Graduate Theses, Dissertations, and Problem Reports

2010

Seasonal trends in separability of leaf reflectance spectra for Ailanthus altissima and four other tree species

Aaron Burkholder
West Virginia University

Follow this and additional works at: https://researchrepository.wvu.edu/etd

Seasonal trends in separability of leaf reflectance spectra for
***Ailanthus altissima* and four other tree species**

Aaron Burkholder

Thesis submitted to the Eberly College of Arts and Sciences
at West Virginia University
in partial fulfillment of the requirements
for the degree of

Master of Arts
in
Geography

Approved by

Timothy A. Warner, Ph.D., Chair
Rick Landenberger, Ph.D.
Brenden McNeil, Ph.D.
James B. McGraw, Ph.D.
Mark Culp, Ph.D.

Department of Geology and Geography

Morgantown, West Virginia
2010

Keywords: Remote sensing, *Ailanthus altissima,* invasive species, leaf spectral
reflectance, least angle regression, random forests

Abstract

Seasonal trends in separability of leaf reflectance spectra for *Ailanthus altissima* and four other tree species

Aaron Burkholder

This project investigated the spectral separability of the invasive species *Ailanthus altissima*, commonly called tree of heaven, and four other native species. Leaves were collected from *Ailanthus* and four native tree species from May 13 through August 24, 2008, and spectral reflectance factor measurements were gathered for each tree using an ASD (Boulder, Colorado) FieldSpec Pro full-range spectroradiometer. The original data covered the range from 350-2500 nm, with one reflectance measurement collected per one nm wavelength. To reduce dimensionality, the measurements were resampled to the actual resolution of the spectrometer's sensors, and regions of atmospheric absorption were removed. Continuum removal was performed on the reflectance data, resulting in a second dataset. For both the reflectance and continuum removed datasets, least angle regression (LARS) and random forest classification were used to identify a single set of optimal wavelengths across all sampled dates, a set of optimal wavelengths for each date, and the dates for which *Ailanthus* is most separable from other species. It was found that classification accuracy varies both with dates and bands used. Contrary to expectations that early spring would provide the best separability, the lowest classification error was observed on July 22 for the reflectance data, and on May 13, July 11 and August 1 for the continuum removed data. This suggests that July and August are also potentially good months for species differentiation. Applying continuum removal in many cases reduced classification error, although not consistently. Band selection seems to be more important for reflectance data in that it results in greater improvement in classification accuracy, and LARS appears to be an effective band selection tool. The optimal spectral bands were selected from across the spectrum, often with bands from the blue (401-431 nm), NIR (1115 nm) and SWIR (1985-1995 nm), suggesting that hyperspectral sensors with broad wavelength sensitivity are important for mapping and identification of *Ailanthus*.

Acknowledgments

TABLE OF CONTENTS

Contents

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Over time, humans have spread to the far reaches of the globe, taking other species with them in the process. The extent to which biological invasions have taken place in recent history is thought to be contributing significantly to the breakdown of important natural barriers that sustain biodiversity around the world (Pimentel et al. 2005, Vitousek et al. 1997). The movement of humans between continents has been such a substantial factor in the introduction of species to new habitats that it can be considered a noteworthy component of global change on its own (Vitousek 1994). In fact, more than twenty percent of continental species and more than fifty percent of island species can be considered invasive (Asner et al. 2008b, Vitousek et al. 1997). The eastern deciduous forest of North America has been subject to invasions by non-native insects and pathogens that have caused the near-disappearance of the American chestnut and the American elm in the eastern forest and affected other species such as the American beech, eastern hemlock, flowering dogwood, and sugar maple (Niemelä & Mattson 1996, Vitousek 1994). Research suggests that when allowed to spread unchecked, invasive species can even play a role in the extinction of native species (Vitousek et al. 1997), although this conclusion has been recently brought into question (Sax & Gaines 2008).

Remote sensing is often used to map and monitor the spread of invasive species, and it can potentially be of assistance in management (Asner et al. 2008b, Landenberger et al. 2009, Lawrence et al. 2006). Mapping the location of different species reliably using satellite-based and aerial imagery is challenging, and this study aims to establish a foundation for remotely identifying *Ailanthus altissima*, an invasive tree often seen in the eastern deciduous forest, by finding the optimal spectral wavelengths and time of year to differentiate *Ailanthus* from other tree species.

1.1 *Ailanthus* altissima

The invasive tree species *Ailanthus altissima*, or tree of heaven, was purposely introduced to the United States from its native China in the late eighteenth century, and in the years since it has become a notorious invader in the North American eastern deciduous forest (Kota et al. 2007). The tree was originally planted in urban areas of the United States because of traits that

allow it to survive under stressful urban growing conditions, such as its tolerance for infertile, compacted soils and drought, its capability to grow rapidly and mature early, and its resistance to insect predation (Kowarik & Säumel 2007, Landenberger et al. 2009). In addition to the properties that allow it to survive in urban areas *Ailanthus* has other traits associated with highly invasive plant species that allow it to move rapidly into suburban and rural environments, especially when disturbance is present (Call & Nilsen 2005). *Ailanthus* is dioecious, and females generate large numbers of seeds: as many as three hundred thousand per tree that are dispersed by wind from late summer to spring (Kowarik & Säumel 2007, Landenberger et al. 2007). This prolonged dispersal period and large number of seeds put *Ailanthus* at an advantage by increasing the probability that many seeds will land on suitable sites and successfully germinate. The tree's seeds have also been found to be capable of germinating in a wide variety of habitats, but especially in disturbed areas such as highway or power line corridors and sites where timber has been harvested (Kota et al. 2007, Kowarik & Säumel 2007). In addition to the large number of seeds each female tree produces *Ailanthus* is able to reproduce prolifically by sprouting from roots and stumps (Kowarik 1995, Kowarik & Säumel 2007), and when combined with the allelopathic compounds found in its leaves and roots (Heisey 1990, Kowarik & Säumel 2007), this allows the tree to rapidly dominate an area, persist and exclude native vegetation. *Ailanthus* is also tolerant of air pollutants such as ozone, sulfur dioxide and dust, and this characteristic makes the tree a very strong competitor along highway edges, in urban and suburban areas with high traffic and other areas with significant local air pollution (Kowarik & Säumel 2007, Landenberger et al. 2007, Rentch et al. 2005).

  Ailanthus was largely confined to urban areas and localized disturbed sites through the nineteenth and twentieth centuries, but it is currently spreading from metropolitan areas into suburban and rural environments (Kowarik & Säumel 2007, Landenberger et al. 2009). In the mid-Atlantic region *Ailanthus* can be found throughout the interstate highway system, and it is quickly becoming a harmful invasive pest in a variety of ecosystem types. Since the tree is no longer confined solely to urban areas, roadsides and habitat edges, it is becoming established in places like clear-cut and partially cut forest stands, mature second-growth forest and other relatively intact forest communities in rural areas (Kowarik & Säumel 2007, Landenberger et al. 2007). *Ailanthus* has been found to be a very aggressive competitor with native species, especially in disturbed areas, and it is increasing in local abundance and geographic extent as it

continues to disperse along highway and power line corridors and into adjacent habitats. In the near future *Ailanthus* is expected to become even more common in the mid-Atlantic region as development continues and a predicted increase in timber harvesting creates large openings that will almost surely provide ample habitat for *Ailanthus* to invade and colonize (Landenberger et al. 2009, Kota et al. 2007).

1.2 Spectral Reflectance Measurements

Radiometry can be defined as the measurement of optical radiation in the range from 0.1 to 1000 micrometers on the electromagnetic spectrum (Schaepman-Strub et al. 2009). In this study we focus on the visible (400-700 nm), the near infrared (NIR; 700-1400 nm) and the shortwave infrared (SWIR; 1400-2500 nm). Radiometry can be performed both in the field and in a laboratory environment. There are certainly benefits to either approach, but making spectral reflectance measurements in the laboratory environment allows control over illumination and minimizes problems associated with atmospheric transmission and weather (Curtiss & Goetz 1994). Studies of the spectral properties of vegetation have shown that although nearly all healthy leaves have a generally similar spectral response, differences in pigments such as chlorophyll, as well as differences in leaf structure, water content and biochemical composition can yield spectral signatures that can be effective for discriminating among species (Asner et al. 2008a). On the other hand, it has also been suggested that there is so much variation in spectral reflectance among individuals of the same species that discriminating among different species may not be possible in all situations (Price 1994).

A wide variety of studies have used laboratory and occasionally field spectra to discriminate between species, including grassland species (Schmidt & Skidmore 2001), wetland species (Schmidt & Skidmore 2003), forest species (van Aardt & Wynne 2001, Pinard & Bannari 2003), plants in an arid region of Australia (Lewis 2002) and for highly targeted surveys, such as differentiating marijuana plants from other species (Daughtry & Walthall 1998).

Unlike laboratory spectra, which are normally collected as reflectance, aerial imagery is usually calibrated to radiance. Converting radiance imagery to reflectance can be a challenge, and as an alternative continuum removal (Clark & Roush 1984) is sometimes applied. Continuum removal is a technique for normalizing data in which a convex hull is fitted to the reflectance curve. The convex hull can be imagined as an elastic band stretched over the spectral

reflectance curve, touching at several points and never crossing the spectral curve. Reflectance is then divided by the convex hull, resulting in a curve with values falling between zero and one, emphasizing absorption features and reducing differences between spectral curves due to brightness or illumination (Schmidt & Skidmore 2001). This technique is often employed in geological remote sensing studies and has been used to varying degrees of success in differentiating vegetation (Clark & Roush 1984, Psomas et al. 2005, Schmidt & Skidmore 2001, 2003). The benefit of using continuum removal is that it effectively suppresses illumination differences, thus reducing variability within a scene and facilitating comparison with spectral library data in reflectance format that have also been transformed with continuum removal. A potential disadvantage is that continuum removal eliminates albedo information that can be important in differentiating spectral signatures. (Chen et al. 2010).

Another important preprocessing step is the decision of whether to apply feature selection to reduce the number of variables used in the classification (Warner 2010). The observation that using fewer bands often results in decreased classification error is known as the Hughes Phenomenon (Hughes 1968). Although it might be supposed that additional bands would provide more information, increased dimensionality often reduces classifier performance (Warner & Nerry 2009). Spectral reflectance measurements of vegetation are very often highly correlated and datasets usually must be simplified prior to performing statistical analyses, for example through stepwise discriminant analysis (SDA) (Palacios-Orueta & Ustin 1996, van Aardt & Wynne 2001), or principal components analysis (PCA) either on its own or in combination with SDA (Lewis 2002, Palacios-Orueta & Ustin 1996). A recently developed technique is least angle regression (LARS) (Efron et al. 2004), which to our knowledge has not yet been applied to remotely sensed data. LARS is a linear model that identifies a subset of the input variables according to their importance in the regression. It is a type of forward selection method that is computationally efficient, intuitive and elegant, and it has been reported as handling high dimensionality well (Efron et al. 2004). Another advantage LARS has over techniques like PCA and SDA is that while the latter are based on Euclidean distances, LARS draws upon the correlation of the input variables to the response variables. Potential disadvantages of LARS include susceptibility to high correlation and especially the extreme case of co-linearity (Efron et al. 2004, Hastie & Efron 2007). This issue is a possible concern for remote sensing data, which tend to be highly correlated.

Statistical methods used to evaluate spectral separability vary considerably, but there are several common approaches for species differentiation. Original reflectance data and continuum removed reflectance data are often subjected to the Mann-Whitney U-test, which is a non-parametric pairwise comparison (Psomas et al. 2005, Schmidt & Skidmore 2001, 2003). Other studies use the *D* discrimination metric, or the average root-mean-square difference over the spectral range being observed, and the θ discrimination metric, which examines the difference in shape between spectra according to the angle between two vectors (Cochrane 2000, Price 1994), and still others use ordinary discriminant analysis (Lewis 2002, Palacios-Orueta & Ustin 1996, van Aardt & Wynne 2001).

A relatively new classification technique that has been used recently with some success in remote sensing is the random forest algorithm (Breiman 2001, Bunn et al. 2005, Ghimire et al. 2010, Lawrence et al. 2006, Pal 2005). It is a bootstrap aggregation procedure where multiple classification trees are grown, with each based on a different random subset of training data (Breiman 2001, Lawrence et al. 2006, Liaw & Wiener 2002). A classification tree attempts to classify the dataset by recursively splitting it into groups, with each split termed a node. Nodes are split based on subsets of the input data that are randomly chosen at each node. In the random forest algorithm many of these trees are grown, and the algorithm chooses the best classification from all of the classification trees (Lawrence et al. 2006, Liaw & Wiener 2002). Advantages of this approach are that it performs well compared to techniques like discriminant analysis, it works with non-parametric data and has been described as robust (Breiman 2001, Lawrence et al. 2006, Liaw & Wiener 2002).

1.3 Research Objectives

Designing a vegetation remote sensing project involves many challenges, particularly those related to image aquisition (Warner 2010). There are questions of the optimal dates of image acquisition, the wavelengths of the imagery that will be collected and the analysis techniques that will be applied. For example, vegetation changes through the growing season and these changes can help to differentiate species at specific times of year. For image acquisition there are choices involving multispectral versus hyperspectral data, the choice of the specific bands to acquire for programmable sensors (such as the Digital Compact Airborne Spectrographic Images sensor (CASI), ITRES, Calgary, Canada) and whether band selection

leads to more efficient processing. There are also choices involved in analysis, including how to normalize for atmospheric effects.

To address these challenges and lay a foundation for subsequent analysis focused on mapping *Ailanthus altissima* from remotely sensed imagery, the overarching aim of this project was to characterize seasonal trends in separability of leaf reflectance spectra for *Ailanthus* and four other tree species, with the specific goal of finding the best wavelengths for differentiating *Ailanthus* individuals from other species during the spring and summer. *Ailanthus* is often distinctive to the human eye in the early spring when the leaves have a red hue due to the increased expression of anthocyanins relative to chlorophyll (Gitelson et al. 2001), and although autumn is a distinctive time of year for many Eastern deciduous forest species, *Ailanthus* leaves tend to have a nondescript brown color at that time. Because of this it was hypothesized that *Ailanthus* would be most separable in the spring with a decline in separability through the growing season. The project also sought to address issues relating to preprocessing, particularly the effects of continuum removal and the benefits of band selection. The following specific questions were posed:

1. Is there a single, consistent set of wavelengths that is best for differentiating *Ailanthus* from other species across all of the sampled dates?

2. What are the optimal wavelengths for differentiating *Ailanthus* individuals from other species for each sampled date?

3. What is the date (or dates) at which *Ailanthus* is most spectrally separable from the surrounding forest community when using i) all available wavelengths, ii) using the optimal wavelengths for any date, and iii) using the optimal wavelengths across all dates?

4. How different are the results for the previous three questions for the original reflectance data compared to the continuum removed data?

5. What difference, if any, does band selection make as a preprocessing step prior to application of random forest classification?

2. METHODS

2.1 Study area

The city of Morgantown, in Monongalia County in north-central West Virginia, provides a useful area for observing the spread of *Ailanthus*, because the area contains a variety of land uses and habitat types where the tree has either been found or could potentially spread. The region is experiencing extensive changes in land use associated with growth such as sprawling development at the city's edges and the conversion of farmland and forested areas both within the city and in its suburban areas to make space for housing and commercial activities. These activities combine with timber harvesting and other types of disturbance in the region to contribute to increasingly fragmented natural habitats that have been shown to be receptive to invasion by non-native species. *Ailanthus* trees have been found in a variety of habitats in the area, from developed urban environments with poor soils and extremes of light and temperature to highly disturbed forest stands and edges of intact forest communities (Landenberger et al. 2009, Kota et al. 2007).

2.2 Field data collection

Seventeen trees along a section of public rail-trail in downtown Morgantown were identified and tagged for the purpose of collecting and spectrally analyzing leaves throughout the sampling period. This sampling occurred on at least a biweekly basis throughout the period from mid-May through late August of 2008, resulting in eleven total collection dates. Leaf samples were collected from *Ailanthus* and four other tree species: staghorn sumac (*Rhus typhina*), silver maple (*Acer saccharinum*), black locust (*Robinia pseudoacacia*), and box elder (*Acer negundo*). The four native species were chosen because they are typically found in association with *Ailanthus* in disturbed habitats in the study area, and therefore typically comprise the background against which we wish to differentiate *Ailanthus*. Five leaves per tree from three individuals of each of the four non-*Ailanthus* species, and five leaves per tree from each of five individual *Ailanthus* trees were collected at random from sunlit portions of each tree's canopy. For each tree random sampling was achieved by generating five random numbers between 0 and 180° that were used as compass bearings from each tree's trunk. The 180° range was used because the ground sloped steeply away from the rail-trail and only half of the canopy of each tree was easily accessible. The leaves were clipped from the tree using a pole pruner with a 5.5 m extension.

Samples from each tree were kept separate and viable using plastic bags containing damp towels until the samples could be analyzed within a few hours. Each tree and its tag were photographed in the field on each collection date for documentation.

2.3 Spectral measurement and preprocessing

Upon returning to the laboratory the samples from each tree were photographed using a Sigma SD9 digital SLR, and then spectral reflectance factor measurements were made for each leaf collected (total of 85 leaves), resulting in five complete spectra for each of the seventeen trees. The spectral measurements were made with an ASD (Boulder, Colorado) FieldSpec Pro full-range (350 – 2500 nm) spectroradiometer with a spectral resolution of 3 nm in the 350 – 1000 nm range and 10 nm in the 1000 – 2500 nm range and a spectral sampling interval of 1.4 nm in the 350 – 1000 nm range and 2 nm in the 1000 – 2500 nm range. These data are then interpolated to a one nanometer sampling interval across the spectrum. The spectra were collected using an ASD contact and plant probe with a 10 mm spot size and an attached leaf clip assembly (ASD 1997). A "white reference" reflectance measurement was acquired using the Gortex white background standard mounted on the leaf clip assembly of the plant probe, and 25 spectra were averaged for each recorded leaf spectrum (ASD 1997).

The raw spectral data exhibited a step offset feature centered near 986 nm, apparently due to calibration differences between the NIR and the first of the SWIR detectors in the spectrometer.   This was suppressed using the ASD ViewSpec Parabolic Correction program (ASD 1997), and the bands between 985 and 987 nm were removed. As the spectral data were collected in the laboratory, problems with atmospheric absorption were not evident, but it was decided to remove wavelengths where atmospheric transmission is sufficiently low to limit aerial and satellite applications. The USGS Spectral Viewer website (USGS 2008) was used to identify the wavelengths of major atmospheric absorption features where atmospheric transmittance is less than 40%, and as a result the following wavelength ranges were removed: less than 400 nm, 1121-1160 nm, 1341-1470 nm, 1781-1980 nm, and greater than 2400 nm.

In order to work at the native resolution of the spectrometer and reduce dimensionality of the dataset the data were sub-sampled to the original resolution by averaging over a 3 nm range from 400 to 1000 nm and over a 10 nm range from 1001 to 2400 nm. This resulted in a total of 302 bands that were used in all subsequent analysis.

In order to produce the continuum removed dataset the ENVI continuum removal function was applied to the raw reflectance data (ITT Visual Information Solutions 2007).

2.4 Statistical Analysis

2.4.1 Identification of a single, optimum set of wavelengths across all sampled dates

For the analysis of a single, optimum set of wavelengths for the combined dataset across all dates and all statistical analyses to follow two specific packages (LARS and randomForest) within the R environment were used (R Development Core Team 2008). In order to identify a single optimal set of wavelengths from the entire multitemporal dataset the LARS algorithm (Hastie & Efron 2007) was applied to the reflectance data to generate a ranked list of the ten best bands, which were selected based on their ability to separate *Ailanthus* from the four other species. The choice of ten for the number of bands was chosen arbitrarily, but is partly based on previous hyperspectral band selection work which suggested that ten bands could represent the major spectral features for vegetation communities (Warner et al. 1999). This analysis was repeated for the continuum removed dataset.

2.4.2 Identification of optimal wavelengths for each date

In order to identify a subset of ten bands for each date individually, the LARS algorithm was applied separately to the reflectance data for each date. This resulted in 11 sets of ten ranked bands – one for each date. This analysis was repeated for the continuum removed dataset.

2.4.3 Identification of dates for which *Ailanthus* is most separable

To address the question of the date for which *Ailanthus* is spectrally most separable, separability was defined according to the performance of the random forest classification algorithm in differentiating *Ailanthus* from the other four species. Three approaches were used in answering this question: making all bands available to the random forest classification algorithm (Liaw & Wiener 2002), using the ten best bands across all dates, and using the ten best bands selected for each date individually. Beginning with the classification using all available bands, the number of classification trees to be grown was set to 500 and the number of variables

used at each node was left at the default value, which is the floor of the square root of the total number of the available spectral variables (302 total spectral variables, 17 used at each node). A bootstrap procedure was used to evaluate performance. 500 replications were run using thirteen *Ailanthus* spectra and thirty spectra from the other four species for training, leaving forty-two spectra for evaluation. This resulted in a classification error value for each of the 500 replications, and from these 500 error values the average error value for each date was calculated. To quantify the relative accuracy of classification using the ten best bands across all dates only the ten bands selected by the LARS algorithm were made available to the random forest classifier, and to quantify the accuracy of using the ten best bands for each date only the ten bands selected by the LARS algorithm for each date were made available to the classifier. The analysis was repeated for the continuum removed dataset.

3. RESULTS AND DISCUSSION

Before evaluating the statistical results a brief qualitative exploration of the variability of the spectra in both the reflectance and continuum removed datasets is presented. Figure 1 shows the variability present in the mean *Ailanthus* spectra across the eleven collection dates. In figure 1a the green-to-yellow (530-630 nm) region and the NIR plateau exhibit relatively high variability across the growing season. In contrast, the SWIR shows relatively limited variability for the majority of the growing season, but shows a large change for the last two dates (August 12 and 24). Continuum removal reduces variability in the NIR as well as in the SWIR around 2200 nm, but it increases the variability in other areas, especially in the green around 500 nm (Figure 1b).

Variability of *Ailanthus* spectra is also present within each date. Figure 2 shows variability within two selected dates (May 29 and July 22). It is apparent from the reflectance dataset (Figure 2a) that the relative amount of variability changes both across the spectrum and between dates, and in some places variability between dates is larger than variability within one date. Continuum removal (Figure 2b) selectively suppresses both within and between date variability at certain wavelengths, especially in the NIR and around 2200 nm.

Figure 3 compares spectra of the five species for the same two dates shown in Figure 2. The most important feature of this figure is that it shows that the relative separability of

Ailanthus from the other four species changes over time. Continuum removal (Figures 3b and 3d) appears to enhance variability in the visible ranges and reduce differences between species in the NIR and in the SWIR around 2200 nm.

In the following sections quantitative statistical results are presented.

3.1 Single, optimum set of wavelengths across all sampled dates

Table 1 lists the ten best bands across all sampling dates for both reflectance data and continuum removed data, and Figure 4 displays the same information graphically. In Figure 4 and other similar graphs to follow the height of each bar represents its rank, and there are ten bars per graph, although some graphs appear to have fewer because the bars are very close together in some cases. The ranks of individual wavelengths are displayed against a reflectance curve or continuum removed curve to show their locations in the spectrum. For both graphs in Figure 4 wavelengths are well-distributed across the spectrum with bands in the visible, near infrared and shortwave infrared. In Figure 4a (reflectance) there are fewer bands in the visible range and no red wavelengths, but there are two wavelengths at the red edge of the NIR plateau. In Figure 4b (continuum removed reflectance) the visible range is well represented, with five total bands and two in the red. In both figures SWIR wavelengths are highly ranked, suggesting that multispectral sensors that do not include these wavelengths are potentially missing important information.

3.2 Optimal wavelengths for each date

Tables 2 and 3 list the ten best bands for each date for both reflectance data and continuum removed data, respectively. Figure 5 reproduces this information graphically. The most striking aspect of the reflectance data (Figure 5a, Table 2) is that the location of the best bands changes over the growing season. Wavelengths are generally well-distributed throughout the spectrum, but the NIR region becomes less important over the last three dates (August 1, 12 and 24). It is notable that a band at 1985 nm is present for all dates, and in all but June 8 and August 12, 1995 nm is also present. Also, a blue wavelength band from the region 401-431 nm is present in all dates except the first (May 13), and bands from the red edge of the NIR plateau are selected for all dates except August 24.

For the continuum removed data (figure 5b) the near infrared region is generally less important on the first two dates (May 13 and 20) and becomes more important over the growing season, in contrast to the pattern seen in the reflectance data, where near infrared is less important in late summer. In addition, bands from the red edge are only selected for six of the eleven dates. The wavelength region 1985-2065 nm is consistently important and highly ranked in the continuum removed dataset. There is at least one, and often many, bands from this region on each date. For example, there are five bands present from this region on May 13 and four on May 20. Also, more than half of the dates do not have any wavelengths from the lower range of the SWIR (1400-1900 nm).

3.3 Dates for which *Ailanthus* is most separable

3.3.1 All available wavelengths

Figure 6 shows the mean classification error for the 500 random forest classifications as well as the standard error of the mean at a 95% confidence interval for each date for the reflectance (Figure 6a) and continuum removal (Figure 6b) data. Three curves are shown on each graph representing classification error when the random forest algorithm uses all available wavelengths, the best wavelengths for each date, and the best wavelengths across all dates, respectively. Starting with the classification error when all wavelengths are available it is apparent in the reflectance data (Figure 6a) that the overall pattern of the change in accuracy over time is complex. It was anticipated that the lowest classification error would be found in the spring, when *Ailanthus* leaves tend to have a distinctive red color, but instead it was found that error was generally higher in spring and lower in late summer, with the most separable date falling on July 22. The least separable dates were May 29 and June 8, and their classification error is significantly higher than all dates in July and August.

In the continuum removed data seen in Figure 6b the lowest mean classification error is found on the first date (May 13), as expected, but it is not significantly different from the error on July 22. The highest error values are seen on June 18, August 12 and August 24.

3.3.2 Best wavelengths for each date

When the best bands are selected for each date independently July 22 is the date when *Ailanthus* reflectance is most separable from other species, and in general, mid-to-late summer is more separable than early in the growing season, though on August 1 and 24 separability is not significantly better than when all bands are made available to the classifier. (Figure 6a).

For the continuum removed data (Figure 6b) the dates with the lowest mean classification errors are May 13, May 20 and August 1. May 29, mid-to-late June and mid-to-late August have significantly higher error values.

3.3.3 Best wavelengths across all dates

When the ten best bands are selected across all dates *Ailanthus* reflectance is most separable from the four other species on July 22, and in general, mid-to-late summer is more separable than early in the growing season (Figure 6a).

The dates in Figure 6b (continuum removed data) with the lowest mean classification errors are May 13, July 11 and August 1. Late May, early June and mid-to-late August have significantly higher error values.

3.4 Effects of continuum removal

As previously discussed, the application of continuum removal resulted in the selection of a different set of wavelengths than the bands selected for the original reflectance data. This is true for a single set of bands across all dates (Figure 4), and for bands selected for each date individually (Figure 5). This result is consistent with Psomas et al. (2005), who found that with continuum removal a different set of bands were selected in their multi-temporal analysis of grassland separability. In this study the application of continuum removal resulted in different patterns of error values as well as different dates when *Ailanthus* is most and least separable. The changing importance of the NIR and SWIR between the reflectance and continuum removed data in Figures 2 and 3 appears to be the result continuum removal's suppression of albedo, especially in the areas where the convex hull is close to or touching the spectral curve.

Applying continuum removal in many cases reduced error, although not consistently. This is somewhat surprising considering the fact that applying continuum removal suppresses albedo information without providing any benefit due to normalizing for illumination, because

data were collected in reflectance format. Schmidt and Skidmore (2003) noted mixed results when applying continuum removal, with improved separability in the chlorophyll absorption regions and generally decreased separability in the NIR and SWIR. However, Schmidt and Skidmore (2001), who found an increased number of species were statistically separable after applying continuum removal, suggested that the benefit of continuum removal is that it reduced the high within species variance in the NIR plateau. Psomas et al. (2005) also found a general trend towards increased accuracy after applying continuum removal. The improved results obtained through the use of continuum removal are encouraging because they imply that continuum removal is a viable alternative to more complex methods of converting aerial imagery radiance data to reflectance.

3.5 Effects of band selection

Band selection has a notable effect in reducing classification error for the reflectance data, in that for every date the mean classification error is significantly lower for the ten best bands selected across all dates compared to making all bands available to the random forest algorithm. When all bands are made available to the random forest classifier the algorithm performs a type of band selection as part of the analysis. Therefore, it can be interpreted that band selection using LARS is preferable to the band selection inherent within the random forest algorithm for these data. In contrast to the results for the reflectance data band selection has little effect for the first five dates for the continuum removed dataset. However, from June 30 onward the ten best bands selected for each date resulted in significantly lower error with the exception of July 22, when the mean classification error is significantly higher. Since one would expect that bands selected for each date individually would have lower classification error than a uniform set of bands selected across all dates, it is surprising that for the reflectance data the lowest classification error in most cases was observed when bands were selected across all sampling dates (Figure 6a), although this difference is only statistically significant on June 8, August 1 and August 24. In the continuum removed data selecting bands by date is generally more accurate than selecting a single set of bands across all dates, but this difference is only statistically significant in late August. In summary it appears that LARS is an effective band selection technique, particularly for reflectance data.

14

3.6 General discussion

For both reflectance and continuum removal datasets the single set of optimal bands selected from the across the growing season (figure 4) included bands selected from each of the major regions of the spectrum that are associated with major structural and chemical properties of plants. Specifically, the selected bands included the visible range, which is dominated by leaf pigments and especially chlorophyll; the NIR, which is affected by leaf structure; and the SWIR, where the overall reflectance is governed by leaf water content. In addition, the red edge of the NIR plateau appears to be important when a single set of bands is selected across all dates for both the reflectance data and the continuum removed data. Bands chosen from the red edge are also generally important for the reflectance data and less important for the continuum removed data when a set of bands is selected for each date. The importance of the red edge in this work is similar to findings by Schmidt and Skidmore (2001) and Daughtry and Walthall (1998).

Looking more closely at the individual bands chosen in the lists of optimal bands it seems likely that many of the bands identified as important were chosen because they fall within absorption features related to certain plant chemicals. For example, the short wavelength end of the blue (401-431 nm) among the most consistently identified optimal spectral regions in this study, is a region of chlorophyll absorption (Tucker & Garratt 1977, Woolley 1971). This focus on the blue wavelengths differs from the findings of Schmidt and Skidmore (2001), who found that red (550-680 nm) was most important in their reflectance data. In fact, red is notable in this work for its general absence in the bands selected from the reflectance data and only occasional presence in the bands selected from the continuum removed dataset. The band centered at 1115 nm, which is near an absorption feature associated with lignin (Curran 1989), is important for both the reflectance and continuum removed data when choosing a set of wavelengths across the growing season and for each individual date. Absorption features associated with protein and nitrogen are found throughout the SWIR, particularly near 1980 nm and 2060 nm (Curran 1989), and in this work the band centered on 1985 nm is among the most important bands for both datasets, whether one chooses a set of wavelengths for the entire set of sampling dates or for each individual date. One or both of the bands centered on 2055 nm and 2065 nm also appear in five of eleven dates when a set of bands is selected for each date from the continuum removed data. The wavelength bands centered at 1985 nm and 1995 nm stand out for their near-consistent selection and high rank in the single set of bands selected across all dates as well as the set of

bands selected for each date in both the reflectance and continuum removed datasets. Both of these bands are associated with leaf water (Curran 1989).

The results of this research have implications for resources managers and scientists choosing sensors for image acquisition for mapping invasive species. Judging by the broad distribution across the spectrum of optimal bands selected it seems important to choose a sensor that is sensitive across the full spectrum, from blue to the SWIR, and that is not limited to only the visible and NIR, for example. When a single set of bands was selected across all sampling dates four bands were common for both the reflectance and continuum removed datasets as seen in Table 1: 413 nm, 1115 nm, 1985 nm and 1995 nm. The same general pattern is apparent when choosing a set of wavelengths for each date individually for the reflectance data, with bands often selected from the blue (401 to 431 nm), the NIR (1115 nm) and SWIR (1985 nm and 1995 nm). When bands are selected from the continuum removal dataset for each date the same wavelengths are usually selected, with the addition of 1995 nm to 2070 nm and 2285 nm to 2350 nm in the SWIR.

While there is a degree of commonality in bands selected, as discussed above, there is also some variation, suggesting that limiting a collection to specific wavelength bands may not be the best approach. Instead, working on a scene-by-scene basis by collecting all bands using a sensor with broad wavelength sensitivity and then performing feature selection after data collection is advisable.

Mapping individual trees generally implies that high spatial resolution imagery is needed. However, satellite-borne high spatial resolution sensors are generally limited to four multispectral bands that do not include SWIR wavelengths. The high spatial resolution satellite-borne sensor with the largest number of spectral bands currently is WorldView-2, which has eight bands that extend from 400 nm to 1040 nm (DigitalGlobe 2009). The limited spectral range of high spatial resolution sensors suggests that hyperspectral aerial sensors will continue to be optimal for mapping and identification of invasive species.

Finding the date (or dates) when *Ailanthus* is most spectrally separable from the four other species was one of the main objectives of this work, and it was hypothesized that *Ailanthus* would be most separable in the spring. In contrast, the data show that late summer had the best separability in general with the exception of May 13 for the continuum removed data. Thus, there appears to be a window in the month of July for the continuum removed dataset and in late

July and August for the reflectance data (when band selection is applied) where image acquisition would potentially be most useful. The presence of bands selected from the edges of leaf water absorption features might suggest that the high separability in the summer was related to drought. However, weather and climate data indicate that the study area had four to nine centimeters above average precipitation in the months of May, June and July 2008, although August precipitation was four centimeters below the average (National Climatic Data Center 2008). A review of the photographs of the leaves from each date showed no obvious visual change in the leaves during the late summer either.

4. CONCLUSIONS

This study investigated the potential for discriminating the invasive species *Ailanthus altissima* from four native tree species in the spring and summer. The accuracy of identification of *Ailanthus* was found to vary both with date and combination of bands used in the classification, and the optimal subset of bands tended to be drawn from each of the major regions of the spectrum that are important for the remote sensing of vegetation, including the visible, red edge, NIR and SWIR. For certain dates some regions have reduced importance. However, this pattern of reduced importance is not consistent between continuum removal and the original reflectance data. For example, in the reflectance data, SWIR bands became more important in the reflectance data towards the end of the growing season (from August 1 onward), whereas in the continuum removed data the date with the most bands drawn from the SWIR is May 13.

Our expectation was that, based on anecdotal visual observations, early spring would be best for separability and error would increase into the summer. The observed pattern, however, is much more complex. For reflectance data the lowest error was found on July 22, and error in the spring was significantly higher. This is true both when all available bands are available to the classifier and when the ten best bands across all dates are used. The continuum removed data behaved more as expected, with low error in the early spring but also showed comparatively low error in July and early August before increasing again for the last two dates (August 12 and 24).

Continuum removal as a data transformation has a notable impact on the bands selected. Continuum removal generally results in lower classification error than seen with the original reflectance data when all spectral bands are made available to the random forest classifier.

Because methods of processing image data to reflectance are often complex, continuum removal would appear to be acceptable as an alternative and simple preprocessing step.

LARS was found to be an effective band selection tool, and it should be investigated further for remote sensing applications. Band selection seems to be more important for reflectance data in that it results in greater improvement in classification accuracy, though band selection can also have a positive effect for continuum removed data, but this is not always the case. In summary, these findings suggest that hyperspectral sensors with broad wavelength sensitivity are important for mapping and identification of *Ailanthus.*

5. REFERENCES

ASD (1997). ASD Technical Guide, 3rd Edition. Analytical Spectral Devices, Inc. Boulder, Colorado.

Asner, G.P., Jones, M.O., Martin, R.E., Knapp, D.E. & Hughes, R.F. (2008a). Remote sensing of native and invasive species in Hawaiian forests. *Remote Sensing of Environment* 112, 1912-1926.

Asner, G.P., Knapp, D.E., Kennedy-Bowdoin, T., Jones, M.O., Martin, R.E., Boardman, J. & Hughes, R.F. (2008b). Invasive species detection in Hawaiian rainforests using airborne imaging spectroscopy and LiDAR. *Remote Sensing of Environment,* 112, 1942-1955.

Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5– 32.

Bunn, A. G., Goetz, S. J., & Fiske, G. J. (2005). Observed and predicted responses of plant growth to climate across Canada. *Geophysical Research Letters*, 32, L16710.

Call, L.J. & Nilsen, E.T. (2005). Analysis of interactions between the invasive tree-of-heaven (*Ailanthus altissima*) and the native black locust (*Robinia pseudoacacia*). *Plant Ecology* 176, 275-285.

Chen, X., Warner, T.A. & Campagna, D. (2010). Integrating visible, near-infrared and short-wave infrared hyperspectral and multispectral thermal imagery for geological mapping at Cuprite, Nevada: A rule-based system. *International Journal of Remote Sensing*, 31(7), 1733-1752.

Clark, R.N. & Roush, T.L. (1984). Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. *Journal of Geophysical Research,* 89 (B7), 6329-6340.

Cochrane, M.A. (2000). Using vegetation reflectance variability for species level classification of hyperspectral data. *International Journal of Remote Sensing,* 21 (10), 2075-2087.

Curran, P.J. (1989). Remote sensing of foliar chemistry. *Remote Sensing of Environment*, 30, 271-278.

Curtiss, B. & Goetz, A.F.H. (1994). Field spectrometry: Techniques and instrumentation. *International Symposium of Spectral Sensing Research*, 31-40.

DigitalGlobe. (2009). WorldView-2. DigitalGlobe, Longmont, CO. http://www.digitalglobe.com/digitalglobe2/file.php/786/WorldView2-DS-WV2.pdf (last date accessed: 7/15/2010).

Daughtry, C.S.T. and C.L. Walthall. (1998). Spectral discrimination of *Cannibis sativa* L. leaves and canopies. *Remote Sensing of Environment*, 64, 192-201.

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, 32 (2), 407-499.

Ghimire, B., Rogan, J. & Miller, J. (2010). Contextual land-cover classification: incorporating spatial dependence in land-cover classification models using random forests and the Getis statistic. *Remote Sensing Letters*, 1 (1), 45-54.

Gitelson, A.A., Merzlak, M.N. & Chivkunova, O.B. (2001). Optical properties and nondestructive estimation of anthocyanin content in plant leaves. *Photochemistry and Photobiology*, 74 (1), 38-45.

Hastie, T. & Efron, B. (2007). LARS: Least angle regression, lasso and forward stagewise. R package version 0.9-7. http://www-stat.stanford.edu/~hastie/Papers/#LARS

Heisey, R.M. (1990). Allelopathic and herbicidal effects of extracts from tree of heaven. *American Journal of Botany,* 77 (5), 662-670.

Hughes, G.F. (1968). On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Informational Theory*, IT-14, pp. 55-63.

ITT Visual Information Solutions. (2007). ENVI User's Guide, ENVI 4.4. ITT Visual Information Solutions. Boulder, CO.

Kota, N.L., Landenberger, R.E. & McGraw, J.B. (2007). Germination and early growth of *Ailanthus* and tulip poplar in three levels of forest disturbance. *Biological Invasions,* 8, 197-211.

Kowarik, I. (1995). Clonal growth in *Ailanthus altissima* on a natural site in West Virginia. *Journal of Vegetation Science,* 6, 853-856.

Kowarik, I. and Säumel, I. (2007). Biological flora of Central Europe: *Ailanthus altissima* (Mill.) Swingle. *Perspectives in Plant Ecology, Evolution and Systematics,* 8, 207-237.

Landenberger, R.E., Kota, N.L. & McGraw, J.B. (2007). Seed dispersal of the non-native invasive tree *Ailanthus altissima* into contrasting environments. *Plant Ecology,* 192 (1), 55-70.

Landenberger, R.E., Warner, T.A. & McGraw, J.B. (2009). Spatial patterns of female *Ailanthus altissima* across an urban-to-rural land use gradient. *Urban Ecosystems,* 12 (4), 437-448.

Lawrence, R.L., Wood, S.D. & Sheley, R.L. (2006). Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (RandomForest). *Remote Sensing of Environment*, 100, 356-362.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2/3, 18– 22.

Lewis, M. (2002). Spectral characterization of Australian arid zone plants. *Canadian Journal of Remote Sensing,* 28 (2), 219-230.

National Climatic Data Center. (2008). West Virginia 2008. *Climatological Data Annual Summary,* 116 (13). 24 pp.

Niemalä, P. & Mattson, W.J. (1996). Invasion of North American forests by European phytophagous insects. *BioScience*, 46 (10), 741-753.

Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26, 217–222.

Palacios-Orueta, A. & Ustin, S.L. (1996). Multivariate statistical classification of soil spectra. *Remote Sensing of Environment*, 57, 108-118.

Pimentel, D., Zuniga, R. & Morrison, D. (2005). Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics*, 52, 273-288.

Pinard, V. & Bannari, A. (2003). Spectroradiometric analysis in a hyperspectral use perspective to discriminate between forest species. *Geoscience and Remote Sensing Symposium, IGARSS '03. Proceedings. 2003 IEEE International*, 7, 4301-4303.

Price, J.C. (1994). How unique are spectral signatures? *Remote Sensing of Environment,* 49, 181-186.

Psomas, A., Zimmermann, N.E., Kneubuhler, M., Kellenberger, T. & Itten, K. (2005). Seasonal variability in spectral reflectance for discriminating grasslands along a dry-mesic gradient in Switzerland. *Proceedings of the 4th EARSEL Conference on Imaging Spectroscopy in Warsaw, Poland 2005.*

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rentch, J.S., Fortney, R.H., Stephenson, S.L., Adams, H.S., Grafton, W.N. & Anderson, J.T. (2005). Vegetation – site relationships of roadside plant communities in West Virginia, USA. *Journal of Applied Ecology*, 42, 129-138.

Sax, D.F. & Gaines, S.D. (2008). Species invasions and extinction: The future of
native biodiversity on islands. *Proceedings of the National Academy of Sciences,*
105, 11490-11497.

Schaepman-Strub, G., Schaepman, M.E., Martonchik, J.V., Painter, T.H. & Dangel, S. (2009).
Radiometry and reflectance: From terminology concepts to measured quantities.   In
T.A. Warner, M.D. Nellis & G.M. Foody (Eds.), *SAGE Handbook of Remote Sensing*
(Chapter 15, p. 215-228).   London, UK: SAGE.

Schmidt, K.S. & Skidmore, A.K. (2001). Exploring spectral discrimination of grass
Species in African rangelands. *International Journal of Remote Sensing,* 22 (17),
3421-3434.

Schmidt, K.S. & Skidmore, A.K. (2003). Spectral discrimination of vegetation types in
a coastal wetland. *Remote Sensing of Environment,* 85, 92-108.

Tucker, C.J. & Garratt, M.W. (1977). Leaf optical system modeled as a stochastic process.
Applied Optics, 16 (3), 635-642.

USGS (2008). Spectral Characteristics Viewer.
http://landsat.usgs.gov/tools_spectralViewer.php (last date accessed: 6/16/2010).

van Aardt, J.A.N. & Wynne, R.H. (2001). Spectral separability among six southern
tree species. *Photogrammetric Engineering & Remote Sensing,* 67 (12),
1367-1375.

Vitousek, P.M. (1994). Beyond global warming: Ecology and climate change.
Ecology, 75 (7), 1861-1876.

Vitousek, P.M., D'Antonio, C.M., Loope, L.L., Rejmanek, M. & Westbrooks, R. (1997).
Introduced species: A significant component of human-caused global change. *New
Zealand Journal of Ecology,* 21 (1), 1-16.

Warner, T.A. (2010). Remote sensing analysis: From project design to implementation. In J.D.
Bossler, R.B. McMaster, C. Rizos & J.B. Campbell (Eds.), *Manual of Geospatial
Sciences (2nd Edition)* (Chapter 17, 301-318).   London, UK: Taylor and Francis.

Warner, T. A. & Nerry, F. (2009). Does a single broadband or multispectral thermal data add
information for classification of visible, near- and shortwave infrared imagery of urban
areas? *International Journal of Remote Sensing*, 30 (9), 2155-2171.

Warner, T., Steinmaus, K. & Foote, H. (1999). An evaluation of spatial autocorrelation-based
feature selection. *International Journal of Remote Sensing*, 20 (8), 1601-1616.

Woolley, J.T. (1971). Reflectance and transmittance of light by leaves. *Plant Physiology*, 47,
656-662.

6. FIGURES
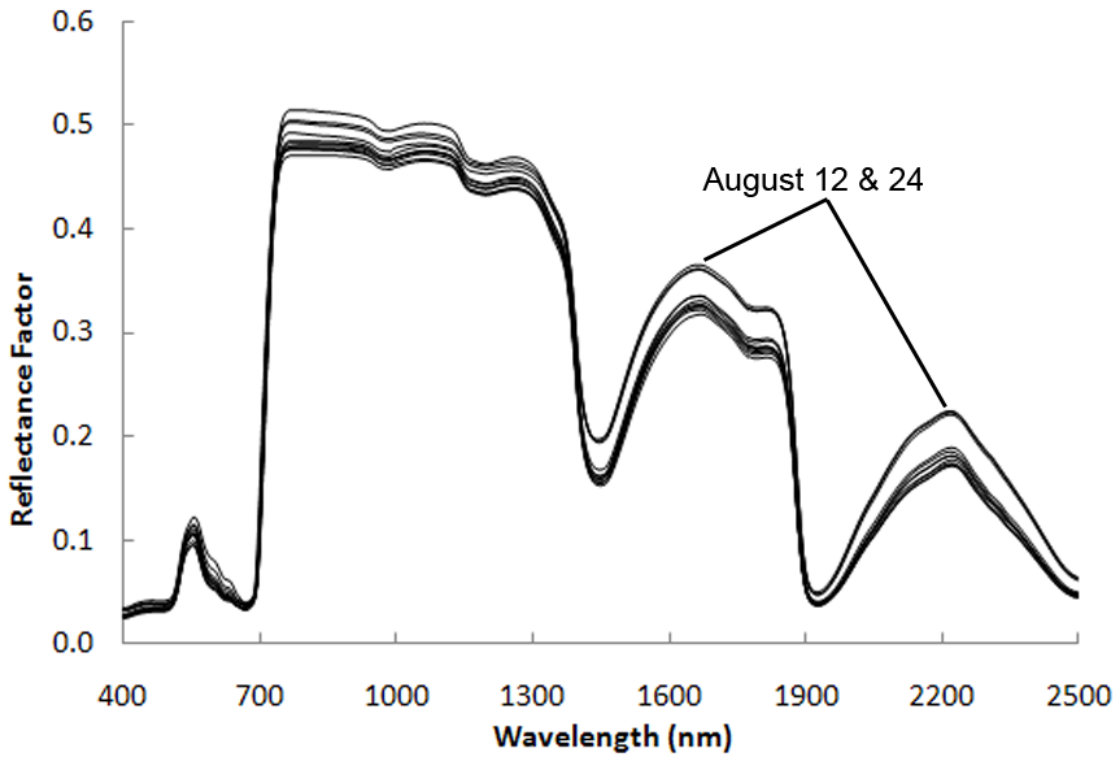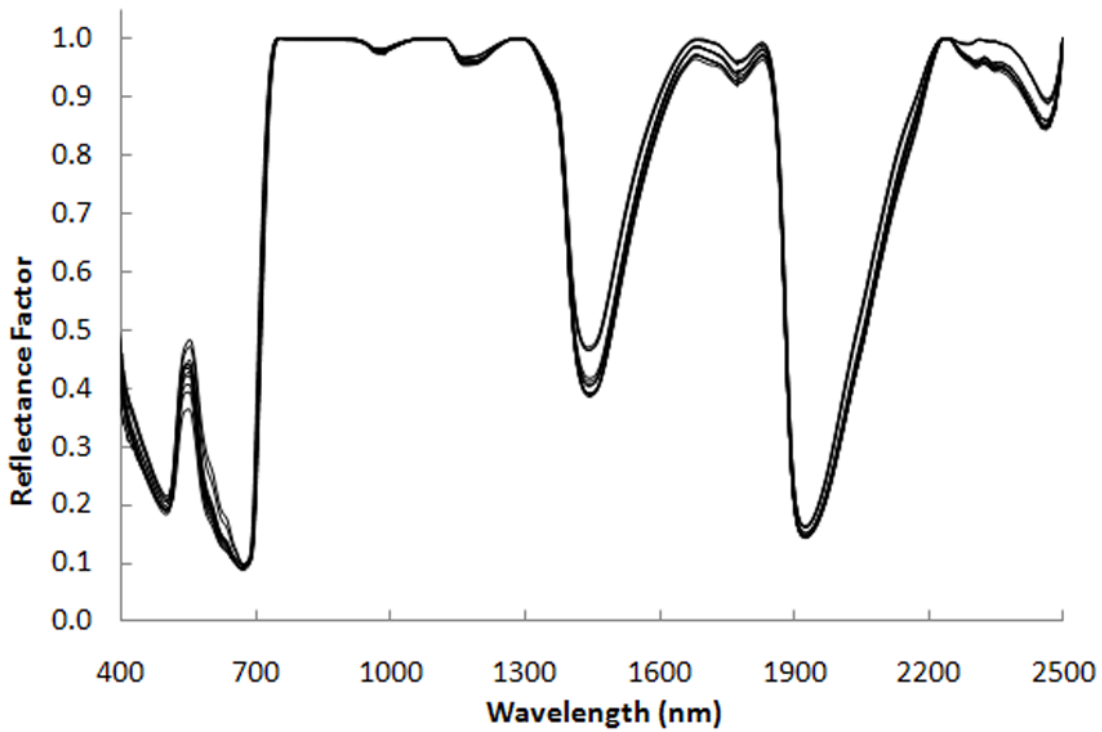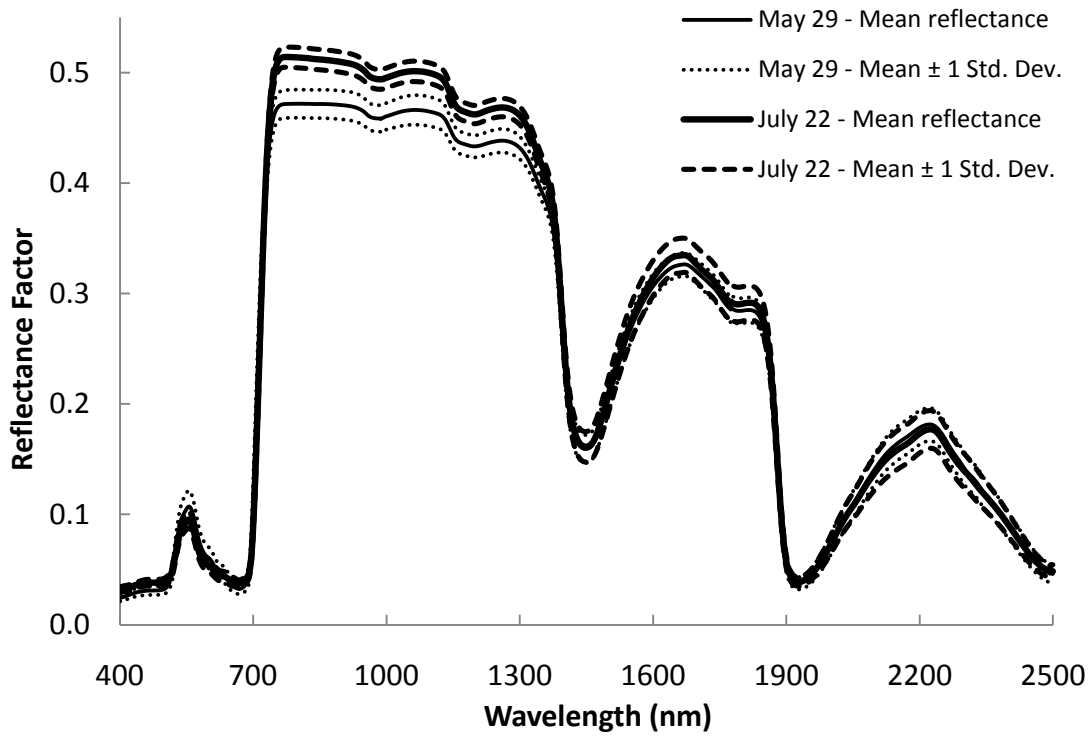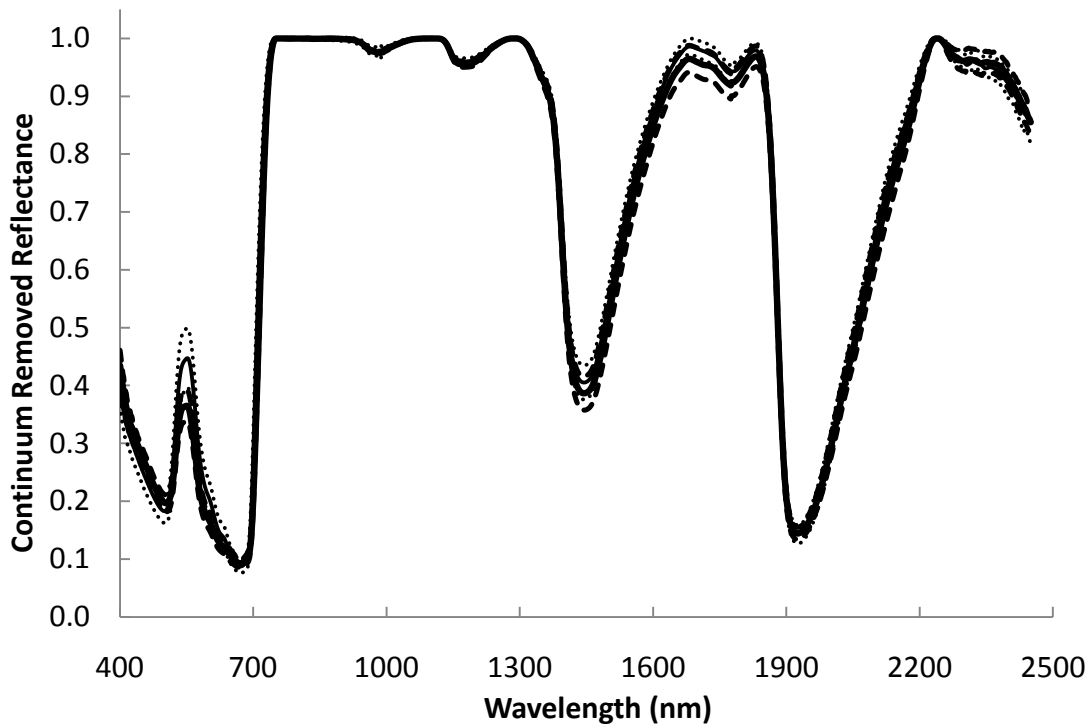

Figure 1. Mean *Ailanthus* spectra for the eleven collection dates for (a) reflectance data and (b) continuum removed reflectance data.
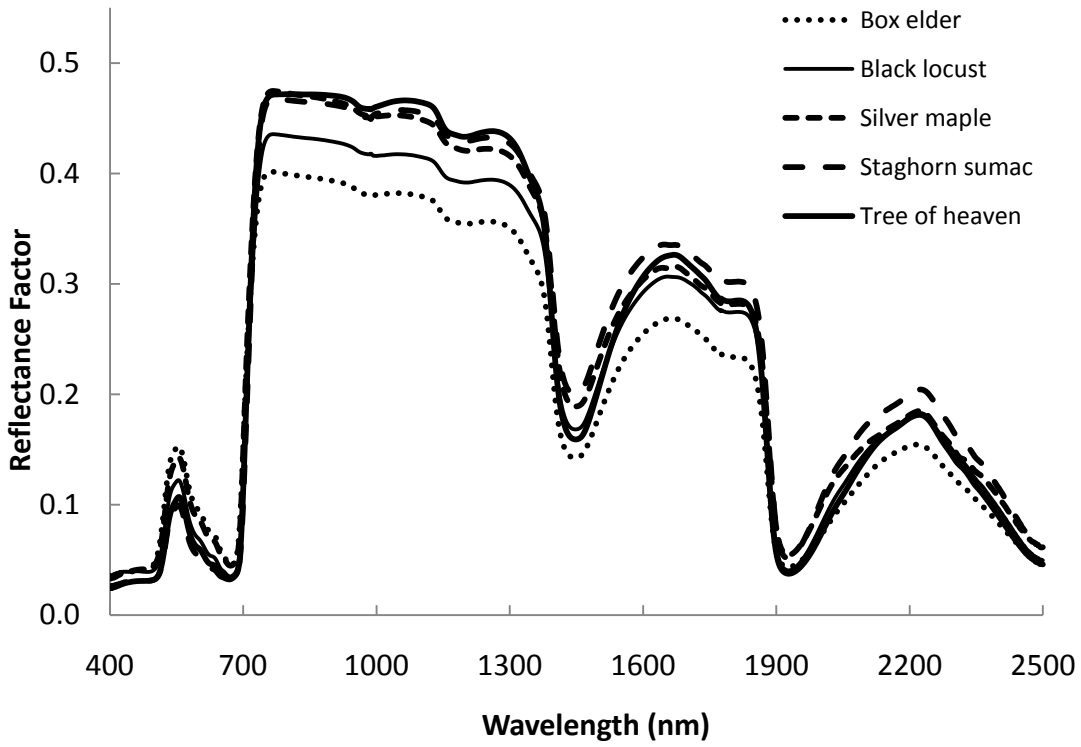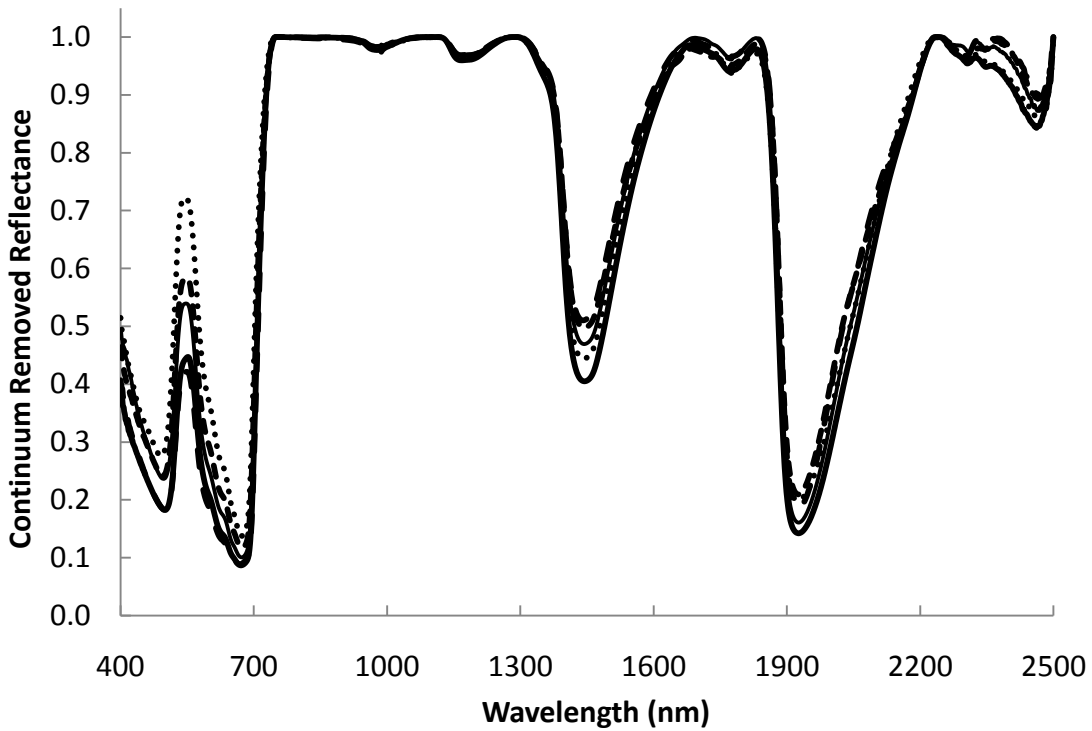
Figure 2. Variability within two dates (May 29 and July 22) for (a) reflectance data and (b) continuum removed reflectance data.
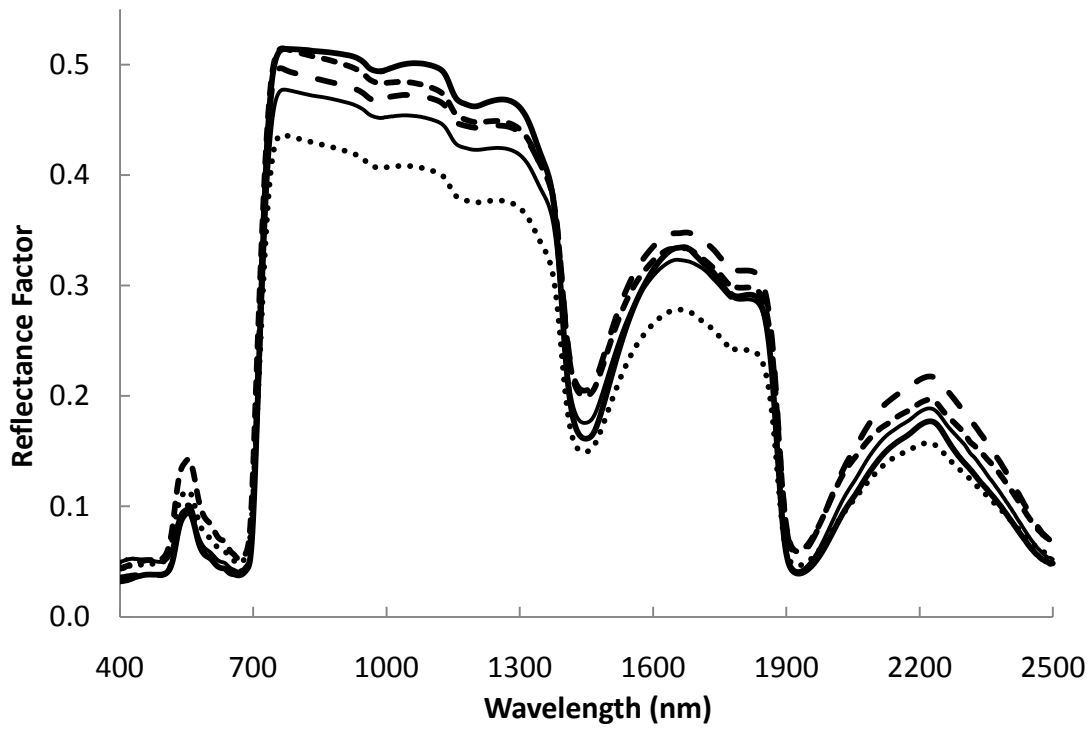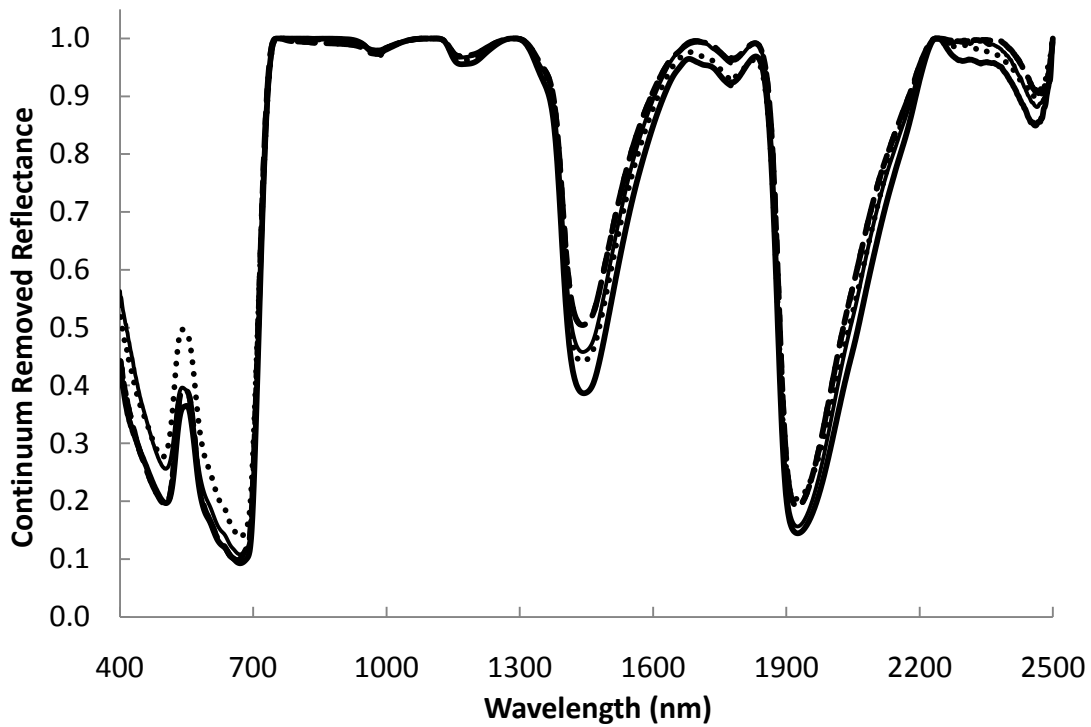
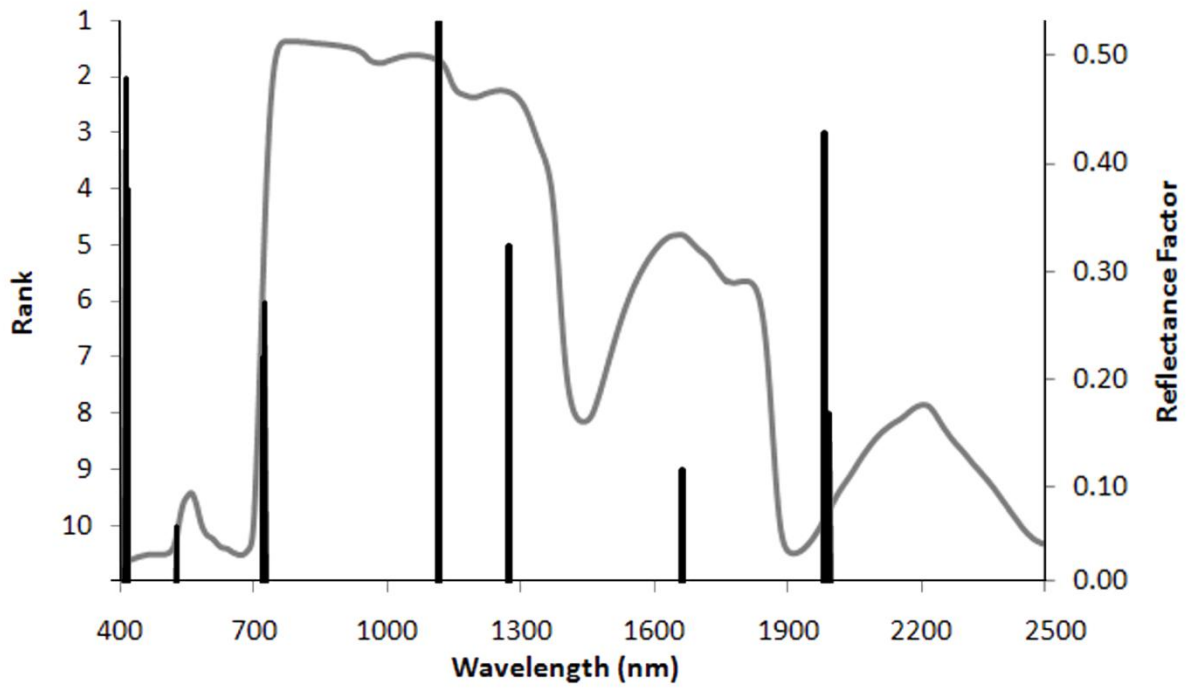Figure 3a & b. Spectra of the five species for May 29 for (a) reflectance data and (b) continuum removed data.
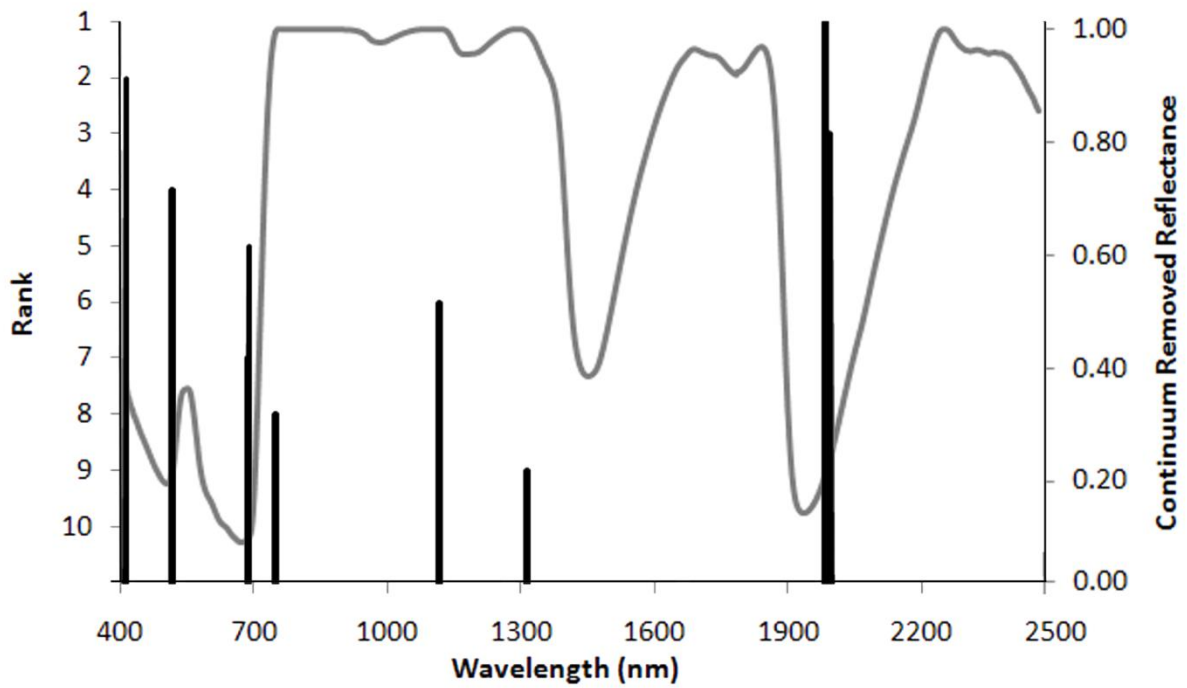
c

d

Figure 3c & d. Spectra of the five species for July 22 for (c) reflectance data and (d) continuum removed data.

Figure 4. The ten best bands across all sampling dates for both (a) reflectance data and (b) continuum removed data.
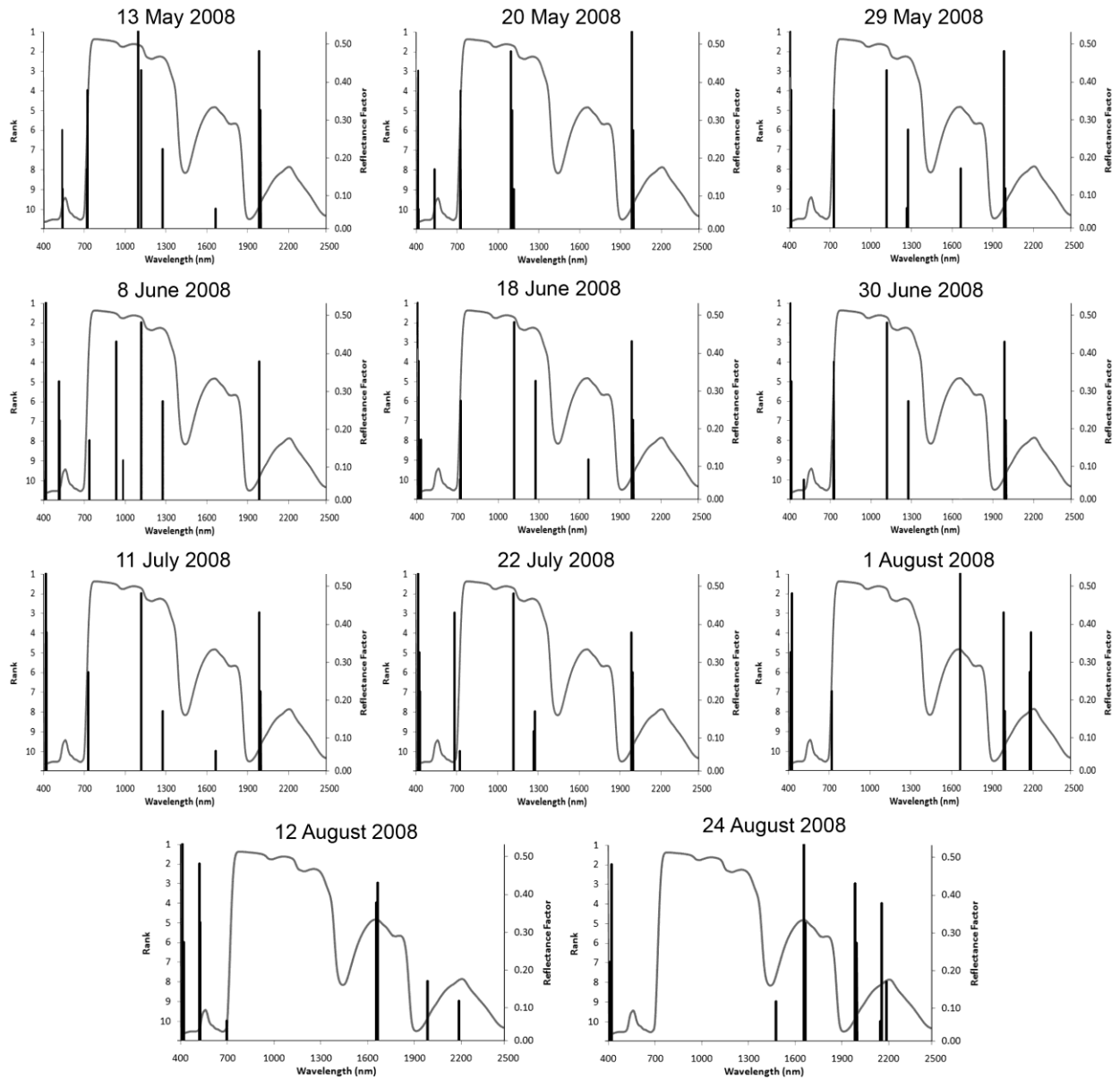
Figure 5a. Ten best bands for each date for the reflectance data.
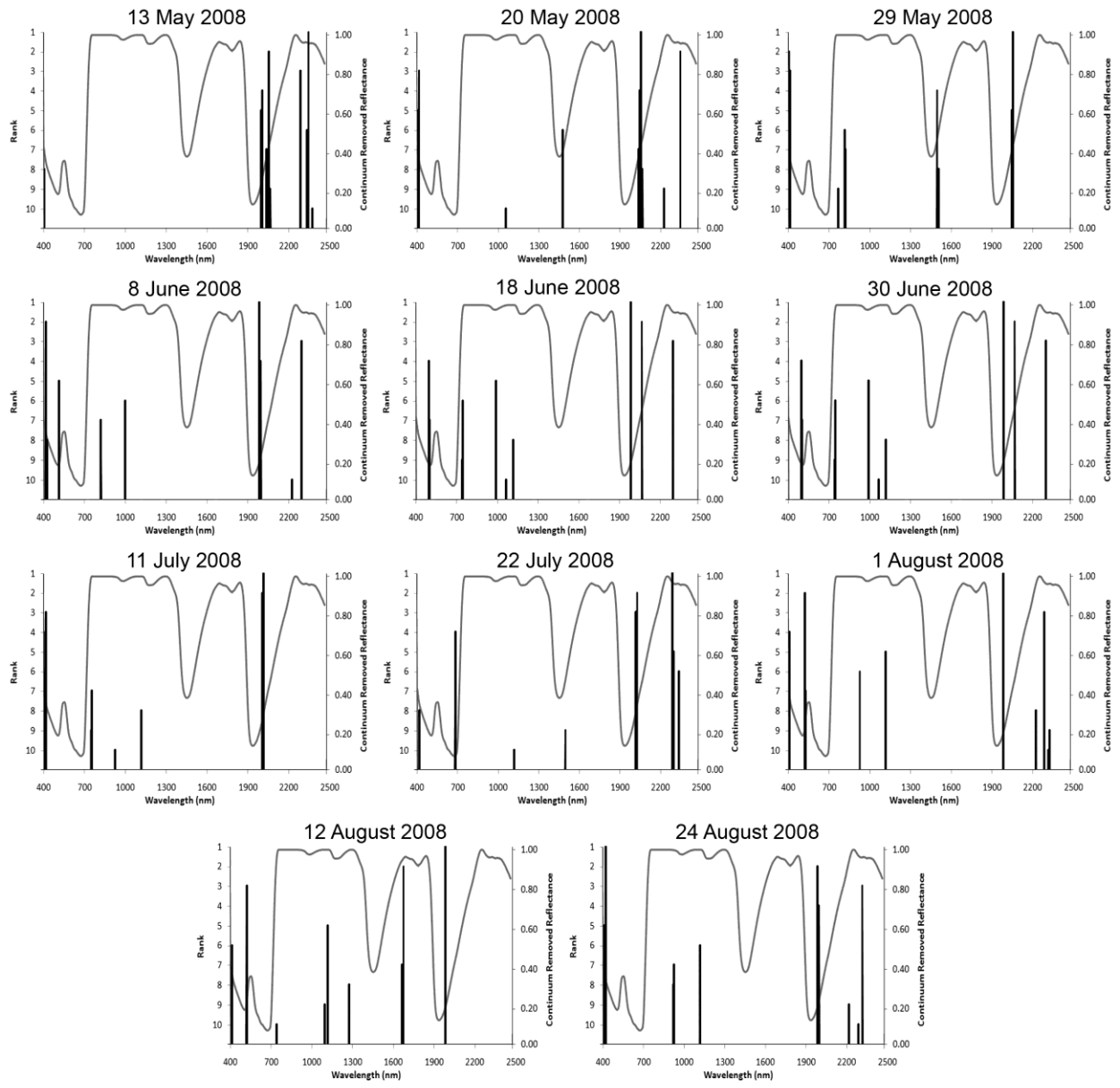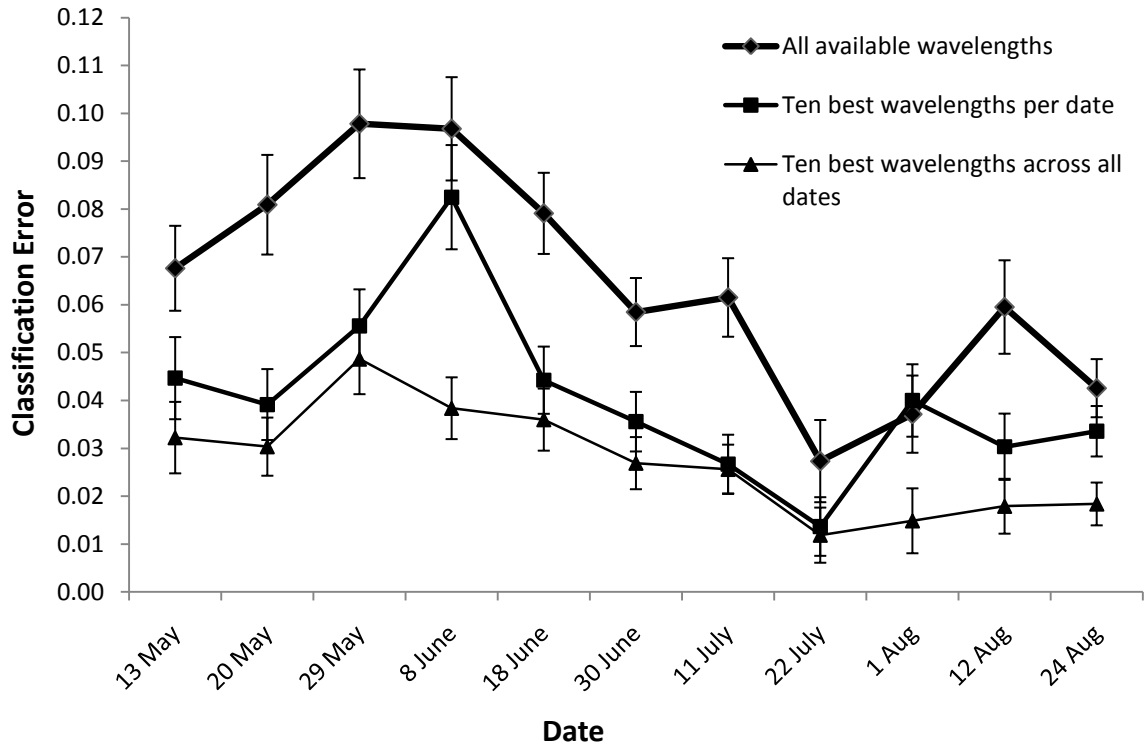
Figure 5b. Ten best bands for each date for the continuum removed data.

a

b

Figure 6. Average classification error per date for the random forest algorithm for (a) reflectance data and (b) continuum removed data.

7. TABLES

Table 1. Ten best bands across all sampling dates for both reflectance data and continuum removed data. The data are ordered in increasing wavelengths to facilitate comparison between the two datasets. Figure 4 displays this information graphically.

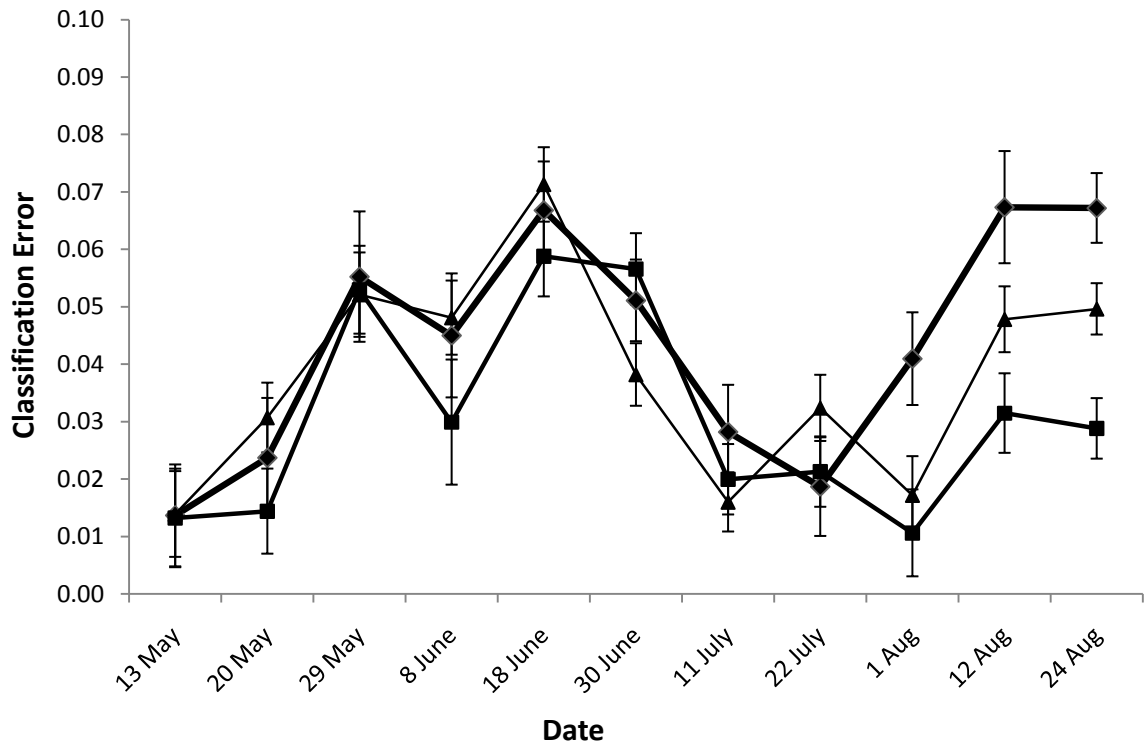| Reflectance		Continuum Removed	
Wavelength (nm)	Rank	Wavelength (nm)	Rank
413	2	413	2
416	4	515	4
527	10	518	10
722	7	686	7
725	6	689	5
1115	1	749	8
1275	5	1115	6
1665	9	1315	9
1985	3	1985	1
1995	8	1995	3

Table 2. Ten best bands for each date for the reflectance data. The data are ordered in increasing wavelengths to facilitate comparison between this table and Table 3. Figure 5a displays this information graphically.

					Date					
13-May	20-May	29-May	8-Jun	18-Jun	30-Jun	11-Jul	22-Jul	1-Aug	12-Aug	24-Aug
536	413	404	410	404	401	416	416	413	410	407
539	416	410	416	410	404	419	425	416	419	419
719	533	725	512	428	410	422	431	419	422	1475
722	722	728	515	719	503	725	683	425	521	1655
1095	725	1115	734	722	722	728	722	719	524	1665
1115	1095	1265	932	1115	725	1115	1115	1665	695	1985
1275	1105	1275	983	1275	1115	1275	1265	1985	1655	1995
1665	1115	1665	1115	1665	1275	1665	1275	1995	1665	2145
1985	1985	1985	1275	1985	1985	1985	1985	2175	1985	2155
1995	1995	1995	1985	1995	1995	1995	1995	2185	2185	2185

Table 3. Ten best bands for each date for the continuum removed data. The data are ordered in increasing wavelengths to facilitate comparison between this table and Table 2. Figure 5b displays this information graphically.

					Date					
13-May	20-May	29-May	8-Jun	18-Jun	30-Jun	11-Jul	22-Jul	1-Aug	12-Aug	24-Aug
404	401	404	416	494	497	404	416	407	410	407
1995	413	410	425	500	500	410	680	521	518	419
2005	1055	413	512	740	674	413	683	524	521	920
2035	1475	767	818	743	989	416	1115	926	740	923
2055	2035	815	821	989	992	749	1495	1115	1095	1115
2065	2045	818	998	1065	1985	752	2015	1985	1115	1985
2285	2055	1495	1985	1115	2055	923	2025	2225	1275	1995
2335	2065	1505	1995	1985	2065	1115	2285	2285	1665	2215
2345	2225	2045	2225	2065	2295	2005	2295	2315	1675	2285
2375	2345	2055	2295	2295	2335	2015	2335	2325	1985	2315