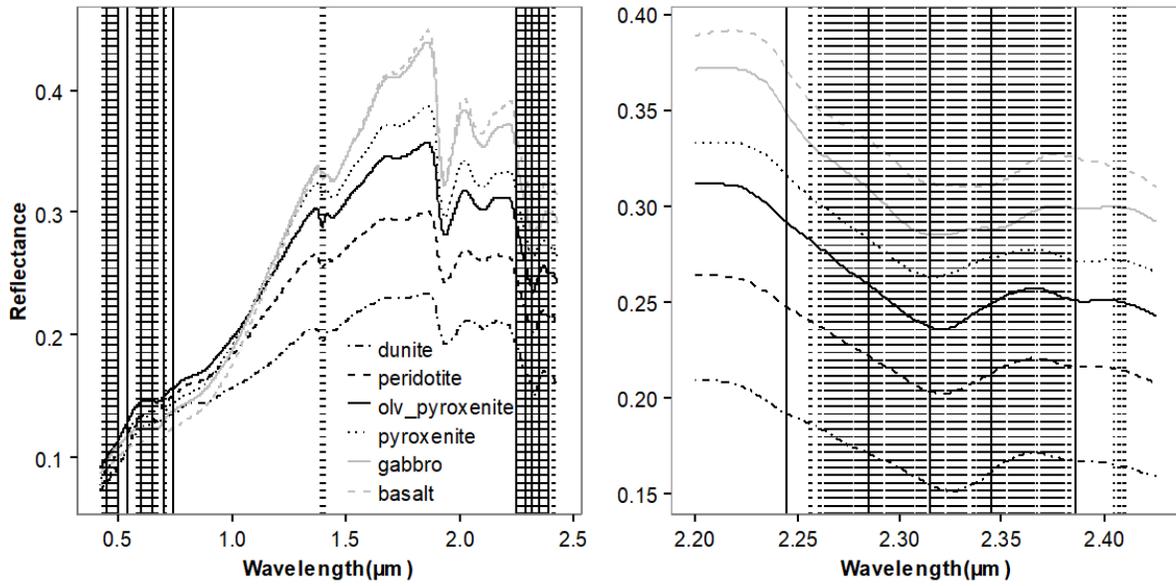
Figure 2.8 Average spectrum of each rock type (RockAverages) and bands selected with the NSSA method shown as vertical dashed lines. (a) full wavelength (0.35-2.50 μm), and (b) the 2.20-2.50 μm region.
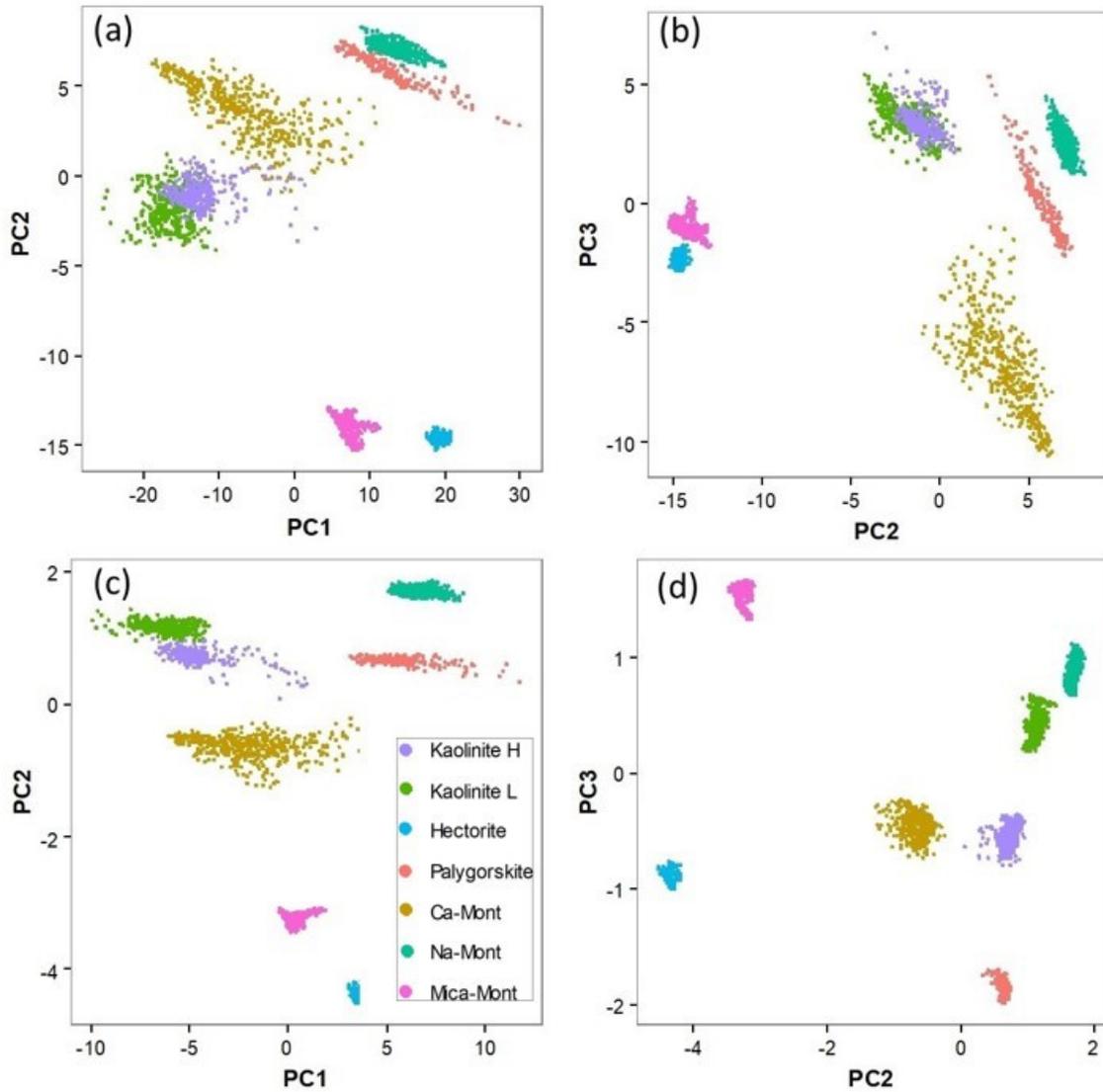
Figure 2.9 Class separation from scatterplots of PCA loadings obtained for clay minerals (245 bands and an average of 460 pixels for each of the seven mineral sample). (a) and (b) using all bands, (c) and (d) using the bands selected with the NSSA method. Data from the seven minerals are represented by different colors.
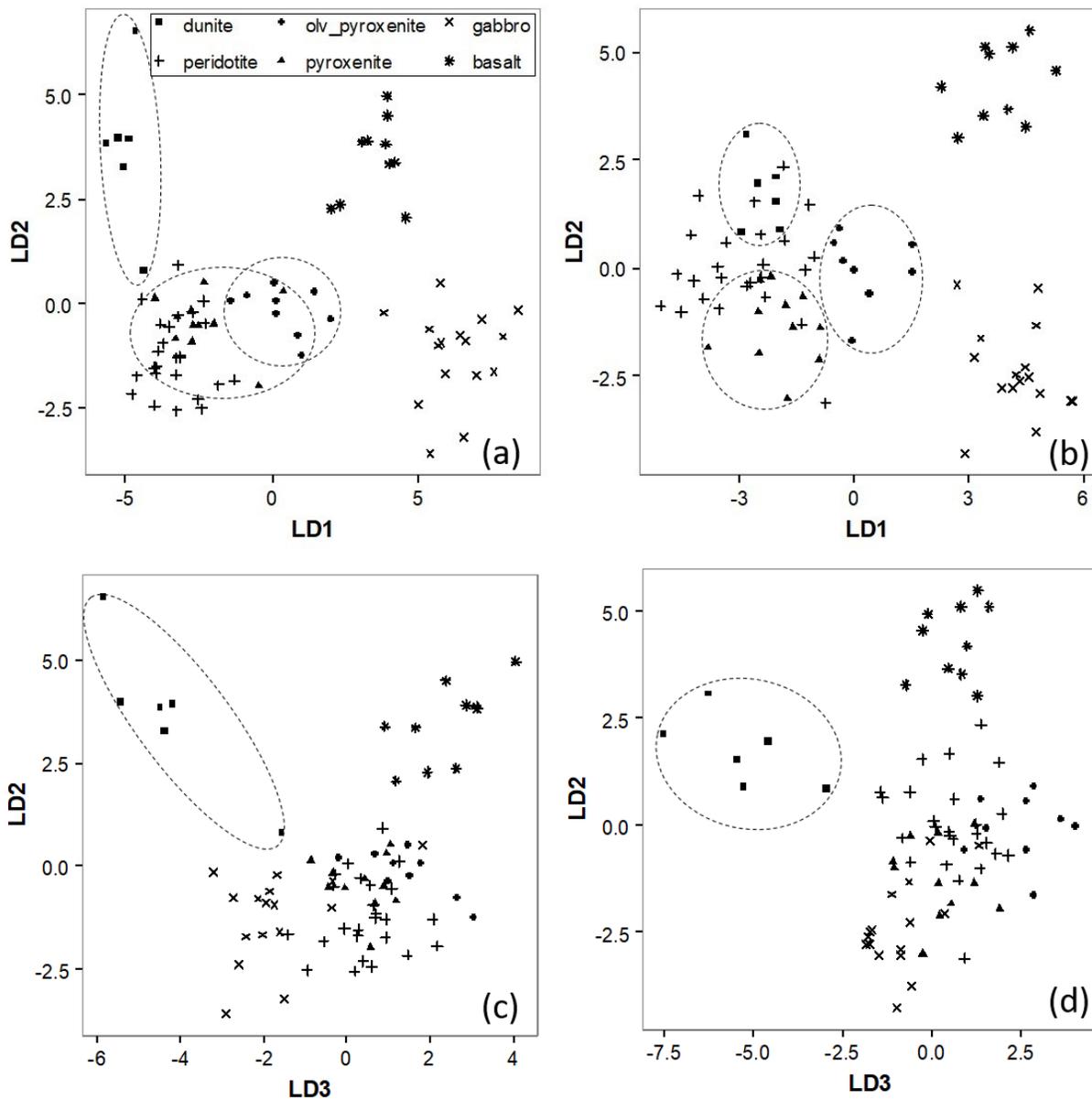
Figure 2.10 Class separation from scatterplots of LDA loadings obtained for rock types. (a) and (c) using all bands, and (b) and (d) using the bands selected by the NSSA method. The field occupied by dunite is highlighted in all plots but that of olivine-pyroxenite and pyroxenite are highlighted in the upper plots because of the enhanced separation of these rock types with band selection.
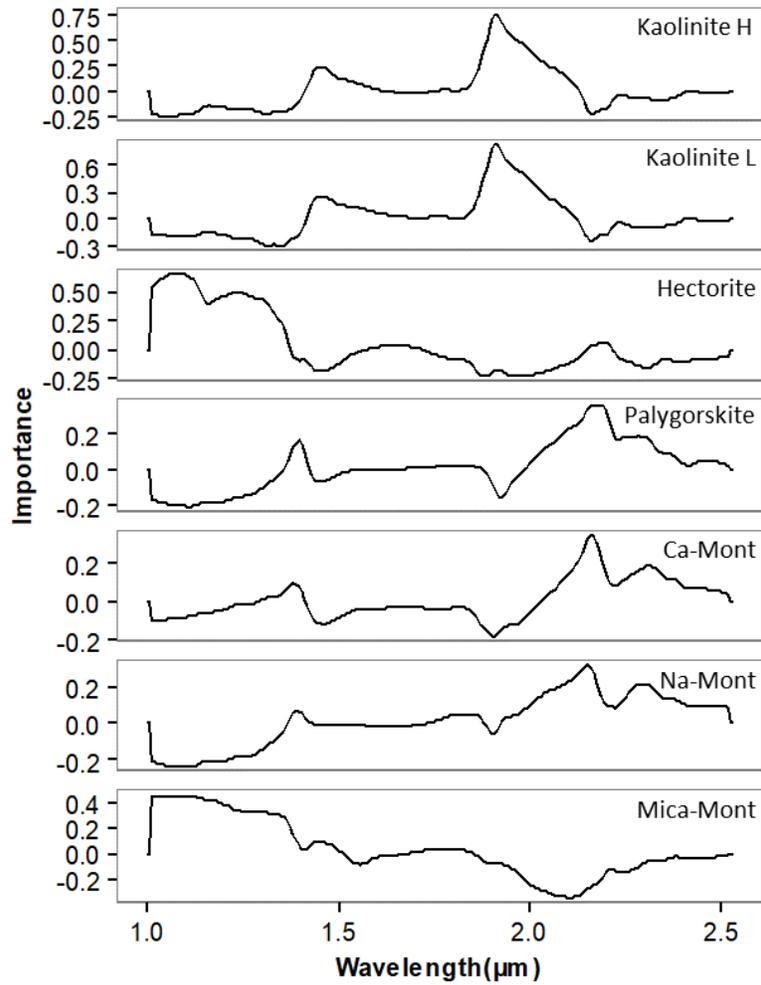
Figure 2.11 Profiles of band importance for each clay mineral obtained using the VNVBS method. Band importance corresponds to the Pr1 score output of the VNVBS method (Wang and Chang, 2007).
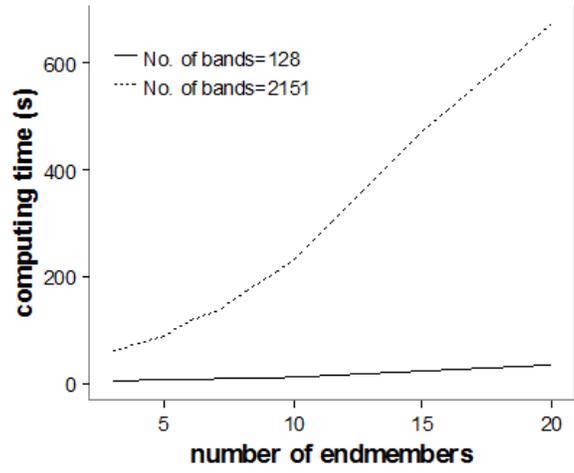
Figure 2.12 Computing time as a function of the number of bands and endmembers (e.g., data size).

References

Bajcsy, P., and Groves, P., 2004. Methodology for hyperspectral band selection. Photogrammetric Engineering & Remote Sensing 70(7): 793-802.

Bechtel, R., Rivard, B., and Sanchez, A., 2002. Spectral properties of foliose and crustose lichens based on laboratory experiments. Remote Sensing of Environment 82: 389-396.

Bioucas-Dias, J. M., Plaza, A., Camps-Valls, G., Scheunders, P., Nasrabadi, N., and Chanussot, J., 2013. Hyperspectral remote sensing data analysis and future challenges. IEEE Geoscience and Remote Sensing Magazine 1(2): 6-36.

Bishop, J., Lane, M., Dyar, M., and Brown, A., 2008. Reflectance and emission spectroscopy study of four groups of phyllosilicates: smectites, kaolinite-serpentines, chlorites and micas. Clay Minerals 43: 35-54.

Chang, C-I., 2013. Hyperspectral data processing: algorithm design and analysis. Hoboken, NJ: John Wiley & Sons.

Clarke, R.N., 1999. Spectroscopy of rocks and minerals, and principles of spectroscopy, Remote Sensing for the Earth sciences. Manual of Remote Sensing 3: 3-58.

Clark, R., King, T., Klejwa, M., and Swayze, G., 1990. High spectral resolution reflectance spectroscopy of minerals. Journal of Geophysical Research 95: 12653-12680.

Costanzo, P. A., and Guggenheim, S., 2001. Baseline studies of the Clay Minerals Society source clays: preface. Clays and Clay Minerals 49(5): 371-371.

Costanzo, P. M., 2001. Baseline studies of the clay minerals society source clays: Introduction. Clays and Clay Minerals 49(5): 372-373.

Goetz, A. F., 2009. Three decades of hyperspectral remote sensing of the Earth: A personal view. Remote Sensing of Environment 113: S5-S16.

Harsanyi, J. C., and Chang C-I., 1994. Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach. IEEE Transactions on Geoscience and Remote Sensing 32(4): 779-785.

Hastie, T., Tibshirani, R. and Friedman, J., 2008. The Elements of Statistical Learning; Data Mining, Inference and Prediction (second edition). Springer, Berlin: Springer series in statistics.

Hunt, G., 1977. Spectral signatures of particulate minerals in the visible and near infrared. Geophysics 42(3): 501-513.

Khoder, J., Younes, R., Obeid, H. and Khalil, M., 2015, June. Dimension reduction of hyperspectral image with rare event preserving. In Iberian Conference on Pattern Recognition and Image Analysis (pp. 621-629). Springer, Cham.

Keshava, N., 2004. Distance metrics and band selection in hyperspectral processing with applications to material identification and spectral libraries. IEEE Transactions on Geoscience and Remote Sensing 42(7): 1552-1565.

Lesher, C. M., 2007. Ni-Cu-(PGE) deposits in the Raglan area, Cape Smith Belt, New Quebec. Mineral Deposits of Canada: A Synthesis of Major Deposits-Types, District Metallogeny, the Evolution of Geological Provinces, and Exploration Methods. Geological Association of Canada, Mineral Deposits Division Special Publication 5: 351-386.

Peng, H., Long, F., and Ding, C., 2005. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(8): 1226-1237.

Rock, G., Gerhards, M., Schlerf, M., Hecker, C., and Udelhoven, T. 2016. Plant species discrimination using emissive thermal infrared imaging spectroscopy. International Journal of Applied Earth Observation and Geoinformation 53: 16-26.

Rogge, D. M., Rivard, B., Zhang, J., Sanchez, A., Harris, J., and Feng, J., 2007. Integration of spatial-spectral information for the improved extraction of endmembers. Remote Sensing of Environment 110: 287-303.

Rogge, D., Rivard, B., Segl, K., Grant, B., and Feng, J., 2014. Mapping of NiCu–PGE ore hosting ultramafic rocks using airborne and simulated EnMAP hyperspectral imagery, Nunavik, Canada. Remote Sensing of Environment 152: 302-317.

Somers, B., Asner, G. P., Tits, L., and Coppin, P., 2011. Endmember variability in spectral mixture analysis: A review. Remote Sensing of Environment 115(7): 1603-1616.

Speta, M., Gingras, M.K., and Rivard, B., 2016. Shortwave Infrared Hyperspectral Imaging: A Novel Method For Enhancing the Visibility of Sedimentary And Biogenic Features In Oil-Saturated Core. Journal of Sedimentary Research 86 (7): 830-842.

Sun, K., Geng, X., Ji L., and Lu, Y., 2014. A new band selection method for hyperspectral image based on data quality. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7(6): 2697-2703.

Tian, M., Feng, J., Rivard, B., and Zhao, C., 2016. A method to compute the n-dimensional solid spectral angle between vectors and its use for band selection in hyperspectral data, International Journal of Applied Earth Observation and Geoinformation 50: 141-149.

Van der Meer, F. D., and Jia, X., 2012. Collinearity and orthogonality of endmembers in linear spectral unmixing. International Journal of Applied Earth Observation and Geoinformation 18: 491-503.

Wang, S., and Chang, C.I., 2007. Variable-number variable-band selection for feature characterization in hyperspectral signatures. IEEE Transactions on Geoscience and Remote Sensing 45(9): 2979-2992.

Chapter 3 Hierarchical Band Selection Using the N-Dimensional Solid Spectral Angle Method to Address Inter- and Intra- Class Spectral Variability

3.1 Introduction

The ever increasing spectral resolution of hyperspectral sensing data enables a detailed discrimination or mapping of ground targets such as minerals and vegetation (Goetz et al., 1985). However, the high dimensionality of the data and the strong correlation between contiguous spectral bands creates obstacles for accurate mapping of target materials. A careful selection of wavelength bands, i.e. band selection, can alleviate the problem by focusing on significant features that maximize spectral differences between targets (Van der Meer and Jia, 2012). This process not only reduces the high spectral dimensionality but also does not transform the original data, which is required in specific applications.

However, the intrinsic spectral variability within a specific class of material (intra-class variability) and the spectral similarity among different classes (inter-class variability) can introduce challenges that are not typically taken into account in the band selection process and material discrimination (Kruse and Fairbarn, 2013). Most mainstream band selection methods apply if the materials are quite spectrally distinct. However, some applications in geology (Rogge et al, 2014), vegetation (Zhang et al., 2006) or urban environments (Wetherley et al., 2017) are often confronted with the challenge of discriminating targets with very similar spectral characteristics. Two such examples are covered in this study. First in a mining exploration context set in lithologies hosting Ni-Cu-PGE mineral deposits (Rogge et al., 2014), mafic and ultramafic rocks endmembers present almost identical spectral shapes while other classes in the scene, namely

vegetation, snow and water are more distinctive. As a result, the mapping process required the careful selection of a subclass of rock endmembers analyzed separately in order for them to be differentiated while the broader classes could be distinguished readily. A second example is provided for the spectral discrimination of tree species using the longwave infrared region (8-12um) where leaf spectra at the species-level can present great similarity. Thus species-level discrimination can be obscured by the relative high contrast of species group-level spectra (Harrison et al., 2018).

In such cases, the reduction of data dimensionality must proceed using an effective band selection and consider intra- and inter- class variability. Here we report on a hierarchical band selection resting on the N-dimensional Solid Spectral Angle (NSSA) method (Tian et al., 2016). The NSSA method is a band selector based on a spectral similarity measurement that extends the spectral angle into n-dimensions. Unlike conventional spectral angle measurements that are calculated between two spectra, NSSA measures the similarity among multiple spectra and selects wavelength band regions with high contrast (maximum difference) for improved separation (Tian et al., 2016). NSSA's capability in capturing the most distinctive feature from similar targets makes it advantageous in coping with the situation described above. This method is also superior to many other supervised methods because it can be applied to a spectral library collected from a limited number of measurements (i.e. from rocks, leaves). Details on the algorithm and its use are available in (Tian et al., 2016).

To address the challenge of band selection presented by the spectral variability issues introduced above (i.e. large contrast for some classes and low contrast for others), band selection with NSSA should be conducted separately for categorized endmembers of higher hierarchy (inter-class) and lower hierarchy (intra-class). Thus this paper focuses on the hierarchical strategy of

band selection (here termed HBS-NSSA) in two case studies of geology and vegetation. The first study involves mafic and ultramafic rocks where little spectral variation is observed among these and other geologic endmembers. The second study deals with leaf spectra of tree species in a Tropical Dry Forest (TDF), Spectra were collected in the longwave spectral range (8-12 um) and display low contrast and were sorted into eight spectral groups based on plant physiology. Three main objectives are pursued: 1) to identify key spectral features for geological and ecological applications using HBS-NSSA; 2) to assess the effectiveness of the method; and 3) to investigate the applicability of the method and its potential use in mapping techniques (Somers et al., 2011) for future needs.

3.2 Hyperspectral Data

3.2.1 Mafic and ultramafic rocks

The first data is a set of 18 endmembers previously generated from an airborne hyperspectral image of the Cape Smith Belt in northern Canada (Rogge et al., 2014). The airborne image was acquired using the Specim AisaDUAL Imaging Spectroscopy System (AISA) spanning a spectral region between 0.4-2.5 µm and containing 128 bands. Endmembers were generated from the image using an automatic endmember extraction tool and have been validated through comparison with field spectra as well as field mapping (Rogge et al., 2014). These include endmembers for 9 geological units, 6 vegetation covers, 2 types of snow packs and 1 for water. The geo-endmembers are dominated by mafic and ultramafic rock, the focus of mining exploration efforts (Rogge et al., 2014).

3.2.2 TDF tree species

The second data is a set of average leaf spectra for 26 tree species that were collected from a typical TDF in the Santa Rosa National Park-Environmental Monitoring Supersite, Guanacaste (SRNP-EMSG), Costa Rica. The leaf samples were collected in-situ during the dry season in 2013 and 2014 (Harrison et al., 2018) using an Agilent 4100 ExoScan FTIR (Fourier Transform Thermal Infrared) spectrometer with a range of 2.5 to 16 µm and a spectral resolution of $4cm^{-1}$. In a prior study, the spectra of the 26 species were grouped into 8 spectral classes based on spectral features attributed to major leaf compounds that include cutin, cellulose, and xylan (Harrison et al., 2018). The features in the longwave infrared region (8-11 µm) that could potentially discriminate tree species were focused.

3.3 Method

This section describes the principle of the NSSA method for band selection and then the hierarchical strategy using the NSSA method, followed by the description of its utilization and evaluation in the rock and tree species datasets.

3.3.1 NSSA method

NSSA refers to the N-dimensional solid angle in the spectral domain. Similar to the spectral angle (SA) describing similarity between two spectra, NSSA measures the similarity among multiple (more than three) spectra (Tian et al., 2016). According to the geometric definition of NSSA, the angle is only measured within a square matrix where the number of spectra should be equal to the number of bands. As a result, a sliding window with varying band intervals (denoted by "K") is moved across the available spectral range (Tian et al., 2016), generating a continuous

profile of NSSA values as a function of band wavelength position that is used as an indicator of band importance. Bands of lower NSSA values are less important because they reflect great similarity among the spectra for the given spectral regions. A fixed number of bands can be selected once a "Threshold" of NSSA values is determined.

3.3.2 Hierarchical band selection with NSSA

A direct application of NSSA without taking account the spectral variability within and between classes would negatively impact band selection if different hierarchies exist amongst the targets of interest. The geological case contains four main classes that have high contrast and the related band selection likely would obscure the subtle difference among materials within the geological class that includes gossan-encrusted rock, mafic and ultramafic rocks. Accordingly, it is desirable to separate the materials in different hierarchies and apply NSSA individually for each level of the hierarchies.

In the examples of this study, the hierarchical selection starts with the categorization of all spectra into two levels of classes, each class being represented by an averaged spectrum. The upper level examines classes with high spectral contrast. The lower level examines classes with low spectral contrast. The NSSA tool was run on each level for band selection and the last step integrates all bands selected from these two steps.

3.3.3 Utilization and evaluation of the method

Subsequently, the approach of HBS-NSSA was applied to the given datasets in this paper. Both datasets have been well categorized based on a priori knowledge described in prior geologic mapping (Rogge et al., 2014) and tree species discrimination (Harrison et al., 2018) studies. In the

geology study, the upper hierarchical level consisted of rock, vegetation, water and snow and was represented by 7 selected endmembers (out of 18 available) and the lower hierarchical level consisted of geo-class represented by 8 endmember spectra. For the tree species, the 8 spectral groups of tree species were regarded as the upper hierarchical level; and the first spectral group, which contains the largest number of tree species at 10, was selected as the low hierarchical level. The HBS-NSSA was conducted on the categorized spectral data based on the procedure described in the above section. To assess the accuracy of the band selection, both case studies used a comparison between obtained bands and spectral features identified by expert knowledge. The fidelity of the expert selected features is supported by extensive literature and for the geological study, by field validation of mapping efforts.

3.4 Results

Below is a brief summary of the results obtained by applying the HBS-NSSA method to the two hyperspectral datasets. Selected bands are shown in Figure 3.1 and 3.2 (vertical lines) and wavelength regions encompassing these bands are listed in Table 3.1.

3.4.1 Band selection on endmembers for geologic mapping

Figure 3.1 displays the bands selected (shown as dotted vertical lines) when HSB-NSSA is applied at an upper level to 7 endmembers representing 4 broad classes of surface materials (1a) and at a lower level to 8 geologic endmembers (1b). As seen in Figure 3.1, the spectral features identified from the two hierarchical levels partially overlap. The selected bands in Figure 3.1a largely reflect the inter-class differences between vegetation, snow and water. Diagnostic bands for vegetation (0.45, 0.5, 0.6, 1.1 μm), water (0.5 μm) and snow (0.7, 1.4 μm) are identified by the

band selection. These would be readily captured by many band selection methods. The intra-class variability amongst geologic materials is largely captured by bands concentrated in the shortwave infrared region (2.2-2.5 μm), consistent with known hydroxyl features that discriminate mafic and ultramafic rocks.

3.4.2 Band selection on longwave spectra for tree species discrimination

Figure 3.2 displays the bands selected at an upper level for the 8 groups encompassing all tree species (2a) and at a lower level for the spectral group (group 1) containing the most species (2b). For the 8 groups, major features attributable to cellulose, silica, cutin and oleanolic acid have been identified in a prior study from a comparison to spectra of driving compounds (Harrison et al., 2018). Our bands selected fall into 3 spectral regions (Figure 3.2a) consistent with cellulose feature around 9 μm, silica features at 9.40-9.70 μm, and ocleanolic acid features around 8.6, 9.00, 9.70 μm.

For the species-level analysis based on the spectral group spectra (Figure 3.2b), the selected features fall primarily into 4 spectral regions that are not obvious from a visual inspection. These are consistent with known features at 8.62, 9.01, and 9.44 μm related to asymmetric C-O-C bridge stretching, anhydroglucose ring asymmetric stretching, and C-O stretching of cellulose. In general, the HBS-NSSA method highlights the importance of bands around 9 μm to capture the intra-class variability in this group.

3.5 Conclusion

This paper presents a hierarchical strategy for band selection based on the NSSA method. The bands selected aim to maximize the spectral difference among multiple targets and the method

addressed inter- and intra- class variability in rock and leaf spectral classes for band selection. The results illustrate that it is feasible to select meaningful bands for the given datasets and that the hierarchical process results in different band sets for different levels. These can be combined to ensure that the discrimination of spectrally similar targets can also be addressed. A particular benefit of this band selection method compared to its peers is that it can be applied to a small number of labeled samples that is of practical value for field-based studies. The proposed band selection captured the variability between high-contrast materials in an upper hierarchy level as well as the variability within low-contrast materials in the low hierarchy level. The combined band sets for each of the two examples captured the same spectral regions identified by expert users in their efforts to capture the most significant bands to drive mapping efforts as reported in Rogge et al. (2014) and Harrison et al. (2018). The proposed method does seem promising for the automation of this band selection process resulting in substantial time savings in data analysis.

Table 3.1 Bands used to evaluate the NSSA band selection for the rock datasets.

Figure	Wavelength regions (µm) of selected bands
1a	0.43-0.76, 1.08-1.13, 1.31-1.37, 1.77-1.79, 2.01-2.02, 2.3-2.43
1b	0.47-0.62, 0.85-1.09, 1.29, 1.80-2.05, 2.12-2.38
2a	8.55-8.87, 8.93-9.17, 9.53-10.13
2b	8.48-8.63, 8.84-9.11, 9.19-9.42, 9.64-9.69
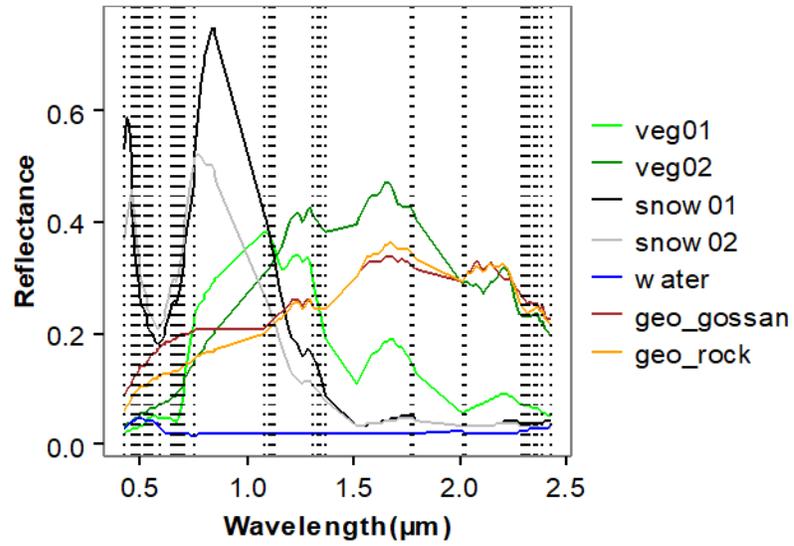
(a)
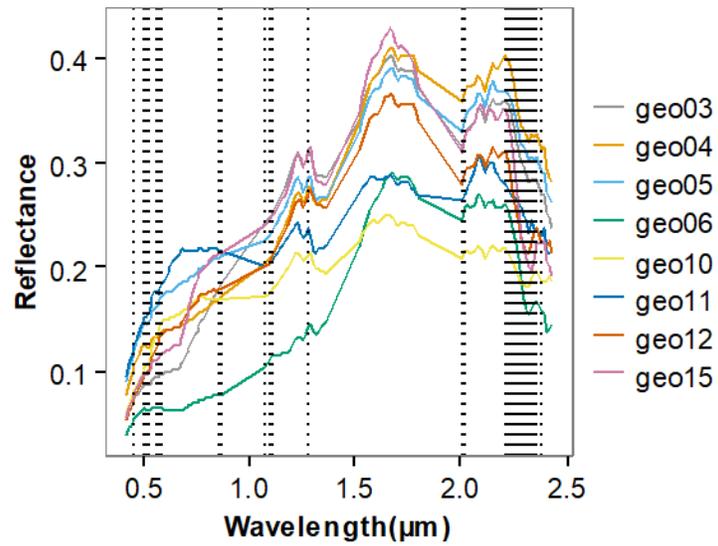
(b)

Figure 3.1 Bands (dotted vertical lines) selected by HBS-NSSA for (a) broad classes of all mapping materials; (b) geological materials.
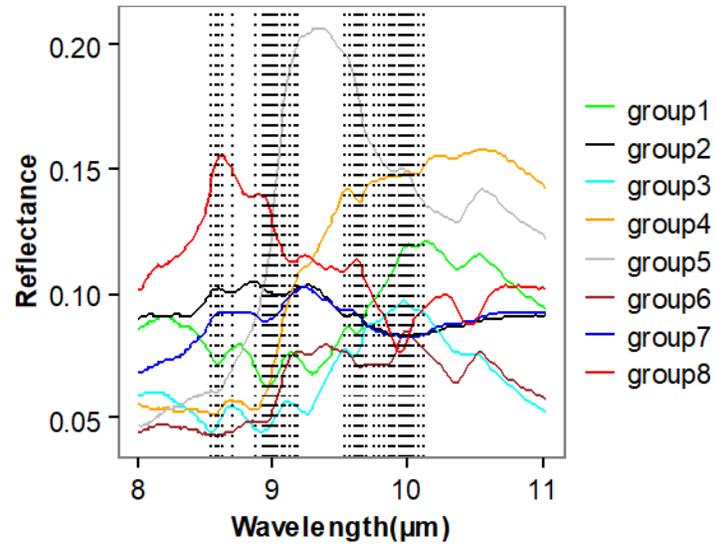
(a)
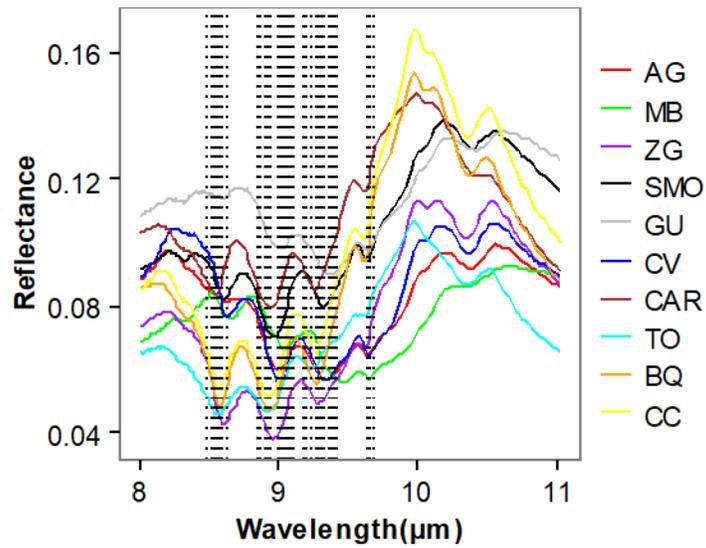
(b)

Figure 3.2 Bands (dotted vertical lines) selected by HBS-NSSA for (a) 8 groups of leaf spectra;

(b) 10 species in group

References

Goetz, A.F., Vane, G., Solomon, J.E., and Rock, B.N., 1985. Imaging spectrometry for earth remote sensing. Science 228(4704): 1147-1153.

Harrison, D., Rivard, B., and Sanchez-Azofeifa, A., 2018. Classification of tree species based on longwave hyperspectral data from leaves, a case study for a tropical dry forest. International Journal of Applied Earth Observation and Geoinformation 66: 93-105.

Kruse, F.A. and Fairbarn, K.G., 2013, May. Spectral variability constraints on multispectral and hyperspectral mapping performance. In Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIX (Vol. 8743, p. 87431O). International Society for Optics and Photonics.

Rogge, D., Rivard, B., Segl, K., Grant, B., and Feng, J., 2014. Mapping of NiCu–PGE ore hosting ultramafic rocks using airborne and simulated EnMAP hyperspectral imagery, Nunavik, Canada. Remote Sensing of Environment 152: 302-317.

Somers, B., Asner, G.P., Tits, L., and Coppin, P., 2011. Endmember variability in spectral mixture analysis: A review. Remote Sensing of Environment 115(7): 1603-1616.

Tian, M., Feng, J., Rivard, B., and Zhao, C., 2016. A method to compute the n-dimensional solid spectral angle between vectors and its use for band selection in hyperspectral data. International Journal of Applied Earth Observation and Geoinformation 50: 141-149.

Van der Meer, F.D., and Jia, X., 2012. Collinearity and orthogonality of endmembers in linear spectral unmixing. International Journal of Applied Earth Observation and Geoinformation 18: 491-503.

Wetherley, E.B., Roberts, D.A., and McFadden, J.P., 2017. Mapping spectrally similar urban materials at sub-pixel scales. Remote Sensing of Environment 195: 170-183.

Zhang, J., Rivard, B., Sánchez-Azofeifa, A., and Castro-Esau, K., 2006. Intra-and inter-class spectral variability of tropical tree species at La Selva, Costa Rica: Implications for species identification using HYDICE imagery. Remote Sensing of Environment 105(2): 129-141.

Chapter 4 Identification of Spectral Features in the Longwave Infrared (LWIR) Spectra of Leaves for the Discrimination of Tropical Dry Forest Tree Species

4.1 Introduction

Hyperspectral remote sensing has been widely used for classifying plant species (Adam & Mutanga, 2009; Belluco et al., 2006; Pu, 2009; Ustin & Xiao, 2001). Compared to conventional floristic research, which require specialized knowledge and extensive field work, the hyperspectral technique shows advantages in reducing time, cost and labor (Ribeiro da Luz, 2006; Ullah 2012). Hyperspectral sensors acquire reflectance measurements across numerous and contiguous bands of the electromagnetic spectrum, providing a fine spectral resolution for the detection of diagnostic spectral features across different tree species.

In the hyperspectral discrimination of tree species, one essential step is to clarify the relations between spectral information and leaf traits (Clark et al., 2005; Chan & Paelinckx, D, 2008). Spectral features of leaves in the Visible to Shortwave Infrared (VIS-SWIR; 0.4-2.5 μm) have been well identified and understood (Asner,1998; Govender et al., 2007; Ustin and Gamon, 2010). For example, the green peak at 0.45-0.55 μm is due to light interactions with the chlorophyll in the leaf cell, and the Near-Infrared (NIR) plateau (strong reflectance) between 0.8 and 1 μm is caused by volume scattering in the leaf mesophyll. These features and water absorptions in the SWIR are frequently exploited to discern species (Clark et al., 2005; Fassnacht et al., 2014).

To date, however, spectral features of leaves in the Thermal Infrared (TIR; 3-14 μm) wavelength region have been seldom studied for the following reasons. First, the TIR domain has been perceived as featureless (Fassnacht et al., 2016). The reflectance of leaves in this region is

comparatively low; moreover, the spectral characteristics are subtle and harder to detect. The lack of attention may reflect a lesser understanding of plant physiology and organic chemistry by the remote sensing community. The key reason is the relative scarcity of hyperspectral TIR data for vegetation. Field, space-borne, and airborne thermal spectrometers are not prevalent, so the data is not easily accessible. Nevertheless, a few studies from the 1980's, have shown the applicability and potential of TIR technology for species discrimination. Salisbury (1986) was the first to suggest that TIR features are possibly dominated by the chemical compounds of the leaf surface such as wax in the upper epidermis (Salisbury 1986; Salisbury & Milton, 1988). More recent research has demonstrated that leaves show distinctive spectral features in the thermal infrared region, and it is possible to discriminate vegetation types at the species level (Acevedo et al., 2017; da Luz & Crowly, 2007; Gilles et al., 2016; Buitrago et al., 2018; Meerdink et al., 2016; Ullah et al., 2012). A recent study used longwave infrared spectral data of leaves for the classification of liana and tree species and achieved a higher accuracy than the classification based on visible-near infrared reflectance data (Guzman et al., 2018). Therefore, the need to identify diagnostic vegetation spectral features in the TIR region becomes clear.

The detailed detection of spectral features in leaf spectra can be achieved manually, or automatically using statistical approaches. For example Luz et al. (2006, 2007) identified characteristic absorption features in leaf spectra collected from a diversity of tree species and agricultural plant species. In their studies, leaves displayed spectral variations in absorption band position and intensity or depth, and similar features were displayed in spectra of leaf chemical compounds (e.g., cellulose, cutin, xylan, silica and oleanolic acid). Thus, they reported that chemical bonds including hydroxyl (OH), carbonyl (C=0), methyl (CH_3) and methylene (CH_2) involved in those compounds were the driving factor of leaf spectral features. This fundamental

study of leaf and compound spectroscopy established foundations for the statistical band analysis, which was later explored (Clark et al., 2005; Zhang et al., 2006). Recently, a series of statistical methods, such as distance measure, Quadratic Discriminant Analysis (QDA), and Partial Least Square Regression (PLSR) models have been exploited to measure the importance of bands or select a subset of bands for the classification of broadleaf, herbaceous to wood species (Gilles et al., 2016; Meerdink et al., 2016; Ullah et al., 2012). These band analyses used a broad spectral range, which includes SWIR, Mid-wave Infrared (MWIR; 3-5 µm) and Long-wave Infrared (LWIR; 7-14 µm), in their classification models.

As features in these three spectral regions are driven by different factors, a recent study on the species classification of 19 broadleaf herbaceous and woody species (Buitrago et al. 2018) conducted band selection separately for each region and reported good classification in the SWIR and LWIR using as little as 5 bands. Prior studies have revealed that the MWIR is more sensitive to leaf water content and is suitable to detect plant stress (Acevedo et al., 2017), and LWIR features are mainly driven by chemical compounds in the cell wall or the cuticle from the leaf surface. Using all bands from the SWIR to the LWIR can make the spectral analysis biased, because the compound-driven features of relatively low contrast can be obscured by the overwhelming water features (Ullah et al., 2014). In addition, the evaluation of band selection should be focusing on the interpretation of selected features, rather than relying exclusively on the classification accuracy. As indicated in Salisbury (1986), it is important to examine the possible origins or contributing factors of the TIR features (i.e., spectroscopic interpretation).

Tree species in certain ecosystems such as in Tropical Dry Forests (TDF) can present great spectral similarity (low Spectral contrast), presenting a challenge for band selection and stimulating the need for more advanced methods (Harrison et al., 2018). Thus, this study employs

a variety of band selection methods to identify key features from the long-wave infrared spectra of tree leaves from a TDF, and compares the selected bands with known spectral features (vibrational bonds) associated with leaf constitutional compounds. Three methods named Variable Selection with Random Forest (VSRF), minimum Redundancy Maximum Relevance (mRMR) and N-dimensional Solid Spectral Angle (NSSA) and an ensemble of them are applied to the classification of LWIR spectral data collected from twenty-six tree species in a tropical dry forest ecosystem. The results of the band selection are evaluated based on agreement with known features and classification accuracy. Three questions are explored in this study: Within the LWIR, what are the key spectral regions or bands that best discern the selected tree species? Which method captures the most discriminating and important features for species discrimination? How is the species classification impacted by the selected features?

4.2 Study site and data

4.2.1 Study site and tree species

Leaves used in this study were collected from a typical Tropical Dry Forest (TDF) located in the Santa Rosa Reservation National Park-Environmental Monitoring Supersite, Guanacaste (SRNP-EMSG) in northwestern Costa Rica (Figure 4.1). TDF is a vegetation type typically dominated by deciduous trees with over 50% of trees being drought deciduous, where the mean annual temperature is $\geqslant 25$ °C and the total annual precipitation ranges between 700 and 2000 mm, and where there are three or more dry months when the precipitation is scarce (<100 mm/month) (Sanchez-Azofeifa, 2005). The study site spans an area of approximately 108 km2 (85 ° 34′ 40″ -85° 43′ 08″ E, 10° 44′ 05″ -10° 54′ 10″ N) and consists of relatively flat terrain (at around 600 m). The TDF at the SRNP-EMSG suffered from human disturbances until

the 1970s. Now the SRNP-EMSG presents a landscape composed of pasture and secondary forest, the later featuring diversity in tree species. A series of ecological and remote sensing investigations have been conducted in this area for the past twenty years (Arroyo-Mora et al., 2005; Castillo et al., 2011& 2012; Guzman et al., 2018; Harrison et al., 2018; Kalacska et al., 2004 & 2007).

Over 159 tree species are known to exist in this area (Kalacska et al., 2004), and this study focused on 26 species abundant in this TDF. The botanical names and families of these species are listed in Table 4.1. Trees of these 26 species have an average age of 50-60 years (Harrison et al., 2018) and encompass 19 families with most species belonging to a different family. However some species, including *Albizia adinocephala* (AA) and *Ateleia herbert-smithii* (AH); *Astronium graveolens* (AG) and *Spondias mombin* (SMO); *Casearia argute* (CAR), *Casearia sylvestris* (CS) and *Zuelania Guidonia* (ZG); *Lonchocarpus minimiflorus* (LM) and *Machaerium biovulatum* (MB); and *Luehea candida* (LC) and *Luehea speciose* (LS) belong to the same family potentially increasing the difficulty in discriminating them based on spectral observations.

4.2.2 Spectral data

4.2.2.1 Leaf spectra

The leaf samples and spectral data were collected as part of a study by Harrison et al. (2018). Leaf samples were acquired and measured in-situ during the wet season (May) in 2013 and 2014. For each species, leaf samples were collected from different canopy layers (sun, mid-canopy and shade leaves) of 2-3 individual trees. The spectra of the samples were measured using an Agilent 4100 ExoScan FTIR (Fourier Transform Thermal Infrared) spectrometer with a spectral range of 2.5 to 16 µm and a spectral resolution of 4 cm^{-1}. Attenuated Total Reflectance (ATR) was measured using a probe enabling photons to penetrate into leaf samples at a small depth (around

20-50 μm) so that the inner leaf structure could be probed. The ATR probe views a leaf footprint of approximately 1.5 cm. A spectral region from 8 to 11 μm was selected for analysis to exclude lower signal to noise beyond this region (Harrison et al., 2018). A total of 656 leaf spectra were obtained, which included approximately 20-30 spectra per species. Each leaf sample and related spectrum was labeled by its species name.

Harrison et al. (2018) assigned the leaf spectra of 21 species (Figure 4.2; Table 4.1) to five Spectral Types (ST), while species of the Tiliaceae family (LC and LS) were assigned a unique spectral type and three single species (AH, LM and RT) were unassigned. The definition of the spectral types was based mainly on the similarity of spectral features amongst species that were then attributed to major leaf compounds, as described in the next section. Harrison et al. (2018) manually identified a set of key features from the leaf spectra and documented their location as listed in the first column of Table 4.2. Each feature can be linked to a corresponding feature observed in compound spectra, but the exact position of features may not be a match. The positions of the features were used as a reference band set to assist in the evaluation of the band selection results of this study. The organization of the species in spectral types highlights commonality in features within a type and contrasts differences amongst spectral types. Here we chose to display the spectra of species in reflectance without any pre-processing. However, as shown by Harrison et al. (2018), similarities and differences in spectra of species are enhanced when viewed in the wavelet representation.

4.2.2.2 Compound spectra and links to species spectral types

An online spectral library of leaf constitutional compounds compiled by da Luz & Crowley (2007) was used in this study and that of Harrison et al. (2018). The library includes spectra of cellulose, xylan, cutin, and oleanolic acid that were collected from vegetative materials such as

tomatoes and wool pulp as well as spectra of silica (da Luz & Crowley, 2007; 2010). The Directional Hemisphere Reflectance (DHR) was measured from the compounds using a Nexus 670 spectrometer and an integrating sphere coated internally with InfragoldTM. The spectral range of the data is 2.5-14 μm and the spectral resolution is 4 cm^{-1}. This study used a range of 7-12 μm for analysis. Each compound contains multiple (4-20) spectra (Figure 4.3), and the averaged spectrum per type was obtained (da Luz & Crowley, 2007; 2010).

The location and possible origin of the features observed in spectra of the five compounds were described by Harrison et al. (2018) and are summarized in Table 4.2. Here we briefly highlight compound features and some of the driving compounds influencing the leaf spectral types (ST). Cellulose, mainly found in the external wall of epidermal plant cells, is the most abundant leaf constitutional compound. Cellulose shows features at a number of locations (8.18, 8.58, 8.62, 8.84, 9.01, 9.3, 9.44, 9.63 μm in Figure 4.3). Some of these features are attributed to C-O or C-O-C stretching (Table 4.2), while others remain unknown. At least three cellulose features are pronounced (8.62, 9.01, 9.44 μm) and diagnostic to the identification of tree species from ST1 and ST2 (Figure 4.2) as the strength and detailed position of the features is variable across species. The specific location and depth of cellulose features vary across species, possibly reflecting the amount and thickness of cellulose near the leaf surface (Luz et al., 2007). Xylan, also found in cell wall, is a type of hemicellulose. The spectra of xylan are characterized by a prominent increase in reflectance at 9.40-9.70 μm that is attributed to C-OH bending. This feature is present in leaf spectra of species from ST3 (e.g., MT, CAM, BS). CAM is a species reported to have near-surface silica (Bjorn et al. 2011; Siegel and Paguaga, 1991) and Harrison et al. (2018) has suggested that silica, with a broad feature centered at 9.08 μm, may be contributing to the spectra of this species. Cutin comes from the cuticle which is the waterproof outer layer covering

the epidermal cells of leaves. Spectra of cutin show two diagnostic features at 8.65 and 9.14 μm identified as C-O-C stretching. These features are observed in leaf spectra from ST5 (species CS, PG, AA and SG). Oleanolic acid, a type of terpene, is known to exist in cuticular waxes. The spectra display features at 8.67, 9.00, 9.27, 9.7, 10.36, and 10.75 μm. The band at 9.00 μm is assigned to carbonyl (C=O) stretching in acids. Features of this compound are seen in spectra of the RT species.

4.3 Method

This section describes three methodological aspects of this study. First, a band selection was conducted on the spectra of constitutional compounds. Then, bands were selected from leaf spectra for discriminating the tree species. Lastly, the selection results were evaluated.

4.3.1 Band analysis on constitutional compounds

The NSSA band selection method, described in Tian et al. (2016), was used to select bands that discriminate the spectra of the leaf compounds. The NSSA method is a band selector based on the geometric difference in spectral shape amongst a suite of spectra. It uses the spectral angle in n-dimensional space to reflect the similarity among multiple targets; the greater the contrast amongst spectra, the more important the band (Tian et al., 2016). NSSA is one of the few methods that can be applied to a limited number of spectra (e.g. average spectrum for each of the 5 compounds). To properly exploit this method, we adjusted two parameters inherent to the NSSA method. K defines the band interval from which the NSSA values are calculated to capture features of different widths. The threshold defines the minimum NSSA value above which bands are selected and thus defines the number of bands selected from a profile of NSSA values as a function

of wavelength. Based on the guidelines on parameter definition defined in Long et al. (2019), discrete values of k (0-9) were considered. The threshold for each k was estimated based on a graph of ranked NSSA values (Long et al., 2019). An individual band set was therefore derived for each k and these were combined in a single band repository. This final band set was compared with features that discriminate the compounds as identified in a prior study (Harrison et al., 2018) to assess the validity of the selected features.

4.3.2 Band analysis on spectra of leaves

To identify spectral features for the discrimination of the 26 species, bands were selected from the leaf spectra using three separate methods and also an ensemble of these methods. The method of Variable Selection with Random Forest (VSRF), minimum Redundancy Maximum Relevance (mRMR) and NSSA were chosen, as they are based on different principles and show distinct advantages (Belgiu & Dragut, 2016; Chan & Paelinckx, 2008; Ding & Peng, 2005; Tian et al., 2016). The VSRF method, which is based on classification accuracy, is applicable to the given dataset (625 spectra) in which a considerable number of labeled samples were provided for each species (20-30). This method also shows robust performance in processing the visible and near-infrared spectral data of vegetation (Chan & Paelinckx, 2008; Feilhauer et al., 2015). The mRMR method, which is based on information theory, has demonstrated excellent performances in a variety of data exploitation applications including in the fields of biology and computing science and thus was adopted in hyperspectral analysis (Ding & Peng, 2005; Peng et al., 2005; Jiang & Li, 2015). The NSSA method, which was used in the analysis of the compound spectra, has advantages in detecting subtle features from spectrally similar materials. Each method

generated a set of bands and a profile of band importance that were combined in a model to form the ensemble selection. Therefore, four sets of bands in total were produced in this analysis.

4.3.2.1 Band selection using a single method

(1) Variable Selection with Random Forest (VSRF)

The first band set was obtained by applying the VSRF method to the 625 spectral samples that were labeled by species names. The band selection process is imbedded in the Random Forest (RF) classification, which builds many binary decision trees using bootstrap training samples and randomly chooses a subset of variables at each node of the tree. The RF contains two key parameters that control the selection process. The parameters of "mtry" and "ntree" represent the number of randomly selected variables and the total number of trees for the RF model. We adjusted these parameters to achieve the highest classification accuracy. The accuracy is calculated by the built-in RF model that automatically splits the data into training and testing subsets. All processes were implemented using the package "VarSelRF" in the R software (Diaz-Uriarte, 2007)

The VSRF method generated two profiles of band importance. The importance is either measured by the increased misclassification error when one variable is randomly permuted in the out of bag (OOB) samples, or by the Gini index based on the average loss of entropy for growing classification trees (Genuer et al., 2010). This study chose the OOB error as the measuring criterion of band importance because previous studies showed that the Gini-importance was not very suitable for predictor variables with many categories while the data in this study encompassed 26 categories (species classes) (Belgiu & Drăguţ, 2016).

(2) minimum Redundancy Maximum Relevance (mRMR)

The second band set was obtained by applying the mRMR method to the same labeled spectral data as used in VSRF. The band selection by mRMR attempts to maximize the dependency

between variables and classes while minimizing the redundancy between contiguous variables (Peng et al., 2005). This method requires that the number of selected bands be defined before selection. For an equivalent comparison, we defined this number to equal the one derived from the VSRF method. All these processes were implemented using a "mRMRe" package in the R environment.

This method also generated band importance that was measured by building the correlation between predictors and target classes. The band importance can be calculated based on two criteria: a classic model (Ding & Peng, 2005) that is proposed by the developer of the method and an ensemble model that modifies the model by adding more evaluation criteria such as variable complementarity. The complementary between two random features X_i and X_j and the output Y is the gain obtained from using the joint mutual information (I) of the two variables. This criterion can be calculated by $I(X_{i,j}; Y)-(I(X_i;Y)+I(X_j;Y))$. The ensemble model was chosen because it is more effective for a large number of variables (184 bands) with a few samples (20-30 spectra per species) (Meyer et al., 2008).

(3) N-dimensional Spectral Solid Angle (NSSA)

The third band set was derived by applying the NSSA method to the data set comprising the average spectrum of each of the 26 tree species. Similar to the band analysis of the compounds described in section 4.3.1, we adjusted the k parameter to capture features of varying width and the threshold parameter to determine a fixed number of bands, the latter defined from graphical analysis.

4.3.2.2 Band selection using an ensemble of three methods

Based on each of the profile of band importance obtained from the three methods, a fourth band set was generated utilizing an ensemble approach as described by Feilhauer et al. (2015). The

ensemble selection of multiple methods has been found to achieve more reliable results than single methods because of reduced bias (Feilhauer et al., 2015; Neumann et al. 2017). The ensemble approach compiled measures of band importance from the three methods and aggregated the normalized values multiplied by weights. The weights were defined based on how much the variance is explained by each method (Feilhauer et al., 2015). Then, a threshold was applied to the compiled importance profile using a value of mean plus one standard deviation. The process of ensemble selection was implemented in the R software.

4.3.3 Classification with selected bands

The band selection results were evaluated by classifying the leaf spectra using the four band sets derived above. For classification, the method of Logistic Regression (LR) was selected because it does not include a variable selection within the model. The LR is a straightforward method and more importantly, it is applicable to the balanced input data (similar number of samples for each species) in this study. The method builds a linear function between input variables (or bands) and the outcome (species class for this study) based on the posterior probability, and it commonly uses a maximum likelihood to fit the model (Friedman et al., 2001). To reduce the strong correlation between adjacent variables in the given spectral data, we imposed regularized coefficients to the LR model.

To avoid over fitting, we conducted a leave-one-out Cross Validation (CV) in the classification process. Other indicators, in addition to classification accuracy, were calculated to evaluate the performance such as Kappa statistic, Mean Absolute Error (MAE), and Relative Absolute Error (RAE). The kappa statistic measures the agreement of prediction with the true class. The MAE measures the average magnitude of the errors in a set of predictions. It is the

average based on the testing samples of the absolute values of the differences between prediction and observation. RAE is the total absolute error relative to a simple predictor. The value is normalized by dividing by the total absolute error of the simple predictor (bin Othman & Yau, 2007; Viera & Garrett, 2005). To further evaluate the classification performance of the ensemble band set, a confusion matrix was generated that enables the analysis of species-level accuracy and relations between species.

4.4 Results

4.4.1 Selected bands

4.4.1.1 Bands selected from compounds' spectra

Figure 4.4 shows four NSSA profiles that depict band importance as a function of wavelength for varying feature widths (i.e., different k values). As revealed by the profiles, peaks, which represent bands of greater importance (i.e., high NSSA value), are mostly present in the region between 8.45 and 9.97 μm. This observation is consistent with the findings of Harrison et al. (2018) and features listed in Table 4.2. Based on the results shown on Figure 4.4, the profiles for the first three k values (k=0, 1, 3) were used, and profiles for k greater than 3 (k=7) were omitted, because they offer largely redundant information to that of the first three k values. For the three selected k values, a threshold value was defined (Figure 4.4, horizontal dotted line) to derive a number of bands; then, the three band sets obtained were combined in a single band repository listed in Table 4.3.

4.4.1.2 Bands selected for leaves

(1) Using single methods

Each band selection method (VSRF, mRMR and NSSA) generates a profile of band importance, as depicted in Figure 4.5. The three profiles show distinctive patterns. In the case of VSRF, the mean decrease of accuracy was chosen to measure band importance with bands with a higher value of mean decrease of accuracy deemed most important. The wavelength profile of this variable presents the most fluctuations of all three profiles as seen in Figure 4.5a and the most important region is located between 8-9 μm. In the case of mRMR, the score of causality was used to measure band importance (Figure 4.5b). This importance profile is unique in that a number of bands at 8.0-8.3 μm show high importance values, and a few singlet and doublet bands are highlighted and dispersed at longer wavelengths (e.g. peaks). In the case of NSSA, the profile for k=0 is selected to measure the band importance, since it captures features at various locations. As was seen in the profile generated by mRMR, several prominent peaks in NSSA value are observed at discrete wavelength regions with a predominance from 8.5-9.2 μm and 9.5-10.2 μm.

Each method generates a band set as represented by the dotted vertical lines in Figure 4.6. These can be examined next to the average leaf spectra of each of the 26 tree species. The band distributions seen in Figure 4.6 reveal disparities across the three band selection methods. The NSSA method selects two main regions (8.5-9.2, 9.5-10.2 μm). Only six of the 57 bands (10%) selected are identified at wavelengths shorter than 8.2 μm and no bands are selected beyond 10.2 μm. The VSRF method selects a few bands short of 8.5 μm and five discrete clusters of bands between 8.61-9.65 μm followed by a group of bands between 10.30-10.42 μm. The mRMR method selects most of its bands (34 out of 47) short of 8.51 μm with the remaining bands defining 8 narrow regions (8.62, 8.70, 8.94, 9.35, 9.80, 10.10, 10.50, 11.02μm). Although different regions are highlighted by the three methods, consistent regions can be observed. For example, bands

identified by RFVS partially overlap with those selected by NSSA at specific regions. All three methods also select bands at 8.6-9.2 µm.

(2) Using the ensemble method

A set of 43 bands (Table 4.4), generated by the ensemble selection, is shown in Figure 4.7. This set captures the spectral regions that were selected by the aforementioned three single methods. A large percentage of bands (79%, 34/43) are located between 8.5 µm and 9.67 µm. The remaining bands are distributed in narrow regions near 8.2 µm, 10 µm and 10.5 µm.

4.4.2 Classification accuracy

4.4.2.1 Overall Accuracy

A series of criteria including overall accuracy, kappa statistic and errors, that evaluate the classification of tree species based on bands selected by the different methods, are listed in Table 4.5. The classification was conducted using the regularized Logistic Regression (LR) method and the first observation based on the results listed in Table 4.5 is that a high overall classification accuracy (>0.86) is generally achieved regardless of the input band set. Other classification methods were examined, such as Random Forest and Multiclass classifier but the accuracy was lower. Upon closer examination amongst the three individual methods, VSRF performed more poorly with mRMR and NSSA showing similar accuracies. The bands by the ensemble selection achieve the highest accuracy (0.94), greatest kappa statistic (0.94) and smallest errors (e.g., RAE=0.06). The increase in accuracy by the ensemble model is 4-8 %, and is an improvement to an already high overall accuracy obtained from band sets from individual methods.

4.4.2.2 Species accuracy: confusion matrix

A confusion matrix shown in Table 4.6 enables the analysis of the species-level classification performance and the relation between species. The numerous off diagonal zero

values reveal that very few samples were misclassified, in accordance with the high overall classification accuracy. The maximum number of misclassified samples is four (LS); LC and CS have three misclassified samples while AH, PG, and SG have 2. Twelve species have only one sample misclassified (e.g., AA, BQ, CV); and eight species have none misclassified (CAR, SMO, ZG, ES, SME). The misclassification could be ascribed to the spectral similarity of species. Species from different Spectral Types (ST) tend to be less confused or misclassified, while species from the same spectral type are more easily confused, for example, CS and PG that belong to ST5, and LC and LS that belong to the *Luchea* sp. ST (these are also from the same family).

4.4.3 Comparison of selected bands with known features

The bands selected from leaf spectra using the ensemble approach were compared with bands selected by Harrison et al. (2018) (Figure 4.7), the later having been compared to features documented in the literature as described in section 4.2.2.1 and shown in Table 4.2. Overall the spectral regions selected by the ensemble approach (Figure 4.7a dotted line) show consistency with the bands identified by experts (solid line) though the band set of the ensemble approach encompasses more contiguous bands as expected, because the bands selected by "the expert" only report the band center of features observed from leaf spectra. Most bands selected lie between 8.5-9.6 µm, a region encompassing spectral features of cellulose. This observation also applies to bands selected for compounds, as reflected by dotted lines in Figure 4.7b. Cellulose is the most influencing driver in the spectral signatures for a third of the species (Harrison et al., 2018). For leaves few bands are selected near 10 µm and at 10.5 µm (Figure 4.7a), though two features attributed to oleanolic acid are located at 10.24 µm and 10.75 µm (Table 4.2, Figure 4.7a).

However, oleanolic acid only exerts a control on the spectra of a single species (RT), thus the band selection based on 26 species cannot be expected to capture these two features.

4.5 Discussion

4.5.1 LWIR features for the discrimination of TDF tree species

The band analysis in this study reveals that a spectral region near 9 μm, specifically spanning 8.5-9.6 μm, is important in every means of feature identification, manually by experts or automatically using statistical band selection tools, based on either compounds or leaf data. In the band selection results (Table 4.4 and Figure 4.7), features that best discern the TDF tree species are mainly located at or near discrete spectral regions seen for the compounds and centered at 8.2, 8.6, 9, and 9.6. These features are very consistent with the carbonyl bond in the cellulose compound, which is reasonable given the large number of species that have spectra with cellulose features. Ten species in ST1 and two species from ST2, twelve in total, are driven by a readily visible triplet feature (near 8.78 μm, 9.12 μm, 9.55 μm) of cellulose. The two prominent features at 8.6 and 9.2 can also be ascribed to the existence of cutin. However, reported features of oleanolic acid (Table 4.2) are not all identified during feature selection, likely because a single species (RT) is driven by this compound as reported by Harrison et al. (2018).

The spectral regions or bands selected in this study are consistent with findings in other studies for species found outside of tropical environments. Ullah et al. (2012) addressed the importance of the spectral ranges 9.27-9.48 and 9.74-10.00 μm whose central wavelengths are 9.36 and 9.87 μm. Rock et al. (2016) included several bands (8.19, 8.4, 8.49, 9.7, 10.19, 10.51 and 10.81 μm) in their classification models. These studies relied solely on statistical models for band selection and did not offer an interpretation on the origins of the features. However, in a recent

study, Buitrago et al. (2018) identified diagnostic LWIR features at 8.54 and 9.79 µm related to cellulose and lignin for woody and herbaceous plant species. The LWIR is proving to be a spectral region rich in discriminating features despite the low spectral contrast displayed by most vegetation species. In this study the maximum reflectance is approximately 15% for all but three species. Because of this low spectral contrast, band analysis should be conducted separately for the LWIR range versus the SWIR and MWIR in order to specifically select features associated with leaf chemical compounds. Focusing on the LWIR also minimizes the impact of leaf water content.

4.5.2 Choice of method for band selection

Classification accuracy of the 26 TDF species without band selection was 91%, nominally the same as that achieved by two of the three single band selection methods, VRSF performing worst with 87%. A noticeable improvement of 3% in classification accuracy was obtained using the ensemble band selection method as compared to the accuracy obtained without band selection. The 3% increase was achieved using a small proportion of the total available bands (23% or 43/187). In this strategy, a band is considered important by aggregating band importance values of multiple methods, the low values in a single method being omitted. A single strong method can cancel out a weak method so that the influence from different methods is balanced (Feilhauer et al., 2015). The difference in the band selection principles among the three methods provides an understanding of the selection results and of the merit in merging their results in the ensemble method. The NSSA method is based on the contrast in spectral shape, thus the band selection result is dependent on the differences amongst the average spectrum of each species. The highest contrast observed in this study occurs near 9 µm, where some species (not all of them) show significant

spectral differences. In the VSRF method, the band selection is highly dependent on the classification accuracy of all samples, and a minimum classification error is obtained by maximizing the separation of each two species. Thus, multiple regions (discrete clusters of bands) that best differentiate two classes of species are selected by this method. While, in the mRMR method, the correlation between bands is critical, and one attempts to minimize the redundancy in the data, therefore, the bands selected are dispersed across the whole spectral region.

4.5.3 Implications for future image mapping

The feature identification based on field spectra of leaves is significant to the mapping of plant types at a large scale, with the access to airborne imaging spectrometers (e.g., HyTES, SEBASS, Hyper-Cam, OWL). The band analysis in this study can be used to analyze thermal hyperspectral imaging data. For example, as this study highlights the importance of a spectral region near 9 um, one might expect such diagnostic features to be seen in image endmembers. In a preliminary analysis of an airborne HyTES (Hook et al., 2013) image collected for the Huntington Botanical Gardens in California, different vegetation types presented detectable spectral variations at bands between 8.5-9.5 μm.

The band selection and classification in this study provide some insights to derive spectral classes from airborne imaging in future research. Species that present features driven by different compounds are more likely to be discriminated using image data. Species that are more readily confused (e.g. LC and LS) may then be combined into one spectral class for image mapping. Furthermore, the relation between species clarified by the confusion matrix support the plausibility of Spectral Types (ST) as suggested by Harrison et al. (2018). Generally, species from different spectral types are more easily distinguished and classified than the ones from the same spectral

type. This indicates that the spectral variability between spectral types is greater than the variability within spectral types.

This study identifies feature for species classification at the leaf level and the classification accuracy was high. However, mapping at the canopy or crown level will have to take account several additional factors including leaf angle and shading that may further decrease spectral contrast. In addition, the analysis of airborne imagery requires pre-processing steps such as atmospheric correction, shadow masking and spatial filtering (Laybros et al., 2019) that will further impact the classification.

4.6 Conclusions

This study applied statistical approaches of band selection to longwave infrared spectral data of leaves collected in a tropical dry forest for the discrimination of twenty-six tree species. The bands selected using an ensemble of multiple methods improved the Logistic Regression classification performance by 3% in comparison to a result without band selection. The ensemble method encompassed the random forest, minimum redundancy maximum relevance and n-dimensional spectral solid angle methods. Results of this study corroborate previous findings for studies in other ecosystems that LWIR spectral features of tree species are driven by leaf constitutional compounds. Since spectral features in the LWIR range have not been extensively exploited for large scale remote sensing mapping, the identified bands or features in this study provide valuable information to guide the future mapping of tree species with the emergence of longwave hyperspectral imaging systems. The ensemble-based band selection method combines multiple methods and can be adapted to such applications for the robust discrimination of spectrally similar tree species.

Table 4.1 Lists of tree species and their botanical family, and spectral types as classified on the basis of dominating features by Harrison et al. (2018).

Species	Abbreviation	Family	Spectral type
Albizia adinocephala	AA	*Leguminosae*	5
Astronium graveolens	AG	*Anacardiaceae*	1
Ateleia herbert-smithii	AH	*Leguminosae*	Single species
Bombacopsis quinate	BQ	*Bombacaceae*	1
Bursera simaruba	BS	*Burseraceae*	3
Byrsonima crassifolia	BC	*Malpighiaceae*	2
Calycophyllum candidissimum	CC	*Rubiaceae*	1
Casearia argute	CAR	*Fagaceae-Flacourtiaceae*	1
Casearia sylvestris	CS	*Fagaceae-Flacourtiaceae*	5
Cedrela odorata	CO	*Meliaceae*	4
Cochlospermum vitifolium	CV	*Cochlospermaceae*	1
Curatella Americana	CAM	*Dilleniaceae*	3
Eugenia solanensis	ES	*Myrtaceae*	2
Guazuma ulmifolia	GU	*Sterculiaceae*	1
Lonchocarpus minimiflorus	LM	*Fabaceae-Papolionoideae*	Single species
Luehea candida	LC	*Tiliaceae*	Luehea sp. ST
Luehea speciose	LS	*Tiliaceae*	Luehea sp. ST
Machaerium biovulatum	MB	*Fabaceae-Papolionoideae*	1
Maclura tinctorial	MT	*Moraceae*	3

Psidium guajava	PG	*Myrtaceae*	5
Rehdera trinervis	RT	*Verbenaceae*	Single species
Semialarium mexicanum	SME	*Hippocrateaceae*	4
Simarouba glauca	SG	*Simaroubaceae*	5
Spondias mombin	SMO	*Anacardiaceae*	1
Tabebuia ochracea	TO	*Bignoniaceae*	1
Zuelania Guidonia	ZG	*Fagaceae-Flacourtiaceae*	1

Table 4.2 Lists of spectral features reported in the literature for leaf compounds and corresponding locations of leaf spectral features reported by Harrison et al. (2018). The wavelengths listed in the first and second column correspond to the center of absorption features seen in leaf and compound spectra respectively.

Band of Leaf (µm)	Band of Compounds (µm)	Probably origin
8.19	8.18 (Cellulose)	Unknown (Poletto et al., 2013)
8.74	8.62 (Cellulose)	asymmetric COC bridge stretching (Poletto et al., 2013)
9.12	9.01 (Cellulose)	anhydroglucose ring asymmetric stretching (Poletto et al., 2013)
9.55	9.44 (Cellulose)	the CO stretching of primary alcohol functional groups (Ribeiro da Luz and Crowley, 2007)/CO stretching (Poletto et al., 2013)
8.19	8.58 (Cellulose)	Unknown (Poletto et al., 2013)
8.74	8.84 (Cellulose)	asymmetric COC bridge stretching (Poletto et al., 2013)
9.12	9.3 (Cellulose)	anhydroglucose ring asymmetric stretching (Poletto et al., 2013)
9.55	9.63 (Cellulose)	the CO stretching of primary alcohol functional groups (Ribeiro da Luz and Crowley, 2007)/CO stretching (Poletto et al., 2013)
9.17−9.40	9.40−9.70 (xylan?)	COH bending (Ribeiro da Luz and Crowley, 2007); (most likely: Xyloglucan, present in many matrix glycan very strong band; Kacuráková 2000)
9.17-9.40	9.08 band max (silica)	Si-O bond identified from fumed silica (Luz & Crowley, 2007) or opaline silica (Bjorn and Li, 2011)
8.61	8.65 (Cutin)	asymmetrical stretching modes of ester COC bonds (Ribeiro da Luz and Crowley, 2007)
9.25	9.14 (Cutin)	symmetrical stretching modes of ester COC bonds (Ribeiro da Luz and Crowley, 2007)
8.61	8.67 (Acid)	Unknown
8.93	9 (Acid)	Stretching vibrations of CO group of carbonic acid (Vyas and Argal 2014)
9.31	9.27 (Acid)	Unknown
9.65	9.7 (Acid)	Unknown
10.24	10.36 (Acid)	Unknown
10.75	10.75 (Acid)	Unknown

Table 4.3 Bands obtained using the NSSA method applied to the compound spectra.

Dataset	Selected bands (µm) using NSSA method
Compounds spectra	8.19, 8.2, 8.24, 8.26, 8.27, 8.28, 8.3, 8.45, 8.46, 8.47, 8.49, 8.51, 8.57, 8.61, 8.83, 8.85, 8.86, 8.88, 8.89, 8.91, 8.93, 8.96, 8.97, 8.99, 9.03, 9.05, 9.08, 9.1, 9.11, 9.26, 9.28, 9.34, 9.36, 9.38, 9.39, 9.41, 9.43, 9.45, 9.46, 9.48, 9.5, 9.51, 9.53, 9.55, 9.57, 9.58, 9.6, 9.62, 9.67, 9.69, 9.97, 10.39, 10.41, 10.43, 10.45, 10.48, 10.5, 10.52, 10.54, 10.56

Table 4.4 Bands selected from the leaf spectra by the ensemble method.

Dataset	Selected bands (µm) using ensemble approach
Leaf spectra	8.19, 8.21, 8.23, 8.26, 8.62, 8.63, 8.64, 8.66, 8.68, 8.70, 8.71, 8.73, 8.74, 8.76, 8.77, 8.89, 8.90, 8.92, 8.93, 8.95, 8.96, 9.07, 9.08, 9.24, 9.25, 9.27, 9.29, 9.30, 9.32, 9.33, 9.35, 9.53, 9.55, 9.58, 9.60, 9.62, 9.64, 9.65, 9.98, 9.99, 10.01, 10.03, 10.50

Table 4.5 Classification evaluation for bands selected by single methods and the ensemble model.

Evaluation criteria (Cross-Validation)	bands by VSRF	bands by mRMR	bands by NSSA	bands by ensemble
Overall accuracy	0.87	0.90	0.90	**0.94**
Kappa statistic	0.86	0.90	0.90	**0.94**
Mean absolute error	0.01	0.01	0.01	**0.00**
Relative absolute error	0.14	0.10	0.10	**0.06**

Table 4.6 The confusion matrix of the classification based on the ensemble band set.

AA	PG	SG	AG	BQ	CAR	CC	CV	GU	LC	LS	SMO	TO	ZG	BS	CAM	BC	CS	ES	MB	LM	CO	MT	AH	SME	RT	
15	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	AA
0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	PG
0	0	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	SG
0	0	0	19	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	AG
0	0	0	0	17	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	BQ
0	0	0	0	0	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	CAR
0	0	0	0	0	0	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	CC
0	0	0	0	0	0	0	14	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	CV
0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	GU
1	0	0	0	0	0	0	0	0	22	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	LC
0	0	0	0	0	0	0	0	0	2	13	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	LS
0	0	0	0	0	0	0	0	0	0	0	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	SMO
0	0	0	0	0	0	0	0	0	0	0	0	29	0	0	0	0	0	0	0	0	0	0	0	0	0	TO
0	0	0	0	0	0	0	0	0	0	0	0	0	13	0	0	0	0	0	0	0	0	0	0	0	0	ZG
0	0	0	1	0	0	0	0	0	0	0	0	0	0	18	0	0	0	0	0	0	0	0	0	0	0	BS
0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	18	1	0	0	0	0	0	1	0	0	0	CAM
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	BC
0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	CS
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22	0	0	0	0	0	0	0	ES
0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	21	0	0	0	0	0	0	MB
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	14	0	0	0	0	0	LM
0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	27	0	0	0	0	CO
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	18	2	0	0	MT
0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17	1	0	AH
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	29	0	SME
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28	RT
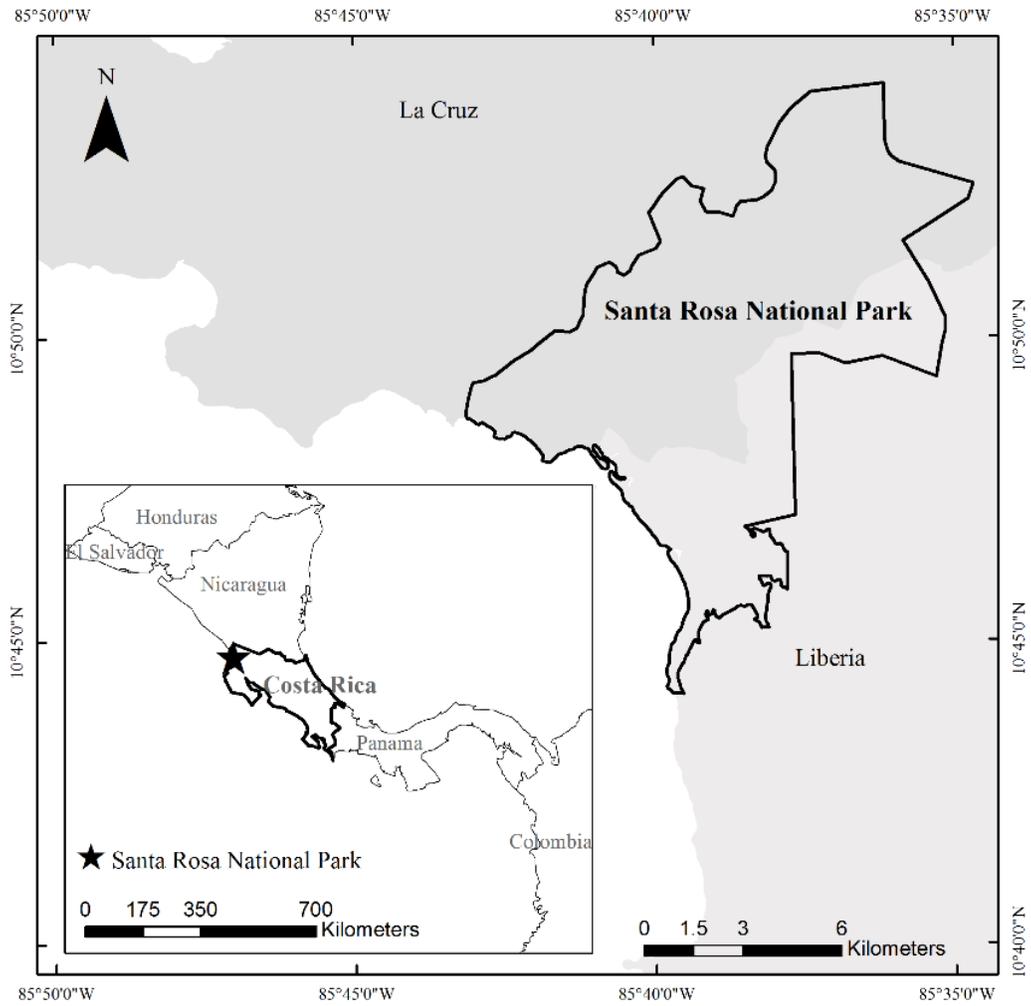
Figure 4.1 Location of the Santa Rosa National Park in Costa Rica.
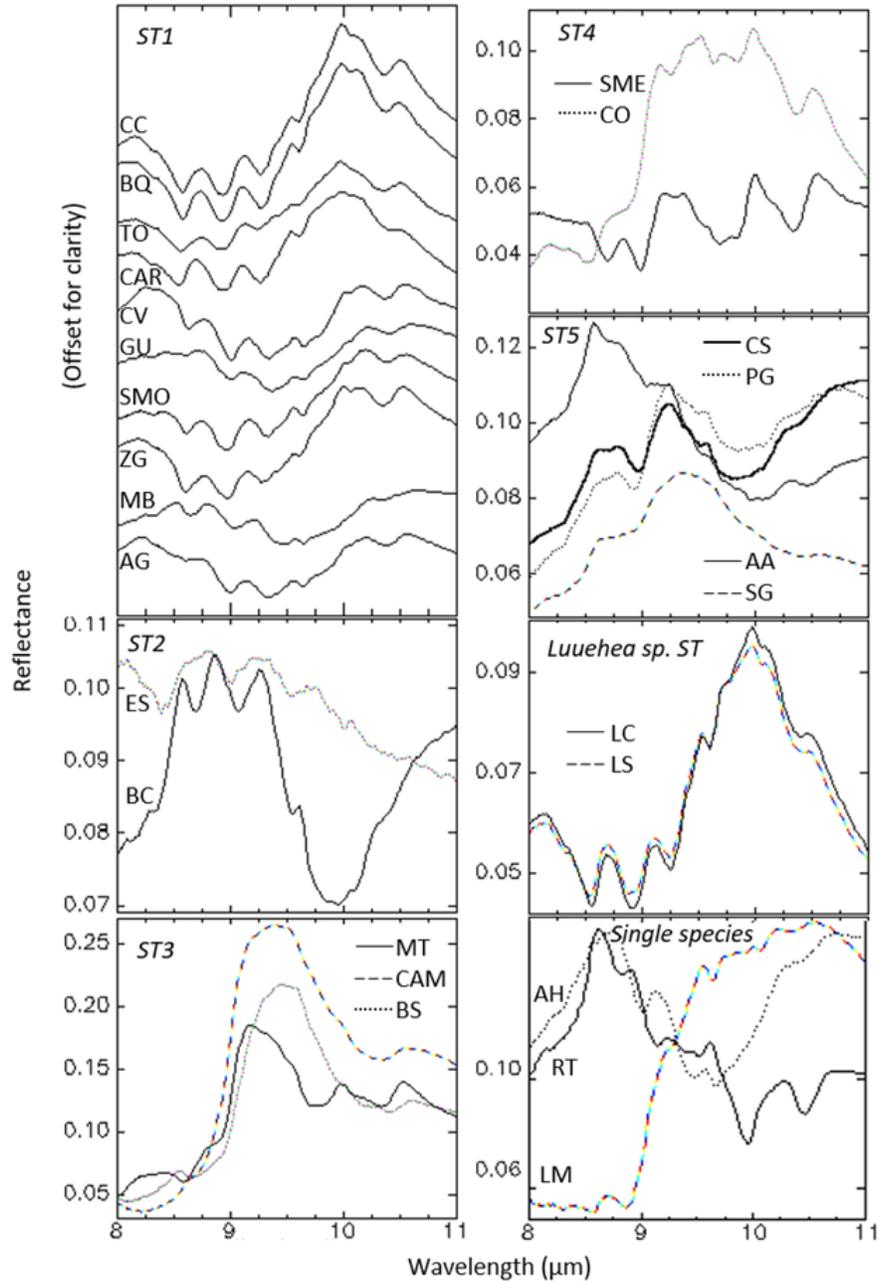
Figure 4.2 Average spectrum of each of 21 tree species assigned to five Spectral Types (ST) by Harrison et al. (2018) based on the similarity of spectral features observed in their leaf spectra. Also shown in separate plots are species LC and LS, part of the *Luehea* spectral type, and two single species (AH, RT and LM).
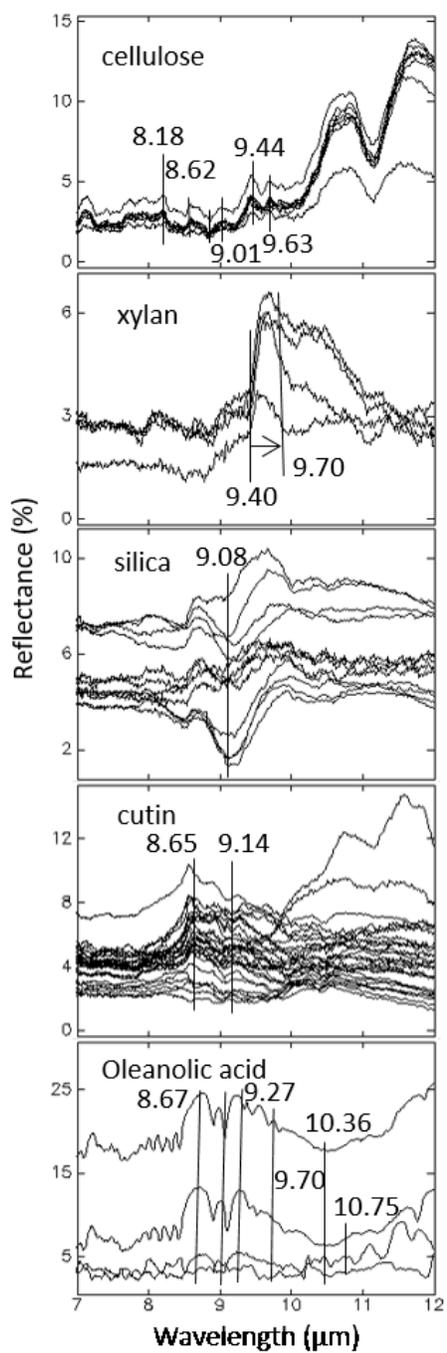
Figure 4.3 Reflectance spectra of constitutional compounds: cellulose, xylan, silica, cutin, and oleanolic acid. Vertical lines denote the location of features in compound spectra identified by Harrison et al. (2018) based on comparison with features reported in the literature (labeled in Table 4.2 as band of compounds).
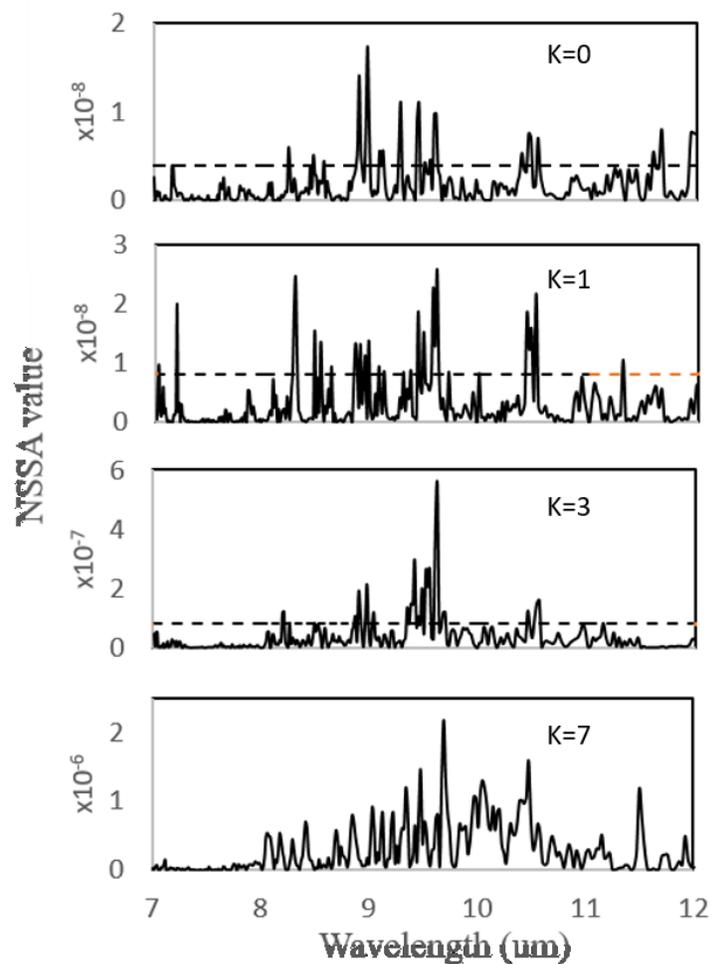
Figure 4.4 Band selected from the average spectra of the five compounds using NSSA profiles of the NSSA values as a function of wavelength for discrete k values (0, 1, 3, and 7) thus capturing narrow and broad spectral features. The horizontal dotted line in each profile marks the threshold value above which the bands are retained.
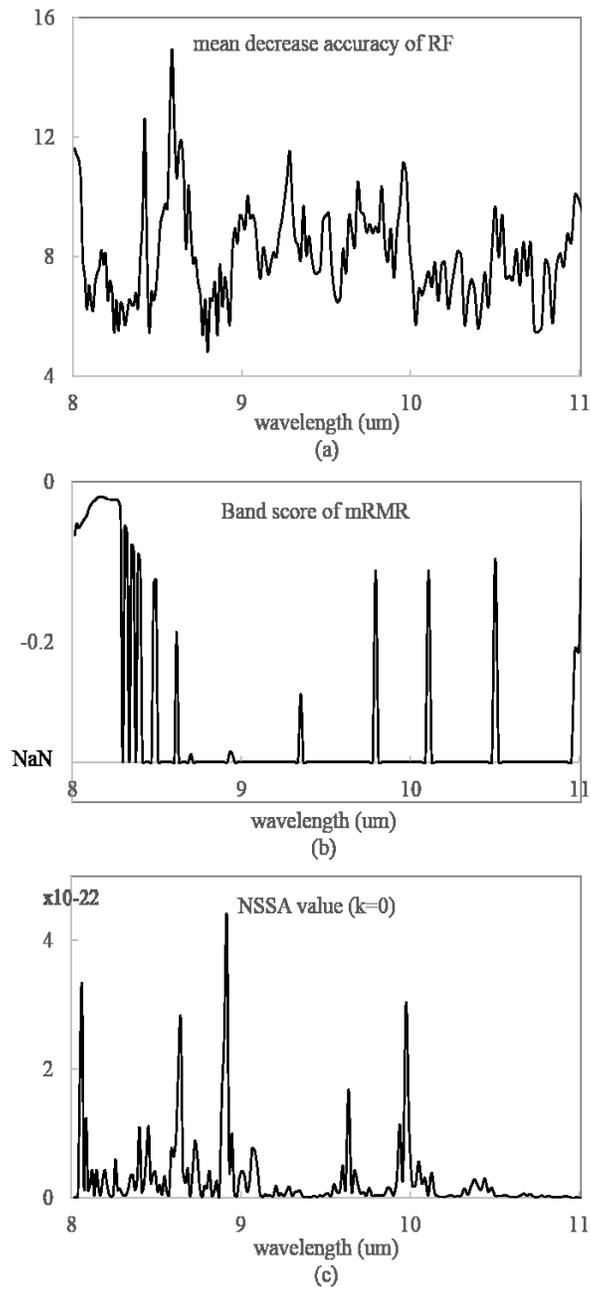
Figure 4.5 Band importance for leaf spectra generated by the Variable Selection with Random Forest (VSRF) (a), minimum Redundancy Maximum Relevance (mRMR) (b), and N-dimensional Spectral Solid Angle (NSSA) (c) methods.
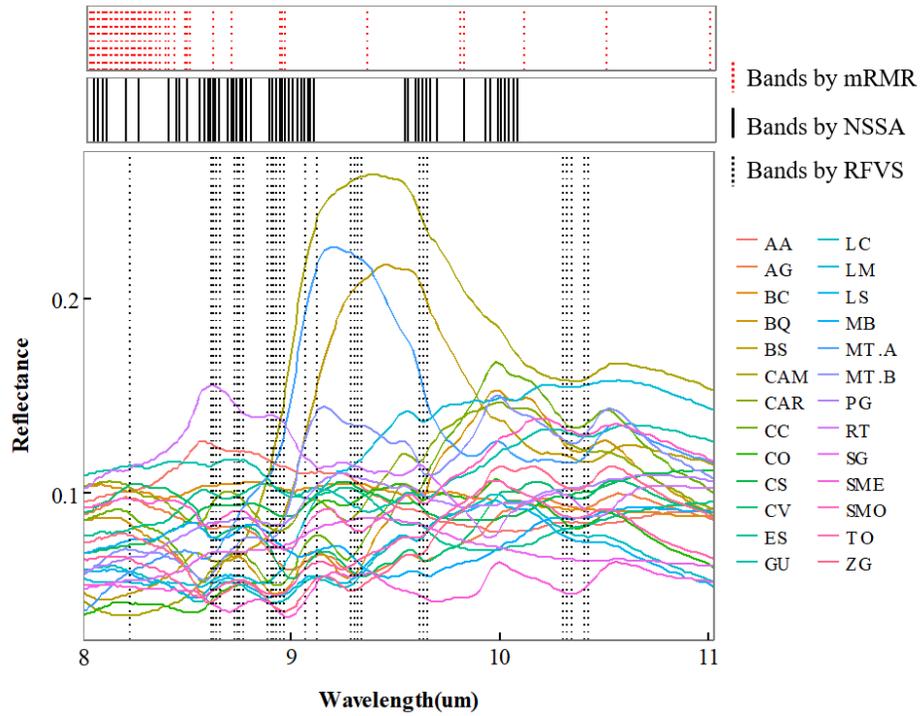
Figure 4.6 Average spectrum of each tree species and bands selected by the VSRF, mRMR and NSSA methods shown as vertical lines.
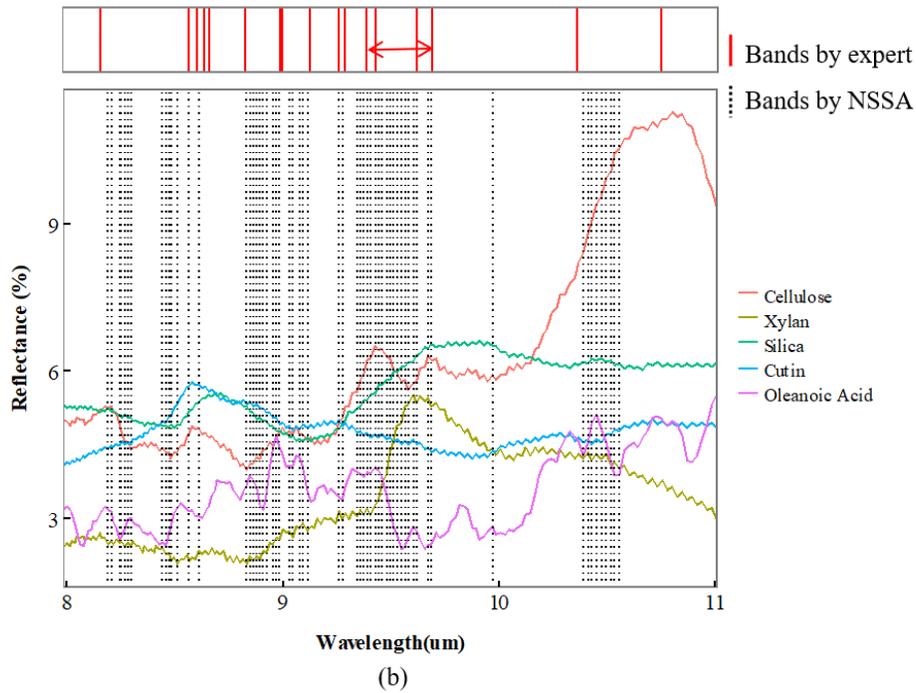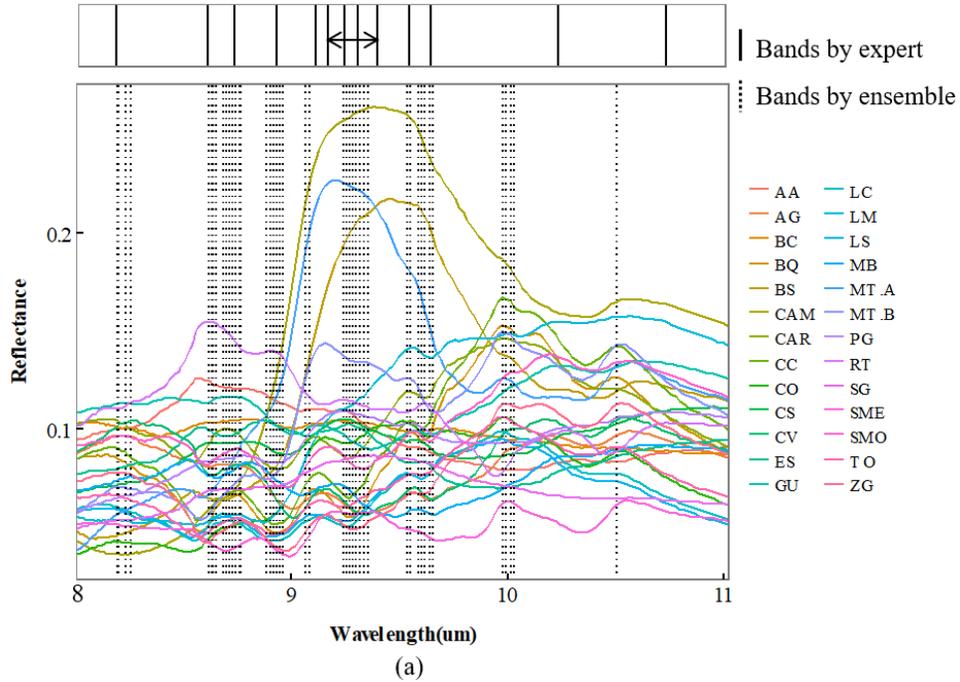
Figure 4.7 Comparison of the bands selected in this study with those selected by Harrison et al. (2018) and here labeled "by expert". Bands selected for the 26 tree species by the ensemble method and by Harrison et al. (a), and for the compounds using the NSSA method and by Harrison et al. (b).

References

Acevedo, M. F. B., Groen, T. A., Hecker, C. A., and Skidmore, A. K., 2017. Identifying leaf traits that signal stress in TIR spectra. ISPRS journal of Photogrammetry and Remote Sensing 125: 132-145.

Asner, G. P., 1998. Biophysical and biochemical sources of variability in canopy reflectance. Remote Sensing of Environment 64(3): 234-253.

Belgiu, M., and Drăguţ, L., 2016. Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing 114: 24-31.

bin Othman, M.F. and Yau, T.M.S., 2007. Comparison of different classification techniques using WEKA for breast cancer. In 3rd Kuala Lumpur International Conference on Biomedical Engineering 2006 (520-523). Springer, Berlin, Heidelberg.

Buitrago, M. F., Skidmore, A. K., Groen,T. A., and Hecker, C. A., 2018. Connecting infrared spectra with plant traits to identify species. ISPRS Journal of Photogrammetry and Remote Sensing 139: 183-200.

Castro-Esau, K. L., Sánchez‐Azofeifa, G. A., Rivard, B., Wright, S. J., and Quesada, M., 2006. Variability in leaf optical properties of Mesoamerican trees and the potential for species classification. American Journal of Botany 93(4): 517-530.

Chan, J. C. W., and Paelinckx, D., 2008. Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. Remote Sensing of Environment 112(6): 2999-3011.

Clark, M. L., Roberts, D. A., and Clark, D. B., 2005. Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. Remote sensing of environment 96(3-4): 375-398.

da Luz, B. R., 2006. Attenuated total reflectance spectroscopy of plant leaves: a tool for ecological and botanical studies. New Phytologist 172(2): 305-318.

da Luz, B. R., and Crowley, J. K., 2007. Spectral reflectance and emissivity features of broad leaf plants: Prospects for remote sensing in the thermal infrared (8.0–14.0 μm). Remote Sensing of Environment 109(4): 393-405.

da Luz, B. R., and Crowley, J. K., 2010. Identification of plant species by using high spatial and spectral resolution thermal infrared (8.0–13.5 μm) imagery. Remote sensing of Environment 114(2): 404-413.

Diaz-Uriarte, R., 2007. GeneSrF and varSelRF: a web-based tool and R package for gene selection and classification using random forest. BMC bioinformatics 8(1): 328.

Ding, C., and Peng, H., 2005. Minimum redundancy feature selection from microarray gene expression data. Journal of Bioinformatics and Computational Biology 3(02): 185-205.

Fassnacht, F.E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L.T., Straub, C., and Ghosh, A., 2016. Review of studies on tree species classification from remotely sensed data. Remote Sensing of Environment 186: 64-87.

Fassnacht, F.E., Neumann, C., Förster, M., Buddenbaum, H., Ghosh, A., Clasen, A., Joshi, P.K., and Koch, B., 2014. Comparison of feature reduction algorithms for classifying tree species with hyperspectral data on three central European test sites. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7(6): 2547-2561.

Feilhauer, H., Asner, G. P., and Martin, R. E., 2015. Multi-method ensemble selection of spectral bands related to leaf biochemistry. Remote Sensing of Environment 164: 57-65.

Friedman, J., Hastie, T. and Tibshirani, R., 2001. The elements of statistical learning (Vol. 1, No. 10). New York: Springer series in statistics.

Gamon, J. A., and Surfus, J. S., 1999. Assessing leaf pigment content and activity with a reflectometer. The New Phytologist 143(1): 105-117.

Genuer, R., Poggi, J-M., and Tuleau-Malot, C., 2010. Variable selection using Random Forests. Pattern Recognition Letters, Elsevier 31 (14): 2225-2236.

Gerber, F., Marion, R., Olioso, A., Jacquemoud, S., Da Luz, B. R., and Fabre, S., 2011. Modeling directional–hemispherical reflectance and transmittance of fresh and dry leaves from 0.4 μm to 5.7 μm with the PROSPECT-VISIR model. Remote Sensing of Environment 115(2): 404-414.

Gong, P., Pu, R., and Yu, B., 1997. Conifer species recognition: An exploratory analysis of in situ hyperspectral data. Remote sensing of Environment 62(2): 189-200.

Guzman, J.A., Rivard, B., and Sánchez-Azofeifa, G.A., 2018. Discrimination of liana and tree leaves from a Neotropical Dry Forest using visible-near infrared and longwave infrared reflectance spectra. Remote Sensing of Environment, 219:135-144.

Harrison, D., Rivard, B., and Sánchez-Azofeifa, A., 2018. Classification of tree species based on longwave hyperspectral data from leaves, a case study for a tropical dry forest. International Journal of Applied Earth Observation and Geoinformation 66: 93-105.

Hook, S.J., Johnson, W.R. and Abrams, M.J., 2013. NASA's hyperspectral thermal emission spectrometer (HyTES). In Thermal Infrared Remote Sensing (pp. 93-115). Springer, Dordrecht.

Kalacska, M., Bohlman, S., Sanchez-Azofeifa, G. A., Castro-Esau, K., and Caelli, T., 2007. Hyperspectral discrimination of tropical dry forest lianas and trees: Comparative data reduction approaches at the leaf and canopy levels. Remote Sensing of Environment 109(4): 406-415.

Kalacska, M., Sanchez-Azofeifa, G. A., Calvo-Alvarado, J. C., Quesada, M., Rivard, B., and Janzen, D. H., 2004. Species composition, similarity and diversity in three successional stages of a seasonally dry tropical forest. Forest Ecology and Management 200(1-3): 227-247.

Laybros, A., Schläpfer, D., Féret, J. B., Descroix, L., Bedeau, C., Lefevre, M. J., and Vincent, G., 2019. Across date species detection using airborne imaging spectroscopy. Remote Sensing 11(7): 789.

Meerdink, S. K., Roberts, D. A., King, J. Y., Roth, K. L., Dennison, P. E., Amaral, C. H., and Hook, S. J., 2016. Linking seasonal foliar traits to VSWIR-TIR spectroscopy across California ecosystems. Remote Sensing of Environment 186: 322-338.

Meyer, P. E., Schretter, C., and Bontempi, G., 2008. Information-theoretic feature selection in microarray data using variable complementarity. IEEE Journal of Selected Topics in Signal Processing 2(3): 261-274.

Peng, H., Long, F., and Ding, C., 2005. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(8): 1226-1238.

Pu, R., 2009. Broadleaf species recognition with in situ hyperspectral data. International Journal of Remote Sensing 30(11): 2759-2779.

Quesada, M., Sanchez-Azofeifa, G.A., Alvarez-Anorve, M., Stoner, K.E., Avila-Cabadilla, L., Calvo-Alvarado, J., Castillo, A., Espírito-Santo, M.M., Fagundes, M., Fernandes, G.W., and Gamon, J., 2009. Succession and management of tropical dry forests in the Americas: Review and new perspectives. Forest Ecology and Management 258(6): 1014-1024.

Ravikumar, P., Wainwright, M. J., and Lafferty, J. D., 2010. High-dimensional Ising model selection using $\ell 1$-regularized logistic regression. The Annals of Statistics 38(3): 1287-1319.

Salisbury, J. W., 1986. Preliminary measurements of leaf spectral reflectance in the 8-14 µm region. International Journal of Remote Sensing 7(12): 1879-1886.

Salisbury, J. W., and Milton, N. M., 1988. Thermal infrared (2.5-to 13.5-µm) directional hemispherical reflectance of leaves. Photogrammetric Engineering and Remote Sensing 54(9): 1301-1304.

Sánchez-Azofeifa, G.A., Castro, K., Wright, S.J., Gamon, J., Kalacska, M., Rivard, B., Schnitzer, S.A. and Feng, J.L., 2009. Differences in leaf traits, leaf internal structure, and spectral reflectance between two communities of lianas and trees: Implications for remote sensing in tropical environments. Remote Sensing of Environment 113(10): 2076-2088.

Sims, D. A., and Gamon, J. A., 2002. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. Remote Sensing of Environment 81(2-3): 337-354.

Tian, M., Feng, J., Rivard, B., and Zhao, C., 2016. A method to compute the n-dimensional solid spectral angle between vectors and its use for band selection in hyperspectral data. International Journal of Applied Earth Observation and Geoinformation 50: 141-149.

Ullah, S., Groen, T. A., Schlerf, M., Skidmore, A. K., Nieuwenhuis, W., and Vaiphasa, C., 2012. Using a genetic algorithm as an optimal band selector in the mid and thermal infrared (2.5–14 µm) to discriminate vegetation species. Sensors 12(7): 8755-8769.

Ullah, S., Schlerf, M., Skidmore, A. K., and Hecker, C., 2012. Identifying plant species using mid-wave infrared (2.5–6 µm) and thermal infrared (8–14 µm) emissivity spectra. Remote Sensing of Environment 118: 95-102.

Ullah, S., Skidmore, A. K., Ramoelo, A., Groen, T. A., Naeem, M., and Ali, A., 2014. Retrieval of leaf water content spanning the visible to thermal infrared spectra. ISPRS Journal of Photogrammetry and Remote Sensing 93: 56-64.

Viera, A.J. and Garrett, J.M., 2005. Understanding interobserver agreement: the kappa statistic. Fam med, 37(5): 360-363.

Zhang, J., Rivard, B., Sánchez-Azofeifa, A., and Castro-Esau, K., 2006. Intra-and inter-class spectral variability of tropical tree species at La Selva, Costa Rica: Implications for species identification using HYDICE imagery. Remote Sensing of Environment 105(2): 129-141.

Chapter 5 Incorporating Band Selection in the Spatial Selection of Spectral Endmembers

5.1 Introduction

5.1.1 Spectral Mixture Analysis

Hyperspectral remote sensing has been used for regional mapping in a range of studies including land cover, biophysical, water, agricultural, urban and geologic investigations (Adams, 1993; Powell et al., 2007; Van der Meer & Jia, 2012). Collecting data in narrow and continuous bands, hyperspectral sensors enable the detailed detection of constituent materials in a scene. While common mapping techniques classify each pixel, Spectral Mixture Analysis (SMA) aims to address the per-pixel mixture of pure spectral signatures (i.e. endmembers) due to the presence of multiple targets in each pixel. With knowledge of the endmember spectra in the scene, one can retrieve the fractional abundance of each endmember in each pixel through linear or nonlinear modeling, a process known as "Spectral Unmixing (SU)" (Keshava & Mustard, 2002).

The accuracy of the predicted fractional abundances is greatly impacted by the quality and contrast of endmembers detected from the scene. For endmembers with low contrast it is important to highlight the specific bands that best discriminate them. A variety of methods have been proposed to extract representative and realistic endmembers, for example, the Pixel Purity Index (PPI), N-finder and the Iterative Error Analysis (IEA) (Zortea & Plaza, 2009). However, these prevailing methods focus exclusively on the spectral nature of the data and neglect the spatial context that exists in the image data (Plaza et al., 2002). They consider all adjacent pixels as random individuals; however, the pixels are in fact spatially correlated, and this correlation is an important information source that can guide the exploitation of spectral information (Zortea &

Plaza, 2009). To date, few attempts have been made to include spatial information in endmember determination. Plaza et al. (2002) first proposed a spectral/spatial approach for endmember extraction using multidimensional morphological operations. Later, they implemented a spatial preprocessing for endmember extraction based on a predefined set of spatially representative image regions (Martin & Plaza, 2011). Earlier Rogge et al. (2007) proposed the Spatial-Spectral Endmember Extraction (SSEE) approach that involved dividing the image into equal and non-overlapping spatial subsets to select more endmembers that are spectrally similar but spatially independent.

5.1.2 Band selection and SMA

Selecting a subset of available bands can improve the selection of endmembers and the estimation of abundance for the selected endmembers (Rogge et al., 2014). Spectral unmixing approaches, such as singular value decomposition, that are used to solve the linear equations predicting endmember abundances using an endmember matrix, assume the orthogonality of input endmembers (perpendicular vectors). However, the endmember matrix is mathematically non-orthogonal as endmembers can be highly correlated (defined as 'collinearity' in SU). The collinearity makes the abundance estimation sensitive to random errors (Van der Meer & Jia, 2012). An effective band selection can decorrelate endmembers or increase their decorrelation by focusing on the most discriminating spectral features (Somers et al., 2011). Thus, the collinearity is reduced, and spectral unmixing becomes more accurate with these uncorrelated and spectrally distinct endmembers.

Band selection has been effectively used to improve spectral mixture analysis in a series of practical studies. A majority of these studies rely on expert knowledge of ground materials and their spectral properties to select bands. Asner and Lobel (2000) proposed the AutoSWIR technique for spectral unmixing of plant cover through a careful selection of bands. Their work focused on the SWIR region of 2.05-2.50 □m to highlight important features linked to the optical properties of leaves, litter, and soils. In a geologic study, Rogge et al. (2014) manually selected a subset of thirty-seven bands from hyperspectral imagery to optimize endmember extraction and spectral unmixing. More bands were selected in the SWIR region to cover key absorption features that discriminate mafic and ultramafic rocks that have low spectral contrast. Without band selection, the discrimination of these rocks was biased toward differences in the broad continuum of endmember spectra that, in their study, was largely controlled by extensive lichen cover on rock surfaces (Rogge et al., 2014). The above studies showed improvements in SMA results when expert knowledge was used to select bands that were applied to the entire scene. Studies that have reported the use of a statistical approach to select bands for unmixing include that of Miao et al. (2006) that applied band selection unconstrained linear spectral mixture models to assess the abundance of an invasive weed species in California. Somers et al. (2010) automatically selected spectral features for SMA that are least sensitive to spectral variability based on a criterion examining the within and between endmember class variability.

In most applications, band selection is applied as a preprocessing step that is independent of the endmember selection process. In such instance researchers select bands associated with spectral features that have been identified from visual analysis of spectral libraries or obtained from statistical methods that ignore the correlation of endmembers. The impact of band selection on endmember selection is seldom explored (Iordache et al, 2015; Tane et al., 2018). In one such

study, Iordache et al. (2015) examined band selection before scene-based endmember selection as compared to band selection following endmember selection. Improvements in unmixing results were obtained when band selection was performed after finalizing the endmember selection. We know of no studies that have examined band selection conducted simultaneously with the endmember selection process.

5.1.3 The incorporation of band selection in endmember extraction

This study integrates band selection in the extraction of endmembers and examines the impact of band selection on the final selection of endmembers as well as on unmixing results. Thus, the processes of band selection and endmember extraction are simultaneously rather than sequentially employed as explored by Iordache et al. (2015), for example. We integrate two established methods to accomplish this task namely the N-dimensional Spectral Solid Angle (NSSA) band selection method and the above mentioned Spatial-Spectral Endmember Selection method (SSEE).

The NSSA band selection method was proposed by Tian et al. (2016) to select the most dissimilar spectral regions amongst endmembers. Using both synthetic and real data, it was shown to capture subtle spectral features that improve the separation of similar spectra (Tian et al., 2016; Long et al., 2019). The NSSA method requires a spectral library of endmembers as input from which bands are selected. The SSEE method determines candidate endmembers for spatial subsets of the image prior to determining a global endmember set. Thus, it is feasible to integrate band selection directly into the first stage of SSEE. Such integration implies that the spatial information at the local scale (e.g. the local targets and their mixtures) drives the band selection, an approach

to band selection that is distinct from traditional methods of band selection that solely use spectral information.

The premise of this study is that the incorporation of band selection in the spatial and spectral endmember selection process will result in a band set that enhances the differentiation among endmembers that are spectrally similar but spatially independent. In turn, this will impact the nature of the final endmember set used for unmixing and the unmixing results. Therefore the objectives of this paper are to describe a new methodology to conduct band selection by integrating the NSSA method within the SSEE process (defined as "NSSA-SSEE) and examine the impact of band selection on endmember selection and spectral unmixing. The paper is organized as follows: Section 5.2 describes the study site and the hyperspectral data collected to address the objectives and for the evaluation of the results. Section 5.3 describes the method in two aspects: the conceptual framework to integrate NSSA in the SSEE process and the utilization of this tool for analysis of hyperspectral imagery. The results address the bands selected, and their impact on endmember selection and on abundances obtained from unmixing. Particular attention is paid to determine if the method aids in the detection of geologic spectral subclasses with low spectral contrast as these are challenging to retrieve in spectral analysis.

5.2 Study site and hyperspectral data

5.2.1 Study site

The study area is located in the eastern part of the 1.9 Ga Proterozoic Cape Smith Belt in Nunavik, northeast Canada (nominally 62 latitude, -73 longitude) (Parrish 1989; Machado et al. 1993). The Cape Belt is one of the largest magmatic Ni-Cu resource in Canada (Lydon, 2007).

Here deposits occur in ultramafic sills and intrusions (dunite, peridotite, olivine-pyroxenite, pyroxenite, gabbro) (50-200m thick) exposed discontinuously and these have been metamorphosed at lower greenschist facies (Barnes and Lightfoot, 2005; Lesher 2007). These rocks can also include gossans, oxidized surfaces formed from the weathering of sulfides. Vegetation in the study area is sparse comprising grasses, mosses, and shrubs, and rock-encrusting lichens are predominant on bedrock surfaces (Laakso et al., 2016). The presence of rock encrusting lichens and the similarity of the mineralogy for the rock types relevant to mineral exploration (mafic to ultramafic rocks) presents a challenge for detection and mapping using remote sensing data. Though early exploration in the belt dates to 1898 (Lesher, 2007), exploration and development remain at an early stage in many parts of the belt. This led to the deployment in 2008 of an expansive hyperspectral survey, part of which is used in this study.

Rogge et al. (2014) analyzed an area of 10x20 km comprising 20 flight lines of this survey and encompassing the specific study area of our investigation. Their investigation targeted the use of image-derived spectral endmembers validated using spectral measurements from samples collected during a field mapping campaign. This information supported their spectral unmixing analysis of the airborne data for regional mapping. In a complementary study, Laakso et al. (2016) reported on the spectroscopy and mapping of gossans in the area. Our band selection investigation builds on the work of Rogge et al. (2014) making use of data described in the next section.

5.2.2 Hyperspectral data

5.2.2.1 Spectra of rock samples

The spectral characteristics of the predominant rock types in the study area were reported in a prior study (Rogge et al. 2014) using reflectance spectra collected from 77 rock samples obtained during the 2009 and 2010 summer field campaigns. These include samples of ultramafic rocks namely dunite (6), peridotite (25), olivine-pyroxenite (9), and pyroxenite (11), and mafic rocks namely gabbro (16) and basalt (10). As shown in Figure 5.1a, these samples display variable and typically extensive lichen cover.

Rogge et al. (2014) reported that these metamorphosed rocks are comprised of similar mineralogy but differ in the relative abundance of three dominant minerals namely antigorite $[(Mg,Fe)_3(Si_2O_5)(OH)_4]$, actinolite $[(Ca_2(Mg,Fe)_5Si_8O_{22})(OH)_2]$, and clinochlore $[(Mg_6(Mg_4Al_2)[Si_6Al_2O_{22}](OH)_{16}]$ (refer to Table 5.2 in Rogge et al. 2014). Plagioclase is also present in gabbro and basalt. The most mafic rocks, dunite and peridotite, have the highest abundance of antigorite. Olivine-pyroxenite and pyroxenite contain a mixture of antigorite, actinolite, and clinochlore. The least mafic rocks, basalt and gabbro, have actinolite and clinochlore. The presence of lichen on the weathered rock surfaces implies that band selection must not only consider the mineralogy of the samples but the presence of lichen coatings (Rogge et al., 2014). Features seen in the spectra of the rock samples (Figure 5.1b) at wavelengths shorter than 2.0 μm are commonly controlled by lichen (Bechtel et al. 2002). At longer wavelengths, the spectra of the six rock types can be discerned on the basis of their detailed shape that is controlled by the presence of three minerals as discussed in detail by Rogge et al. (2014) (see Figure 5.7 herein). Antigorite, actinolite, and clinochlore present diagnostic features at [2.285, 2.325], [2.245, 2.315, 2.386] and 2.345 μm, respectively caused by overtone and combination tones of the OH stretch (Clark et al. 1990; Clark, 1999). The discrimination of the four ultramafic rock types is mainly based on the subtle shift from 2.315 to 2.325 μm in the absorption feature observed as a

function of the change in the abundance of antigorite as well as the progressive increase of an absorption at 2.386 μm. The mafic rocks contain more clinochlore and can be distinguished from ultramafic rocks based on a relatively broad absorption extending from 2.285-2.345 μm.

5.2.2.2 Airborne imagery

This study makes use of airborne imagery covering an area of 3.3 x 5.3 km (Figure 5.2) that represents a subset of a larger mosaic of airborne hyperspectral data assembled by Rogge et al. (2014). The reader is referred to that publication for detail on the preprocessing procedures used to generate the georeferenced at-surface reflectance data including across-track and along-track radiometric corrections to assemble the mosaic. The airborne imagery was acquired at a ground sampling distance of 2m with the Specim AisaDUAL imaging system that comprises 178 bands (0.4-2.45 μm). From this band set, a number of wavelength regions were removed due to poor signal quality including bands short of 0.45 and long of 2.4 μm, bands from 0.87-1.08 μm, and bands within water absorption features from 1.37-1.52 μm and 1.8-2.0 μm. The remaining band set encompassed 128 bands. The specific image subset of this study shown in Figure 5.2 was selected because it encompasses bedrock exposures of the six dominant rocks mentioned above within a single large fold structure and because fieldwork conducted in the area can be part of the basis to assess mapping results obtained from the spectral analysis.

In their analysis of the larger mosaic, Rogge et al. (2014) extracted a suite of image endmembers that were used for spectral unmixing. The resulting abundance maps were then compared with the regional geological survey of Canada (GSC) 1: 50,000 geological map. Of relevance to this study, Rogge et al. (2014) recognized the need for band selection to capture the mineralogical information required to discriminate rock types. Based on the study of image and sample spectra, bands were selected throughout the available spectral range to discriminate broad

classes such as vegetation, snow, and water, and to capture spectral features of lichens. Then all bands from 2.238-2.427 μm were selected to cover key mineral features for the discrimination of the six dominant rock types. The resulting band set encompassed 37 bands defined by expert knowledge on the basis of field investigations and X-ray diffraction analysis (Table 5.1) that are used for comparison with band selection results of this study. The image endmembers they obtained spanned four broad classes including vegetation, snow, water, and rock types. Several geologic endmembers were associated with the mafic and ultramafic lithologies, and each broad class contained a number of subclasses that described the variability of such materials.

5.3 Method

This section describes three methodological aspects of this study. The first involves the description of the conceptual framework to incorporate the process of band selection into the process of spatial and spectral endmember selection. This is followed by a description of the experimental design enabling the comparison of two datasets, one resulting in unmixing results generated from endmembers without band selection, the other with endmembers using a spectral subset. For the last aspect, the process of band selection is evaluated at three levels via i) an examination of the bands selected, ii) their impact on the endmembers used for unmixing, and iii) on the abundance maps resulting from unmixing.

5.3.1 Conceptual framework to incorporate band selection in the endmember selection process

5.3.1.1 Description of the Spatial-Spectral Endmember Extraction algorithm (SSEE)

The SSEE algorithm described in Rogge et al. (2007, 2012) was one of the first endmember extraction tool taking advantage of both spectral and spatial information for the search of endmembers in a hyperspectral scene. It was selected for this study because it was designed specifically to utilize spatial information and extract endmembers of low spectral contrast.

The SSEE process comprises three steps. In the first step, the hyperspectral image is divided into equal-sized non-overlapping subset regions (here 25 pixels to the side), and a set of eigenvectors that explains the majority of the spectral variance in each subset is determined. In the second step, local candidate endmembers are selected based on projecting the pixels within the given subset onto the local eigenvectors. Then the local candidate pixels are projected on the eigenvectors compiled from all subsets and pixels lying at either end of the eigenvectors are selected as global candidate endmembers. In the third step global candidate pixels that are similar based on a spectral angle tolerance and that occur with a defined spatial window are averaged. SSEE typically finds a large number of endmembers compared to methods that are only spectrally based (e.g., PPI), but its advantage is that it retains endmembers that are spectrally similar but spatially distinct and that otherwise would typically be averaged.

5.3.1.2 Description of the N-dimensional Solid Spectral Angle (NSSA) band selection algorithm

Tian et al. (2016) proposed a band selection method using the N-dimensional Solid Spectral Angle (NSSA) that can be used to select the most dissimilar spectral regions amongst target endmembers. This band selection method was chosen for this study for two reasons. It is suitable for selecting bands from a small number of endmembers (e.g. local candidate SSEE endmembers), and it can enhance the separation of spectrally similar materials.

The NSSA method measures the solid angle between two or more spectra for a given spectral region encompassing a number of bands equal to the number of spectra under comparison. Since this number is typically small, the NSSA is computed by moving a sliding window with varying band intervals (denoted by "k") over the available spectral range. At each position of the sliding window, bands with the highest NSSA values indicate a high contrast between the spectra. Varying the band interval "k" enables the sliding window to encompass spectral features of varying widths. With this method, a continuous profile of NSSA values as a function of wavelength can be obtained and thresholded to define a selected band set. The reader is referred to Tian et al. (2016) for a graphical representation of the method.

5.3.1.3 Band selection using NSSA incorporated in SSEE

The NSSA band selection process can be incorporated in the SSEE process. As illustrated schematically in Figure 5.3, the NSSA band selection is conducted in each image subset after completion of the second step of the SSEE process when a small number (usually 3-6) of locally distinctive candidate endmembers are generated. An examination for the NSSA values as a function of wavelength for a number of subsets led us to select a threshold of 50 bands to be selected in each subset, where these bands corresponding to the highest NSSA values. Then, a record of the bands selected for all subset is assembled to generate a frequency histogram of the bands and this histogram can be thresholded to obtain a final set of selected bands (Figure 5.3). With the incorporation of the NSSA method into the SSEE process, bands are selected to contrast the materials in each subset.

5.3.2 Description of the experimental design

This section describes the experimental design, illustrated as a flowchart in Figure 5.4, to produce three sets of results namely: 1) selected bands from the integration of NSSA into the SSEE process, 2) ensuing endmember clusters, and 3) abundance maps following unmixing. The methods of evaluation of these datasets are described in section 5.3.3.

5.3.2.1 Endmember extraction and band selection

Using the AISA imagery shown in Figure 5.2 as input, an endmember set was obtained with SSEE using all bands available. Then two different processing paths were followed, labeled by A and B on the flowchart (Figure 5.4), for endmember clustering to define a final endmember set that is used for unmixing. Path A retains all bands during these steps but in path B, NSSA is integrated into the SSEE process to generate a band set. This band set is used to spectrally resample the endmembers which influences the clustering results, the definition of the final endmembers set and the unmixing results obtained from this spectral subset.

5.3.2.2 Endmember clustering and spectral unmixing

The endmember set derived from SSEE was clustered and labeled to derive final endmember sets for unmixing (Path A and B, Figure 5.4b). For clustering, we used a tree cluster (de Hoon et al., 2003) that recursively merges a pair of clusters based on a similarity measurement. To start, each endmember was treated as an individual cluster; endmembers that are most similar were successively merged. In this study, the Spectral Angle (SA) between two endmembers was the measure of similarity. A threshold of minimum SA was defined to stop the merging process that took place when all pairwise clusters had a similarity greater than the threshold. To address the spectral variability of the extracted endmembers, the tree cluster tool was applied twice on the given data. The first time using all endmembers and a SA threshold of 0.2 (radians) producing clusters that capture the broad material classes, namely vegetation snow, rocks and water (Figure

5.4c). The next level of clustering focused on the geological class to capture subclasses and define multiple geological endmember clusters (Figure 5.4c). In this case, a smaller value of the SA threshold (0.07 in radians) was used because these endmembers are more spectrally similar. Clustered endmembers were then averaged to obtain an individual endmember representing a given class contributing to the final endmember set used for unmixing. The clustering process was generally data-driven, though some endmember classes were removed or integrated with others based on expert knowledge.

Unmixing was conducted using the Iterative Spectral Mixture Analysis (ISMA) method of Rogge et al. (2006). The ISMA, often combined with the SSEE, has been used in several geologic investigations (Rogge et al., 2014; Laakso et al., 2016; Feng et al., 2018). The ISMA was designed to define an optimal per-pixel endmember set that is then used during unmixing.The ISMA can make use of a large input endmember set including endmembers that are spectrally similar as is the case in this study. The optimal per-pixel endmember set is obtained through an iterative unmixing process, is an unconstrained method, and importantly, it can address the per-pixel variability in the type and number of endmembers across the scene.

5.3.3 Evaluation of selected bands and impacts on endmember selection and unmixing results

5.3.3.1 Evaluation of selected bands

The bands selected by NSSA-SSEE were compared to three other band sets. The first set reported by Rogge et al. (2014), and shown in Table 5.1, relied on an expert interpretation of endmember spectra obtained from their AISA image mosaic. The second set consists of bands selected by the Maximum Variance Principal Component Analysis (MVPCA) (Chang et al. 1999)

for the AISA airborne mosaic of this study. This method is selected for comparison because, it is based solely on spectral information and it is unsupervised, meaning that there is no need for collecting labeled samples. It is also easily implemented (Chang & Wang, 2006; Jia et al., 2012; Torbick & Becker, 2009). In MVPCA, bands are ranked by the loadings of eigenvectors calculated from the principal component analysis. The last band set was reported by Long et al. (2019) and obtained by applying the NSSA band selection method to the average spectra obtained from field samples of the six mafic and ultramafic rock types presented in section 5.2.2.1. Of the 2151 bands available with the field spectrometer, 2001 bands had been retained spanning the spectral range of 0.426-2.426 μm, the rest excluded due to poorer system responsivity. From this number, a set of 265 bands encompassing three primary spectral regions were selected to best discriminate these rocks (Long et al., 2019): 0.436-0.493, 0.586-0.677, 0.700-0.713; 1.391-1.404; 2.256-2.386, 2.405-2.411 μm. Due to the high spectral resolution of the data, we only list the selected regions. The selected bands encompass the spectral features of the key rock-forming minerals described in section 5.2.2.1. Note that these bands were selected from data with a 1 nm band sampling interval, a resolution considerably higher than that of the AISA airborne imagery, therefore disparities in the fine detail of band selection results can be expected. The three band sets were compared for agreement with an expectation that they would capture features that discriminate key materials prevailing in the scene, including the key rock types.

5.3.3.2 Comparison of endmember clusters

To evaluate the impact of band selection on endmember clustering, differences in broad classes and subclasses of the two clustered endmember sets (with and without band selection) were compared. As the band selection attempts to enhance the discrimination of spectrally similar targets, particular attention is given to the changes in endmember clusters representing rock types.

5.3.3.3 Comparison of unmixing results

To assess the impact of band selection on spectral unmixing, differences in abundance maps were examined. A priori geologic knowledge on the spatial distribution of key rock types was used to assist the interpretation of these abundance maps. This knowledge came in the form of an existing geological map (Figure 5.2), exploration maps not available for publication, and observations collected in the field during two field campaigns.

5.4 Results

5.4.1 Bands selected by NSSA-SSEE

Figure 5.5 displays the frequency histograms of the tabulated bands selected using the NSSA-SSEE process for k values of 0 and 1. The patterns displayed by both histograms are similar, and thus they highlight similar spectral regions though differences are apparent in the specific spectral range of regions. Note that similar patterns were observed but not used for histograms of higher k values examined namely values of 3 and 7. A final band set of 67 bands (Table 5.2) for the AISA data was obtained by combining all bands above the threshold in each histogram. The threshold was defined as the mean value of selected times plus 50% of the standard deviation of the mean. Bands above the threshold, which is represented by a dotted line on Figure 5.5 (4089 for k=0 and 2687 for k=1), were retained.

Figure 5.6 was assembled to aid in the evaluation of the selected bands. The top portion of the figure incorporates type spectra of the dominant surface types in the AISA imagery including water, vegetation, and rock to which was added a spectrum of gossan. Below these are the band sets selected from the AISA imagery by Rogge et al. (2014) (Table 5.1), by MVPCA, and by the

NSSA SSEE method. Also shown in the lower portion of Figure 5.6 are the average spectra of the six dominant rock types determined from rock samples (Figure 5.1) and the corresponding band sets determined using the NSSA method.

As seen in Figure 5.6, the bands selected by NSSA-SSEE occur primarily from 0.4-0.65 μm in the VNIR and 2.0-2.4 μm in the SWIR. The regions encompass almost all of the bands selected from the average spectra of lichen covered rock samples by NSSA. This indicates that absorption features and variations of the continuum, associated with pigments in lichen and key rock-forming minerals are captured by the NSSA-SSEE band set. The VNIR NSSA-SSEE band set is more extensive than that of the field samples also capturing diagnostic features of non-geological materials in the AISA imagery such as vegetation (0.4, 0.5, 0.6, 1.1 μm) and water (0.5 μm). This band set also includes a few contiguous bands near 1.3 and 1.7 μm. Overall there is good concordance between the NSSA-SSEE band set and that selected by Rogge et al. (2014) with the intent to differentiate broad cover classes and distinguish the main rock units in the presence of lichen. Note the expanded set of NSSA-SSEE SWIR bands towards shorter wavelengths in comparison with the band set obtained from rock samples. This likely reflects the variability in lichen abundance in the scene and the need to capture the broad feature of lichens centered near 2.1 μm as well as variability in water content in vegetation that would influence the slope of the continuum.

Lastly the band set selected by MVPCA encompasses four spectral regions: 0.42-0.60, 0.71-0.74, 1.52-1.79, 2.18-2.30 μm. It misses three important spectral features of key minerals beyond 2.30 μm, the selection process being biased by the presence of lichen and an associated broad feature short of 2.3 μm. However, most of the bands selected by MVPCA occupy the region of 1.52-1.79 μm. This specific result stands out and is not consistent with the band sets obtained

from other methods where only a few bands in this region are selected to capture the continuum of spectra. In the context of geology, an extensive suite of bands here is difficult to explain and to assign value since there are no spectral features related to mineralogy in this wavelength region.

5.4.2 Endmember clusters with and without band selection

Two endmember sets, each with 221 endmembers, one without band selection (Path A, Figure 5.4a) and the other with NSSA-SSEE band selection (Path B), were clustered and labeled to obtain final endmember sets. Clustering of the endmembers without band selection resulted in 18 clusters including 12 labeled as geological and 6 for non-geological materials. Clustering of endmembers with band selection resulted in 21 clusters including 14 manually labeled as geological and 7 for non-geological materials.

5.4.2.1 Clusters of non-geological materials

Clusters for snow and water remained unchanged following band selection. Clustering for the vegetation class varied slightly after band selection resulting in one additional vegetation cluster. Thus, more spectral variability was captured for vegetation endmembers using band selection.

5.4.2.2 Clusters of geological materials

In the subsequent re-clustering of geological endmembers, eight clusters were not impacted by band selection (e.g., identical clusters, Figure 5.7). These clusters represent rock types for ultramafic (2), mafic (3), and gossan (3). Figure 5.8 provides example endmember spectra to illustrate the labeling of geological clusters. Gossan endmembers (e.g., C1 and C3) were

characterized by a rapid increase in visible reflectance to approximately 0.8 μm (Figure 5.8a) followed by relatively uniform reflectance to about 2 μm. The removal of bands in AISA data from 0.87-1.08 μm due to poor signal quality precludes the clear detection of a broad feature due to the presence of iron oxides although the beginning of a feature can be seen short of 0.87 μm. However, these gossans are also characterized by the lack of spectral features in the SWIR in comparison to mafic and ultramafic rocks (e.g., C7 and C8 Figure 5.8b), which also indicates that they are thick gossans and not simply iron oxide coatings on a rock substrate. The abundance maps show that C1 and C4 highlight exposures along drainage or along cliffs while that of C3 marks the location of known gossans that occur in the area.

Endmembers for mafic (C5, C8, C9) and ultramafic (C2, C7) rocks were labeled on the basis of their distinct spectral shapes that can be linked to their primary mineralogy (cf, section 5.2.2.1). Specifically, an observed singlet absorption, nominally centered near 2.31 μm for ultramafics, and a doublet absorption nominally centered at 2.28 and 2.33 μm for mafic rocks, resulting in flatter spectra in this region (Figure 5.8b and 5.8c). The exact position of the features is constrained by the spectral resolution of the data as seen from the comparison of ultramafic and mafic rock spectra C7 and C8 on Figure 5.8b with laboratory spectra of samples for a full range of mafic and ultramafic rock compositions (Figure 5.8c). For the latter, a subtle absorption shift from 2.325 μm to 2.315 μm can be observed as one proceeds systematically from dunite, peridotite, olivine-pyroxenite and pyroxenite reflecting the progressive increase in the abundance of actinolite at the expense of antigorite. In contrast, mafic rocks (basalt and gabbro) show additional absorptions at 2.285-2.345 μm reflecting the presence of actinolite and clinochlore.

The remaining clusters (C6, C10, C11, C12) were altered by band selection in two ways as detailed in the next section: 1) a cluster was split into two clusters (e.g., split clusters, Figure 5.7)

or, 2) a re-arrangement of endmembers took place between two clusters (e.g. re-arranged clusters, Figure 5.7).

5.4.3 Characteristics of clusters altered by band selection and related abundance maps

5.4.3.1 Split clusters and their abundance maps

Figure 5.9 presents the first of two cases where a cluster is split into two clusters following band selection, namely the separation of two endmembers from cluster C10 into clusters C'5 and C'7. These represent ultramafic rocks with spectra showcasing a displacement of their primary SWIR feature from 2.325 for C'5 to 2.315 μm for C'7 (inset in Figure 5.9). This is consistent with the range of band position displacement observed for a range of ultramafic rock composition (Figure 5.8c) and is significant for the detailed distinction of such rocks during mapping. Figure 5.10 displays the abundance maps for the three cluster endmembers and two detailed inset images. These show that the two endmembers obtained after band selection result in abundances that define continuous patterns that are spatially complementary. This is consistent with the observation in the field of compositional variability often consisting of layering or the formation of border zones in ultramafic sills.

For the second case, the four endmembers of cluster C11 were split into clusters C'8 and C'9, each containing two endmembers (Figure 5.11). These represent mafic gossans showcasing typically flat spectra beyond 0.8 um attributable to iron oxides in gossans. Band selection captures a key distinction shown in the SWIR inset of Figure 5.11 whereby endmembers of cluster C'8 are typical of mafic rocks (see Figure 5.8b) whereas endmembers of C'9 likely signify mafic rocks with deeper gossan weathering. Figure 5.12 displays the abundance maps for the three cluster

endmembers and two detailed inset images. The insets illustrate that the two clusters obtained after band selection result in abundances that define continuous patterns and likely reflects the variable development of gossan surfaces. In most instances, the distribution of these mafic gossans is along drainage and is thus of lesser significance for geological exploration.

5.4.3.2 Rearranged endmember clusters and their abundance maps

Figure 5.13 presents a re-arrangement of endmembers between two clusters after band selection. All endmembers belonging to cluster C6 (no band selection) are assigned to cluster C'13 after band selection and are supplemented by a number of endmembers from cluster C12 while the remaining endmembers of C12 are assigned to C'11.

Prior to band selection, the differing spectral characteristic of the two clusters (C6 and C12) is primarily the slope of the spectra between 0.4-0.7 µm and to some extent to 1µm best explained by variability in the species composition of the lichen community (e.g., color) and lichen cover on the ultramafic rock substrate. In this respect, spectra in C12 have lower reflectance, however, the SWIR inset for C12 spectra shows a mixed population consisting of two groups that are better differentiated by clusters after band selection (Figure 5.13). Spectra for C'11 all display a singlet absorption characteristic of the most ultramafic rocks while several spectra of C'13 display two absorptions and band centers at shorter wavelengths indicative of a less ultramafic character (refer to Figure 5.8c for a continuum of examples). Together these capture ultramafic map units in regions where we have no field observations and thus the significance of the observed map patterns cannot be assessed.

5.4.4 Compilation of mafic and ultramafic rock endmembers following band selection

Following the band selection with NSSA-SSEE, a final set of five mafic and five ultramafic geologic endmembers was compiled, among a total of twenty-one endmembers. Figure 5.14 displays the absorption features observed in the SWIR region of 2.2-2.4 μm (with continuum removal) for the two sets of endmembers. Mafic rocks, as they contain clinochlore, have an overall flat and broad feature with common absorptions centered at 2.288 μm and 2.326 μm (Figure 5.14a). Ultramafic rocks, however, are characterized by a deep and narrower absorption with a center that shifts from 2.309, 2.314 to 2.326 μm (Figure 5.14b). Similar displacements are observed for spectra of field samples (2.315-2.325 μm) (Figure 5.14c) indicating that the compositional variability seen in rock samples is also captured by image endmembers after band selection. These subtle spectral differences relate to changes in the modal abundance of antigorite and actinolite and were detected via band selection that enables the fine discrimination of the various ultramafic rocks.

The results are significant for mapping and capture lithologies that have spatial continuity as one would expect from the prior mapping of differentiated ultramafic sills in the area. This can be seen in Figure 5.15 showing a composite of abundance maps for C'5-C'7-C10 (RGB) that represents the most ultramafic composition (C'5) to progressively less ultramafic compositions as inferred from spectra shown in Figure 5.14b. The detailed discrimination of mafic and ultramafic rocks is an outcome of the automated band selection and endmember extraction using the NSSA-SSEE method. Note that C'7 delineates the largest abundance of ultramafics reported in prior maps (c.f. Figure 5.2) and C'5 adds detail to this map. However C'5 does not exclusively highlight ultramafic rocks and also highlights some drainage, as seen in the eastern portion of the map (Figure 5.15). This is perhaps not surprising given the more complex spectral character of this endmember that encompasses a wider spectral region as seen on Figure 5.14b.

5.5 Discussion

5.5.1 The incorporation of NSSA into SSEE

This study presents a conceptual framework and then demonstrates the use of a band selection tool incorporated into the process of spatial and spectral endmember selection. The incorporation aims to define a band set that improves the spectral contrast between endmembers at each step of the spatial-spectral endmember search and ultimately captures key features for discriminating spectrally similar materials. Most band selection methods in the literature are designed for classification purposes, and they are often based on classification accuracy or similarity measures. The NSSA-SSEE is a unique band selection method that is specifically designed to improve the detailed determination of endmembers and thus spectral mixture analysis.

The incorporation has resulted in an improvement in the selection of detailed endmembers that are similar and significant for mapping. This can be because: (1) bands are selected during the successive analysis of spatial subsets and thus reflect many endmember sets across the scene; and (2) bands are selected from endmembers that are locally distinctive, and thus bands can be selected that contrast similar endmembers. The incorporation of NSSA into SSEE is feasible because both methods are well suited for this process. NSSA is one of the few methods of band selection that is suitable for the analysis of a small number of endmembers. SSEE provides such endmember sets via spatial subsetting. While the incorporation of other methods into the approach given here is possible, NSSA and SSEE are particularly suited for this task.

5.5.2 Guidelines for the use of the NSSA-SSEE method

From the experiences in this study, we provide some recommendations for defining the three key parameters that are inherited from the NSSA and SSEE methods namely: the band interval (k), which captures features of varying width; the number of selected bands (n) from each spatial subset; and the size of the spatial subset window (w). Limits on the range of the first two parameters can be established by testing several band selections using NSSA on randomly collected endmember sets (e.g., a sample of spatial subsets). Also, our prior work with NSSA has shown that the k value can be constrained based on the maximum width of spectral features of interest in the data analyzed, and intervals with doubling window size (0,1,3,7,15…) are recommended to start the selection process (Long et al. 2019). The n value can be determined by analyzing the NSSA profiles as a function of wavelength for the sample of spatial subsets. The w parameter is inherited from the SSEE method of Rogge et al. (2007), and the author notes that the number of pixels (width) for a given subset must be greater than the square root of the number of bands. The maximum size is optional and can be chosen based on the spatial distribution of materials in the scene.

The final band set selected using the NSSA-SSEE method is accomplished from the analysis of the band frequency histogram. Understanding the output of the band frequency histogram can help the users to better exploit this method in that the appearance of the histogram reveals the spectral and spatial information of the hyperspectral image. The selected bands (high-frequency values) from the histogram relate to the material make-up and their abundance in the image. For instance, the image in this study encompasses many pixels dominated by rock and vegetation, thus more bands were selected that capture spectral features of these two materials. However, if targets of interest are sparsely distributed in the image, for example, if they constitute

less than 5% of the total pixels, the band selection is unlikely to capture their key spectral features. This is simply because the decision process for the selection of the final band set rests on a threshold applied to the band frequency histogram. A simple threshold may not work for objects with minimal spatial distribution and this is a problem that needs to be considered when using the method.

Another key aspect for users also relates to the spectral characteristics of their input image data. The method is of lesser value for scenes that encompasses high-contrast targets since the NSSA-SSEE was designed with the intent to capture bands that enhance the differentiation among spectrally similar endmembers. In this study, the impact of band selection by NSSA-SSEE on the clustering of broad classes and geologic subclasses is evidence of such a capability. The band selection had a minor effect on the clustering of vegetation, snow, and water but impacted the clustering of the geologic targets.

5.6 Conclusions

This study incorporates the N-dimensional Spectral Solid Angle (NSSA) band selection tool into the Spatial-Spectral Endmember Extraction (SSEE) tool to define a band set that can be used to better define endmembers classes used in Spectral Mixture Analysis (SMA). The proposed method (NSSA-SSEE) was evaluated for lithological mapping using a hyperspectral image encompassing a range of spectrally similar mafic and ultramafic rock units. The band selected by NSSA-SSEE showed a good agreement with known features of scene components identified by experts, and it improved the selection of detailed geological endmembers and thus the estimation of their abundance maps. These results demonstrated the feasibility, effectiveness of NSSA-SSEE

and its advantages over a band selection method that does not use the spatial context and endmember variability in hyperspectral imagery. The method is particularly useful for the discrimination of spectrally similar materials. With the NSSA-SSEE method, the band selection process becomes automatic which makes the spectral mixture analysis less dependent on expert knowledge for feature identification. This is significant for applications that lack field-based information and prior knowledge of the study area.

Acknowledgments

Table 5.1 Bands selected in Rogge et al. (2014)

Range	Band location (µm)
VNIR	0.454, 0.500, 0.546, 0.603, 0.651, 0.698, 0.746
SWIR	1.659, 1.710, 1.735, 1.772, 2.112, 2.150, 2.188, 2.200, 2.213, 2.225, 2.238, 2.251, 2.263, 2.276, 2.288, 2.301, 2.314, 2.326, 2.339, 2.351, 2.364, 2.376 2.389, 2.402, 2.414, 2.427

Table 5.2 The final band set obtained from the NSSA-SSEE (combined band sets for k=0 and 1).

Range	Bands (µm) selected by NSSA-SSEE
VNIR	0.445, 0.454, 0.463, 0.472, 0.482, 0.491, 0.5, 0.509, 0.518, 0.528, 0.537, 0.546, 0.555, 0.565, 0.574, 0.584, 0.594, 0.603, 0.613, 0.622, 0.632, 0.641, 0.651, 0.66, 0.814, 0.824, 0.834, 0.853, 1.081, 1.093, 1.106, 1.118, 1.131, 1.307, 1.332
SWIR	1.772, 1.785, 1.798, 2.012, 2.024, 2.037, 2.049, 2.062, 2.074, 2.087, 2.1, 2.112, 2.125, 2.137, 2.15, 2.163, 2.175, 2.188, 2.2, 2.213, 2.225, 2.238, 2.251, 2.263, 2.276, 2.288, 2.301, 2.314, 2.326, 2.339, 2.351, 2.364
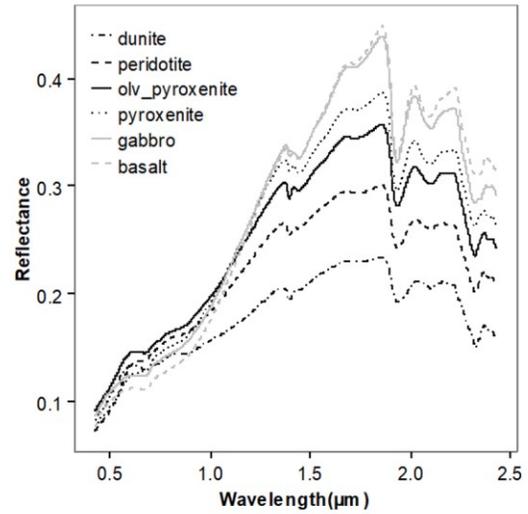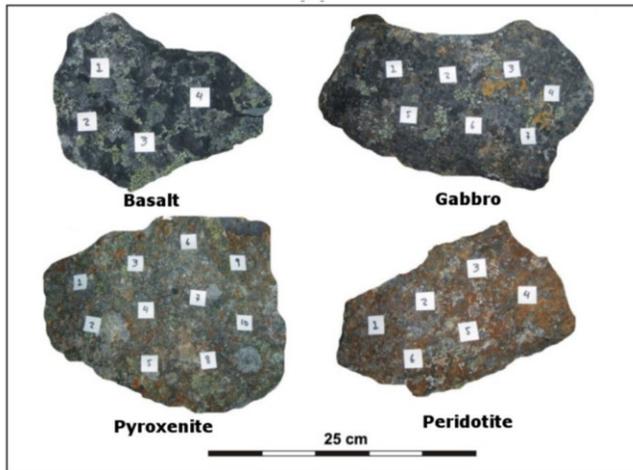
**Figure 5.1** Photographs of rock samples for four of the six dominant rock types (note the extensive lichen cover) (a), and average spectrum derived from point spectra for each of the six rock type (b). Modified from Rogge et al. (2014).
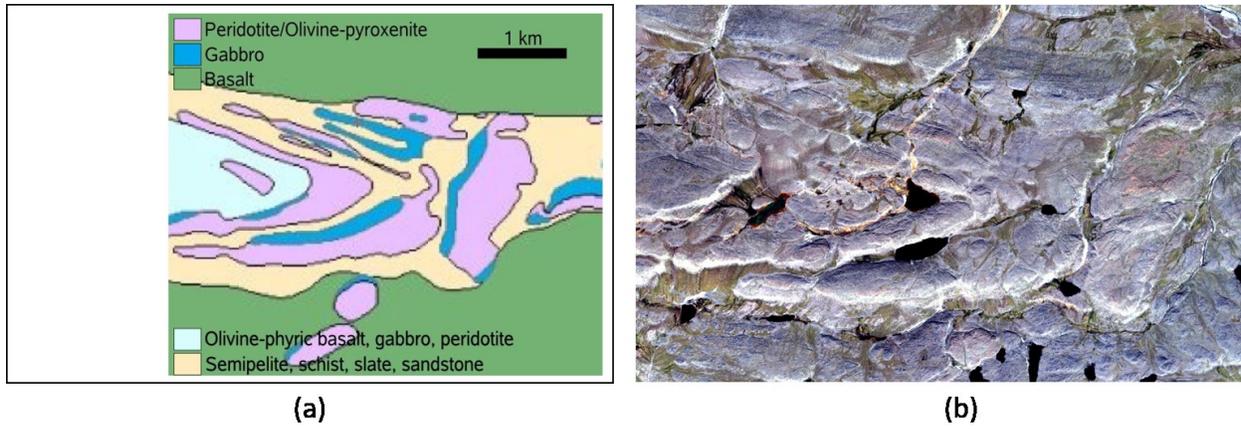
Figure 5.2 Geologic map from the Geological Survey of Canada (St-Onge and Lucas, 1993) (left), and AISA imagery true color RGB (b). Modified from Rogge et al. (2014).
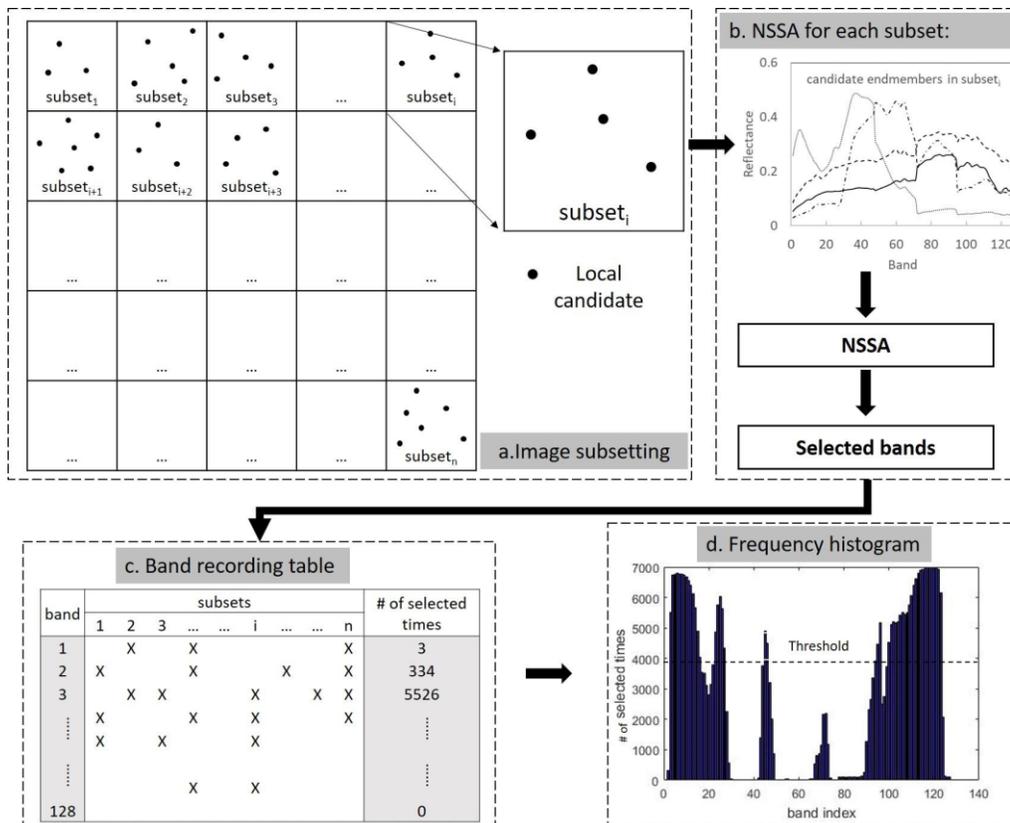
Figure 5.3 The conceptual framework of NSSA incorporated in SSEE. (a) The image is spatially divided into equal size non-overlapping subsets (1 to n), and a small number of local candidate endmembers are generated for each subset. (b) A band selection using the NSSA method is applied to the endmembers in each subset to obtain a fixed number of bands (from $subset_1$ to $subset_n$). (c) The bands selected from all subset are recorded in a table where the x marks selected bands. (d) A frequency histogram of bands is thresholded to obtain a final set of bands.
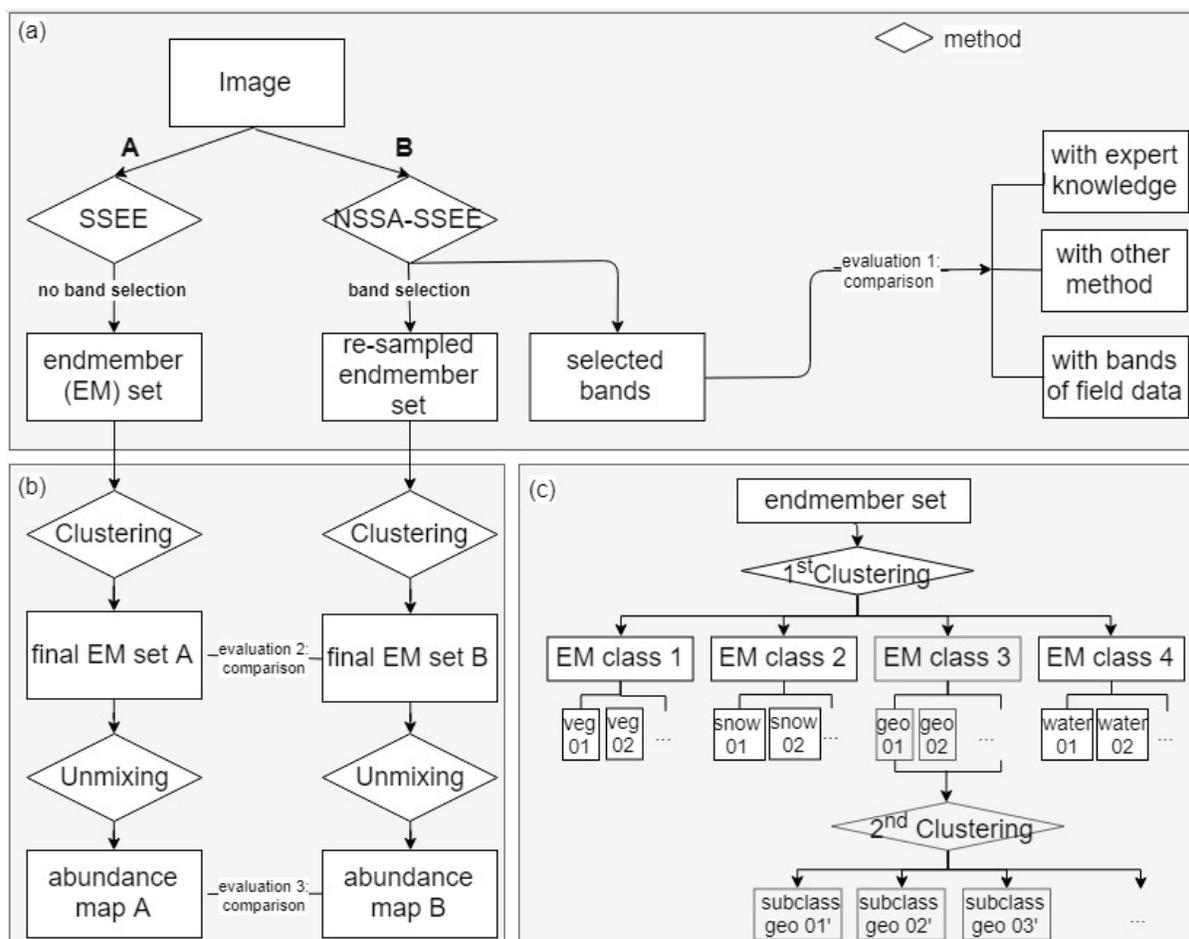
Figure 5.4 Flowchart of the experimental design. With the AISA imagery as input two processing paths are followed (A and B) involving the extraction of the same endmember set with SSEE using all bands available (part a), followed by endmember clustering to define a final endmember set that is then used for unmixing (part b). However, in path B, NSSA is integrated into the SSEE process to generate a band set. This band set is used to resample the endmembers thus influencing the clustering results, the definition of the final endmembers set and the unmixing results. A hierarchical clustering process is used to define the broad classes and geological subclasses (part c).
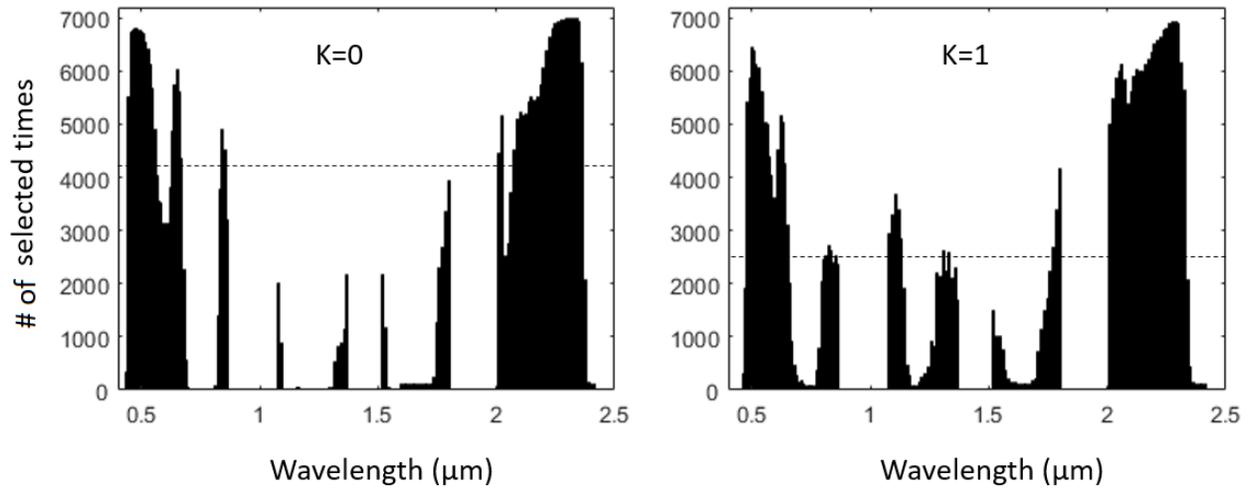
Figure 5.5 Frequency histogram of bands obtained from the NSSA SSEE process for all spatial subsets for k=0 and 1. These k values capture spectral features of different widths from the endmembers. The horizontal dotted line in each graph marks the threshold value for band selection that was defined as the mean value of selected times plus 50% of the standard deviation of the mean. Bands above the threshold (4089 for k=0 and 2687 for k=1) were retained.
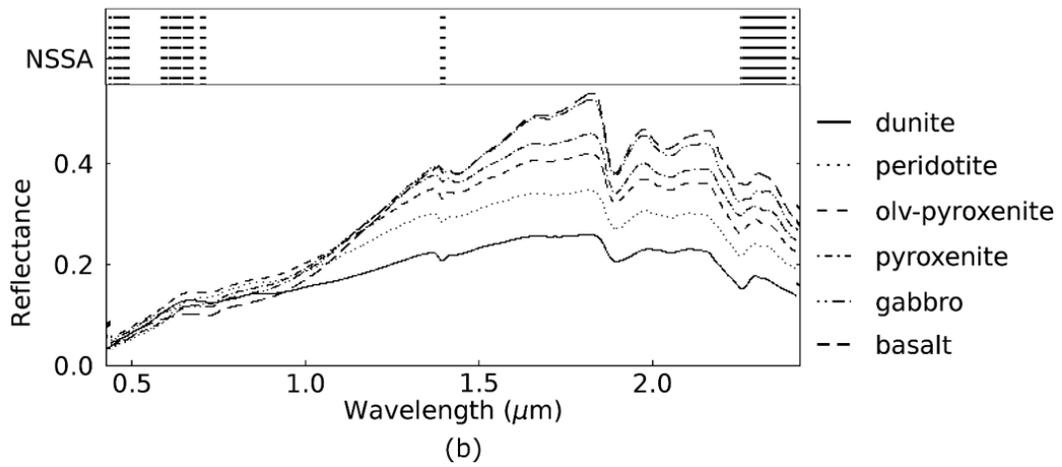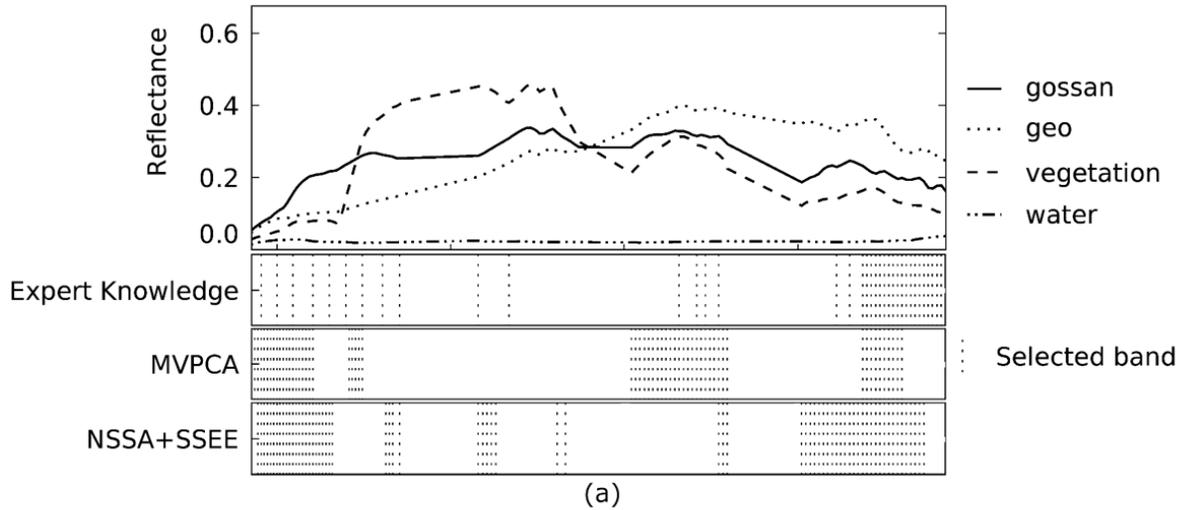
Figure 5.6 Comparison of the AISA imagery band set selected by the NSSA-SSEE method with those selected by Rogge et al. (2014) (expert knowledge) and the MVPCA method. Spectra shown at the top are image endmembers for typical surfaces including gossan (C3 cluster spectra), a geological substrate (C6 ultramafic cluster spectra), the spectrum of one vegetation class and water. Spectra shown in the lower part are the average spectra of the six dominant rock types determined from rock samples (cf Figure 5.1), and the corresponding band sets determined using the NSSA method.
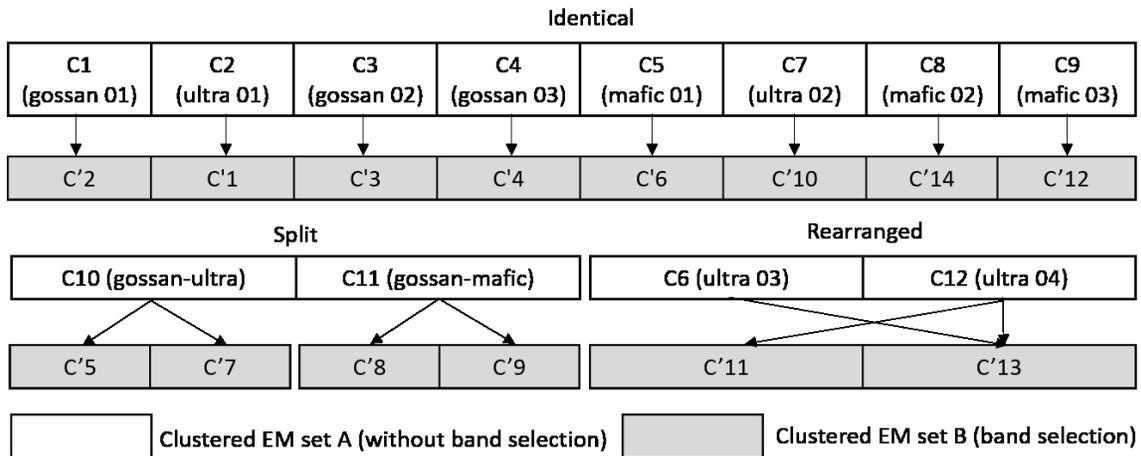
Figure 5.7 Clustering of candidate geological endmembers before and after band selection. About half of the geological clusters are identical in that they contain the same endmembers and are thus unaffected by band selection. Two clusters are split into subclasses, and two clusters undergo a rearrangement into two new clusters. Clusters obtained after band selection are labeled as C'.
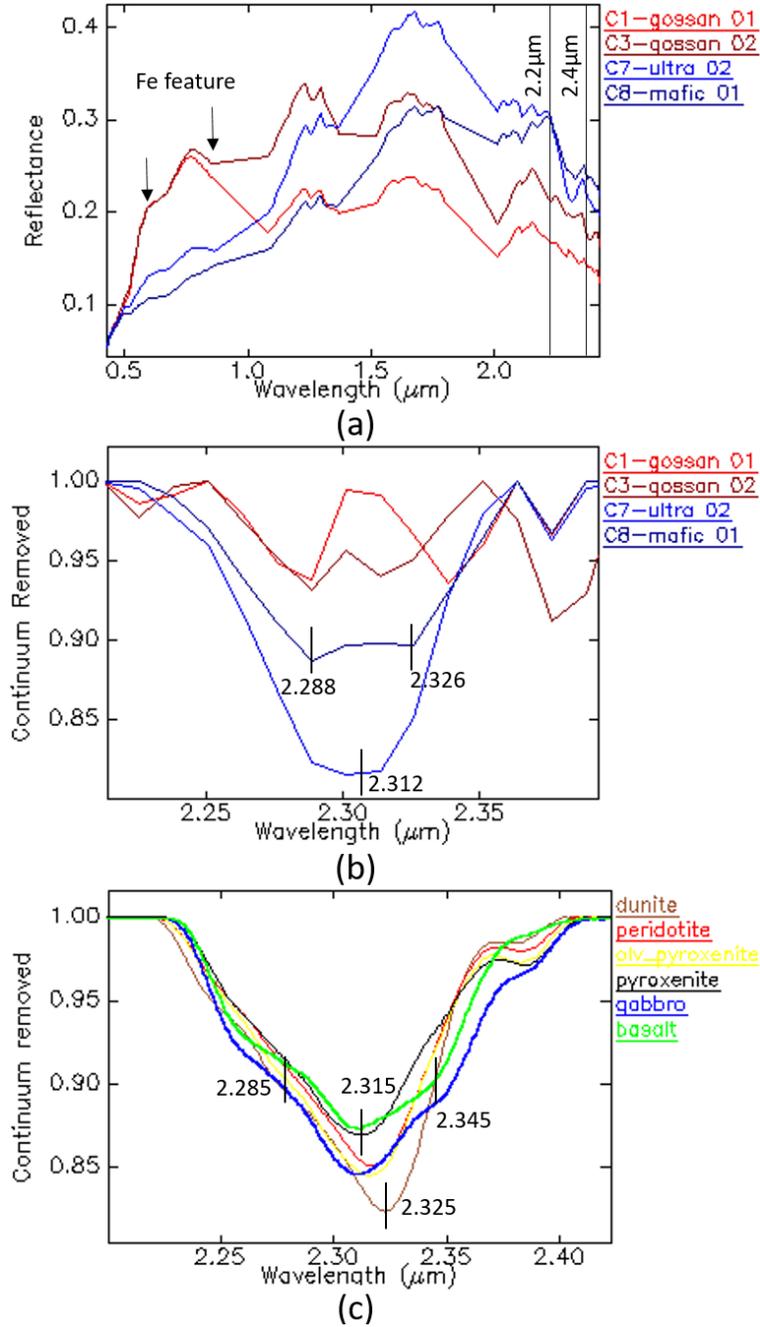
Figure 5.8 Endmember spectra for clusters (Figure 5.7, first row, C1, C3, C7, and C8) obtained from path A (without band selection) and unchanged by band selection (a); enlargement for 2.2-2.4 um showing SWIR features that discriminate mafic and ultramafic endmembers (b); enlargement for type spectra shown in Figure 5.1 obtained from mafic and ultramafic rock samples (c).

145
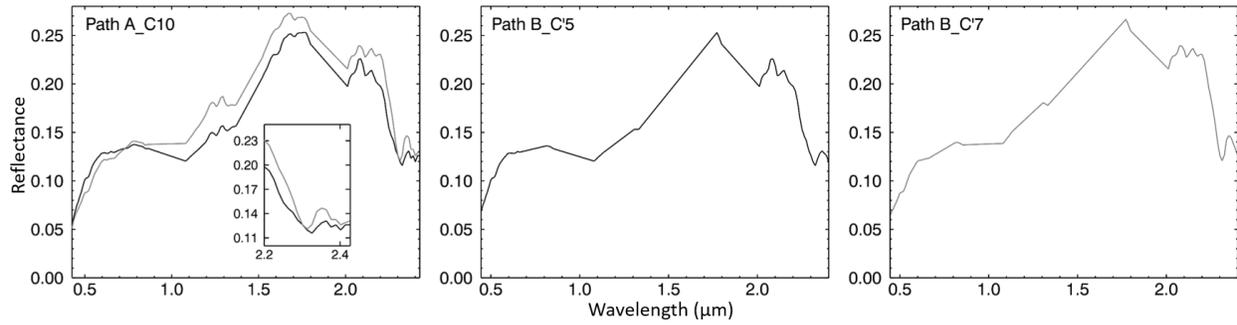
Figure 5.9 Splitting of ultramafic gossan endmembers from Cluster C10 into clusters C5 and C7 after band selection. Grey and black lines are used to facilitate visualization of spectra into their separate clusters.
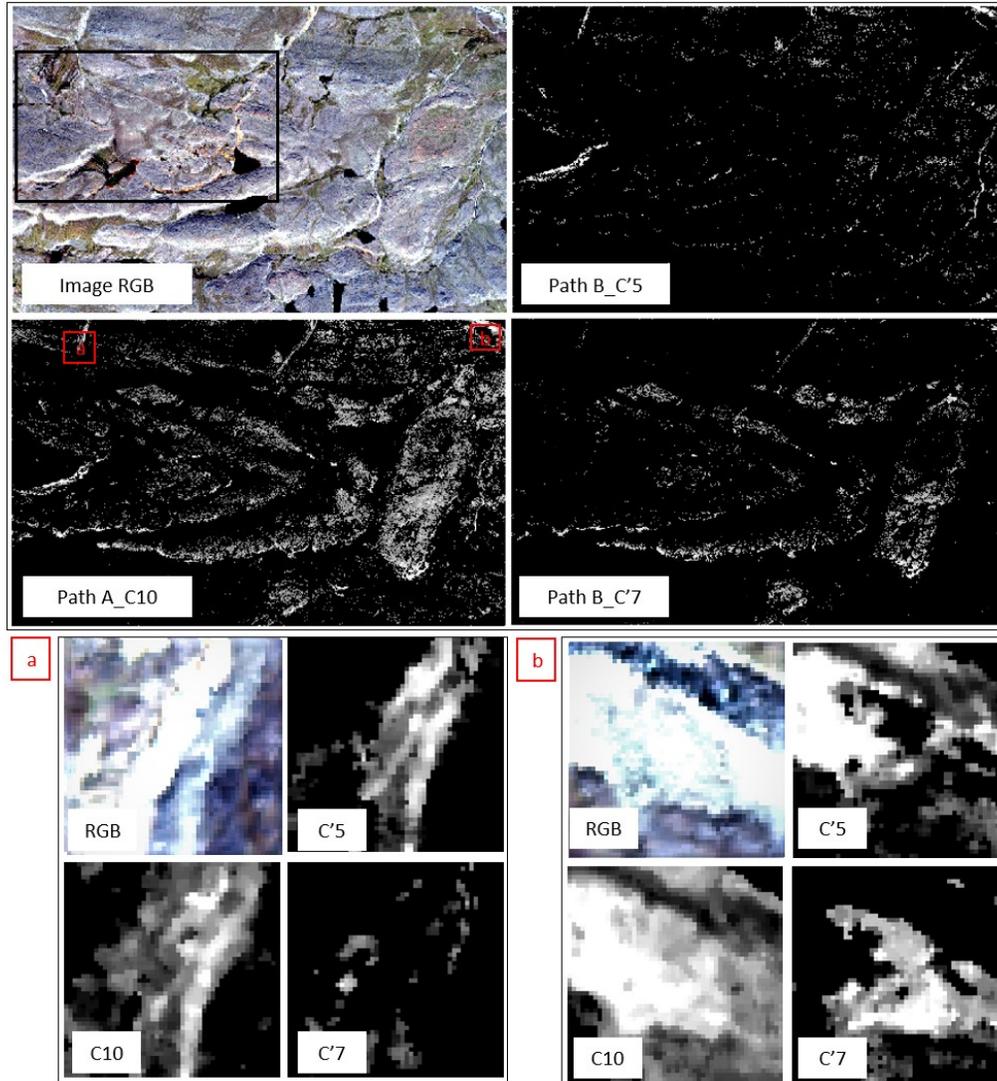
Figure 5.10 Abundance maps for the endmember resulting from cluster C10 and corresponding endmembers resulting from its splitting into clusters C'5 and C'7 following band selection. A true color image in the upper left provides spatial context. The location of insets a and b are shown on the abundance map of cluster C10. The insets illustrate that the two endmembers obtained after band selection result in abundances that define continuous patterns that are spatially complementary consistent with the observation in the field of layering or border zones in ultramafic sills. The black box in the upper left RGB image defines the area covered by Figure 5.12.
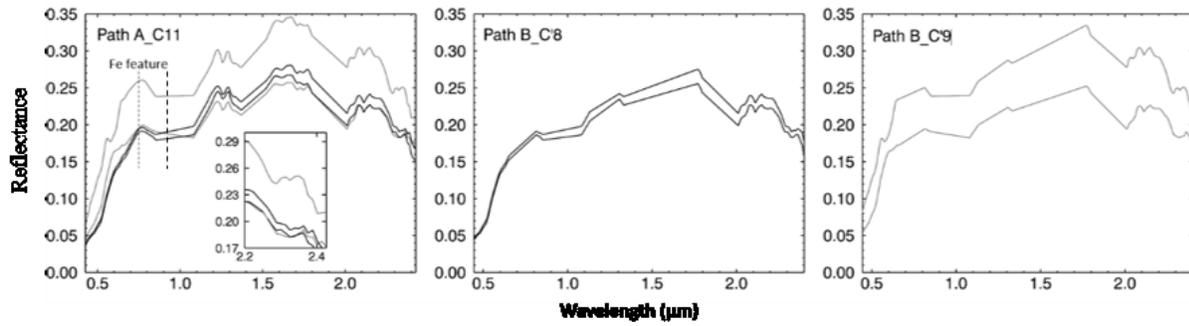
Figure 5.11 Splitting of mafic gossan endmembers from Cluster C11 into clusters C8 and C9 after band selection. Grey and black lines are used to facilitate visualization of spectra into their separate clusters.
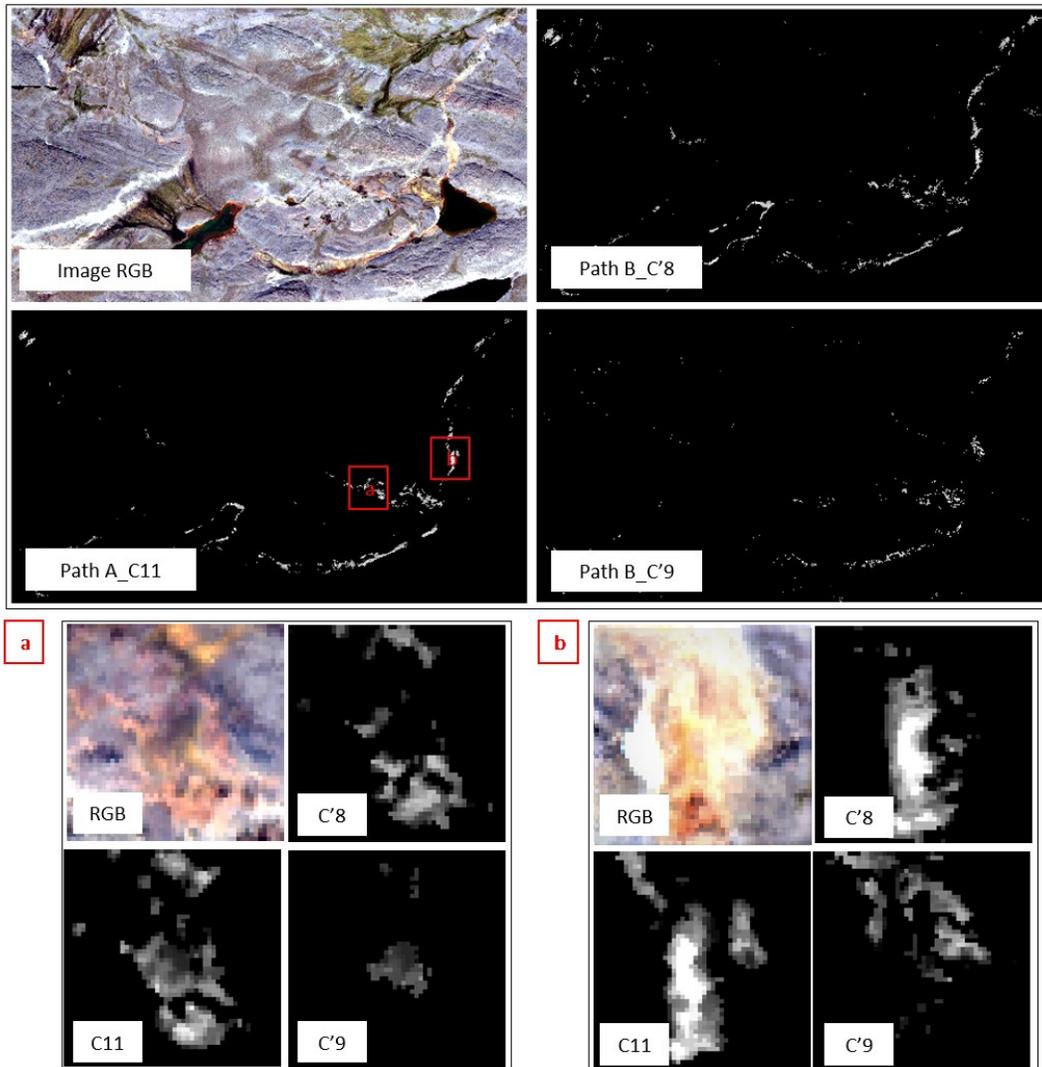
Figure 5.12 Abundance maps for the endmember resulting from cluster C11 and corresponding endmembers resulting from its splitting into clusters C'8 and C'9 following band selection. A true color image in the upper left provides spatial context. The location of insets a and b are shown on the abundance map of cluster C11. The insets illustrate that the two endmembers obtained after band selection result in abundances that define continuous patterns and likely reflects the variable development of gossan surfaces. In most instances, the distribution of these mafic gossans is along drainage and is thus of lesser significance for exploration.
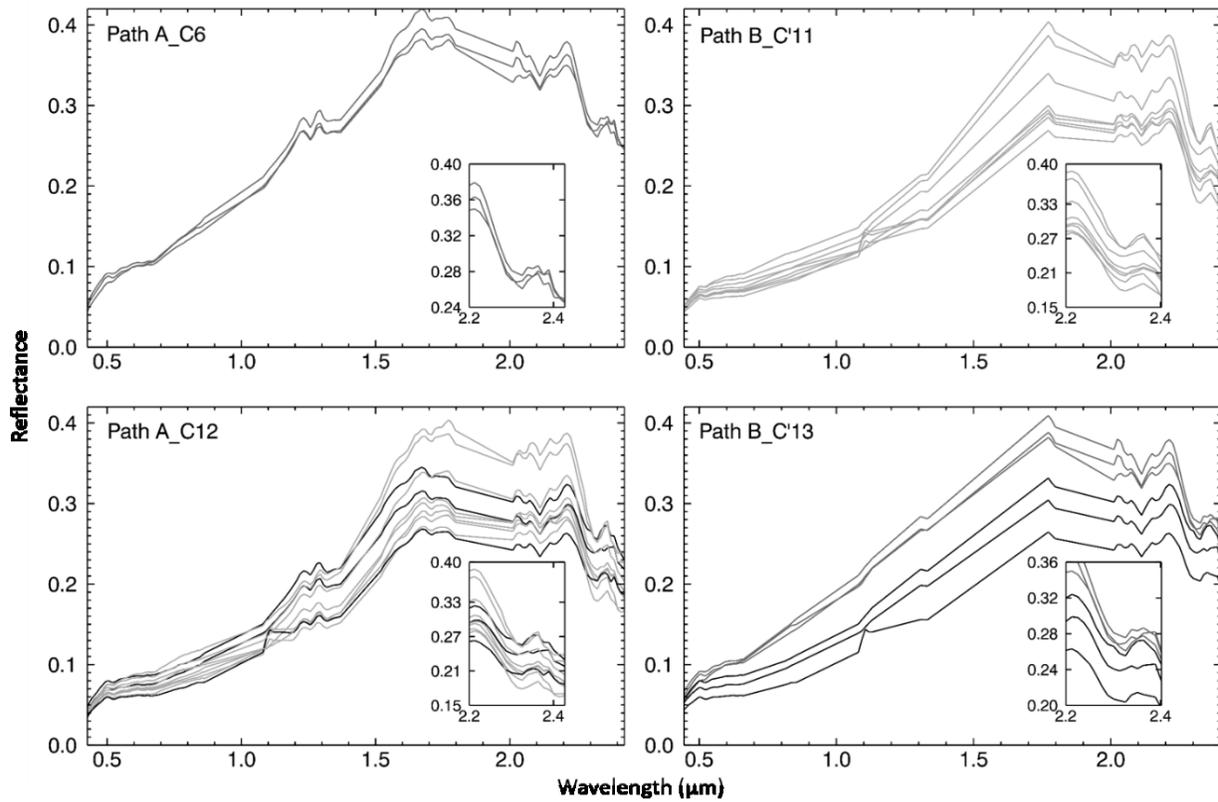
Figure 5.13 Endmembers of rearranged clusters (c.f. Figure 5.7). Clusters C6 and C12 were obtained without band selection, and their endmembers are rearranged into clusters C11 and C13 after band selection. In each case, grey and black lines are used to facilitate the visualization of spectra into their separate clusters.
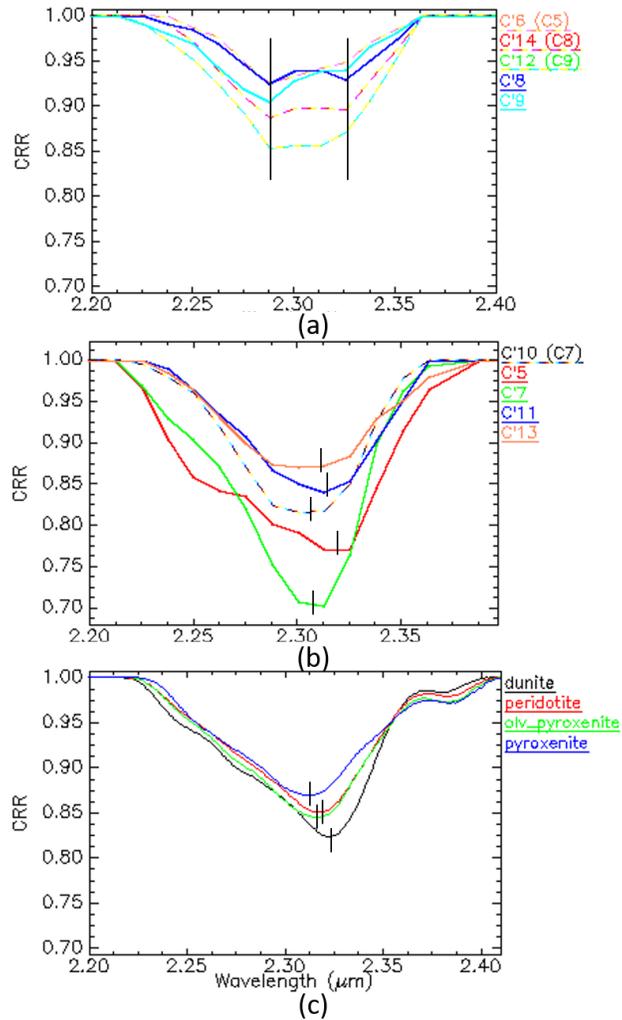
Figure 5.14 Compilation of cluster spectra for mafic (a) and ultramafic (b) lithologies and their key absorption features. Spectra of mafic rocks show features centered at 2.288 μm and 2.326 μm while those of ultramafic rocks show a relatively narrow absorption shifting from 2.309 to 2.326 μm. The latter compares well with the absorption displacement observed from spectra of field samples (2.315-2.325 μm) indicating that the compositional variability seen in rock samples is also captured by image endmembers after band selection (c). The vertical lines label the center of the absorption. Thicker lines represent clusters obtained after band selection. CRR = continuum removed reflectance.
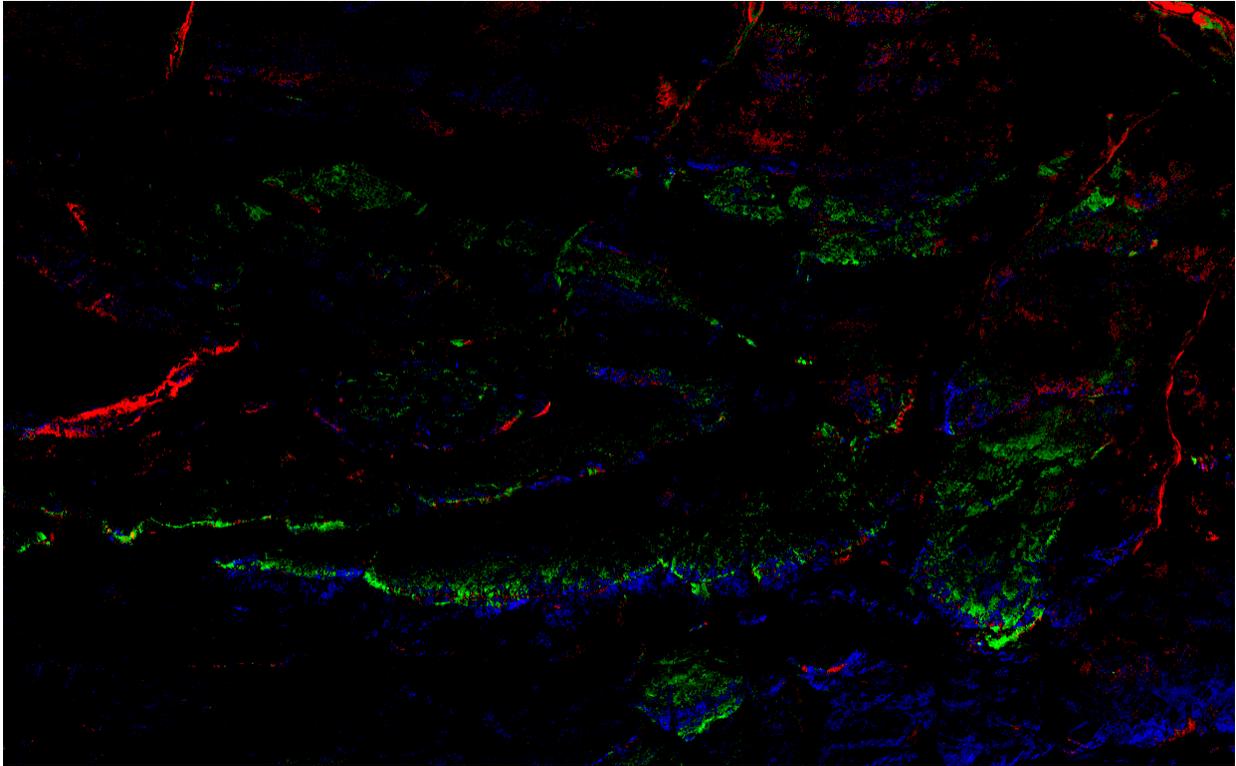
Figure 5.15 Composite of abundance maps for C'5-C'7-C'10 (RGB) representing the most ultramafic composition (C'5) to progressively less ultramafic compositions as inferred from spectra shown in Figure 5.14b.

References

Adams, J. B., 1993. Imaging spectroscopy: Interpretation based on spectral mixture analysis. Remote geochemical analysis: Elemental and mineralogical composition, 145-166.

Asner, G. P., and Heidebrecht, K. B., 2002. Spectral unmixing of vegetation, soil and dry carbon cover in arid regions: comparing multispectral and hyperspectral observations. International Journal of Remote Sensing 23(19): 3939-3958.

Asner, G. P., and Lobell, D. B., 2000. A biogeophysical approach for automated SWIR unmixing of soils and vegetation. Remote Sensing of Environment 74(1): 99-112.

Barnes, S.J. and Lightfoot, P.C., 2005. Formation of magmatic nickel-sulfide ore deposits and processes affecting their copper and platinum-group element contents. Economic Geology 100th Anniversary: 179-213.

Bechtel, R., Rivard, B. and Sánchez-Azofeifa, A., 2002. Spectral properties of foliose and crustose lichens based on laboratory experiments. Remote Sensing of Environment 82(2-3): 389-396.

Chang, C. I., Du, Q., Sun, T. L., and Althouse, M. L., 1999. A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing 37(6): 2631-2641.

Chang, C. I., and Wang, S., 2006. Constrained band selection for hyperspectral imagery. IEEE Transactions on Geoscience and Remote Sensing 44(6): 1575-1585.

Clarke, R.N., 1999. Spectroscopy of rocks and minerals, and principles of spectroscopy, Remote Sensing for the Earth sciences. Manual of Remote Sensing 3: 3-58.

Clark, R., King, T., Klejwa, M., and Swayze, G., 1990. High spectral resolution reflectance spectroscopy of minerals. Journal of Geophysical Research 95: 12653-12680.

Feng, J., Rogge, D., and Rivard, B., 2018. Comparison of lithological mapping results from airborne hyperspectral VNIR-SWIR, LWIR and combined data. International Journal of Applied Earth Observation and Geoinformation 64: 340-353.

Jia, S., Ji, Z., Qian, Y., and Shen, L., 2012. Unsupervised band selection for hyperspectral imagery classification without manual band removal. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 5(2): 531-543.

Iordache, M.D., Bioucas-Dias, J.M. and Plaza, A., 2015, June. Potential and limitations of band selection and library pruning in sparse hyperspectral unmixing. In 2015 7th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS) (pp. 1-4). IEEE.

Keshava, N., and Mustard, J. F., 2002. Spectral unmixing. IEEE signal processing magazine 19(1): 44-57.

Laakso, K., Rivard, B., and Rogge, D., 2016. Enhanced detection of gossans using hyperspectral data: Example from the Cape Smith Belt of northern Quebec, Canada. ISPRS Journal of Photogrammetry and Remote Sensing 114: 137-150.

Lesher, C. M., 2007. Ni-Cu-(PGE) deposits in the Raglan area, Cape Smith Belt, New Quebec. Mineral Deposits of Canada: A Synthesis of Major Deposits-Types, District Metallogeny, the Evolution of Geological Provinces, and Exploration Methods. Geological Association of Canada, Mineral Deposits Division Special Publication 5: 351-386.

Long, Y. Q, Rivard, B., and Rogge, D., 2019. Hyperspectral band selection using the N-dimensional Spectral Solid Angle method for the improved discrimination of spectrally similar targets. International Journal of Applied Earth Observation and Geoinformation 79: 35-47.

Lydon, J.W. and Goodfellow, W.D., 2007. An overview of the economic and geological contexts of Canada's major mineral deposit types. Geological Association of Canada, Mineral Deposits Division: 3-48.

Machado, N., David, J., Scott, D., Lamothe, D., Philippe, S., and Gariepy, C., 1993. U-Pb geochronology of the western Cape Smith belt, Canada: new insights on the age of initial rifting and arc magmatism. Precambrian Research 63: 211-223.

Miao, X., Gong, P., Swope, S., Pu, R., Carruthers, R., Anderson, G.L., Heaton, J.S. and Tracy, C.R., 2006. Estimation of yellow starthistle abundance through CASI-2 hyperspectral imagery using linear spectral mixture models. Remote Sensing of Environment 101(3): 329-341.

Parrish, R., 1989. U-Pb geochronology of the Cape Smith belt and Sugluk block, northern Quebec. Geoscience Canada 16(3): 126-130.

Plaza, A., Martinez, P., Pérez, R., and Plaza, J., 2002. Spatial/spectral endmember extraction by multidimensional morphological operations. IEEE Transactions on Geoscience and Remote Sensing 40(9): 2025-2041.

Powell, R. L., Roberts, D. A., Dennison, P. E., and Hess, L. L., 2007. Sub-pixel mapping of urban land cover using multiple endmember spectral mixture analysis: Manaus, Brazil. Remote Sensing of Environment 106(2): 253-267.

Rogge, D. M., Rivard, B., Zhang, J., and Feng, J., 2006. Iterative spectral unmixing for optimizing per-pixel endmember sets. IEEE Transactions on Geoscience and Remote Sensing 44(12): 3725-3736.

Rogge, D. M., Rivard, B., Zhang, J., Sanchez, A., Harris, J., and Feng, J., 2007. Integration of spatial-spectral information for the improved extraction of endmembers. Remote Sensing of Environment 110: 287-303.

Rogge, D. M., Bachmann, M., Rivard, B., and Feng, J., 2012. Spatial sub-sampling using local endmembers for adapting OSP and SSEE for large-scale hyperspectral surveys, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 5(1): 183-195

Rogge, D. M., Rivard, B., Segl, K., Grant, B., and Feng, J., 2014. Mapping of Ni-Cu-PGE ore hosting ultramafic rocks using airborne and simulated EnMAP hyperspectral imagery, Nunavik, Canada. Remote Sensing of Environment 152: 302-317.

Somers, B., Delalieux, S., Verstraeten, W. W., Van Aardt, J. A. N., Albrigo, G. L., and Coppin, P., 2010. An automated waveband selection technique for optimized hyperspectral mixture analysis. International Journal of Remote Sensing 31(20): 5549-5568.

Somers, B., Asner, G. P., Tits, L., and Coppin, P., 2011. Endmember variability in spectral mixture analysis: A review. Remote Sensing of Environment 115(7): 1603-1616.

St-Onge, M. R., and Lucas, S. B., 1993. Geology of the eastern Cape Smith Belt: parts of the Kangiqsujuaq, cratere du Nouveau-Quebec, and Lacs Nuvilik map areas, Quebec (No. 438). Geological Survey of Canada.

Tane, Z., Roberts, D., Veraverbeke, S., Casas, Á., Ramirez, C., and Ustin, S., 2018. Evaluating endmember and band selection techniques for multiple endmember spectral mixture analysis using post-fire imaging spectroscopy. Remote Sensing 10(3): 389.

Tian, M., Feng, J., Rivard, B., and Zhao, C., 2016. A method to compute the n-dimensional solid spectral angle between vectors and its use for band selection in hyperspectral data. International Journal of Applied Earth Observation and Geoinformation 50: 141-149.

Torbick, N., and Becker, B., 2009. Evaluating principal components analysis for identifying optimal bands using wetland hyperspectral measurements from the Great Lakes, USA. Remote Sensing 1(3): 408-417.

Van der Meer, F. D., and Jia, X., 2012. Collinearity and orthogonality of endmembers in linear spectral unmixing. International Journal of Applied Earth Observation and Geoinformation 18: 491-503.

Zortea, M., and Plaza, A., 2009. Spatial preprocessing for endmember extraction. IEEE Transactions on Geoscience and Remote Sensing 47(8): 2679-2693.

Chapter 6 Conclusions and Future Work

6.1 Summary and key contributions

This research explores the use of band selection for improving the discrimination of spectrally similar targets, in applications of increasing degrees of complexities from chapter 2 to 5. A band selection method named the N-dimensional Spectral Solid Angle (NSSA) was first applied to identify features from real datasets of geologic relevance; through the experiments, the performance of the method was evaluated, and guidelines for the use of the method were defined (chapter 2). Following the guidelines, the NSSA band selection method was used in a hierarchical fashion to address the endmember variability among geologic and vegetation materials (chapter 3), then compared and combined with other band selection methods to form an ensemble selection strategy (chapter 4), and lastly incorporated in the process of endmember extraction so that it could be applied to hyperspectral imagery for improved mapping of ground targets (chapter 5). The four chapters all highlight the significance of band selection but from different data analysis and application perspectives.

Chapter 2 titled "Hyperspectral band selection using the N-dimensional Spectral Solid Angle method for the improved discrimination of spectrally similar targets" emphasized the utilization of the NSSA band selection method for realistic applications. It filled a research gap in that few band selection methods were specifically designed for spectrally similar targets and for dataset encompassing a limited number of labeled spectra. A key outcome of this paper was the guidelines that constrain the selection of two key parameters in the NSSA method, which allows non-expert users to exploit this novel band selection tool. Another main result was to demonstrate the effectiveness of the method in detecting subtle features from mineral and rock hyperspectral

datasets. Based on the results, I concluded that the NSSA method is a valuable tool to assist experts in feature identification and can be used to analyze spectral libraries built from the collection of spectra in the field or endmembers from imagery.

Based on the outcomes of chapter 2, chapter 3 titled "Hierarchical band selection using the n-dimensional solid spectral angle method to address inter- and intra-class spectral variability" applied the NSSA method in a hierarchical manner to select bands from targets that showed spectral variability. The hierarchical strategy was proposed because materials of broad classes (e.g., vegetation vs. water) in the upper hierarchy level showed greater spectral contrast than the materials (e.g., different lithologies) in the low hierarchy level. The hierarchical band selection combined the bands selected from the two levels thus capturing meaningful features for the discrimination of all materials. This study encourages peers to pay attention to the characteristics of targets for band selection.

Then, chapter 4 titled "Identification of spectral features in the Longwave Infrared (LWIR) spectra of leaves for the discrimination of tropical dry forest tree species" used an ensemble band selection strategy, which combined the NSSA method with the random forest and mRMR methods, to identify spectral features in LWIR spectra of leaves. This is one of the few studies that applied statistical approaches to select and characterize LWIR features for species-level classification. The method captured leaf features related to constitutional compounds and improved the discrimination of twenty-six tree species in terms of classification accuracy. With the development of longwave hyperspectral imaging systems, the selected features can guide the future mapping of tree species on regional scales. The ensemble strategy for band selection can be adapted to other vegetation studies (e.g., liana and tree species in the tropical dry forest, plant species other ecosystems) for the robust discrimination of spectrally similar species.

Lastly, chapter 5 titled "Incorporating band selection in the spatial selection of spectral endmembers" investigated how band selection could be incorporated at the pixel level in spectral mixture analysis. The NSSA method was incorporated into the endmember extraction tool SSEE to define a band set that improved the selection of detailed endmembers thus improving the lithological detail offered by maps generated from spectral unmixing. The main innovation of this paper was to include the spatial context in band selection while considering the endmember variability in hyperspectral imagery. The NSSA-SSEE method contributes to the automation of the band selection process in spectral mixture analysis. This is a particular benefit for applications of SSEE and ISMA that are highly dependent on expert knowledge for feature identification.

6.2 Future Work

6.2.1 Expand the application scope of NSSA and NSSA-SSEE

This research has demonstrated the effectiveness of the NSSA method for the discrimination of geologic (minerals and rocks) and vegetation (tree species) materials. Future work shall test the method in other fields such as urban and agricultural environments. Furthermore, I would argue that any study involving spectrally similar targets can benefit from this band selection method, as indicated in the discussion of chapter 2. My undergoing tests on two other datasets showed promising results. One is from Faba beans scanned in the laboratory using the SisuROCK hyperspectral imaging system for the purpose of detecting Lygus bug damage. Through feature identification using the NSSA method, the damaged beans were easily differentiated from the healthy beans (Smith et al.). The other is for oil sands samples, whose spectral variations are barely detectable by the human eyes due to their strong spectral similarity (Entezari et al., 2017). The NSSA method showed the potential capability of capturing subtle

features, but the noise in the data must first be minimized which means intervention is required as a preprocessing step. The robustness of the NSSA method for a broad range of applications still requires further evaluations, but thus far all tests have yielded useful results.

The use of the NSSA-SSEE method, which is proposed for band selection from imagery, can also be expanded into fields outside of geology. The NSSA-SSEE has only been evaluated using the hyperspectral imagery from a geologic study with prospects for mineral exploration and regional mapping. Imagery from a diversity of earth surface materials should be used to test the functionality of the method. This is important because the results of the NSSA-SSEE method can be impacted by the material make-up and their abundance in the image, as deduced in chapter 5. The free high-resolution imaging data provided by the National Ecological Observatory Network (NEON) project (Kampe et al., 2010) could be a good candidate for evaluating NSSA-SSEE in ecologic applications.

6.2.2 NSSA method integrated with the spectral unmixing method

Future research can involve the integration of band selection with spectral unmixing to further improve the spectral mixture analysis. The accuracy of the predicted fractional abundances may be improved with the use of the NSSA band selection during the per-pixel selection of endmembers. Therefore, I will examine the Iterative Spectral Mixture Analysis (ISMA) tool (Rogge et al. 2006) that optimizes the endmember sets on a per-pixel basis and reduces the fractional abundance errors. The NSSA band selection can be incorporated into the ISMA process, allowing for endmembers and their bands varying from pixel to pixel for unmixing. This is distinct from the work conducted in this thesis that only addressed the per pixel band selection and ultimately a scene-based band set for the selection of endmembers. Since both the NSSA and the

ISMA methods address the spectral variability among endmembers, their integration can potentially achieve great improvements in the accuracy of spectral unmixing predictions. To test the performance of the integrated tool, the airborne AISA imagery described in chapter 5 can be used. Hyperspectral imagery from a variety of application in agriculture, urban, hydrology and ecology may also take advantage of this integrated tool. To quantitatively evaluate the results of spectral unmixing by the NSSA-ISMA method, an Abundance Map Reference Data (Williams et al., 2017), that offers per-pixel abundance values of each component endmembers, would ideally be available.

References

Entezari, I., Rivard, B., Geramian, M. and Lipsett, M.G., 2017. Predicting the abundance of clays and quartz in oil sands using hyperspectral measurements. International journal of applied earth observation and geoinformation, 59:1-8.

Kampe, T.U., Johnson, B.R., Kuester, M.A. and Keller, M., 2010. NEON: the first continental-scale ecological observatory with airborne remote sensing of vegetation canopy biochemistry and structure. Journal of Applied Remote Sensing, 4(1): 043510.

Kruse, F.A., Kierein-Young, K.S. and Boardman, J.W., 1990. Mineral mapping at Cuprite, Nevada with a 63-channel imaging spectrometer. Photogrammetric Engineering and Remote Sensing, 56(1): 83-92.

Rogge, D. M., Rivard, B., Zhang, J., and Feng, J., 2006. Iterative spectral unmixing for optimizing per-pixel endmember sets. IEEE Transactions on Geoscience and Remote Sensing, 44(12): 3725-3736.

Smith, A.M., Rivard, B., Feng, J., and Carcamo, H., 2019. Quantifying Lygus damage in Faba bean seeds using near-infrared imaging. Canadian Entomologist, in press.

Williams, M., Parody, R., Fafard, A., Kerekes, J., and van Aardt, J., 2017. Validation of abundance map reference data for spectral unmixing. Remote Sensing 9(5): 473.

Literature Cited (Thesis)

Abdel-Rahman, E. M., Ahmed, F. B., and Ismail, R., 2013. Random forest regression and spectral band selection for estimating sugarcane leaf nitrogen concentration using EO-1 Hyperion hyperspectral data. International Journal of Remote Sensing 34(2): 712-728.

Acevedo, M. F. B., Groen, T. A., Hecker, C. A., and Skidmore, A. K., 2017. Identifying leaf traits that signal stress in TIR spectra. ISPRS journal of Photogrammetry and Remote Sensing 125: 132-145.

Asner, G. P., 1998. Biophysical and biochemical sources of variability in canopy reflectance. Remote Sensing of Environment 64(3): 234-253.

Bajcsy, P., and Groves, P., 2004. Methodology for hyperspectral band selection. Photogrammetric Engineering & Remote Sensing 70(7): 793-802.

Bateson, C. A., Asner, G. P., and Wessman, C. A., 2000. Endmember bundles: A new approach to incorporating endmember variability into spectral mixture analysis. IEEE Transactions on Geoscience and Remote Sensing 38(2): 1083-1094.

Bechtel, R., Rivard, B., and Sanchez, A., 2002. Spectral properties of foliose and crustose lichens based on laboratory experiments. Remote Sensing of Environment 82: 389-396.

Belgiu, M., and Drăguţ, L., 2016. Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing 114: 24-31.

bin Othman, M.F. and Yau, T.M.S., 2007. Comparison of different classification techniques using WEKA for breast cancer. In 3rd Kuala Lumpur International Conference on Biomedical Engineering 2006 (520-523). Springer, Berlin, Heidelberg.

Bioucas-Dias, J. M., Plaza, A., Camps-Valls, G., Scheunders, P., Nasrabadi, N., and Chanussot, J., 2013. Hyperspectral remote sensing data analysis and future challenges. IEEE Geoscience and Remote Sensing Magazine 1(2): 6-36.

Bishop, J., Lane, M., Dyar, M., and Brown, A., 2008. Reflectance and emission spectroscopy study of four groups of phyllosilicates: smectites, kaolinite-serpentines, chlorites and micas. Clay Minerals 43: 35-54.

Borengasser, M., Hungate, W. S., and Watkins, R., 2007. Hyperspectral remote sensing: principles and applications. CRC press.

Buitrago, M. F., Skidmore, A. K., Groen,T. A., and Hecker, C. A., 2018. Connecting infrared spectra with plant traits to identify species. ISPRS Journal of Photogrammetry and Remote Sensing 139: 183-200.

Castro-Esau, K. L., Sánchez‑Azofeifa, G. A., Rivard, B., Wright, S. J., and Quesada, M., 2006. Variability in leaf optical properties of Mesoamerican trees and the potential for species classification. American Journal of Botany 93(4): 517-530.

Chan, J. C. W., and Paelinckx, D., 2008. Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. Remote Sensing of Environment 112(6): 2999-3011.

Chang, C-I., 2000. An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis. IEEE Transactions on Information Theory 46(5): 1927-1932.

Chang, C-I., 2013. Hyperspectral data processing: algorithm design and analysis. Hoboken, NJ: John Wiley & Sons.

Clark, M. L., Roberts, D. A., and Clark, D. B., 2005. Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. Remote sensing of environment 96(3-4): 375-398.

Clark, R., King, T., Klejwa, M., and Swayze, G., 1990. High spectral resolution reflectance spectroscopy of minerals. Journal of Geophysical Research 95: 12653-12680.

Clarke, R.N., 1999. Spectroscopy of rocks and minerals, and principles of spectroscopy, Remote Sensing for the Earth sciences. Manual of Remote Sensing 3: 3-58.

Costanzo, P. A., and Guggenheim, S., 2001. Baseline studies of the Clay Minerals Society source clays: preface. Clays and Clay Minerals 49(5): 371-371.

Costanzo, P. M., 2001. Baseline studies of the clay minerals society source clays: Introduction. Clays and Clay Minerals 49(5): 372-373.

da Luz, B. R., 2006. Attenuated total reflectance spectroscopy of plant leaves: a tool for ecological and botanical studies. New Phytologist 172(2): 305-318.

da Luz, B. R., and Crowley, J. K., 2007. Spectral reflectance and emissivity features of broad leaf plants: Prospects for remote sensing in the thermal infrared (8.0–14.0 μm). Remote Sensing of Environment 109(4): 393-405.

da Luz, B. R., and Crowley, J. K., 2010. Identification of plant species by using high spatial and spectral resolution thermal infrared (8.0–13.5 μm) imagery. Remote sensing of Environment 114(2): 404-413.

Diaz-Uriarte, R., 2007. GeneSrF and varSelRF: a web-based tool and R package for gene selection and classification using random forest. BMC bioinformatics 8(1): 328.

Ding, C., and Peng, H., 2005. Minimum redundancy feature selection from microarray gene expression data. Journal of Bioinformatics and Computational Biology 3(02): 185-205.

Du, Q., and Yang, H., 2008. Similarity-based unsupervised band selection for hyperspectral image analysis. IEEE Geoscience and Remote Sensing Letters 5(4): 564-568.

Entezari, I., Rivard, B., Geramian, M. and Lipsett, M.G., 2017. Predicting the abundance of clays and quartz in oil sands using hyperspectral measurements. International journal of applied earth observation and geoinformation, 59:1-8.

Fassnacht, F.E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L.T., Straub, C. and Ghosh, A., 2016. Review of studies on tree species classification from remotely sensed data. Remote Sensing of Environment 186: 64-87.

Fassnacht, F.E., Neumann, C., Förster, M., Buddenbaum, H., Ghosh, A., Clasen, A., Joshi, P.K., and Koch, B., 2014. Comparison of feature reduction algorithms for classifying tree species with hyperspectral data on three central European test sites. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7(6): 2547-2561.

Feilhauer, H., Asner, G. P., and Martin, R. E., 2015. Multi-method ensemble selection of spectral bands related to leaf biochemistry. Remote Sensing of Environment 164: 57-65.

Friedman, J., Hastie, T. and Tibshirani, R., 2001. The elements of statistical learning (Vol. 1, No. 10). New York: Springer series in statistics.

Gamon, J. A., and Surfus, J. S., 1999. Assessing leaf pigment content and activity with a reflectometer. The New Phytologist 143(1): 105-117.

Genuer, R., Poggi, J-M., and Tuleau-Malot, C., 2010. Variable selection using Random Forests. Pattern Recognition Letters, Elsevier 31 (14): 2225-2236.

Gerber, F., Marion, R., Olioso, A., Jacquemoud, S., Da Luz, B. R., and Fabre, S., 2011. Modeling directional–hemispherical reflectance and transmittance of fresh and dry leaves from 0.4 μm

to 5.7 µm with the PROSPECT-VISIR model. Remote Sensing of Environment 115(2): 404-414.

Goetz, A. F., 2009. Three decades of hyperspectral remote sensing of the Earth: A personal view. Remote Sensing of Environment 113: S5-S16.

Goetz, A.F., Vane, G., Solomon, J.E., and Rock, B.N., 1985. Imaging spectrometry for earth remote sensing. Science 228(4704): 1147-1153.

Gong, P., Pu, R., and Yu, B., 1997. Conifer species recognition: An exploratory analysis of in situ hyperspectral data. Remote sensing of Environment 62(2): 189-200.

Guzman, J.A., Rivard, B., and Sánchez-Azofeifa, G.A., 2018. Discrimination of liana and tree leaves from a Neotropical Dry Forest using visible-near infrared and longwave infrared reflectance spectra. Remote Sensing of Environment, 219:135-144.

Harrison, D., Rivard, B., and Sanchez-Azofeifa, A., 2018. Classification of tree species based on longwave hyperspectral data from leaves, a case study for a tropical dry forest. International Journal of Applied Earth Observation and Geoinformation 66: 93-105.

Harsanyi, J. C., and Chang C-I., 1994. Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach. IEEE Transactions on Geoscience and Remote Sensing 32(4): 779-785.

Hastie, T., Tibshirani, R. and Friedman, J., 2008. The Elements of Statistical Learning; Data Mining, Inference and Prediction (second edition). Springer, Berlin: Springer series in statistics.

Hook, S.J., Johnson, W.R. and Abrams, M.J., 2013. NASA's hyperspectral thermal emission spectrometer (HyTES). In Thermal Infrared Remote Sensing (pp. 93-115). Springer, Dordrecht.

Hunt, G., 1977. Spectral signatures of particulate minerals in the visible and near infrared. Geophysics 42(3): 501-513.

Jia, S., and Qian, Y., 2009. Constrained nonnegative matrix factorization for hyperspectral unmixing. IEEE Transactions on Geoscience and Remote Sensing 47(1): 161-173.

Jia, S., Tang, G., Zhu, J., and Li, Q., 2016. A novel ranking-based clustering approach for hyperspectral band selection. IEEE Transactions on Geoscience and Remote Sensing 54(1): 88-102.

Kalacska, M., Bohlman, S., Sanchez-Azofeifa, G. A., Castro-Esau, K., and Caelli, T., 2007. Hyperspectral discrimination of tropical dry forest lianas and trees: Comparative data reduction approaches at the leaf and canopy levels. Remote Sensing of Environment 109(4): 406-415.

Kalacska, M., Sanchez-Azofeifa, G. A., Calvo-Alvarado, J. C., Quesada, M., Rivard, B., and Janzen, D. H., 2004. Species composition, similarity and diversity in three successional stages of a seasonally dry tropical forest. Forest Ecology and Management 200(1-3): 227-247.

Kampe, T.U., Johnson, B.R., Kuester, M.A. and Keller, M., 2010. NEON: the first continental-scale ecological observatory with airborne remote sensing of vegetation canopy biochemistry and structure. Journal of Applied Remote Sensing, 4(1): 043510.

Keshava, N., 2004. Distance metrics and band selection in hyperspectral processing with applications to material identification and spectral libraries. IEEE Transactions on Geoscience and Remote Sensing 42(7): 1552-1565.

Keshava, N., and Mustard, J. F., 2002. Spectral unmixing. IEEE signal processing magazine 19(1): 44-57.

Khoder, J., Younes, R., Obeid, H. and Khalil, M., 2015, June. Dimension reduction of hyperspectral image with rare event preserving. In Iberian Conference on Pattern Recognition and Image Analysis (pp. 621-629). Springer, Cham.

Kruse, F.A. and Fairbarn, K.G., 2013, May. Spectral variability constraints on multispectral and hyperspectral mapping performance. In Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIX (Vol. 8743, p. 87431O). International Society for Optics and Photonics.

Kruse, F.A., Kierein-Young, K.S. and Boardman, J.W., 1990. Mineral mapping at Cuprite, Nevada with a 63-channel imaging spectrometer. Photogrammetric Engineering and Remote Sensing, 56(1): 83-92.

Landgrebe, D., 2002. Hyperspectral image data analysis. IEEE Signal Processing Magazine 19(1): 17-28.

Laybros, A., Schläpfer, D., Féret, J. B., Descroix, L., Bedeau, C., Lefevre, M. J., and Vincent, G., 2019. Across Date Species Detection Using Airborne Imaging Spectroscopy. Remote Sensing 11(7): 789.

Lesher, C. M., 2007. Ni-Cu-(PGE) deposits in the Raglan area, Cape Smith Belt, New Quebec. Mineral Deposits of Canada: A Synthesis of Major Deposits-Types, District Metallogeny, the Evolution of Geological Provinces, and Exploration Methods. Geological Association of Canada, Mineral Deposits Division Special Publication 5: 351-386.

Li, S., Qiu, J., Yang, X., Liu, H., Wan, D., and Zhu, Y., 2014. A novel approach to hyperspectral band selection based on spectral shape similarity analysis and fast branch and bound search. Engineering Applications of Artificial Intelligence 27: 241-250.

Long, Y. Q, Rivard, B., and Rogge, D., 2019. Hyperspectral band selection using the N-dimensional Spectral Solid Angle method for the improved discrimination of spectrally similar targets. International Journal of Applied Earth Observation and Geoinformation 79: 35-47.

Meerdink, S. K., Roberts, D. A., King, J. Y., Roth, K. L., Dennison, P. E., Amaral, C. H., and Hook, S. J., 2016. Linking seasonal foliar traits to VSWIR-TIR spectroscopy across California ecosystems. Remote Sensing of Environment 186: 322-338.

Meyer, P. E., Schretter, C., and Bontempi, G., 2008. Information-theoretic feature selection in microarray data using variable complementarity. IEEE Journal of Selected Topics in Signal Processing 2(3): 261-274.

Peng, H., Long, F., and Ding, C., 2005. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(8): 1226-1237.

Plaza, A., Benediktsson, J.A., Boardman, J.W., Brazile, J., Bruzzone, L., Camps-Valls, G., Chanussot, J., Fauvel, M., Gamba, P., Gualtieri, A. and Marconcini, M., 2009. Recent advances in techniques for hyperspectral image processing. Remote Sensing of Environment 113: S110-S122.

Pu, R., 2009. Broadleaf species recognition with in situ hyperspectral data. International Journal of Remote Sensing 30(11): 2759-2779.

Pu, R., 2017. Hyperspectral remote sensing: fundamentals and practices. CRC Press.

Quesada, M., Sanchez-Azofeifa, G.A., Alvarez-Anorve, M., Stoner, K.E., Avila-Cabadilla, L., Calvo-Alvarado, J., Castillo, A., Espírito-Santo, M.M., Fagundes, M., Fernandes, G.W., and

Gamon, J., 2009. Succession and management of tropical dry forests in the Americas: Review and new perspectives. Forest Ecology and Management 258(6): 1014-1024.

Ravikumar, P., Wainwright, M. J., and Lafferty, J. D., 2010. High-dimensional Ising model selection using $\ell 1$-regularized logistic regression. The Annals of Statistics 38(3): 1287-1319.

Rock, G., Gerhards, M., Schlerf, M., Hecker, C., and Udelhoven, T. 2016. Plant species discrimination using emissive thermal infrared imaging spectroscopy. International Journal of Applied Earth Observation and Geoinformation 53: 16-26.

Rogge, D. M., Rivard, B., Zhang, J., and Feng, J., 2006. Iterative spectral unmixing for optimizing per-pixel endmember sets. IEEE Transactions on Geoscience and Remote Sensing, 44(12): 3725-3736.

Rogge, D. M., Rivard, B., Zhang, J., Sanchez, A., Harris, J., and Feng, J., 2007. Integration of spatial-spectral information for the improved extraction of endmembers. Remote Sensing of Environment 110: 287-303.

Rogge, D., Rivard, B., Segl, K., Grant, B., and Feng, J., 2014. Mapping of NiCu–PGE ore hosting ultramafic rocks using airborne and simulated EnMAP hyperspectral imagery, Nunavik, Canada. Remote Sensing of Environment 152: 302-317.

Salisbury, J. W., 1986. Preliminary measurements of leaf spectral reflectance in the 8-14 µm region. International Journal of Remote Sensing 7(12): 1879-1886.

Salisbury, J. W., and Milton, N. M., 1988. Thermal infrared (2.5-to 13.5-µm) directional hemispherical reflectance of leaves. Photogrammetric Engineering and Remote Sensing 54(9): 1301-1304.

Sánchez-Azofeifa, G.A., Castro, K., Wright, S.J., Gamon, J., Kalacska, M., Rivard, B., Schnitzer, S.A. and Feng, J.L., 2009. Differences in leaf traits, leaf internal structure, and spectral

reflectance between two communities of lianas and trees: Implications for remote sensing in tropical environments. Remote Sensing of Environment 113(10): 2076-2088.

Sims, D. A., and Gamon, J. A., 2002. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. Remote Sensing of Environment 81(2-3): 337-354.

Smith, A.M., Rivard, B., Feng, J., and Carcamo, H., 2019. Quantifying Lygus damage in Faba bean seeds using near-infrared imaging. Canadian Entomologist, in press.

Somers, B., Asner, G. P., Tits, L., and Coppin, P., 2011. Endmember variability in spectral mixture analysis: A review. Remote Sensing of Environment 115(7): 1603-1616.

Speta, M., Gingras, M.K., and Rivard, B., 2016. Shortwave Infrared Hyperspectral Imaging: A Novel Method For Enhancing the Visibility of Sedimentary And Biogenic Features In Oil-Saturated Core. Journal of Sedimentary Research 86 (7): 830-842.

Sun, K., Geng, X., Ji L., and Lu, Y., 2014. A new band selection method for hyperspectral image based on data quality. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7(6): 2697-2703.

Tian, M., Feng, J., Rivard, B., and Zhao, C., 2016. A method to compute the n-dimensional solid spectral angle between vectors and its use for band selection in hyperspectral data. International Journal of Applied Earth Observation and Geoinformation 50: 141-149.

Tuia, D., Flamary, R., and Courty, N., 2015. Multiclass feature learning for hyperspectral image classification: Sparse and hierarchical solutions. ISPRS Journal of Photogrammetry and Remote Sensing 105: 272-285.

Ullah, S., Groen, T. A., Schlerf, M., Skidmore, A. K., Nieuwenhuis, W., and Vaiphasa, C., 2012. Using a genetic algorithm as an optimal band selector in the mid and thermal infrared (2.5–14 µm) to discriminate vegetation species. Sensors 12(7): 8755-8769.

Ullah, S., Schlerf, M., Skidmore, A. K., and Hecker, C., 2012. Identifying plant species using mid-wave infrared (2.5–6 µm) and thermal infrared (8–14 µm) emissivity spectra. Remote Sensing of Environment 118: 95-102.

Ullah, S., Skidmore, A. K., Ramoelo, A., Groen, T. A., Naeem, M., and Ali, A., 2014. Retrieval of leaf water content spanning the visible to thermal infrared spectra. ISPRS Journal of Photogrammetry and Remote Sensing 93: 56-64.

Van der Meer, F. D., and De Jong, S. M. (Eds.), 2011. Imaging spectrometry: basic principles and prospective applications (Vol. 4). Springer Science & Business Media.

Van der Meer, F. D., and Jia, X., 2012. Collinearity and orthogonality of endmembers in linear spectral unmixing. International Journal of Applied Earth Observation and Geoinformation 18: 491-503.

Van Der Meer, F., 2004. Analysis of spectral absorption features in hyperspectral imagery. International Journal of Applied Earth Observation and Geoinformation 5(1): 55-68.

Viera, A.J. and Garrett, J.M., 2005. Understanding interobserver agreement: the kappa statistic. Fam med, 37(5): 360-363.

Wang, S., and Chang, C.I., 2007. Variable-number variable-band selection for feature characterization in hyperspectral signatures. IEEE Transactions on Geoscience and Remote Sensing 45(9): 2979-2992.

Wetherley, E.B., Roberts, D.A., and McFadden, J.P., 2017. Mapping spectrally similar urban materials at sub-pixel scales. Remote Sensing of Environment 195: 170-183.

Williams, M., Parody, R., Fafard, A., Kerekes, J., and van Aardt, J., 2017. Validation of abundance map reference data for spectral unmixing. Remote Sensing 9(5): 473.

Zhang, J., Rivard, B., Sánchez-Azofeifa, A., and Castro-Esau, K., 2006. Intra-and inter-class spectral variability of tropical tree species at La Selva, Costa Rica: Implications for species identification using HYDICE imagery. Remote Sensing of Environment 105(2): 129-141.

Appendix: Description of the NSSA method

1. Definition and calculation of NSSA

The mathematical background described below can be found in Tian et al. (2016). The N-dimensional Solid Spectral Angle (NSSA) is an extension of the traditional Spectral Angle (SA) to the n-dimensional space. The SA measures the similarity between two spectra, a reference spectrum (x) and a target spectrum (y) by using the following formula:

$$SA(x, y) = \cos^{-1} \frac{< x, y >}{||x||||y||}$$

The NSSA measures the solid angle or the similarity among n (more than two) spectra. The solid angle measures the intersection of a polygonal cone C_n and the corresponding unit sphere surface S_{n-1}. Figure A1 describes the solid angle in 2, 3 and n (>3) dimensions.
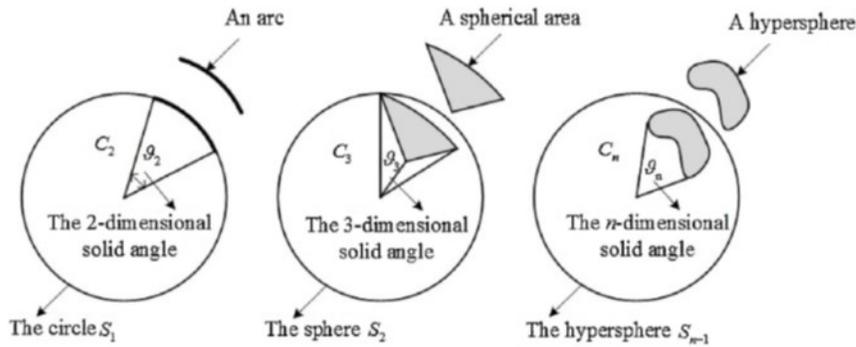

Figure A.1 Illustration of the solid angle in 2, 3, and n(>3) dimensions.

The mathematical definition of the N-dimensional solid angle is:

$$NSSA(\vartheta_n) = |\det(E)| \int_S ||EV||^{-n} ds$$

where E={e_1, e_2, e_3, $\cdots$, e_n} is a vector matrix of given spectra, and this matrix is normalized. $||\cdot||$ represents the Euclidean norm. V=[v_1, v_2, v_3, $\cdots$, v_i, $\cdots$, v_n]^T is the vector of a unit sphere and specifically defined as:

$$
\begin{cases}
v_1 = cos\theta_1, & 0 \le \theta_1 \le \pi/2 \\
v_2 = sin\theta_1 cos\theta_2, & 0 \le \theta_2 \le \pi/2 \\
v_3 = sin\theta_1 sin\theta_2 cos\theta_3, & 0 \le \theta_3 \le \pi/2 \\
\ldots \ldots \\
v_{n-1} = sin\theta_1 sin\theta_2 \ldots sin\theta_{n-2}cos\theta_{n-1}, & 0 \le \theta_{n-1} \le \pi/2 \\
v_n = sin\theta_1 sin\theta_2 \ldots sin\theta_{n-2}sin\theta_{n-1}, & 0 \le \theta_{n-1} \le \pi/2
\end{cases}
$$

The ds refers to the surface area of S_{n-1} and can be expressed as:

$$
ds = sin^{n-2}(\theta_1)sin^{n-3}(\theta_2), \ldots, sin(\theta_{n-2})\, d\theta_1 d\theta_2 \ldots d\theta_{n-1}
$$

The simplified equation for calculating the NSSA is:

$$
NSSA(\vartheta_n) = |\det(E)| \int_s \left(||EV||^2 \right)^{-n/2} ds = |\det(E)| \int_s \left(1 + 2\sum_{i<j} < e_i, e_j > v_i v_j \right)^{-n/2} ds
$$

$$
= |\det(E)| \int_0^{\pi/2} \ldots \int_0^{\pi/2} \left(1 + 2\sum_{i<j} < e_i, e_j > v_i v_j \right)^{-n/2} J(\theta)d\theta_1 d\theta_2 \ldots d\theta_{n-1}
$$

where $J(\theta) = sin^{n-2}(\theta_1)sin^{n-3}(\theta_2), \ldots, sin(\theta_{n-2})$

2. NSSA for band selection

The calculation of the NSSA requires that the data matrix E be square, thus the number of spectra should be equal to the number of bands. To depict the similarity among multiple targets as a function of wavelength, a sliding window with a flexible band interval (represented by the

parameter 'k') was designed for band selection (fig. A2). The NSSA value calculated within the window is assigned to the middle band of the sliding window.
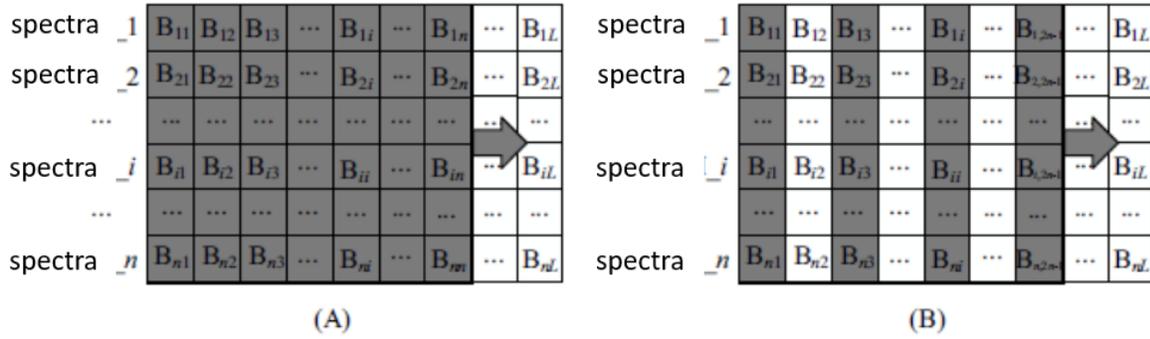

Figure A.2 Data matrix and design of the sliding window for different band interval k=0 (A) and k=1 (B). Bands included in the sliding window for the NSSA computation are shown as shaded. L is the number of bands, and n is the number of spectra.

The spectral regions or bands with the largest NSSA are deemed the most important because they reflect greater differences in spectral shapes amongst the suite of spectra investigated. For a fixed k value, a user-defined parameter 'threshold' was also designed to determine the number of selected bands. Bands with NSSA values above the threshold are retained (fig. A3).
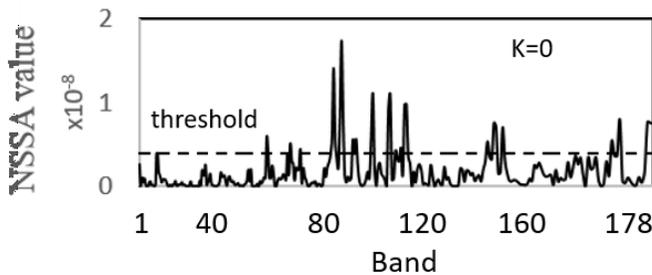

Figure A.3 Example profile of NSSA value as a function of wavelength or band for a given k value. Bands above the threshold (dotted horizontal line) are retained.

Reference

Tian, M., Feng, J., Rivard, B., and Zhao, C., 2016. A method to compute the n-dimensional solid spectral angle between vectors and its use for band selection in hyperspectral data. International Journal of Applied Earth Observation and Geoinformation 50: 141-149.