DISCRIMINATION OF LEAFY SPURGE (*EUPHORBIA ESULA* L.) AND PURPLE LOOSESTRIFE (*LYTHRUM SALICARIA* L.) BASED ON FIELD SPECTRAL DATA

A Paper
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Kathryn Marie Hooge Hom

In Partial Fulfillment of the Requirements
for the Degree of
MASTER OF SCIENCE

Major Program:
Natural Resources Management

December 2018

Fargo, North Dakota

North Dakota State University
Graduate School

Title

DISCRIMINATION OF LEAFY SPURGE (*EUPHORBIA ESULA* L.) AND PURPLE LOOSESTRIFE (*LYTHRUM SALICARIA* L.) BASED ON FIELD SPECTRAL DATA

By

Kathryn Marie Hooge Hom

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

MASTER OF SCIENCE

SUPERVISORY COMMITTEE:

Sreekala Bajwa
Chair

Rodney Lym

John Nowatzki

Approved:

4/8/2019
Date

Edward S. DeKeyser
Department Chair

ABSTRACT

Leafy spurge and purple loosestrife are noxious weeds that displace native vegetation. Herbicides are often applied to these weeds during flowering, making it ideal to identify them early in the season. This paper evaluates the spectral separability of the inflorescences and leaves of these plants from surrounding vegetation. Spectral data of leafy spurge, purple loosestrife, and surrounding vegetation were collected from sites in southeastern North Dakota. Partial least squares discriminant analysis (PLS-DA) was used to separate the spectral signatures of these weeds in the visible and near infrared wavelengths. Using PLS-DA the weeds were discriminated from their surroundings with R^2 values of 0.86 to 0.92. Analysis of the data indicated that the bands contributing the most to each model were in the red and red edge spectral regions. Identifying these weeds by the leaves allows them to be identified earlier in the season, allowing more time to plan herbicide application.

ACKNOWLEDGEMENTS

I would first like to thank the North Dakota Agricultural Experiment Station and the North Dakota Soybean Council for funding this research project. I would also like to thank Dr. Kevin Sedivec and Jamen Windish for permission and coordination to collect data on lands in southeastern North Dakota. I would like to extend a special thanks to Kathy Christianson for sharing her time and knowledge of noxious weeds in North Dakota, as well as Ali Shirzadifar, Brock Zenzen, and Austin Eide for their help with my data collection.

I would also like to thank my advisor, Dr. Sreekala Bajwa, and the members of my committee, John Nowatzki and Dr. Rodney Lym, for their knowledge and patience to help me through my research and writing. Finally, I would like to thank my husband, Kenneth, for his encouragement and support as I have pursued this degree.

DEDICATION

This paper is dedicated to my wonderful husband, Kenneth. He has believed in me and consistently supported me throughout this entire process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AISA	Airborne Imaging Spectroradiometer for Applications.
AVIRIS	Airborne Visible-Infrared Imaging Spectrometer.
B/G	Blue to green ratio.
CASI	Compact Airborne Spectrographic Imager.
CCD	Charge coupled device.
DA	Discriminant analysis.
EO-1	Earth Observing-1.
FWHM	Full width half maximum.
G/R	Green to red ratio.
GNDVI	Green normalized difference vegetation index.
ISODATA	Iterative Self-Organizing Data Analysis Technique.
ML	Maximum likelihood.
MNF	Minimum Noise Fraction.
MTMF	Mixture-tuned matched filtering.
NDDA	North Dakota Department of Agriculture.
NDVI	Normalized difference vegetation index.
NIR	Near-infrared.
OBIA	Object-based image analysis.
PCA	Principal component analysis.
PLSR	Partial least squares regression.
PLS-DA	Partial least squares discriminant analysis.
RGB	Red Green Blue.
SAM	Spectral angle mapper.

SIMCA.....Soft independent modeling of class analogy.

SNV .....Standard normal variate.

SPOT.....Satellite Pour l'Observation de la Terre.

UAV.....Unmanned aerial vehicle.

VI .....Vegetation index.

1. INTRODUCTION

1.1. Background

Invasive species can inhabit rangeland, ditches, and wetlands causing competition with the native vegetation for necessary resources (Goodell and Parker 2017; Levine et al. 2003). Identifying the areas infested with these noxious weeds is necessary for tracking them and controlling the infestations. Previous studies have shown remote sensing to be a possible solution to efficiently and effectively identify noxious weeds (Garcia-Ruiz et al. 2015; Hung et al. 2014; Mitchell and Glenn 2009; Peña et al. 2013). Aerial multispectral or hyperspectral imagery of land where noxious weeds grow can be analyzed to identify the spectral signatures of different plant species in the imagery, thereby identifying the noxious weeds of interest. Small unmanned aerial vehicles (UAV) are beginning to be used for the purpose of collecting imagery over land at a relatively low cost with many current models being user-friendly. Identifying noxious weeds in imagery with small UAVs would reduce the need for ground surveys and could potentially allow interested parties to survey more land than is possible by foot.

1.2. Noxious Weeds in North Dakota

Leafy spurge (*Euphorbia esula* L.) and purple loosestrife (*Lythrum salicaria* L.) are two major invasive plant species found in North Dakota, each with distinctly colored inflorescences that are identifiable from surrounding green vegetation (Bourchier et al. 2006; North Dakota Department of Agriculture 2016; Wilson et al. 2004). The North Dakota Department of Agriculture (NDDA) maintains a list of noxious weeds for the state of North Dakota, with leafy spurge and purple loosestrife being two of the weeds on the list (North Dakota Department of Agriculture 2016). These noxious weeds are invasive plant species that cause problems within the state or have a high potential to do so (Goodell and Parker 2017; Leistritz et al. 2004; North

Dakota Department of Agriculture 2016). If not managed, invasive species can replace native vegetation and take over large areas of land. Leafy spurge was estimated to infest over 400,000 hectares of land in North Dakota in 2017, and over 2 million hectares in the Rocky Mountain West and Northern Great Plains in 2006 (Bourchier et al. 2006; North Dakota Department of Agriculture 2017). Purple loosestrife infests over 1000 hectares in North Dakota, but has the potential to spread quickly in aquatic areas (North Dakota Department of Agriculture 2017). Small infestations of noxious weeds are easier and more economical to manage than large infestations; therefore, identifying these two noxious weeds early and effectively is important.

1.2.1. Leafy Spurge

Leafy spurge is a perennial plant that generally grows about 60 to 100 cm tall and produces yellow bracts in late May to early June and green flowers around mid-June (Bourchier et al. 2006; Hunt et al. 2007; Lym 1998). Many herbicides are often best applied to actively growing plants or during the true flower growth stage (Knezevic et al. 2004; Lym 1998; Mullin 1998). Leafy spurge spreads easily through both seeds and the root system and takes nutrients from the surrounding vegetation, making controlling this weed difficult (Bourchier et al. 2006; Lym and Messersmith 1985). Leafy Spurge often grows in rangeland and pastureland, along ditches, and in other uncultivated areas (Dunn 1979). If found in pastureland or hay eaten by animals, leafy spurge can be toxic (Bourchier et al. 2006; Dunn 1979).

1.2.2. Purple Loosestrife

Purple loosestrife is a perennial plant often found in aquatic environments, such as along rivers and wetlands (Blossey et al. 2001; Thompson et al. 1987; Wilson et al. 2004). During flowering, purple loosestrife has distinct purple flowers organized along spikes (Wilson et al. 2004). In North Dakota, these flowers generally appear in early July to the middle of September

(Thompson et al. 1987; Wilson et al. 2004). If not managed in these areas, this plant can cause damage to the wetland ecosystems by quickly replacing the native vegetation and affecting the habitat of wildlife that depend on the native vegetation, including waterfowl (Weihe and Neely 1997; Wilson et al. 2004). Controlling purple loosestrife can be complicated because any herbicides used in wetlands must have approval to be used near water. Herbicide application is often recommended while the plants are actively growing and blooming (Knezevic et al. 2004; Mullin 1998). Small infestations can be controlled with herbicides, manual removal, and plowing; however, controlling large infestations with these methods can be expensive and difficult (Welling and Becker 1993; Wilson et al. 2004).

1.2.3. Current Survey Methods

North Dakota county and city weed boards collect information about the noxious weeds in their specific areas (CA Penuel, NDDA, personal communication). These data are primarily collected by weed officers with the NDDA who conduct ground surveys and scout the land. They visually locate the weeds and log the GPS coordinates. The noxious weed information collected by NDDA is then shared with the public through an interactive map on the agency's website (CA Penuel, NDDA, personal communication; North Dakota Department of Agriculture 2016). Currently, aerial imagery and surveying with small UAVs has not been implemented as part of these scouting practices.

1.3. Research Significance

Remote sensing methods of detection that focus on specific plant characteristics, such as spectral signatures, shape, or textures, have the ability to distinguish noxious weeds from other vegetation. In recent years, new technology and studies have shown possibilities of making the detection of noxious weeds more efficient and accurate. Sensors and platforms are becoming

more advanced, which allows the spatial resolution of the imagery available for identifying weeds to be much higher than in the past. This has promising applications for detecting small infestations of the invasive species on the North Dakota noxious weed list before the infestations expand. Developing methods of detection and mapping using this technology for North Dakota means that the NDDA and other similar agencies will be able to detect and track noxious weeds more effectively than by visually identifying them through ground surveys and manually logging their locations.

Previous research attempting to identify leafy spurge and purple loosestrife has primarily used imagery collected from satellites or manned aircraft with spatial resolutions of 1.5 m or lower (Glenn et al. 2005; Mitchell and Glenn 2009; Swain et al. 2011). This paper proposes a method of identifying leafy spurge and purple loosestrife, by both the plant inflorescences and leaves, through the spectral signatures. This method has potential for furthering research into using spectral signatures in ultra-high spatial resolution imagery for scouting areas around crop fields and rangeland easily and efficiently.

1.4. Objective

The objective of this research was to classify leafy spurge and purple loosestrife noxious weeds based on spectral data manually collected in the field. To accomplish this objective, the parts of the plants that are most distinct and the best wavelengths to use in data collection to distinguish leafy spurge and purple loosestrife noxious weeds from other vegetation were identified.

1.5. Hypothesis

The main hypothesis of this paper is that both the inflorescences and leaves of leafy spurge and purple loosestrife are spectrally different from the surrounding vegetation and soil materials and can be distinguished by the plant spectral signatures from the surroundings.

1.6. References

Blossey B, Skinner L, Taylor J (2001) Impact and management of purple loosestrife (Lythrum salicaria) in North America. Biodivers Conserv 10:1787–1807

Bourchier R, Hansen R, Lym R, Norton A, Olson D, Bell Randall C, Schwarzländer M, Skinner L (2006) Biology and Biological Control of Leafy Spurge. US Department of Agriculture, Forest Service, Forest Health Technology Enterprise Team. 125 p

Dunn PH (1979) The distribution of leafy spurge (Euphorbia esula) and other weedy Euphorbia spp. in the United States. Weed Sci 27:509–515

Garcia-Ruiz FJ, Wulfsohn D, Rasmussen J (2015) Sugar beet (*Beta vulgaris* L.) and thistle (*Cirsium arvensis* L.) discrimination based on field spectral data. Biosyst Eng 139:1–15

Glenn NF, Mundt JT, Weber KT, Prather TS, Lass LW, Pettingill J (2005) Hyperspectral data processing for repeat detection of small infestations of leafy spurge. Remote Sens Environ 95:399–412

Goodell K, Parker IM (2017) Invasion of a dominant floral resource: Effects on the floral community and pollination of native plants. Ecol 98:57–69

Hung C, Xu Z, Sukkarieh S (2014) Feature learning based approach for weed classification using high resolution aerial images from a digital camera mounted on a UAV. Remote Sens 6:12037–12054

Hunt ER, Daughtry CST, Kim MS, Parker Williams AE (2007) Using canopy reflectance models and spectral angles to assess potential of remote sensing to detect invasive weeds. J Appl Remote Sens 1:013506

Knezevic SZ, Smith D, Kulm R, Doty D, Kinkaid D, Goodrich M, Stolcpart R (2004) Purple loosestrife (Lythrum salicaria) control with herbicides: Single-year application. Weed Technol 18:1255–1260

Leistritz FL, Bangsund DA, Hodur NM (2004) Assessing the economic impact of invasive weeds: The case of leafy spurge (Euphorbia esula). Weed Technol 18:1392–1395

Levine JM, Vilà M, D'Antonio CM, Dukes JS, Grigulis K, Lavorel S (2003) Mechanisms underlying the impacts of exotic plant invasions. Proc R Soc Lond B Biol Sci 270:775–781

Lym RG, Messersmith CG (1985) Leafy spurge control with herbicides in North Dakota: 20-year summary. J Range Manag 38:149–154

Lym, RG (1998) The biology and integrated management of leafy spurge (Euphorbia esula) on North Dakota rangeland. Weed Technol 12:367–373

Mitchell JJ, Glenn NF (2009) Leafy spurge (*Euphorbia esula*) classification performance using hyperspectral and multispectral sensors. Rangel Ecol Manag 62:16–27

Mullin BH (1998) The biology and management of purple loosestrife (Lythrum salicaria). Weed Technol 12:397–401

North Dakota Department of Agriculture (2016) Noxious Weeds. http://www.nd.gov/ndda/plant-industries/noxious-weeds. Accessed: November 1, 2016

North Dakota Department of Agriculture (2017) Weed Survey Report. www.agdepartment.vision-technology.com/weedsurvey/report.asp. Accessed: December 9, 2018

Peña JM, Torres-Sánchez J, de Castro AI, Kelly M, López-Granados F (2013) Weed mapping in early-season maize fields using object-based analysis of unmanned aerial vehicle (UAV) images. PLOS One 8:e77151

Swain S, Narumalani S, Mishra DR (2011) Monitoring invasive species: Detecting purple loosestrife and evaluating biocontrol along the Niobrara River, Nebraska. GIsci Remote Sens 48:225–244

Thompson DQ, Stuckey RL, Thompson EB (1987) Spread, Impact, and Control of Purple Loosestrife (Lythrum salicaria) in North American Wetlands. Washington, DC: US Department of the Interior, Fish and Wildlife Service. 55 p

Weihe PE, Neely RK (1997) The effects of shading on competition between purple loosestrife and broad-leaved cattail. Aquat Bot 59:127–138

Welling CH, Becker RL (1993) Reduction of purple loosestrife establishment in Minnesota wetlands. Wildl Soc Bull 21:56–64

Wilson LM, Schwarzlaender M, Blossey B, Bell Randall C (2004) Biology and Biological Control of Purple Loosestrife. US Department of Agriculture, Forest Service, Forest Health Technology Enterprise Team. 78 p

2. LITERATURE REVIEW

2.1. Introduction

Past research into identifying invasive plant species with aerial imagery has used a variety of spatial and spectral resolutions and classification methods (Bradley 2014; He et al. 2011; He et al. 2015; Thorp and Tian 2004). Platforms and sensors used in these studies have also differed, ranging from satellites to small UAVs, and from multispectral cameras to hyperspectral cameras. The platform and sensor technology used has an impact on the spatial and spectral resolutions of the images. Both the technology and classification methods have certain benefits and limitations for specific applications. Therefore, the best configuration to use for a project depends on the scope of the research and the resources available.

Plant species have spectral signatures, which often allow them to be separated from one another when these signatures are distinct (Bradley 2014). For this reason, much of the previous research has focused on analyzing these spectral signatures and their characteristics. However, some studies have utilized spatial features of the plants in imagery, including texture, shape, and other morphological features such as location with respect to crop rows (Hung et al. 2014; Peña et al. 2013; Pérez-Ortiz et al. 2015; Torres-Sánchez et al. 2013).

2.2. Spatial Resolution and Platforms

The spatial resolution of imagery has a significant impact on what plant characteristics can be identified in the imagery. Spatial resolution is roughly categorized as low, high, and ultra-high. The category is determined by the remote sensing platform used to hold the sensor, as well as the specifications of the sensor itself. Three common remote sensing platforms for collecting imagery used in weed detection are satellites, manned aircraft, and small UAVs, all of which have both benefits and drawbacks, depending on the specific purpose of the application.

2.2.1. Low Spatial Resolution

Low spatial resolution imagery is often collected from sensors on satellite or manned aircraft platforms, which can cover large swaths of land relatively quickly. In this paper, low spatial resolution refers to images with a pixel size of 20 m or more. Several studies have used low resolution imagery from satellite platforms, such as the Landsat, Satellite Pour l'Observation de la Terre (SPOT), and Earth Observing-1 (EO-1), to detect large infestations of invasive species, including saltcedar (*Tamarix* spp.*,* tamarisk) and leafy spurge (Carter et al. 2009; Diao and Wang 2016; Hunt and Parker Williams 2006; Mitchell and Glenn 2009). Low spatial resolution imagery from manned aircraft platforms, such as the Airborne Visible-Infrared Imaging Spectrometer (AVIRIS) sensor, were used to detect invasive species, including leafy spurge and downy brome (*Bromus tectorum* L.) (Hunt and Parker Williams 2006; Hunt et al. 2010; Noujdina and Ustin 2008; Parker Williams and Hunt 2002). Plot sizes used for accuracy assessments were based on the size of the imagery resolution of 20 m. Mitchell and Glenn (2009) obtained 78% accuracy in circular plots with 7.3 m radiuses and greater than 20% leafy spurge coverage, while several other studies used plots of 30 m^2 or greater, measuring the presence or absence of the invasive species in the plots and obtaining up to 88% accuracy (Carter et al. 2009; Hunt and Parker Williams 2006; Hunt et al. 2010). Imagery with spatial resolutions of 20-30 m allowed the researchers to investigate methods of detecting large infestations of these plants, but the resolution was inadequate to detect small infestations. Additionally, mixed pixels − pixels that were an average of the spectral data of the plants of interest and the surrounding vegetation, − affected the classification accuracy.

2.2.2. High Spatial Resolution

High spatial resolution imagery is also collected from sensors on either satellites or manned aircraft, although with different specifications from those that collect low spatial resolution. These systems are generally not able to cover as large an area, or as quickly, as lower resolution sensors on similar platforms. High spatial resolution in this paper refers to images with a pixel size of 1 m or greater but less than 20 m. Studies using high spatial resolution imagery for weed detection have generally used imagery ranging from 1.5 m to 4 m resolution (Casady et al. 2005; Swain et al. 2011). Sensors on both satellite and aerial platforms are able to attain high spatial resolution; however, manned aircraft are often able to collect data at higher spatial resolutions than satellites due to their proximity to the ground (Bradley 2014).

Satellite imagery, such as that from the QuickBird and IKONOS satellites, and aerial imagery from fixed-wing aircraft, such as the HyMap, the Compact Airborne Spectrographic Imager (CASI), and the Airborne Imaging Spectroradiometer for Applications (AISA), were evaluated to detect invasive species, including saltcedar, leafy spurge, spotted knapweed (*Centaurea maculosa* Lam.), and flowering purple loosestrife (Carter et al. 2009; Casady et al. 2005; de Castro et al. 2013; Glenn et al. 2005; Lass et al. 2005; Mirik et al. 2013; Mitchell and Glenn 2009; Swain et al. 2011). Plot sizes used for accuracy assessments were smaller than those used for low spatial resolution imagery. Mirik et al. (2013) validated the presence or absence of flowering musk thistle (*Carduus nutans* L.) with 1 m^2 plots with 91% accuracy, Swain et al. (2011) obtained 82% accuracy, validating with 2 m^2 plots, and Carter et al. (2009) achieved 91% accuracy with 10 m^2 plots. Imagery with high spatial resolutions allows researchers to identify smaller weed infestations with higher accuracy, ranging from 59% to 91%, compared to low

spatial resolution imagery, although this may be at the expense of increased collection time, financial cost, or larger amounts of data required to analyze and store.

2.2.3. Ultra-High Spatial Resolution

Ultra-high spatial resolution images are often gathered with sensors attached to small UAVs (UAVs weighing less than 25 kg), which fly at lower altitudes than either satellites or manned aircraft and generally collect imagery within 125 m above the ground. Ultra-high spatial resolution in this paper refers to pixel size of less than 1 m. This ultra-high spatial resolution data are more appropriate for detecting small patches or even individual plants (Bradley 2014). Small UAVs have become more popular due to the fast development of UAV technology since they were introduced to the civilian market. One benefit of small UAVs over satellite platforms is the versatility and potential to carry a variety of sensor payloads. Many small UAVs are also able to takeoff and land in a variety of locations and are not limited to paved runways, compared to fixed-wing manned aircrafts. However, while cloudy and rainy weather is a limiting factor for remote sensing from all platforms, small UAVs are also limited by high wind speeds more than satellites or manned aircrafts. Nevertheless, for studies requiring higher spatial resolution, such as weed identification, and time-critical applications, small UAVs are an attractive option.

UAV imagery with spatial resolutions of 1.3 mm to 20 cm were used to identify invasive plants, such as Canada thistle (*Cirsium arvense* L., creeping thistle), water hyacinth (*Eichhornia crassipes* Mart.), and other weeds in crop fields (Garcia-Ruiz et al. 2015; Hung et al. 2014; Lu and He 2018; Peña et al. 2013; Pérez-Ortiz et al. 2015; Tamouridou et al. 2017; Torres-Sánchez et al. 2013). Imagery was collected at a variety of altitudes ranging from 2 m to 115 m above ground level, with several researchers collecting data at an altitude of 30 m. Studies utilizing

ultra-high spatial resolution imagery were able to discriminate between invasive species and the surrounding vegetation with up to 95% accuracy.

2.2.4. Comparison of Platforms

Each of the remote sensing platforms has advantages and disadvantages, and the best one to use for a specific application is dependent on the objectives, scope, environment, and funding of the project. Some of the factors to consider while choosing a platform include sensor resolution, timing, and cost.

2.2.4.1. Sensor resolution

The distance of satellites and manned aircraft above the ground allows them to collect data in large swaths, acquiring large amounts of data in short periods of time. Due to the very high altitudes of satellites, the imagery they collect often have relatively lower spatial resolution compared to aerial platforms. However, some of the recent satellite platforms collect images in ultra-high spatial resolution, such as those deployed by Planet, which collect imagery with a resolution of up to 72 cm. The imagery collected from both satellites and manned aircraft can range in spatial resolution, depending on the sensor specifications (Bradley 2014). In addition, imagery collected from satellites and manned aircraft at high altitudes have higher atmospheric interference and must take into consideration atmospheric corrections to the imagery before analysis (Hunt and Parker Williams 2006; Hunt et al. 2010). On the other hand, small UAVs fly at low altitudes, often collecting imagery within 100 m above the ground, which removes the necessity for atmospheric corrections.

2.2.4.2. Timing

The timing of the data collection differs between the three platforms in terms of when and how often the data are collected, and the feasibility of collecting imagery over a specific spot

at a specific time. When identifying plant species in imagery, it is important to know the best point in the growth stage for plant identification, as well as to time the data collection during this growth stage. Improper timing can lead to less accurate results (Carter et al. 2009; Mladinich et al. 2006). For example, Laba et al. (2005) found that purple loosestrife, common reed (*Phragmites australis* Cav.), and cattail (*Typha* spp.) were most easily identifiable using the wavebands of 680 nm to 740 nm in late summer in upstate New York when the flower heads matured to seed, resulting in a distinguishable phenological change.

In addition to phenological and vegetation factors, environmental and technological factors also have an impact on imagery collection timing. Satellite imagery is not always available for the desired data collection time; and even if imagery is available, there is no guarantee cloud cover will not be present over part or the entire area of interest (Mladinich et al. 2006; Rodriguez-Moreno et al. 2017). Satellite sensors often cover large areas of land in regular intervals, allowing researchers to find historical image data, as well as to know in advance future data collection schedules. For example, Diao and Wang (2016) used a series of satellite imagery over several months to identify saltcedar in low spatial resolution imagery. Satellites with larger swath widths and lower spatial resolution have a higher likelihood of collecting imagery over a specific area within the desired time frame than satellites with smaller swath widths. Although the swath widths are narrower, manned aircraft and UAVs offer the ability to have more flexibility and control over the timing of flights, an advantage over satellites (Rodriguez-Moreno et al. 2017). This can alleviate certain issues, including limiting the likelihood of cloud cover in the imagery as well as having data available for a specific time period, such as during the flowering growth stage of a plant species being investigated. Compared to manned aircraft, small UAVs fly at very low altitudes and can fly under cloud cover, although there may still be cloud

shadows in the imagery. Due to the ease of use and versatility, small UAVs can collect imagery over the same land area at frequent time intervals. Small UAVs, however, are slower in covering similar areas compared to satellites or larger aircraft. They also tend to have a limited flight time due to battery life. If very large areas of data are needed, satellites and manned aircraft are more appropriate. However, the flexibility in data collection timing with manned aircraft and small UAVs makes these platforms more ideal when specific time periods or frequent data collections are needed.

2.2.4.3. Cost of imagery

The cost of imagery acquired from various platforms is a major factor affecting the viability for certain applications, such as plant species detection. These costs can vary a great deal and are dependent on a variety of factors, including spatial and spectral resolutions, cost of the platform and sensors, cost of data collection, accessibility, and availability. Imagery from some satellites, such as Landsat, is available free of charge, while other image data are available at a cost (Bradley 2014). Higher spatial resolution image data can be costly, such as WorldView 4, which costs about $25 per km^2 for 30 cm, 4-band imagery (K Nale, eMap International, personal communication). Similarly, while some data previously collected by manned aircrafts are freely available for new research, other data, such as that collected by the AVIRIS sensor, could still be expensive, especially the data at higher spatial resolutions (Bradley 2014). Some sensors on manned aircraft platforms, such as the AISA sensors, are available for a one-time purchase, but aircraft operation costs must then be factored into the overall cost. Similarly, small UAVs and attached sensors can be purchased once and used routinely without a need to purchase the imagery each time. Also, small UAV operating costs are generally lower than those of

manned aircraft. The overall cost of imagery collected with small UAVs is dependent on the cost of the UAV and sensor.

2.3. Spectral Resolution and Sensors

Imaging sensors used to detect plant species can be put into one of two categories: multispectral sensors, including red, green, blue (RGB) or visible light sensors, and hyperspectral sensors; both are able to detect light in a range of wavelengths along the electromagnetic spectrum (Thorp and Tian 2004). Both of these sensor types have benefits and drawbacks in regard to cost, accessibility, image resolution, and image quality, as well as the amount of information required to sort through, analyze, and store. Additional factors that can affect the performance of a sensor in an application include the number of bands, the wavelengths, and bandwidth the sensor is able to collect.

2.3.1. Multispectral

Multispectral sensors collect imagery in the visible light and infrared spectra in multiple spectral bands (Bradley 2014). Many of the studies using multispectral imagery used bands primarily between 400 nm and 1000 nm (Carter et al. 2009; Casady et al. 2005; Garcia-Ruiz et al. 2015; Peña et al. 2013; Pérez-Ortiz et al. 2015; Torres-Sánchez et al. 2013). However, multispectral Landsat imagery includes bands within the range of 400-2400 nm (Diao and Wang 2016; Hunt and Parker Williams 2006; Mitchell and Glenn 2009). The number of multispectral bands generally ranges from four to ten. Multispectral imagery is often acquired from satellite and manned aircraft platforms at lower cost and higher availability than imagery with higher spectral resolutions (Bradley 2014; Hunt and Parker Williams 2006).

Common multispectral sensors on satellite platforms that have been used in plant identification studies include the QuickBird, IKONOS, Landsat, and SPOT sensors, with spatial

resolution ranging from 2-30 m (Carter et al. 2009; Casady et al. 2005; de Castro et al. 2013; Diao and Wang 2016; Hunt and Parker Williams 2006; Mitchell and Glenn 2009). In recent years several studies aimed at plant species detection have used multispectral cameras to collect imagery at ultra-high spatial resolutions with small UAVs (Garcia-Ruiz et al. 2015; Lu and He 2018; Peña et al. 2013; Pérez-Ortiz et al. 2015; Torres-Sánchez et al. 2013). The spatial resolution of the imagery in these studies has ranged from sub-centimeter to 20 cm, and the wavelengths used in these studies were collected in the range of 400-1000 nm. A variety of classification methods were used, including partial least squares discriminant analysis (PLS-DA), which resulted in classification accuracies of up to 95% (Garcia-Ruiz et al. 2015). While many studies involving multispectral imagery did achieve high accuracies, the spectral resolution alone was not enough to attain these notable results and depended a great deal on the high spatial resolutions and classification strategies.

RGB sensors are multispectral sensors that collect images in three bands within the visible light wavelengths, which range from about 400 nm to 700 nm (Thorp and Tian 2004). Many studies investigating weed detection methods utilize visible wavelengths, but generally these wavelengths are included within a wider spectral range that also includes near-infrared (NIR) wavelengths (Bradley 2014; Thorp and Tian 2004). Nevertheless, Hung et al. (2014) used RGB imagery to classify several plants and relied on multiple features, including shape and texture of the plants. By combining the low spectral resolution of an RGB camera with ultra-high spatial resolution of less than 1 cm, water hyacinth and serrated tussock (*Nassella trichotoma* (Nees) Hack.) were classified with 94% and 93% accuracy, respectively. However, only 72% accuracy was attained for tropical soda apple (*Solanum viarum* Dunal) because there was not enough distinction between the invasive species and the surrounding vegetation. Mafanya et al.

(2017) achieved 87.7% accuracy using a maximum likelihood (ML) algorithm to identify an invasive species using ultra-high spatial resolution RGB imagery. Although RGB imagery alone has not been widely utilized in plant detection, Hung et al. (2014) and Mafanya et al. (2017) found that when paired with ultra-high spatial resolution, plant detection with RGB imagery does have potential, depending on which plant species require identification.

2.3.2. Hyperspectral

Hyperspectral sensors collect image data in the same electromagnetic range as multispectral sensors, but in hundreds of narrow spectral bands rather than only a few broad bands (Bradley 2014; He et al. 2011). Because of the high number of spectral bands used in these sensors, the imagery has very high spectral resolution. Due to this high spectral resolution, hyperspectral images can be used to detect subtle differences between the invasive species and the surrounding vegetation or soil (Bradley 2014). The high number of bands allows algorithms to use many more data points along the electromagnetic spectrum to detect differences between spectral signatures. However, more bands also mean more data, which can take longer to analyze and require more storage space. Hyperspectral sensors and imagery can be expensive, and data may not be readily available for a location of interest (He et al. 2011; Narumalani et al. 2009). By contrast, multispectral imagery is more readily available, and image analysis does not require the same level of proficiency as hyperspectral data (Hunt and Parker Williams 2006; Narumalani et al. 2009).

While analyzing large amounts of hyperspectral data can pose challenges, the large database can also provide more precise information, as evidenced by studies attaining accuracies of up to 94% when classifying hyperspectral imagery (Glenn et al. 2005; Narumalani et al. 2009). The number of bands of hyperspectral imagery used in invasive plant detection studies

ranged from 50 bands for the AISA Eagle sensor to 242 bands for the Hyperion sensor on the EO-1 satellite (Carter et al. 2009; Mirik et al. 2013). The hyperspectral sensors generally operate within the range of 350-2500 nm (Carter et al. 2009; Glenn et al. 2005; Hunt and Parker Williams 2006); although many sensors and plant monitoring applications only use bands up to 1000 nm (Lass et al. 2005; Mirik et al. 2013; Parker Williams and Hunt 2002; Swain et al. 2011). In addition to airborne sensors, some studies collected the spectral reflectance in the range of 400-2400 nm of plants using a ground spectrometer (Garcia-Ruiz et al. 2015; Shapira et al. 2010; Shapira et al. 2013; Shirzadifar et al. 2018; Ustin and Santos 2010). Several of these studies attained high classification accuracies; however, in several cases when compared with multispectral imagery of the same spatial resolution, classification of hyperspectral imagery had no clear benefit (Carter et al. 2009; Hunt and Parker Williams 2006; Mitchell and Glenn 2009).

2.3.3. Comparison of Multispectral and Hyperspectral Imagery

Several studies have compared multispectral and hyperspectral imagery for identifying specific plant species and have found that, while providing additional information that can help distinguish spectral signatures, analyzing hundreds of narrow bands can add noise that distracts from important spectral characteristics (Carter et al. 2009; Hunt and Parker Williams 2006; Hunt et al. 2007; Mitchell and Glenn 2009).

Comparisons of multispectral and hyperspectral imagery with varying spatial resolutions for classifying leafy spurge and saltcedar had similar results, even when using different classification methods (Carter et al. 2009; Mitchell and Glenn 2009). Mitchell and Glenn (2009) acquired high spatial resolution hyperspectral imagery over a study area with leafy spurge. They spectrally and spatially degraded the imagery and used the mixture-tuned matched filtering (MTMF) classification method to compare classification accuracy under varying spectral and

spatial resolutions. Accuracies achieved were 91% and 85% for the high spatial resolution multispectral and hyperspectral imagery, respectively. Carter et al. (2009) classified images using normalized difference vegetation index (NDVI) threshold values and ML methods at different spectral and spatial resolutions. The study did not include high spatial resolution hyperspectral imagery; however, based on the three categories that were compared, analysis of high spatial resolution multispectral imagery produced the highest classification accuracy of 91% with NDVI threshold values, similar to the results of Mitchell and Glenn (2009). Based on the results of these studies, high spatial resolution multispectral imagery produced better accuracies than hyperspectral imagery for plant species classification. The 88% accuracy of the low spatial resolution hyperspectral imagery in Carter et al. (2009) could have been lower than that of the 91% accuracy high spatial resolution multispectral imagery in Mitchell and Glenn (2009) in part because of the lower spatial resolution. However, another possible reason for this is the higher number of bands from the hyperspectral imagery, compared to multispectral imagery, can create more noise, causing poorer classification results (Mitchell and Glenn 2009).

While high spatial resolution multispectral imagery achieved better classification accuracies than hyperspectral imagery, comparisons between multispectral and hyperspectral imagery at low spatial resolutions had mixed results. At low spatial resolutions, some studies found higher accuracies with hyperspectral imagery (Carter et al. 2009; Hunt et al. 2007). Hunt et al. (2007) had higher success rates with hyperspectral imagery than multispectral imagery to identify flowering leafy spurge using spectral angle mapper (SAM). They achieved 74% overall accuracy with hyperspectral imagery, but the highest overall accuracy achieved for multispectral imagery was 61% when using only the visible and NIR bands. However, other studies had higher accuracies with multispectral imagery (Hunt and Parker Williams 2006; Mitchell and Glenn

2009). While hyperspectral imagery provides more spectral information that has potential to identify smaller variations in spectra, multispectral imagery may be less likely to add noise that can degrade results.

2.4. Classification Methods Used for Plant Species Identification

Classification methods can be broadly divided into unsupervised and supervised classification. Unsupervised classification methods use a computer to separate the data into different classes with minimal user inputs. These methods divide the image pixels into a specified number of natural groups. These groups are then matched to the desired categories of vegetation (Jensen 2016). Supervised classification methods require user inputs to define classes and training data to determine which pixels belong in which classes. These methods require input data of known categories, such as spectral data or physical characteristics of the plant species of interest (Jensen 2016). Most published studies identify invasive species using the supervised classification methods, but some have used unsupervised classification methods (Mladinich et al. 2006; Stitt et al. 2006; Swain et al. 2011).

2.4.1. Iterative Self-Organizing Data Analysis Technique (ISODATA)

The ISODATA is an unsupervised classification method that divides pixels into clusters depending on which cluster mean is nearest to the pixel mean. After the first iteration, the cluster means are recalculated and the pixels are again sorted into clusters based on the distance between the cluster mean and pixel value. This method continues to run through a specified number of iterations or until there are minimal changes in clusters (Jensen 2016). ISODATA does not require predefined patterns describing the data, which can be beneficial when the patterns are very complex (Ball and Hall 1965).

Mladinich et al. (2006), Stitt et al. (2006), and Swain et al. (2011) have used the ISODATA classification method to identify purple loosestrife and leafy spurge. Mladinich et al. (2006) used the ISODATA method for mapping leafy spurge due to a lack of ground reference data, with 63% classification accuracy. The ISODATA algorithm separated the image pixels into clusters of different vegetation types, soil, and water. The clusters were then matched up with the areas on the ground that were known to contain specific vegetation types. Stitt et al. (2006) reached 66% accuracy when using 30 m multispectral imagery resampled to 10 m to identify leafy spurge. Swain et al. (2011) attained 82.1% accuracy with high spatial resolution hyperspectral imagery when identifying purple loosestrife using ISODATA classification. Most of the purple loosestrife pixels that were misclassified as grassland were located along the border between these two vegetative areas and may be the result of pixel mixing. ISODATA classification is an option when little information is known about the area of interest or when there is a lack of ground data, however, when reference spectra are available, supervised classification methods may be a better option when classifying imagery.

2.4.2. Statistical Methods

Partial least squares regression (PLSR), discriminant analysis (DA), and soft independent modeling of class analogy (SIMCA) are some of the statistical methods used to construct models using the spectral signatures of the plant species of interest and distinguish the spectral signatures from each other or from surrounding vegetation. PLSR creates a linear regression model using X and Y variables and determines the best factors of X that will most accurately predict Y. SIMCA uses principal component analysis (PCA) to develop a model separating the plant species.

Several studies using PLSR, DA, or a combination of these methods have been able to achieve overall accuracies ranging from 87% to 95% (Garcia-Ruiz et al. 2015; Girma et al. 2005;

Shapira et al. 2013). These studies differentiated green vegetation species using the spectral signatures collected with spectrometers both in the field and in the greenhouse. These methods were also used to determine the best wavelengths to use for specific plant species. Garcia-Ruiz et al. (2015) found that some of the pixels at the edges of sugarbeet (*Beta vulgaris* L.) leaves were misclassified as Canada thistle, lowering the overall accuracy, but still achieving 89% and 95% accuracies for sugarbeet and thistle, respectively. A PLS-DA model using spectral reflectance data collected with a spectroradiometer was developed and applied to multispectral imagery with sub-centimeter resolution. Shapira et al. (2013) achieved 87% accuracy using these methods, which may have been influenced by soil mixed in with some vegetation spectral samples.

Shirzadifar et al. (2018) used the SIMCA method to separate kochia (*Kochia scoparia* L.), lamb's-quarters (*Chenopodium album* L.), and water-hemp (*Amaranthus rudis* Sauer) in the greenhouse with 88.9% accuracy using the 400-920 nm wavelengths and 100% overall accuracy using the 920-2500 nm wavelengths. These statistical methods have shown promise in the potential to differentiate green vegetation species with similar leaf shape.

2.4.3. Spectral Angle Mapper

Spectral angle mapper (SAM) is a supervised classification method that has been used to identify weeds based on the leaves and canopies from aerial imagery. SAM requires a reference spectra, such as a spectral signature or endmembers taken from an image. The algorithm compares the angle between a reference spectra and an unknown pixel vector to classify in the spectral space and place the pixel in the nearest class. The use of SAM does not assume specific statistical distribution of the data; therefore, the data does not need to be normally distributed (Petropoulos et al. 2010). In SAM the angle of the spectra is the same regardless of the illumination (i.e. shadows). SAM takes only the direction of the spectra into account and does

not use the length of the spectra in the calculation (Kruse et al. 1993; Rashmi et al. 2014). Smaller spectral angles indicate more similar spectra (Kruse et al. 1993).

SAM has primarily been used with hyperspectral imagery (Hunt et al. 2010; Kloppenburg 2014; Lass et al. 2005; Narumalani et al. 2009); however, this method has also been used with multispectral imagery, although not necessarily with the best results (Rodriguez-Moreno et al. 2017). O'Neill et al. (2000) used a combination of Minimum Noise Fraction (MNF) and SAM to achieve 93% accuracy using low spatial resolution hyperspectral imagery to identify a number of species, including leafy spurge. Hunt et al. (2010) and Narumalani et al. (2009) used SAM with low and high spatial resolution hyperspectral images, to classify invasive species in Wyoming and Nebraska, both obtaining 74% overall accuracy. Overall accuracies are calculated as the number of correctly classified pixels as a percentage of the total number of pixels (Jensen 2016).

When separating out the user and producer accuracies, both Hunt et al. (2010) and Narumalani et al. (2009) had acceptable accuracies for specific categories. User accuracy is a measure of commission error, which is when pixels are inaccurately included in a specific class. A high user's accuracy indicates low commission error so that the map pixels classified as a specific class have a high likelihood of actually representing that class on the ground (Jensen 2016). Producer's accuracy is a measure of the omission error, which is when pixels are inaccurately omitted from the correct class. A high producer's accuracy indicates a low omission error so that the plant species on the ground are correctly classified in the map pixels (Jensen 2016). Hunt et al. (2010) obtained 93% user accuracy when identifying leafy spurge. The lower overall accuracy in this study may have been a result of misclassification of nonflowering leafy spurge.

Narumalani et al. (2009) obtained greater than 90% for both producer's and user's accuracies for saltcedar and Russian olive (*Elaeagnus angustifolia* L.). However, the producer's and user's accuracies were below 75% in areas of mixed invasive species. Lass et al. (2005) used SAM to classify spotted knapweed and babysbreath (*Gypsophila paniculata* L.) from hyperspectral imagery with an overall accuracy rate of 67% for spotted knapweed and 83.5% for babysbreath. One reason for the low accuracy rate of spotted knapweed was the uncertainty caused by high moisture variation in the field (Lass et al. 2005).

While applying SAM to hyperspectral imagery resulted in varying accuracies, SAM applied to multispectral imagery did not consistently perform well for plant species classification. The highest overall accuracy achieved by Hunt et al. (2007) using SAM on low spatial resolution multispectral imagery to identify leafy spurge was 61%, while the user's accuracy was 85%. This same study achieved 74% overall accuracy with low spatial resolution hyperspectral imagery. Rodriguez-Moreno et al. (2017) achieved 49% overall accuracy with SAM applied on sub-centimeter resolution multispectral imagery to identify plants and materials in a field. This low accuracy was attributed to the overlapping of spectral signatures in the classes caused by collecting imagery in uncontrolled settings in order to be representative of real environmental conditions.

2.4.4. Object-Based and Spectral and Spatial Feature Learning Methods

Several recent studies utilizing ultra-high spatial resolution imagery collected from UAV sensors have incorporated the use of spectral and spatial features, such as edge and texture information, into their algorithms (Hung et al. 2014; Mafanya et al. 2017; Peña et al. 2013; Pérez-Ortiz et al. 2015; Torres-Sánchez et al. 2013). Ultra-high spatial resolution imagery allows for more plant features to be used to detect specific plant species or for crop rows to aid in

identifying weeds due to the increase of information available when individual plants are visible. Using ultra-high spatial resolution aerial RGB imagery, Hung et al. (2014) employed spatial and spectral feature learning to identify invasive plant species, obtaining 94% accuracy for water hyacinth and 93% accuracy for serrated tussock. This method was deemed viable only when the plant species of interest could be differentiated from the surroundings, such as during changes in the plant life-cycle. Identifying the crop rows can aid in separating the weed pixels from the crop pixels in images of agricultural fields (Peña et al. 2013; Pérez-Ortiz et al. 2015; Torres-Sánchez et al. 2013). Methods such as Hough transform and object-based image analysis (OBIA) can be used to separate crop rows and identify the weeds between and within the rows. The OBIA uses segmentation to identify objects with spectral and spatial similarity (Jensen 2016). Using OBIA, Peña et al. (2013) achieved 86% accuracy when identifying weeds within a maize field. Lu and He (2018) used geographic OBIA to separate grass species and achieved 83% accuracy with ultra-high spatial resolution multispectral imagery. With ultra-high spatial resolution imagery where spatial characteristics of the plants are clearly visible, spatial feature detection and object-based methods have good potential for accurate plant species identification.

2.4.5. Mixture-Tuned Matched Filtering

Mixture-tuned matched filtering (MTMF) is a partial unmixing supervised classification method that produces matched filtering and a feasibility value for each pixel (Mundt et al. 2007). The MTMF method is performed on an MNF transform file and has been used in several studies to detect leafy spurge, with both high and low spatial resolution imagery (Glenn et al. 2005; Kloppenburg 2014; Mitchell and Glenn 2009; Parker Williams and Hunt 2002). With high spatial resolution hyperspectral imagery, up to 94% accuracy was achieved when there was at least 40% leafy spurge cover (Glenn et al. 2005). When using spectrally degraded hyperspectral

imagery to simulate high spatial resolution multispectral imagery, Mitchell and Glenn (2009) achieved 91% accuracy. Parker Williams and Hunt (2002) noted that MTMF performed better in prairie areas than in woodland areas. The MTMF method has primarily been used with hyperspectral imagery, but has also shown promising results with simulated multispectral imagery.

2.4.6. Maximum Likelihood

Maximum likelihood (ML) classification method uses class probability to determine the likelihood that a pixel belongs to a class in normally distributed data. The target classes are defined prior to calculating the likelihood, or probability, that a pixel belongs in a particular class (Jensen 2016). The ML method has successfully identified plant species in multispectral imagery (Carter et al. 2009; Casady et al. 2005; de Castro et al. 2013; Laba et al. 2010; Mafanya et al. 2017; Tamouridou et al. 2017).

Tamouridou et al. (2017) achieved 81% accuracy with 0.1 m resolution and 87% accuracy with degraded 1 m resolution multispectral imagery with a texture layer included using ML classification. Carter et al. (2009) used ML and NDVI threshold values at differing spectral and spatial resolutions for plant species classification. The highest accuracy produced by ML was 80% for low spatial resolution multispectral imagery; however, better accuracies were achieved with NDVI to classify high spatial resolution multispectral imagery and low spatial resolution hyperspectral imagery.

Casady et al. (2005) used ML to classify multi-date multispectral imagery over two years, which resulted in accuracies of 59% to 87%. Leafy spurge was detected in a mixed grass area at two different sites, with one site having a higher forb content. There was a high error of commission rate, meaning that a high number of pixels were classified as leafy spurge when the

weed did not actually exist on the ground in those locations. The relatively low accuracy range was attributed to misclassification of vegetation as leafy spurge during the ground survey, as well as interference of a higher forb content in that site, indicating that ML may perform better in areas with higher grass content. Laba et al. (2010) used ML to identify purple loosestrife, achieving a user's accuracy of 75% and a producer's accuracy of 78%. De Castro et al. (2013) identified flowering cruciferous weeds within wheat fields with 91.3% accuracy using ML on high spatial resolution multispectral imagery. The ML classification method has shown promising results using multispectral imagery.

2.4.7. Vegetation Indices

Vegetation indices (VI) can identify several aspects about green vegetation, including the health and density of the vegetation. Using VI, Carter et al. (2009), de Castro et al. (2013), and Hunt and Parker Williams (2006) identified saltcedar, cruciferous weeds, and leafy spurge, respectively, with varying results. Carter et al. (2009) compared accuracies of NDVI threshold values and the ML classification method to identify saltcedar and achieved accuracies of up to 91% using NDVI threshold values. NDVI is a ratio using the near infrared and red bands to show green vegetation in an image. Pixels are plotted between -1 and 1, with values falling closer to 1 representing more green vegetation (Jensen 2016; Rouse et al. 1973). Carter et al. (2009) produced higher accuracies with NDVI threshold values than with ML for high spatial resolution multispectral imagery and for low spatial resolution hyperspectral imagery at 91% and 88% accuracy, respectively. Using low spatial resolution multispectral and hyperspectral imagery, Hunt and Parker Williams (2006) compared the results of VI and their correlation with leafy spurge presence. They compared NDVI, green normalized difference vegetation index (GNDVI), and green to red ratio (G/R), with poor correlation results for all three indices with all of the

imagery types. The GNDVI is a ratio that uses the near infrared band, similar to NDVI, but uses the green band instead of the red band (Gitelson et al. 1996). The G/R uses a ratio of the green and red bands to analyze vegetation (Motohka et al. 2010; Tucker 1979). De Castro et al. (2013) identified flowering cruciferous weeds within wheat fields with 89.45% accuracy using a blue to green ratio (B/G) on high spatial resolution multispectral imagery. Previous studies utilizing VI did not consistently provide acceptable results for multispectral imagery, making this a less promising method than other methods discussed in this chapter.

2.5. Conclusion

The general trend of research is moving towards the use of higher spatial resolution imagery when investigating methods of identifying invasive plant species in aerial imagery. Small UAVs allow researchers to collect imagery close to the ground, resulting in ultra-high spatial resolutions. Spatial resolution from sensors on satellites and manned aircraft has increased as well. In regard to spectral resolution, both multispectral and hyperspectral imagery has been used with promising results.

Imagery with spatial resolutions of 1.5 m to 30 m has been collected from satellites or manned aircraft and used in previous research to investigate the ability to detect leafy spurge and purple loosestrife in imagery. The use of small UAVs can result in imagery with much higher spatial resolutions and more flexibility in the timing of data collection than imagery collected using manned aircraft or satellites. Higher resolution imagery will allow smaller patches of weeds to be detected. Many herbicides used to control these weeds should be applied at true flowering, allowing a limited time window to collect and analyze imagery. Small UAVs have the flexibility to collect within a specific time window at an affordable cost. Ultra-high spatial resolution imagery has been used successfully with statistical, object-based, and feature learning

methods of classification, which utilize the larger amount of data about the plant species of interest to achieve high accuracies.

Previously the spectral signatures of the plants, collected by either imagery or spectrometers, have identified specific plant species of interest. Statistical methods have shown promise in separating out these spectral signatures, making these viable and potential methods for future research. The research presented in this paper uses spectral signatures of leafy spurge and purple loosestrife, both of their leaves and inflorescences, to separate these plant species from their surrounding vegetation with PLSR. This research may contribute to future methods where more affordable and accessible options are available to survey land for invasive plant species such as leafy spurge and purple loosestrife.

2.6. References

Ball GH, Hall DJ (1965) ISODATA, a novel method of data analysis and pattern classification. Stanford Research Institute. Menlo Park, California

Bradley BA (2014) Remote detection of invasive plants: A review of spectral, textural and phenological approaches. Biol Invasions 16:1411–1425

Carter GA, Lucas KL, Blossom GA, Lassitter CL, Holiday DM, Mooneyhan DS, Fastring DR, Holcombe TR, Griffith JA (2009) Remote sensing and mapping of tamarisk along the Colorado River, USA: A comparative use of summer-acquired Hyperion, Thematic Mapper and QuickBird data. Remote Sens 1:318–329

Casady GM, Hanley RS, Seelan SK (2005) Detection of leafy spurge (*Euphorbia esula*) using multidate high-resolution satellite imagery. Weed Technol 19:462–467

de Castro AI, López-Granados F, Jurado-Expósito M (2013) Broad-scale cruciferous weed patch classification in winter wheat using QuickBird imagery for in-season site-specific control. Precis Agric 14:392–413

Diao C, Wang L (2016) Temporal partial unmixing of exotic salt cedar using Landsat time series. Remote Sens Lett 7:466–475

Garcia-Ruiz FJ, Wulfsohn D, Rasmussen J (2015) Sugar beet (*Beta vulgaris* L.) and thistle (*Cirsium arvensis* L.) discrimination based on field spectral data. Biosyst Eng 139:1–15

Girma K, Mosali J, Raun WR, Freeman KW, Martin KL, Solie JB, Stone ML (2005) Identification of optical spectral signatures for detecting cheat and ryegrass in winter wheat. Crop Sci 45:477–485

Gitelson AA, Kaufman YJ, Merzlyak MN (1996) Use of a green channel in remote sensing of global vegetation from EOS-MODIS. Remote Sens Environ 58:289–298

Glenn NF, Mundt JT, Weber KT, Prather TS, Lass LW, Pettingill J (2005) Hyperspectral data processing for repeat detection of small infestations of leafy spurge. Remote Sens Environ 95:399–412

He KS, Bradley BA, Cord AF, Rocchini D, Tuanmu MN, Schmidtlein S, Turner W, Wegmann M, Pettorelli N (2015) Will remote sensing shape the next generation of species distribution models? Remote Sens Ecol Conserv 1:4–18

He KS, Rocchini D, Neteler M, Nagendra H (2011) Benefits of hyperspectral remote sensing for tracking plant invasions. Divers Distrib 17:381–392

Hung C, Xu Z, Sukkarieh S (2014) Feature learning based approach for weed classification using high resolution aerial images from a digital camera mounted on a UAV. Remote Sens 6:12037–12054

Hunt ER, Daughtry CST, Kim MS, Parker Williams AE (2007) Using canopy reflectance models and spectral angles to assess potential of remote sensing to detect invasive weeds. J Appl Remote Sens 1:013506

Hunt ER, Gillham JH, Daughtry CST (2010) Improving potential geographic distribution models for invasive plants by remote sensing. Rangel Ecol Manag 63:505–513

Hunt ER, Parker Williams AE (2006) Detection of flowering leafy spurge with satellite multispectral imagery. Rangel Ecol Manag 59:494–499

Jensen JR (2016) Introductory Digital Image Processing: A Remote Sensing Perspective. 4th edn. Glenview, IL: Pearson Education, Inc. Pp 325-569

Kloppenburg C (2014) Detecting Leafy Spurge in Native Grassland Using Hyperspectral Image Analysis. MS thesis. Lethbridge, AB: University of Lethbridge. 93 p

Kruse FA, Lefkoff AB, Boardman JW, Heidebrecht KB, Shapiro AT, Barloon PJ, Goetz AFH (1993) The spectral image processing system (SIPS) − interactive visualization and analysis of imaging spectrometer data. Remote Sens Environ 44:145–163

Laba M, Blair B, Downs R, Monger B, Philpot W, Smith S, Sullivan P, Baveye PC (2010) Use of textural measurements to map invasive wetland plants in the Hudson River National Estuarine Research Reserve with IKONOS satellite imagery. Remote Sens Environ 114:876–886

Laba M, Tsai F, Ogurcak D, Smith S, Richmond ME (2005) Field determination of optimal dates for the discrimination of invasive wetland plant species using derivative spectral analysis. Photogramm Eng Remote Sensing 71:603–611

Lass LW, Prather TS, Glenn NF, Weber KT, Mundt JT, Pettingill J (2005) A review of remote sensing of invasive weeds and example of the early detection of spotted knapweed

(*Centaurea maculosa*) and babysbreath (*Gypsophila paniculata*) with a hyperspectral sensor. Weed Sci 53:242–251

Lu B, He Y (2018) Optimal spatial resolution of unmanned aerial vehicle (UAV)-acquired imagery for species classification in a heterogeneous grassland ecosystem. GIsci Remote Sens 55:205–220

Mafanya M, Tsele P, Botai J, Manyama P, Swart B, Monate T (2017) Evaluating pixel and object based image classification techniques for mapping plant invasions from UAV derived aerial imagery: *Harrisia pomanensis* as a case study. ISPRS J Photogramm Remote Sens 129:1–11

Mirik M, Ansley RJ, Steddom K, Jones DC, Rush CM, Michels GJ, Elliott NC (2013) Remote distinction of a noxious weed (musk thistle: *Carduus nutans*) using airborne hyperspectral imagery and the support vector machine classifier. Remote Sens 5:612–630

Mitchell JJ, Glenn NF (2009) Leafy spurge (*Euphorbia esula*) classification performance using hyperspectral and multispectral sensors. Rangel Ecol Manag 62:16–27

Mladinich CS, Ruiz Bustos M, Stitt S, Root R, Brown K, Anderson GL, Hager S (2006) The use of Landsat 7 Enhanced Thematic Mapper Plus for mapping leafy spurge. Rangel Ecol Manag 59:500–506

Motohka T, Nasahara KN, Oguma H, Tsuchida S (2010) Applicability of green-red vegetation index for remote sensing of vegetation phenology. Remote Sens 2:2369–2387

Mundt JT, Streutker DR, Glenn NF (2007) Partial unmixing of hyperspectral imagery: Theory and methods. *In* Proceedings of the American Society of Photogrammetry and Remote Sensing; 2007. Tampa, FL: American Society of Photogrammetry and Remote Sensing

Narumalani S, Mishra DR, Wilson R, Reece P, Kohler A (2009) Detecting and mapping four invasive species along the floodplain of North Platte River, Nebraska. Weed Technol 23:99–107

Noujdina NV, Ustin SL (2008) Mapping downy brome (*Bromus tectorum*) using multidate AVIRIS data. Weed Sci 56:173–179

O'Neill M, Ustin SL, Hager S, Root R (2000) Mapping the distribution of leafy spurge at Theodore Roosevelt National Park using AVIRIS. Pages 0-18 *in* Proceedings of the Ninth JPL Airborne Earth Science Workshop. Pasadena, California: NASA Jet Propulsion Laboratory

Parker Williams A, Hunt ER (2002) Estimation of leafy spurge cover from hyperspectral imagery using mixture tuned matched filtering. Remote Sens Environ 82:446–456

Peña JM, Torres-Sánchez J, de Castro AI, Kelly M, López-Granados F (2013) Weed mapping in early-season maize fields using object-based analysis of unmanned aerial vehicle (UAV) images. PLOS One 8:e77151

Pérez-Ortiz M, Peña JM, Gutiérrez PA, Torres-Sánchez J, Hervás-Martínez C, López-Granados F (2015) A semi-supervised system for weed mapping in sunflower crops using unmanned aerial vehicles and a crop row detection method. Appl Soft Comput 37:533–544

Petropoulos GP, Vadrevu KP, Xanthopoulos G, Karantounias G, Scholze M (2010) A comparison of spectral angle mapper and artificial neural network classifiers combined with Landsat TM imagery analysis for obtaining burnt area mapping. Sens 10:1967–1985

Rashmi S, Swapna A, Venkat, Ravikiran S (2014) Spectral angle mapper algorithm for remote sensing image classification. Int J Innov Sci Eng Technol 1:201–205

Rodriguez-Moreno F, Kren J, Zemek F, Novak J, Lukas V, Pikl M (2017) Advantage of multispectral imaging with sub-centimeter resolution in precision agriculture: Generalization of training for supervised classification. Precis Agric 18:615–634

Rouse JW, Hass RH, Schell JA, Deering DW (1973) Monitoring vegetation systems in the Great Plains with ERTS. Third Earth Resour Technol Satell Symp 1:309–317

Shapira U, Herrmann I, Karnieli A, Bonfil DJ (2013) Field spectroscopy for weed detection in wheat and chickpea fields. Int J Remote Sens 34:6094–6108

Shapira U, Herrmann I, Karnieli A, Bonfil JD (2010) Weeds detection by ground-level hyperspectral data. Theory to Pract XXXVIII:27–33

Shirzadifar A, Bajwa S, Mireei SA, Howatt K, Nowatzki J (2018) Weed species discrimination based on SIMCA analysis of plant canopy spectral data. Biosyst Eng 171:143–154

Stitt S, Root R, Brown K, Hager S, Mladinich C, Anderson GL, Dudek K, Ruiz Bustos M, Kokaly R (2006) Classification of leafy spurge with earth Observing-1 Advanced Land Imager. Rangel Ecol Manag 59:507–511

Swain S, Narumalani S, Mishra DR (2011) Monitoring invasive species: Detecting purple loosestrife and evaluating biocontrol along the Niobrara River, Nebraska. GIsci Remote Sens 48:225–244

Tamouridou AA, Alexandridis TK, Pantazi XE, Lagopodi AL, Kashefi J, Moshou D (2017) Evaluation of UAV imagery for mapping *Silybum marianum* weed patches. Int J Remote Sens 38:2246–2259

Thorp KR, Tian LF (2004) A review on remote sensing of weeds in agriculture. Precis Agric 5:477–508

Torres-Sánchez J, López-Granados F, de Castro AI, Peña-Barragán JM (2013) Configuration and specifications of an unmanned aerial vehicle (UAV) for early site specific weed management. PLOS One 8:e58210

Tucker CJ (1979) Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens Environ 8:127–150

Ustin SL, Santos MJ (2010) Spectral identification of native and non-native plant species. Pages 1–17 *in* Proceedings of ASD and IEEE GRS; Art, Science and Applications of Reflectance Spectroscopy Symposium, Vol. 2. Boulder, CO

3. CLASSIFICATION OF LEAFY SPURGE AND PURPLE LOOSESTRIFE BASED ON SPECTRAL DATA

3.1. Introduction

Leafy spurge and purple loosestrife are two commonly found noxious weeds in North Dakota (North Dakota Department of Agriculture 2016). North Dakota county weed boards and weed officers monitor the state for noxious weeds, to ensure that they are controlled and do not replace native vegetation. When used, herbicides are often applied to actively growing plants or to the true flowers of these plants making it necessary to identify the plants by the time the flowers are visible (Knezevic et al. 2004; Lym 1998; Mullin 1998). It is important to identify these noxious weeds early and efficiently because small infestations of noxious weeds are easier and less costly to manage than large infestations.

As technology has improved, higher spatial resolutions of imagery have become more readily available. A number of studies have used imagery collected from sensors on satellites and manned aircraft to identify plant species (de Castro et al. 2013; Hunt et al. 2010; Mirik et al. 2013; Noujdina and Ustin 2008; Swain et al. 2011). However, in recent years small UAVs have become available at more affordable costs and have provided an accessible method of acquiring ultra-high spatial resolution imagery at desired locations and times. Several studies using ultra-high spatial resolution imagery have utilized the higher levels of details and information available by using statistical methods and object-based and feature learning methods, many achieving accuracies above 90% (Garcia-Ruiz et al. 2015; Girma et al. 2005; Hung et al. 2014; Mafanya et al. 2017; Peña et al. 2013; Pérez-Ortiz et al. 2015; Shapira et al. 2013; Shirzadifar et al. 2018; Torres-Sánchez et al. 2013). The studies using statistical methods differentiated specific plant species from the surrounding green vegetation, achieving accuracies of up to 95% when

using wavelengths ranging from 400-1000 nm (Garcia-Ruiz et al. 2015; Girma et al. 2005; Shapira et al. 2013; Shirzadifar et al. 2018).

Both hyperspectral and multispectral data have been used to identify plant species with levels of accuracy above 90% (Garcia-Ruiz et al. 2015; Glenn et al. 2005; Hung et al. 2014; Mafanya et al. 2017; Narumalani et al. 2009); however, hyperspectral imagery also has the potential to add in noise, unwanted data that does not provide useful information, that can decrease accuracies (Carter et al. 2009; Hunt and Parker Williams 2006; Hunt et al. 2007; Mitchell and Glenn 2009).

Attempts to identify leafy spurge and purple loosestrife by spectral reflectance of the inflorescences have generally not included the spectral signatures of the plant leaves prior to flowering to identify the plant species (Hunt et al. 2010; Kloppenburg 2014; Laba et al. 2010; Mitchell and Glenn 2009; Swain et al. 2011). Non-flowering plants within the stand can decrease the accuracy of the studies when a plant such as leafy spurge is identified solely by the inflorescences (Hunt et al. 2010). Identifying plants both with and without flowers may increase classification accuracies.

Separation of the spectral signatures of flowering leafy spurge from sweet clover (*Melilotus* spp.), another yellow flowering plant, was difficult (O'Neill et al. 2000; Parker Williams and Hunt 2002; Ustin and Santos 2010). Using hyperspectral data Parker Williams and Hunt (2002) noted that the spectral signatures of leafy spurge were distinguishable from those of yellow sweet clover (*Melilotus officinalis*) because leafy spurge showed less reflectance in the 550-685 nm wavelengths but higher reflectance in the NIR region than sweet clover. Using SAM on field spectra, O'Neill et al. (2000) obtained up to 99% accuracy when separating out leafy spurge, sweet clover, and snowberry (*Symphoricarpos albus*). SAM is a classification method

that compares the angle between a reference spectra and an unknown pixel to classify the spectral space and match the pixel to the nearest class. Ustin and Santos (2010) were able to separate sweet clover from leafy spurge using field spectra; however, they had less success when moving to the image scale. This indicates that sweet clover is spectrally distinct from leafy spurge; however, more work is needed to separate these species in imagery.

Based on the hypothesis that the spectral signatures of leafy spurge and purple loosestrife are unique from other species, and in order to classify these two plant species based on their spectral signatures, this chapter utilized ground-collected spectral data. Using PLS-DA, the spectral signatures of the inflorescences and leaves of leafy spurge and purple loosestrife were separated from the surrounding vegetation. PLS-DA uses the dependent variable, X, and response variable, Y, to create a linear regression model that determines the latent variables of X that will predict Y, and subsequently classifies the samples based on the model. The spectral signatures were collected from sites located in southeastern North Dakota where either leafy spurge or purple loosestrife was visible. The objective of this research was to classify leafy spurge and purple loosestrife by the spectral signatures and the spectral data of the inflorescences and the leaves that was manually collected in the field.

3.2. Methods and Materials

3.2.1. Study Area

Two sites (leafy spurge site: 46.543°N, 97.141°W; purple loosestrife site: 46.963°N, 98.184°W) in southeastern North Dakota were included in the study area. The leafy spurge collection site was rangeland on the Albert Ekre Grassland Preserve in northeast Richland County, North Dakota, that was infested with leafy spurge. The purple loosestrife collection site

was a slough, located next to a soybean field in central Barnes County, North Dakota, west of Valley City that was infested with purple loosestrife.

3.2.2. Data Collection Dates

Data were collected for leafy spurge and purple loosestrife before flowering and during flowering in order to compare the spectral signatures of both the leaves of young plants and the plant flowers or bracts with the surrounding vegetation. Purple loosestrife is most easily identifiable in the near-infrared range when the flowers mature to seed (Laba et al. 2005); however, many herbicides are applied to purple loosestrife when the plants are still flowering (Knezevic et al. 2004; Mullin 1998). Therefore, field data were collected prior to this change in the plant life-cycle in order to identify the plants within a time range that herbicides would normally be applied.

The ground-collected spectral reflectance data for leafy spurge, the surrounding vegetation, and background were collected on 26 May 2017. The ground-collected spectral reflectance data for purple loosestrife, the surrounding vegetation, and background were collected on 6 September 2017.

3.2.3. Spectral Data Collection

Spectral data were collected in order to gather precise and accurate spectral signatures of the plants of interest and to classify these plants accurately. The collected data included spectral reflectance of the inflorescences and leaves of the noxious weeds as well as the surrounding vegetation and soil background from both sites (Table 1). For leafy spurge, the spectral reflectance of the bracts was collected instead of the flowers because the bracts were a more distinct yellow in color than the flowers. Spectral signatures of the leaves were collected from leafy spurge and purple loosestrife plants before flowering.

The spectral data of the vegetation were collected with a spectrometer system consisting of a UV-NIR spectrometer (model USB2000+XR1-ES, Ocean Optics, Largo, Florida, USA) and an NIR spectrometer (model NIRQuest512-2.5, Ocean Optics, Largo, Florida, USA). The USB2000+XR1-ES used a Sony ILX511B detector (2048-element linear silicon charge coupled device (CCD) array), collected data in the wavelength range of 200-1025 nm, and had an optical resolution of about 1.7-2.1 nm full width half maximum (FWHM). The NIRQuest512-2.5 used a Hamamatsu G9208-512W InGaAs linear array detector, collected data in the wavelength range of 900-2500 nm, and had an optical resolution of about 6.3 nm FWHM. The light source for the spectrometer system was an Ocean Optics tungsten halogen lamp (HL-2000).

Table 1. The ten spectral data classes from the leafy spurge and purple loosestrife sites and the sample totals collected at each site.

Class	**Calibration sample #**	**Validation sample #**	**Total samples**
Leafy spurge leaves	72	24	96
Leafy spurge bracts	78	26	104
Grass	73	25	98
Background material	33	11	44
Plant litter	72	23	95
Soil	18	7	25
Purple loosestrife leaves	81	26	107
Purple loosestrife flowers	45	16	61
Background plants	81	27	108
Background material	65	22	87
Totals	**618**	**207**	**825**

The spectral data classes with the calibration and validation sample totals. The calibration group was comprised of 75% of the samples, and the validation group was comprised of 25% of the samples.

The spectrometers were calibrated with a reference measurement taken from a Spectralon Diffuse Reflectance Standards calibration panel (model USRS-99-010, Labsphere, North Sutton,

New Hampshire, USA) with 99% reflectance in the measured wavelengths. The reference and dark spectra were periodically checked throughout the sample collection in order to ensure that the machine was calibrated correctly. About 50-100 plants from each class were gathered from across the study areas and brought to the spectrometers for data collection. The target plant material was generally small; therefore, multiple leaves or petals were layered for data collection to ensure the spectrometer did not collect data from outside the target material. One reflectance spectra was acquired per cluster of plant material. The spectral reflectance of each sample was collected with a fiber optic probe held within 1 cm above the plant material so that data were collected on only a small portion of the plant. This was especially important when collecting data of the inflorescences, which were small in size, in order to obtain pure spectral signatures. Data were collected in the late morning to early afternoon.

3.2.4. Spectral Data Analysis Using Partial Least Squares Discriminant Analysis

Partial least squares discriminant analysis (PLS-DA) was used to analyze the spectral data collected with the spectrometer for leafy spurge and purple loosestrife in order to identify how well the spectral signatures of the plants could be distinguished from the surrounding vegetation and materials, as well as to identify the bands that provided the most information to distinguish these invasive plants from the surroundings. PLS-DA combines PLSR and DA as a classification method. PLSR creates a linear regression model that uses X and Y variables to determine the latent variables of X that will best predict Y. PLS-DA models the material classes using PLSR and subsequently categorizes these classes. This method was chosen because of the ability to classify spectral signatures and the promising results in previous studies (Garcia-Ruiz et al. 2015; Shapira et al. 2010). The data preprocessing and PLS-DA model development of the spectral data were performed using the Unscrambler v10.4.1 (CAMO, Oslo, Norway).

3.2.4.1. Initial setup of data

The spectral data of leafy spurge, the surrounding vegetation, and the soil background included six classes; while the spectral data of purple loosestrife and the surrounding vegetation and materials included four classes (Table 1). At both sites, wavelengths 400-1000 nm were used, which resulted in 1490 bands. Previous studies identifying invasive species utilized wavelengths within this range with positive results (Garcia-Ruiz et al. 2015; Peña et al. 2013; Swain et al. 2011).

The samples were divided into calibration and validation groups. The calibration data were used to develop the PLSR models, and the validation data were used to validate the models and show how well each class was predicted. The calibration group consisted of 75% of the samples and the validation groups consisted of the remaining 25% of the samples. The samples were arranged in the order in which they were collected so that each of the two groups would be representative of the data. Every fourth sample was categorized as validation, and the remaining samples were categorized as calibration. This pattern was used until all samples were labeled as either calibration or validation.

3.2.4.2. Spectral data preprocessing

The raw spectra in the calibration category were preprocessed prior to developing the PLSR models. The second derivative was calculated using a Savitzky-Golay transform with a second order polynomial in order to reduce the additive effects in the spectral data and to highlight the peaks of interest (Savitzky and Golay 1964). The second derivative measures the change in the slope of the data curves and removes linear baseline effects and sloping. This can be beneficial to use because the peaks are kept in the same places as in the original data. The Savitzky-Golay filter smoothed the spectral data and removed excess noise. The total smoothing

points used were 31, with 15 smoothing points on each side. The standard normal variate (SNV) was then used to analyze the second derivative transformed data to remove scatter effects. SNV divides the mean from across the spectrum by the standard deviation across the spectrum, and is calculated:

$$x_{T,k} = \frac{x_k - \bar{x}}{s}$$

where $x_{T,k}$ is the wavelength with the SNV transformation applied, x_k is the observed spectra prior to the SNV transformation, $\bar{x}$ is the mean spectra, and s is the standard deviation of the spectra.

3.2.4.3. Partial least squares regression model development

Four different PLSR models were created using the preprocessed data that was labeled as calibration. Models 1 and 2 were developed using the leafy spurge spectral data in the calibration group, and models 3 and 4 were developed using the purple loosestrife spectral data in the calibration group. For each model, spectral data was divided into either two or three categories. Each category was given a numerical value, which was used as the Y value. The X values consisted of the preprocessed spectra. The models were initially validated using random cross-validation.

- Model 1 was comprised of three target categories, namely leafy spurge bracts, leafy spurge leaves, and background consisting of the four classes of background plants and materials (grass, background material, plant litter, and soil).
- Model 2 was comprised of two categories, namely leafy spurge leaves and background consisting of the four classes of background plants and materials.
- Model 3 was comprised of three categories, namely purple loosestrife flowers, purple loosestrife leaves, and the background plants and materials.

- Model 4 was comprised of two categories, namely purple loosestrife leaves and the background plants and materials.

PLSR was initially run on every tenth band of the spectral signatures; therefore, of the total 1490 bands that were collected, 149 bands were initially used to develop each model.

3.2.4.4. Uncertainty test

Martens' Uncertainty Test was used to determine the significant bands, or the bands that significantly contributed to the development of each model, doing so by using a jack-knifing method to detect the variables that did not significantly contribute to the model (Martens and Martens 2000). The significant bands were determined using the regression coefficients. If the uncertainty limit of the regression coefficient of the band crossed zero, then the band was not a predicting variable and was not significant to the model. In the Unscrambler, this test highlighted the bands that contributed to the model and gave the option to recalculate the model using only these significant bands. Each model was recalculated using only the significant bands until all of the remaining bands were deemed significant according to the uncertainty test.

3.2.4.5. Random cross-validation

Random cross-validation was used to validate the models and indicate how the models performed at predicting samples in the validation group. The number of factors used in each of the four models was reduced from the original number of five, six, or seven factors to four factors. The smaller number of factors, or latent variables, made the models simpler and, therefore, more robust than the five to seven factors used in each initial model. The effectiveness of each of the models was similar when decreased to a 4-factor models. However, reducing the factors past this point decreased the R^2 values and how well the data was explained by the models.

3.2.4.6. Validation using prediction

After developing the PLSR models and running cross-validation, a prediction using the regression model was run on the validation group, using the four suggested components, or factors, of each model and all 1490 bands for each model. This prediction value indicated how well the spectral signatures were able to be differentiated from each other.

3.3. Results

3.3.1. Spectral Signature Analysis

Initial visual analysis of the raw spectra and the second derivative after SNV correction found that purple loosestrife and leafy spurge can be separated from the surrounding vegetation, both by the plant inflorescences and by the leaves (Figures 1 and 2). The raw spectral data needed to be normalized before accurate comparisons could be made about the data. The data normalized with the second derivative and SNV had comparable spectral signatures while still highlighting plant differences. Visible distinctions between the spectral signatures were observed within the range of 500-750 nm, which corresponded with the green, red, and near infrared regions. The leafy spurge bracts were most visually distinguishable starting at 665 nm until about 730 nm, and the leafy spurge leaves were most visually distinguishable between 675 nm and 725 nm (Figure 1). The purple loosestrife flowers were most visually distinguishable between 575 nm and 680 nm, while the purple loosestrife leaves were most visually distinguishable between 680 nm and 750 nm (Figure 2). The second derivative and SNV preprocessed data had a peak for the leafy spurge inflorescence and leaves and purple loosestrife flowers at 685-690 nm and a dip around 710-725 nm. However, the purple loosestrife flowers peaked around 590 nm and dipped around 665 nm.

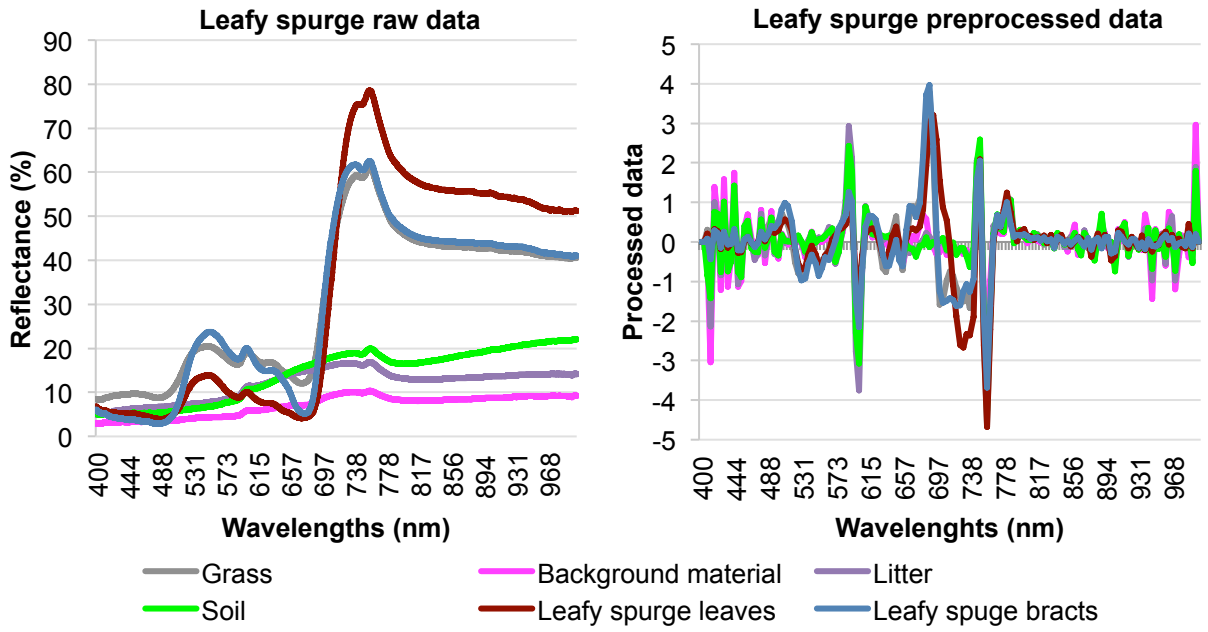

Figure 1. Averaged spectral data for each of the six classes from the leafy spurge collection site.
Averaged raw spectra of each class of materials from the leafy spurge collection site (left).
Averaged preprocessed spectra of each class of materials from the leafy spurge collection site after 2nd derivative and SNV transformation (right).

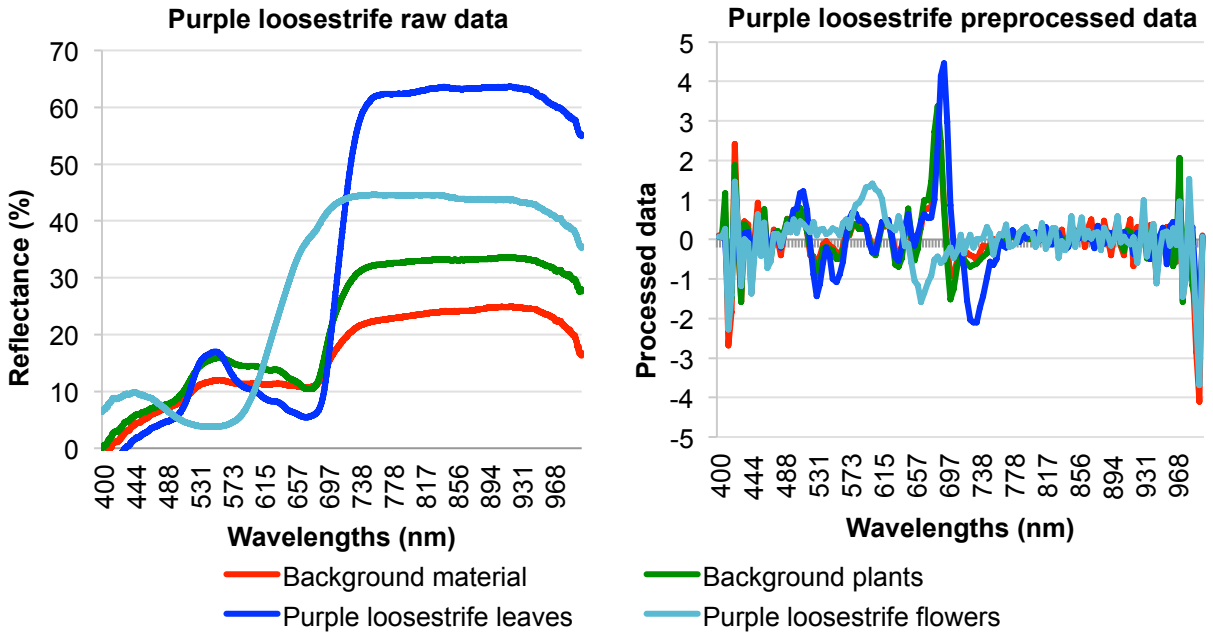

Figure 2. Averaged spectral data for each of the four classes from the purple loosestrife collection site.
Averaged raw spectra of each class of materials from the purple loosestrife collection site (left).
Averaged preprocessed spectra of each class of materials from the purple loosestrife collection site after 2nd derivative and SNV transformation (right).

3.3.2. Model Development and Calibration

The model performance was analyzed based on random cross-validation and the score plots. Analysis of the methods found weaknesses in the models that needed to be modified prior to running the final prediction analysis on the validation group. All four models were reduced to 4-factor models, making them simpler and more robust than the 5- to 7-factor models initially produced, while still explaining the data and attaining R^2 values above 0.85 (Table 2). The number of bands used in each model was reduced as well. Each model was initially developed using 149 bands, which were reduced by the final variation of each model.

3.3.2.1. Random cross-validation

- Model 1 (all of the leafy spurge data and the surrounding background plants and materials) was recalculated five times using only the variables marked as significant by Martens' Uncertainty Test, reducing the model from a 6-factor model to a 4-factor model. The final variation used 73 bands, and both the calibration and validation R^2 values from the random cross-validation were 0.92.
- Model 2 (leafy spurge leaves and the surrounding background plants and materials) was recalculated four times using only the variables marked as significant by Martens' Uncertainty Test, reducing the model from a 7-factor model to a 4-factor model. The final variation used 21 bands, and both the calibration and validation R^2 values from the random cross-validation were 0.88.
- Model 3 (all of the purple loosestrife data and the surrounding background plants and materials) was recalculated five times using only the variables marked as significant by Martens' Uncertainty Test, reducing the model from a 7-factor model to a 4-factor

model. The final variation used 37 bands. The calibration R^2 value from the random cross-validation was 0.91, and validation R^2 value was 0.90.

- Model 4 (purple loosestrife leaves and the surrounding background plants and materials) was recalculated five times using only the variables marked as significant by Martens' Uncertainty Test, reducing the model from a 5-factor model to a 4-factor model. The final variation used 25 bands. The calibration R^2 value from the random cross-validation was 0.89, and validation R^2 value was 0.88.

Table 2. The PLSR model development depicting the number of categories, factors, and bands used in developing each model.

	Noxious weed	**# of categories**	**Initial # of factors**	**Final # of factors**	**Final # of bands**	**Cal. R^2**	**Val. R^2**
Model 1	Leafy spurge	3	6	4	73	0.92	0.92
Model 2	Leafy spurge	2	7	4	21	0.88	0.88
Model 3	Purple loosestrife	3	7	4	37	0.91	0.90
Model 4	Purple loosestrife	2	5	4	25	0.89	0.88

PLSR models with random cross-validation R^2 results for both calibration and validation, showing the number of categories included in each model, the number of factors used to develop the initial models, and the number of factors and bands used in the final models.

3.3.2.2. Score plots

The score plot is a scatter plot that uses two factors. Factors 1 and 2, the first two of the latent variables of X that best predict Y, described more variation in the data than the other factors, making the score plot depicting these two factors the most informative. Factor 1 explained the majority of the variance in the models. Factors 2 and 3 also explained a significant amount of the variance. Factor 4 explained a small, but still important amount of the variance,

while factors 5-7 did not explain enough of the variance to warrant being included in the final models (Figure 3).

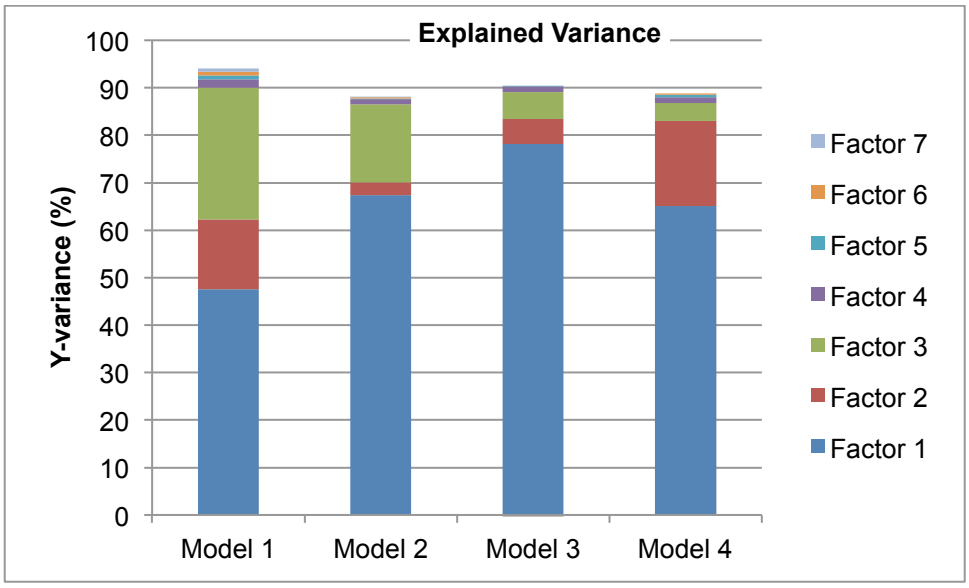

Figure 3. Explained variance of the four models by each of the validation factors, the latent variables of X that best predict Y.
Factors 1-4 explained a significant amount of the variance in the models. Factors 5-7 did not explain enough of the variance and were not included in the models.

The score plots show patterns in the data and sample grouping. The samples depicted close together in the score plots indicate that they have characteristics that are more similar to each other than to samples that are depicted farther away in the score plots (Figure 4). In models 1 and 2, the background material, soil, and litter are concentrated together, far away from leafy spurge leaves and bracts and the grass. The grass and leafy spurge leaves overlap somewhat in both models. In model 1 the leafy spurge bracts appear mostly between the grass and the leafy spurge leaves, overlapping both classes slightly, but overlapping the leaves more than the grass. This indicates that much of the lower accuracy of models 1 and 2 is likely due to the overlapping spectral signatures of the leafy spurge bracts, leaves, and grass.

In models 3 and 4, there is a considerable amount of overlap between the background plants and materials, but the purple loosestrife flower and leaf samples appear in individual

clusters showing distinct separation from each other and from the background plants and materials. Removing one of the classes changed the values of the scores, but the patterns in the data were still visible between models 1 and 2 and models 3 and 4.

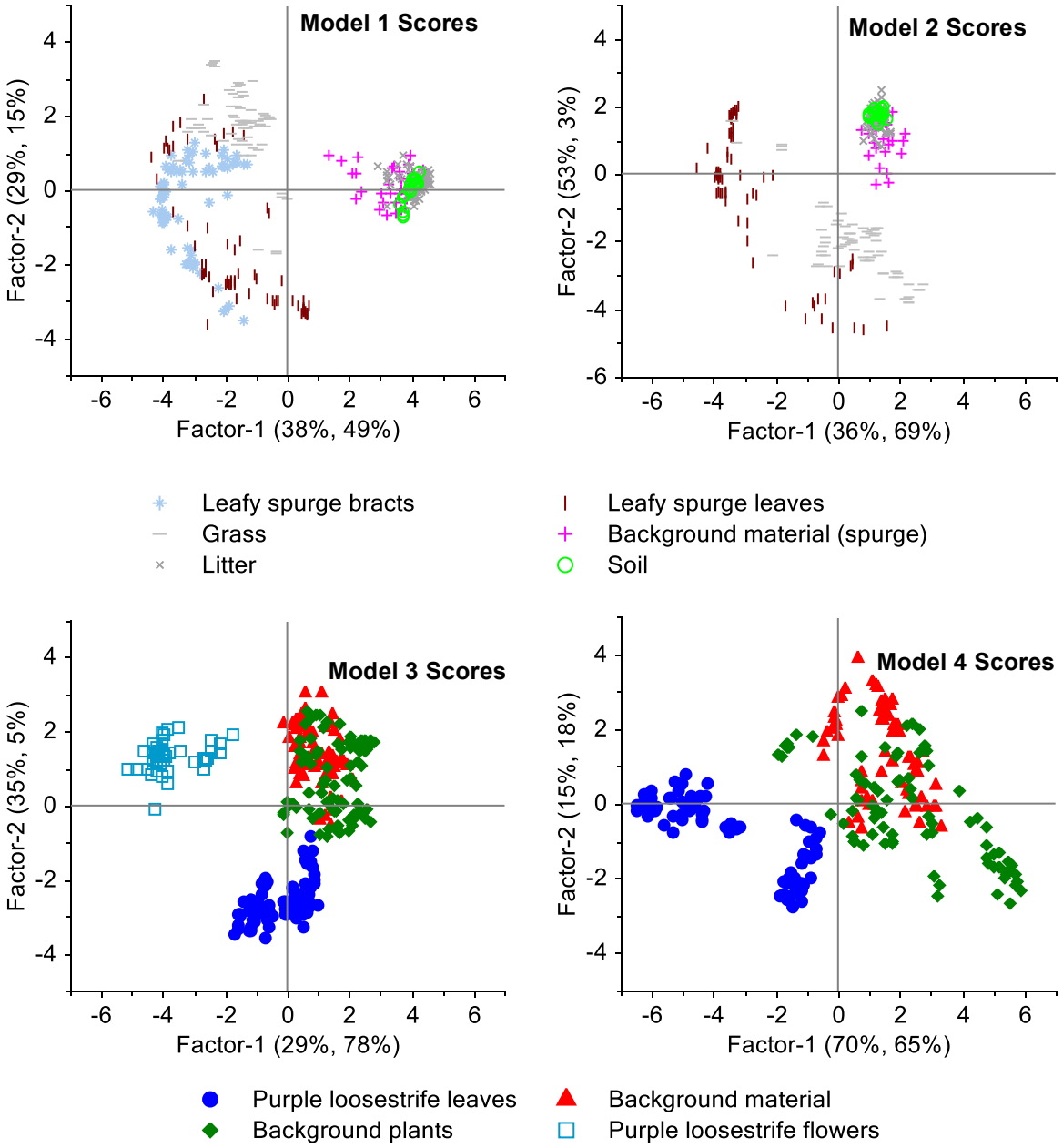

Figure 4. Score plots depicting patterns of plant and background samples in scatterplots. The score plots of factors 1 and 2 for models 1 and 2 show overlap between the samples of leafy spurge bracts, leafy spurge leaves, and grass. The score plots of factors 1 and 2 for models 3 and 4 show almost no overlap between the samples of purple loosestrife flowers and purple loosestrife leaves, either with each other or the other classes.

3.3.3. Significant Bands

Analysis of the spectral bands identified the significant bands, or the bands that contributed to model development, for each model using Martens' Uncertainty Test. These are the bands, or the X variables, that describe the Y variables well and are important for predicting Y. Once the significant bands were identified, each model was recalibrated using only these significant bands. By this method, the number of bands used for each model was reduced from 149 to include only the bands that explained the data well, which ranged from 21 to 73 bands for the four models.

Identifying which bands are useful for each model can determine which bands and sensors are best to use when collecting and analyzing imagery of the plant species in each model. The most important of these wavelengths can also be identified depending on which ones have the largest regression coefficients. While all of the bands used in the final models were important, the X variables with the largest regression coefficients were the variables that contributed the most to the model.

The most important wavelengths generally fell within the visually distinct range of 500-750 nm (Figures 5-8). The single most important or useful wavelengths – those with the largest regression coefficients – for models 1 and 3 were located within the red spectral range at 648.2 nm and 656.5 nm, respectively. The single most useful wavelengths for models 2 and 4 were located within the red edge and NIR ranges at 701.5 nm and 832.4 nm, respectively. Model 3 had the most distinct band clusters of the four models, especially at 560-610 nm, 650-675 nm, and 720-740 nm (Figure 7). Model 2 had only a single loose cluster within 685-720 nm (Figure 6). Model 4 also showed a loose cluster of bands in the range of 635-735 nm (Figure 8). Model 1 was the most difficult of the models to select distinct clusters due to the high number of bands

used; however, the densest areas of significant bands occurred between about 480 nm and 735 nm (Figure 5). Bands with wavelengths at the very edges of the 400-1000 nm spectral range were also included in Model 1. All four of the models had a cluster of bands around 700 nm; therefore, this chlorophyll absorption and red edge region is an important group of wavelengths to include in sensors to detect these weeds.

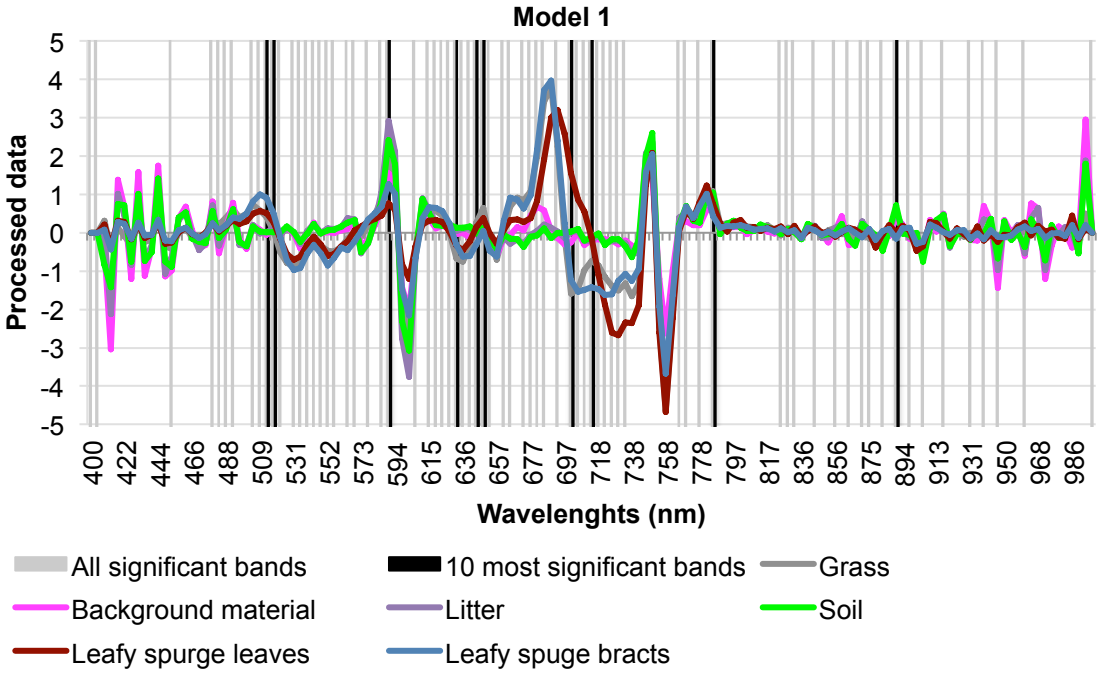

Figure 5. Model 1 preprocessed spectral signatures indicating relative significance of wavebands. Model 1 identified 73 significant bands with the densest areas of significant bands occurring between about 480 nm and 735 nm and the most important wavelength occurring at 648.2 nm. Clusters of significant bands around 700 nm can be seen in all four of the models as chlorophyll absorption in this region is an important factor in plant identification.

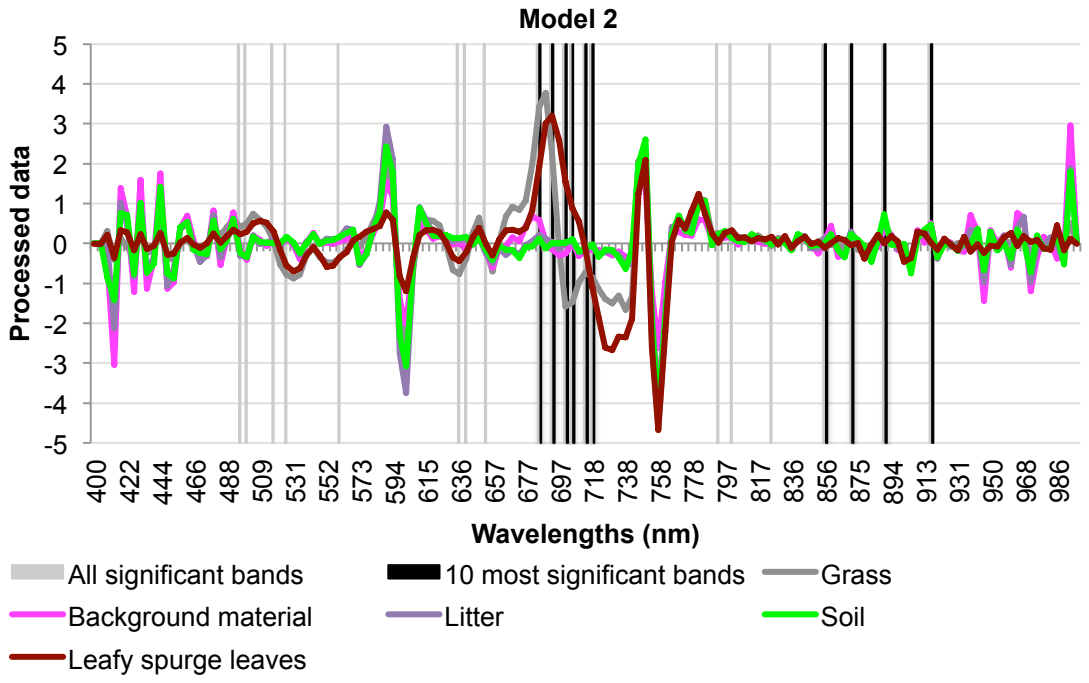

Figure 6. Model 2 preprocessed spectral signatures indicating relative significance of wavebands. Model 2 identified 21 significant bands with the only loose cluster occurring within 685-720 nm and the most important wavelength occurring at 701.5 nm.

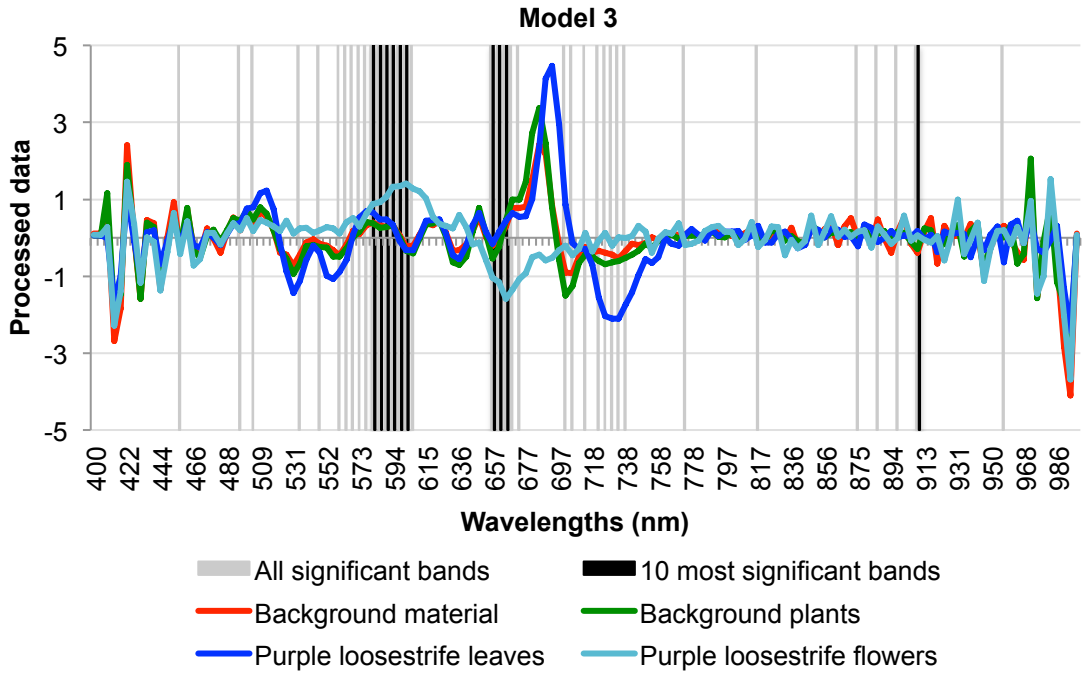

Figure 7. Model 3 preprocessed spectral signatures indicating relative significance of wavebands. Model 3 identified 37 significant bands and had the most distinct band clusters of the four models, which occurred at 560-610 nm, 650-675 nm, and 720-740 nm. The most important wavelength occurred at 656.5 nm.

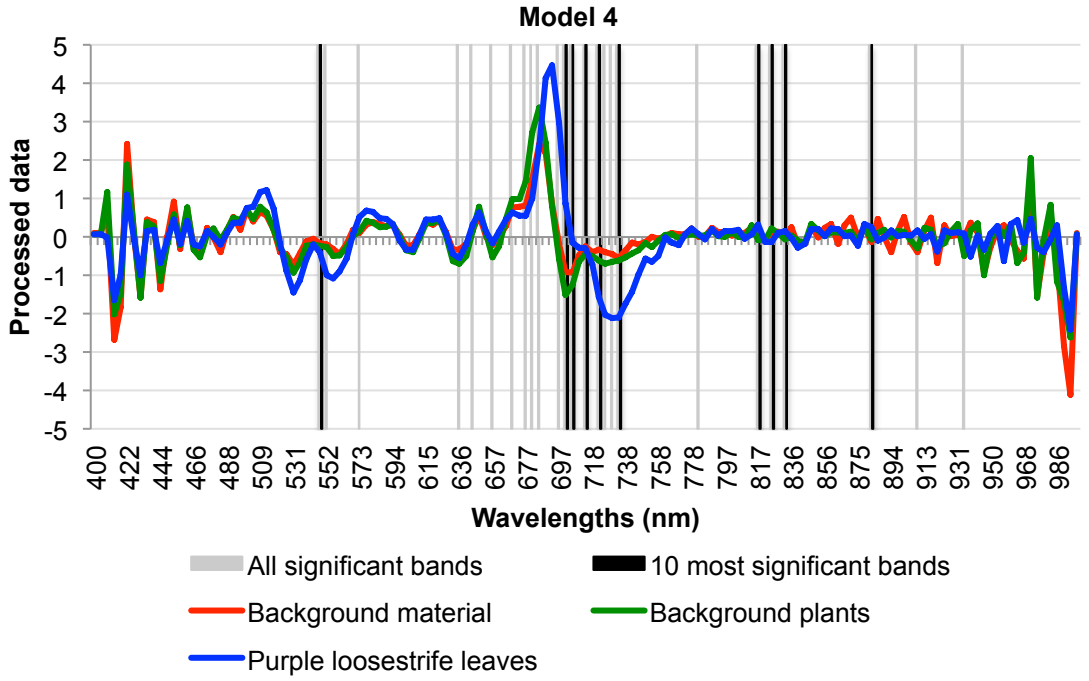

Figure 8. Model 4 preprocessed spectral signatures indicating relative significance of wavebands. Model 4 identified 25 significant bands and showed a loose cluster of bands between 635-735 nm. The most important wavelength occurred at 832.4 nm.

3.3.4. Spectral Model Performance

The results of PLS-DA modeling had accuracies greater than 85% when distinguishing the inflorescences of the plants of interest, as well as their leaves, from the surrounding vegetation. Prediction results using the 4-factor PLSR models were calculated on all 1490 original bands of the validation group, resulting in R^2 values for each of the models (Table 3). The R^2 values indicated how much of the variability in the response variable was explained by the model. Model 1, which compared leafy spurge bracts, leaves, and the background plants and material, had the best prediction results of any of the models, with an R^2 value of 0.92. Models 2 and 4 had very close results, with R^2 values of 0.89 and 0.88, respectively. These two models compared the leaves of leafy spurge and purple loosestrife with surrounding background materials. Model 3, which compared purple loosestrife flowers, leaves, and the background plants and material, had the lowest prediction result, with an R^2 value of 0.86.

The accuracies found in this study are comparable to the accuracies (89-95%) reported by Garcia-Ruiz et al. (2015), when analyzing the spectral signatures of Canada thistle and sugarbeet, and Shirzadifar et al. (2018) when spectral data were used to discriminate three weeds species using SIMCA model (88.89-100%).

Table 3. Performance of PLS-DA models to discriminate two invasive plant species.

	Noxious weed	**# of categories**	**# of factors**	**# of bands**	**Cal. R^2**	**Val. R^2**
Model 1	Leafy spurge	3	4	73	0.92	0.92
Model 2	Leafy spurge	2	4	21	0.88	0.89
Model 3	Purple loosestrife	3	4	37	0.91	0.86
Model 4	Purple loosestrife	2	4	25	0.89	0.88

This table shows the calibration R^2 results from the model development and the final validation R^2 results. Model 1 had the best prediction value of 0.92.

3.3.5. Discussion

While models 2, 3, and 4 reduced the number of significant bands to fewer than 40, model 1 required almost twice as many bands in order to achieve the R^2 value of 0.92. This larger number of bands puts more of a limitation on model 1 than on the other models, making it a more complex and less robust model. Model 3 consisted of nearly half the number of bands as model 1, at 37, while models 2 and 4 consisted of nearly one-third the number of bands, at 21 and 25 bands, respectively. The smaller number of bands allowed clusters of the significant bands to be identified, indicating which groups of wavelengths would be best to include in sensors to identify the weeds. If the number of bands needed for a model can be reduced to less than ten, then a multispectral sensor can be used to analyze the data. Otherwise, a hyperspectral sensor would be necessary, which would increase the cost of the equipment. Model 1 required two to three times as many bands as the other models; therefore, reducing the number of bands to

few enough in order to use a multispectral sensor to identify leafy spurge would be more difficult with model 1.

Additionally, models 1 and 3 consisted of three categories each, while models 2 and 4 consisted of two categories each. The larger number of categories of models 1 and 3 may have contributed to the larger number of bands required to achieve the accuracies of these models.

The inflorescences of leafy spurge and purple loosestrife were distinct yellow and purple colors, respectively, and were visually separable from the surrounding vegetation. Therefore, it is not surprising that the distinctly colored material was separated out in models 1 and 3. However, what is of more interest is that models 2 and 4 were able to achieve decent results when separating the green leaves of leafy spurge and purple loosestrife from the surrounding green vegetation.

The leaves of leafy spurge and purple loosestrife were compared in models 2 and 4 with the surrounding vegetation without including the inflorescences in the models, resulting in only two categories instead of three categories. Good results were achieved with each of these models using fewer bands than models 1 or 3. As mentioned above, if these noxious weeds can be identified by their leaves, then they can be identified earlier in the season before they start flowering. Because many herbicides are applied during flowering, identifying them early would allow for more time to plan and coordinate the herbicide application.

3.4. Conclusion

The objective of this paper was to classify leafy spurge and purple loosestrife noxious weeds based on spectral data manually collected in the field. The spectral data in this study shows promise for separating the spectral signatures of these two noxious weeds from the surrounding green vegetation and materials both by their inflorescences and their leaves. In this

study, the results of PLS-DA when comparing the spectral signatures of leafy spurge and purple loosestrife and the surrounding vegetation and materials indicated that both the inflorescences and leaves of the noxious weeds were spectrally distinct from their surroundings. The red and red edge wavelength regions appeared to be the most important regions for separating these plants from the surroundings. In each of the models, clusters of significant bands around 700 nm can be seen because chlorophyll absorption in this region is important in separating and identifying plants. Bands in these wavelength regions from about 600 nm to 730 nm should be included in sensors used to identify leafy spurge and purple loosestrife.

Determining that the spectral signatures of these noxious weeds were separable from their surroundings was the first step before using the spectral signatures to classify and identify the weeds in aerial imagery. Future research should include separating the spectral signatures of leafy spurge and purple loosestrife in ultra-high spatial resolution imagery, both before and during flowering. Including additional visible aspects of the plants in the imagery, such as texture or shape, could also be included to enhance the accuracy of the classification. Analysis of imagery during flowering, either in the greenhouse or in field conditions, should include yellow and purple flowering plants that grow in similar habits to leafy spurge or purple loosestrife.

There is high potential for identifying leafy spurge and purple loosestrife early in the season when the spectral signatures of the plant leaves or inflorescences are identifiable. Identifying the noxious weeds by their leaves will allow the noxious weeds to be identified earlier than when identifying the plants by their inflorescences. This, in turn, would allow for more time to identify the plants and plan for methods of eradication.

3.5. Reference

Carter GA, Lucas KL, Blossom GA, Lassitter CL, Holiday DM, Mooneyhan DS, Fastring DR, Holcombe TR, Griffith JA (2009) Remote sensing and mapping of tamarisk along the Colorado River, USA: A comparative use of summer-acquired Hyperion, Thematic Mapper and QuickBird data. Remote Sens 1:318–329

de Castro AI, López-Granados F, Jurado-Expósito M (2013) Broad-scale cruciferous weed patch classification in winter wheat using QuickBird imagery for in-season site-specific control. Precis Agric 14:392–413

Garcia-Ruiz FJ, Wulfsohn D, Rasmussen J (2015) Sugar beet (*Beta vulgaris* L.) and thistle (*Cirsium arvensis* L.) discrimination based on field spectral data. Biosyst Eng 139:1–15

Girma K, Mosali J, Raun WR, Freeman KW, Martin KL, Solie JB, Stone ML (2005) Identification of optical spectral signatures for detecting cheat and ryegrass in winter wheat. Crop Sci 45:477–485

Glenn NF, Mundt JT, Weber KT, Prather TS, Lass LW, Pettingill J (2005) Hyperspectral data processing for repeat detection of small infestations of leafy spurge. Remote Sens Environ 95:399–412

Hung C, Xu Z, Sukkarieh S (2014) Feature learning based approach for weed classification using high resolution aerial images from a digital camera mounted on a UAV. Remote Sens 6:12037–12054

Hunt ER, Daughtry CST, Kim MS, Parker Williams AE (2007) Using canopy reflectance models and spectral angles to assess potential of remote sensing to detect invasive weeds. J Appl Remote Sens 1:013506

Hunt ER, Gillham JH, Daughtry CST (2010) Improving potential geographic distribution models for invasive plants by remote sensing. Rangel Ecol Manag 63:505–513

Hunt ER, Parker Williams AE (2006) Detection of flowering leafy spurge with satellite multispectral imagery. Rangel Ecol Manag 59:494–499

Kloppenburg C (2014) Detecting Leafy Spurge in Native Grassland Using Hyperspectral Image Analysis. MS thesis. Lethbridge, AB: University of Lethbridge. 93 p

Knezevic SZ, Smith D, Kulm R, Doty D, Kinkaid D, Goodrich M, Stolcpart R (2004) Purple loosestrife (Lythrum salicaria) control with herbicides: Single-year application. Weed Technol 18:1255–1260

Laba M, Blair B, Downs R, Monger B, Philpot W, Smith S, Sullivan P, Baveye PC (2010) Use of textural measurements to map invasive wetland plants in the Hudson River National Estuarine Research Reserve with IKONOS satellite imagery. Remote Sens Environ 114:876–886

Laba M, Tsai F, Ogurcak D, Smith S, Richmond ME (2005) Field determination of optimal dates for the discrimination of invasive wetland plant species using derivative spectral analysis. Photogramm Eng Remote Sensing 71:603–611

Lym, RG (1998) The biology and integrated management of leafy spurge (Euphorbia esula) on North Dakota rangeland. Weed Technol 12:367–373

Mafanya M, Tsele P, Botai J, Manyama P, Swart B, Monate T (2017) Evaluating pixel and object based image classification techniques for mapping plant invasions from UAV derived aerial imagery: *Harrisia pomanensis* as a case study. ISPRS J Photogramm Remote Sens 129:1–11

Martens H, Martens M (2000) Modified Jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR). Food Qual Prefer 11:5–16

Mirik M, Ansley RJ, Steddom K, Jones DC, Rush CM, Michels GJ, Elliott NC (2013) Remote distinction of a noxious weed (musk thistle: *Carduus nutans*) using airborne hyperspectral imagery and the support vector machine classifier. Remote Sens 5:612–630

Mitchell JJ, Glenn NF (2009) Leafy spurge (*Euphorbia esula*) classification performance using hyperspectral and multispectral sensors. Rangel Ecol Manag 62:16–27

Mullin BH (1998) The biology and management of purple loosestrife (Lythrum salicaria). Weed Technol 12:397–401

Narumalani S, Mishra DR, Wilson R, Reece P, Kohler A (2009) Detecting and mapping four invasive species along the floodplain of North Platte River, Nebraska. Weed Technol 23:99–107

North Dakota Department of Agriculture (2016) Noxious Weeds. http://www.nd.gov/ndda/plant-industries/noxious-weeds. Accessed: November 1, 2016

Noujdina NV, Ustin SL (2008) Mapping downy brome (*Bromus tectorum*) using multidate AVIRIS data. Weed Sci 56:173–179

O'Neill M, Ustin SL, Hager S, Root R (2000) Mapping the distribution of leafy spurge at Theodore Roosevelt National Park using AVIRIS. Pages 0-18 *in* Proceedings of the Ninth JPL Airborne Earth Science Workshop. Pasadena, California: NASA Jet Propulsion Laboratory

Parker Williams A, Hunt ER (2002) Estimation of leafy spurge cover from hyperspectral imagery using mixture tuned matched filtering. Remote Sens Environ 82:446–456

Peña JM, Torres-Sánchez J, de Castro AI, Kelly M, López-Granados F (2013) Weed mapping in early-season maize fields using object-based analysis of unmanned aerial vehicle (UAV) images. PLOS One 8:e77151

Pérez-Ortiz M, Peña JM, Gutiérrez PA, Torres-Sánchez J, Hervás-Martínez C, López-Granados F (2015) A semi-supervised system for weed mapping in sunflower crops using unmanned aerial vehicles and a crop row detection method. Appl Soft Comput 37:533–544

Savitzky A, Golay MJE (1964) Smoothing and differentiation of data by simplified least squares procedures. Anal Chem 36:1627–1639

Shapira U, Herrmann I, Karnieli A, Bonfil DJ (2013) Field spectroscopy for weed detection in wheat and chickpea fields. Int J Remote Sens 34:6094–6108

Shapira U, Herrmann I, Karnieli A, Bonfil JD (2010) Weeds detection by ground-level hyperspectral data. Theory to Pract XXXVIII:27–33

Shirzadifar A, Bajwa S, Mireei SA, Howatt K, Nowatzki J (2018) Weed species discrimination based on SIMCA analysis of plant canopy spectral data. Biosyst Eng 171:143–154

Swain S, Narumalani S, Mishra DR (2011) Monitoring invasive species: Detecting purple loosestrife and evaluating biocontrol along the Niobrara River, Nebraska. GIsci Remote Sens 48:225–244

Torres-Sánchez J, López-Granados F, de Castro AI, Peña-Barragán JM (2013) Configuration and specifications of an unmanned aerial vehicle (UAV) for early site specific weed management. PLOS One 8:e58210

Ustin SL, Santos MJ (2010) Spectral identification of native and non-native plant species. Pages 1–17 *in* Proceedings of ASD and IEEE GRS; Art, Science and Applications of Reflectance Spectroscopy Symposium, Vol. 2. Boulder, CO