REMOTE SENSING BASED DETECTION OF FORESTED WETLANDS: AN EVALUATION OF LIDAR, AERIAL IMAGERY, AND THEIR DATA FUSION

by

Ashley Elizabeth Suiter

B.S., Central Michigan University, 2011

A Thesis
Submitted in Partial Fulfillment of the Requirements for the
Master of Science

Department of Geography and Environmental Resources
in the Graduate School
Southern Illinois University Carbondale
May 2015

THESIS APPROVAL

REMOTE SENSING BASED DETECTION OF FORESTED WETLANDS: AN EVALUATION OF LIDAR, AERIAL IMAGERY, AND THEIR DATA FUSION

By

Ashley Elizabeth Suiter

A Thesis Submitted in Partial

Fulfillment of the Requirements

for the Degree of

Master of Science

in the field of Geography and Environmental Resources

Approved by:

Dr. Guangxing Wang, Chair

Dr. Justin Schoof

Dr. Jonathan Remo

Graduate School
Southern Illinois University Carbondale
March 25th, 2015

AN ABSTRACT OF THE THESIS OF

Ashley Elizabeth Suiter, for the Master of Science degree in Geography and Environmental Resources, presented on March 25th, 2015, at Southern Illinois University Carbondale.

TITLE: REMOTE SENSING BASED DETECTION OF FORESTED WETLANDS: AN EVALUATION OF LIDAR, AERIAL IMAGERY, AND THEIR DATA FUSION

MAJOR PROFESSOR: Dr. Guangxing Wang

Multi-spectral imagery provides a robust and low-cost dataset for assessing wetland extent and quality over broad regions and is frequently used for wetland inventories. However in forested wetlands, hydrology is obscured by tree canopy making it difficult to detect with multi-spectral imagery alone. Because of this, classification of forested wetlands often includes greater errors than that of other wetlands types. Elevation and terrain derivatives have been shown to be useful for modelling wetland hydrology. But, few studies have addressed the use of LiDAR intensity data detecting hydrology in forested wetlands. Due the tendency of LiDAR signal to be attenuated by water, this research proposed the fusion of LiDAR intensity data with LiDAR elevation, terrain data, and aerial imagery, for the detection of forested wetland hydrology. We examined the utility of LiDAR intensity data and determined whether the fusion of Lidar derived data with multispectral imagery increased the accuracy of forested wetland classification compared with a classification performed with only multi-spectral image.

Four classifications were performed: Classification A – All Imagery, Classification B – All LiDAR, Classification C – LiDAR without Intensity, and Classification D – Fusion of All Data. These classifications were performed using random forest and each resulted in a 3-foot resolution thematic raster of forested upland and forested wetland

locations in Vermilion County, Illinois. The accuracies of these classifications were compared using Kappa Coefficient of Agreement. Importance statistics produced within the random forest classifier were evaluated in order to understand the contribution of individual datasets. Classification D, which used the fusion of LiDAR and multi-spectral imagery as input variables, had moderate to strong agreement between reference data and classification results. It was found that Classification A performed using all the LiDAR data and its derivatives (intensity, elevation, slope, aspect, curvatures, and Topographic Wetness Index) was the most accurate classification with Kappa: 78.04%, indicating moderate to strong agreement. However, Classification C, performed with LiDAR derivative without intensity data had less agreement than would be expected by chance, indicating that LiDAR contributed significantly to the accuracy of Classification B.

DEDICATION

For my grandfather, J. Michael Suiter, who had a passion for knowledge, love of nature, and showed me the value in serving others. You were and continue to be my greatest influence. You are missed.

ACKNOWLEDGMENTS

First, I would like to thank my advisor Dr. Guangxing Wang for his insight and guidance during this research. I would also like to express my appreciation for the input of my committee members Dr. Justin Schoof and Dr. Jonathan Remo. To Michael Sertle, thank you for the love, support, and understanding that helped to make this work possible. And finally, to my brother, Liam Grantham, your hard work, and enthusiasm were crucial for the field component of this research. Without you I may still be stuck in a wetland, encased in several feet of mud.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FIGURE PAGE

CHAPTER 1 INTRODUCTION

1.1 BACKGROUND

It is estimated that half of the world's total wetland area has been lost and what wetlands remain have been severely degraded (Zedler and Kercher 2005). Drainage and conversion to agricultural land has been the most significant cause of wetland loss in the United States. In Illinois, where the majority of wetlands are now forested, over 90% of the wetlands have been drained due to agriculture (Dahl and Allord 1996). Although forested wetlands are the most common wetland type in Illinois, and in the United States, they have suffered greater loss in total area than any other wetland type (U.S. Fish and Wildlife Service 2002; Zedler and Kercher 2005; Lang *et al.* 2009). Research suggests that forested wetlands are also the most likely to be lost in the future (U.S. Fish and Wildlife Service 2002; Zedler and Kercher 2005; Lang *et al.* 2009).

Although wetlands currently occupy a relatively small portion of the earth’s land (only 9% worldwide and 5% in the U.S.), ecosystem services provided by wetlands are estimated to be valued at upwards of $70 billion dollars annually (Wilen and Tiner 1989; Zedler and Kercher 2005). Wetlands contribute to ecological diversity, improve air and water quality, sequester carbon, and provide opportunities for recreational activities such as birding, fishing, boating, and hunting. Forested wetlands in particular support biodiversity by providing unique habitat that is vital for the lifecycle of some wildlife (Zedler and Kercher 2005; Töyrä and Pietroniro 2005). Wetlands are retention areas for floodwater storage, which can reduce downstream flooding, and therefor decreasing the cost of flood damages (Zedler and Kercher 2005; Töyrä and Pietroniro 2005; Lang *et al.*

2009). Additionally, forested wetlands have substantial potential for carbon sequestration (Huang *et al.* 2011*).*

Clearly forested wetlands are a crucial component of the landscape. Due to their importance, it is necessary to accurately map and monitor these habitats so that their extent, and quality, can be better understood by natural resource managers. Wetland inventories, in general, can provide context for assessing ecosystem health at regional and landscape levels (Bourgeau-Chavez *et al.* 2008a; Bourgeau-Chavez *et al.* 2008b; Lang *et al.* 2009; Poulin, Davranche, and Lafebvre 2010). These inventories are particularly important for understanding how anthropogenic activities have influenced the abundance and composition of wetlands over time (Bourgeau-Chavez *et al.* 2008b; Poulin, Davranche, and Lafebvre 2010). Wetland inventories typically include the extent and location of wetlands, and also vegetation composition and hydrologic regime; information that may be used to assess potential productivity, determine flood storage potential, or be used to target restoration activities (Töyrä and Pietroniro 2005; Lang and McCarty 2009).

Although many countries do not have wetland inventories, in the United States two federal agencies are tasked with inventorying wetlands, the Army Corps of Engineers (ACE) and the U.S. Fish and Wildlife Service (USFWS). A number of states (i.e. Wisconsin) even complete their own inventories. The U.S. Fish and Wildlife Service performs a wetland inventory, known as the National Wetland Inventory (NWI), to support informed management of wetlands as necessary habitat for numerous plants and animals (Cowardin 1979; Wilen and Tiner 1989). Inventories performed by the USFWS use the Cowardin classification system, a qualitative, hierarchical system, for

classifying wetlands and deep-water habitats (Cowardin 1979, FGDC 1992, FGDC 2009). Within this system wetlands are defined as:

> *...lands transitional between terrestrial and aquatic systems where the water table is usually at or near the surface or the land is covered by shallow water. For purposes of this classification wetlands must have one or more of the following three attributes: (1) at least periodically, the land supports predominantly hydrophytes; (2) the substrate is predominantly undrained hydric soil; and (3) the substrate is nonsoil and is saturated with water or covered by shallow water at some time during the growing season of each year (Cowardin 1979, 3).*

The USFWS uses remote sensing as the primary method for performing the National Wetland Inventory. Remote sensing, through satellite and aerial photography, provides a cost effective method for conducting wetland inventories over a large geographic regions. The goal of updating the National Wetlands Inventory every 10 years was set by the Emergency Wetland Resource Act of 1986 (Wilen and Tiner 1989). Currently, lack of funding has affected the frequency with which the National Wetland Inventory is updated. The planned update interval is every 20 years, with federal funding that allows for less than 2% of the United States to be updated per year (Awl *et al.* 2009). In light of these insufficient resources, the USFWS has developed a strategy for continued updating of wetland maps that includes: 1) Prioritization of mapping and update efforts in the most critical regions; 2) Working with partner agencies (federal, state, local and non-governmental) to share data and funding for mapping efforts; 3) Serving as coordinator for mapping efforts done by partner agencies; and 4) Developing more efficient, and cost effective methods for inventorying wetlands (U.S. Fish and Wildlife Service 2002; Wright and Gallant 2007; Awl *et al.*

2009). This research addresses the 4th strategy by developing new methods for overcoming the inventory difficulty to map forested wetlands in Vermilion County Illinois.

1.2 PROBLEM STATEMENT

Though hydrology is the major abiotic control of wetland location and extent, and one of the three indicators of wetland status based on the definition of the USFWS, few studies have addressed the methods for directly detecting wetland hydrology in forested areas (Lang and McCarty 2009). The intention of this research is to develop an innovative method for identifying forested wetland hydrology, and thus forested wetlands, within a remote sensing classification. Forested wetlands have historically been the most difficult to map due to the tree canopy obscuring wetland hydrology. Traditional optical remote sensors are not adequate for these purposes because they are unable to 'see' below the vegetated tree canopy (Corcoran *et al.* 2013). However, recent research has shown that LiDAR may be the solution to this problem.

LiDAR has the unique ability to penetrate the vegetation canopy and can be used to create high resolution, bare-ground elevation dataset in forested areas. These high-resolution elevation datasets have been shown to identify wetlands with greater accuracy than low and moderate resolution datasets (Hogg and Holland 2007). Due to the absorption of LiDAR infrared signal by water, the LiDAR intensity data, exhibits uniquely low reflectance values in the areas of moist soils and open water (Silva *et al.* 2008). Some researchers have noted this as a negative component of utilizing LiDAR data when in fact it may be useful. The LiDAR signal is unable to be reflected from open water surfaces, making these features evident in LiDAR intensity data. A number of

researchers have shown the utility of using such datasets for forested wetland mapping. Still, more research needs to be done to evaluate the predictive power of LiDAR derived datasets, particularly intensity, compared to multi-spectral aerial imagery.

1.3 RESEARCH QUESTIONS

The objective of this research was to address the following questions:

1. What is the accuracy difference of a forested wetland classification using only multi-spectral imagery compared to a classification performed using LiDAR data?

2. Can the fusion of LiDAR and Aerial Imagery datasets improve forested wetland classification accuracy compared with classification of a single data source?

3. Is LiDAR ground return intensity data useful for forested wetland identification?

CHAPTER 2

LITERATURE REVIEW

2.1 REMOTE SENSING OF WETLANDS

Due to the frequent presence of dense vegetation, saturated soils, and uneven terrain, wetlands are particularly difficult to monitor in situ (Töyrä and Pietroniro 2005; Lang *et al.* 2009). Collecting extensive field observations of wetland conditions is quite time-consuming and thus it is cost-prohibitive to produce landscape scale wetland maps from field observations alone (Zhou *et al.* 2010; Xie *et al.* 2011). Successful wetland mapping efforts use methods that are robust, low-cost, and easily implemented (Bourgeau-Chavez *et al.* 2008a). Remote sensing fulfills these requirements by allowing for the detection of wetland characteristics, over a broad geographic region, without physically going to each wetland; providing a cost-effective method for performing wetland inventories on the landscape scale (Bourgeau-Chavez *et al.* 2008a; Bourgeau-Chavez *et al.* 2008b; Lang *et al.* 2009). In addition to cost-effectiveness, remote sensing provides methods that can be easily and accurately replicated by other researchers (Töyrä and Pietroniro 2005; Zhou *et al.* 2010). Furthermore, remote sensing has the ability to classify the surrounding upland habitat, providing valuable information about factors that potentially affect the quality and quantity of wetlands (Bourgeau-Chavez *et al.* 2008a).

Because of these properties, remote sensing techniques are the most commonly used methods for performing wetland inventories, and the method used by the U.S. Fish and Wildlife Service (Wilen and Tiner 1989; FGDC 1992; FGDC 2009 Klemas 2011). Wetland status, as determined by the USFWS, is based upon evidence of the presence

of one or more of the following wetland characteristics: hydric vegetation, hydric soils, and/or wetland hydrology. To be effective, remote sensing data must provide information about these characteristics. Accuracy of these wetland inventories depends on the type of remotely sensed data used, classification methods, as well as the landscape being classified (Lu and Weng 2007). A review of the literature pertaining to identification of wetlands, particularly forested wetlands, quickly reveals the many factors that must be considered before conducting a wetland inventory. The following discussion will focus on remotely sensed datasets and classification techniques used for the discrimination of wetland indicators - hydrophytic vegetation, hydric soils, and wetland hydrology.

2.1.1. PASSIVE REMOTE SENSING

Remote sensing can be accomplished using either passive or active sensors. Passive sensors rely on electromagnetic energy from the sun reflecting off the feature of interest to be recorded by the sensors. Examples of passive sensors include optical satellite and aerial imaging. Many studies have focused on the use of these sensors for wetland detection. However, a major drawback for the use of passive sensors for forested wetland identification is their inability to receive electromagnetic energy from obstructed features. Passive sensors may be able to identify forest tree species based on their signature, or unique reflectance in each channel of the remote sensing data. However, if foliage is present, the passive sensor cannot receive reflected radiation from below the tree canopy, making forested wetlands difficult to map with passive

remotely sensed data alone (Bourgeau-Chavez *et al.* 2008a; Lang and McCarty 2009; Lang *et al.* 2013).

2.1.1.1 MULTI-SPECTRAL IMAGERY

Multi-spectral imagery, sometimes referred to as visible/infrared, or color-infrared imagery, typically senses radiation within the visible (red, green, blue) and infrared spectrum, though they may also collect data from the ultra violet and thermal regions. Sensors that collect multi-spectral imagery are deployed by satellite or aircraft. These sensors record the spectral reflectance (and sometimes emittance) of surface features, including vegetation. This spectral information can then be used to detect wetland indicators such as hydrophytic vegetation (i.e. vegetation that is adapted to wet conditions) or vegetation that has undergone some stress due to inundation (Schmidt and Skidmore 2003; Bourgeau-Chavez *et al.* 2008b; Silva *et al.* 2008; Lang *et al.* 2013). Vegetation stress may provide information indicating soil moisture or inundation in locations that are not wetlands and do not have vegetation adapted to these conditions (Bourgeau-Chavez *et al.* 2008b; Lang *et al.* 2013).

Vegetation monitoring using remote sensing can be accomplished using visible infrared satellite remote sensing systems such as Satellite Pour L'Observation de la Terre (SPOT), IKONOS, Advanced Spaceborne Thermal Emission and Reflection Radiometer **(**Aster), Moderate-resolution Imaging Spectroradiometer (MODIS), Landsat, and aerial color-infrared sensor systems. Optical sensors deployed on earth-orbiting satellites are useful due to their capacity for monitoring very large areas at regular intervals, providing both the spatial and temporal coverage needed to detect and

classify wetlands (Jensen 1996; O'Hara 2002; *T*öyrä and Pietroniro 2005; Corcoran, Knight, and Gallant 2013). The key factors to consider when acquiring satellite data are obtaining cloud free scenes, pre-processing to account for atmospheric influence, and the spectral and spatial resolution of the potential data source (Töyrä and Pietroniro 2005). Low and moderate resolution satellite imagery may have difficulty in spatially heterogeneous areas in which a single pixel receives the spectral reflectance of multiple cover types, resulting in classification confusion (Töyrä and Pietroniro 2005). This is known as the mixed pixel problem and may be reduced as the spatial resolution of the imagery is increased. Zhou *et al.* (2010) indicates that satellite spectral information tends to be noisy, and that noise increases with higher spatial resolutions, partially due to interference with the atmosphere. Satellite images require pre-processing procedures to account for atmospheric interference as well as to convert digital numbers to spectral reflectance values (Chandler, Markham, and Helder 2009).

SPOT-5 is a French commercial satellite able to capture spectral information at 2.5 to 20-meter resolution (Lillesand, Kiefer, and Chipman 2008; Klemas 2011). Unfortunately the expense of SPOT imagery makes it impractical for large-scale wetland mapping efforts (*T*öyrä and Pietroniro 2005; Klemas 2011). IKONOS imagery is a commercial high-resolution satellite optical sensor system whose 4 meter multispectral (red, green, blue, and infrared) bands can be pansharpened using the 1 meter panchromatic band (Lillesand, Kiefer, and Chipman 2008; Zhou *et al.* 2010). A major benefit of IKONOS imagery is its fine spatial resolution (0.6 to 4 meters) (Klemas 2011). However, IKONOS is expensive and implementation on a large project scale

may not be worth the cost, particularly where other sensors such as Landsat, can be used (Klemas 2011).

ASTER is a moderate (15 to 90 meter) resolution commercial satellite optical sensor system that is not frequently used for wetland classifications (Pantaleoni *et al.* 2009). However, Pantaleoni *et al.* (2009) was able to use this data to classify wetlands with some success. MODIS is a medium/coarse resolution multispectral sensor that provides full earth coverage almost every day (Pflugmacher, Krankina, and Cohen 2007). This sensor consists of 36 spectral bands with resolution ranging from 250 meters to 1 kilometer. The spectral, temporal, and spatial coverage of MODIS makes it an attractive choice for performing wetland assessments. However, due to the low spatial resolution of the sensor it is not ideal for monitoring smaller features such as forested wetlands (Pflugmacher, Krankina, and Cohen 2007).

Gritzner (2009) showed the utility of Landsat 7 infrared bands for the detection of open water in the prairie pothole region. This was done by applying a thresholding technique to the near-infrared channel (band 5.) This technique identifies regions of low reflection in the infrared band of the Landsat 7 optical satellite sensor. This is useful for identifying surface water because the infrared frequency of electromagnetic radiation is attenuated by water ((Töyrä *et al.* 2002; Gritzner 2006; Lang *et al.* 2009). However, optical remote sensing methods have been shown to be ineffective for the identification of wetland hydrology when hydrologic characteristics are obscured by a vegetated canopy, as is the case with most forested wetlands in leaf-on conditions. (Tiner 1990; Töyrä and Pietroniro 2005; Lang *et al.* 2013). Landsat is the most commonly used satellite optical sensor system for land use and land-cover (LULC) classifications

(Jensen 1996; Klemas 2011). Landsat images are freely available and provide moderate resolution (30 meter) imagery every 16 days (Jensen 1996; Pflugmacher *et al.* 2007; Klemas 2011). It is because of the cost effectiveness, moderate spatial resolution, and multispectral capabilities that Landsat is so widely used for land cover mapping.

Aerial imagery is obtained by mounting visible/infrared cameras on low flying aircraft. Aerial imagery is typically collected with spectral resolution that is comparable to satellite imagery and finer spatial resolution than satellite sensors. Although satellite images are useful for wetland classification, aerial imagery is the most common optical sensor used for wetland inventories in the United States (Lang *et al.* 2009). This is a result of moderate and low-resolution satellite images being outperformed by aerial imagery's ability to detect fine-scale wetland details needed for visual identification and classification of wetlands according to the Cowardin Classification System (Nielsen, Prince, and Koeln 2008; Klemas 2011). The Federal Geographic Data Committee (FGDC), Wetlands Mapping Standards Subcommittee requires imagery used for mapping wetlands in the contiguous United States to have a minimum of 1ft spatial resolution and requires the inclusion of visible and infrared bands (FGDC 2009). These standards are easily met by multi-spectral aerial imagery. These standards have been established to ensure that wetlands a half acre in size (the target minimum mapping unit for the National Wetland Inventory) can be adequately mapped and included in wetland mapping efforts (FGDC 2009).

2.1.1.3 HYPERSPECTRAL IMAGERY

The primary difference between multi-spectral and hyperspectral imagery is the number of channels, and the width of the spectral region in which reflected energy can be recorded. Hyperspectral imagery contains many more bands than multi-spectral imagery, with each band collecting information in only a narrow region of the electromagnetic spectrum. Although multi-spectral aerial imagery is the most common optical sensor used for wetland inventories in the United States, hyperspectral sensors provide the unique ability to distinguish very small differences in spectral reflectance that may be useful for discerning otherwise spectrally similar vegetation (*T*öyrä and Pietroniro 2005; Lang *et al.* 2009). Hyperspectral datasets, due to their high spectral resolutions, are excellent for identification of wetlands. They may even aid in estimating the carbon sequestration potential of wetlands (Klemas 2011). NASA's Airborne Visible/Infrared Imaging Spectrometer (AVRIS) system is a satellite-deployed passive remote sensing system that collects 224 separate spectral channels of data. However, hyperspectral imaging is expensive to acquire, and requires complex processing, making operational use of this technology impractical (Lang *et al.* 2009).

2.1.1.4 IMAGE INDICES

Spectral information contained in multi-spectral and hyperspectral images can be analyzed on a channel-by-channel basis or by visualizing some combination of these bands. However, it is frequently useful to calculate image indices and band ratios that combine information from two or more channels to evaluate vegetation, soil, and moisture characteristics. Image indices can be used for detecting open water, moist

soils, or hydrophytic vegetation based high leaf moisture content. Common image indices used for this purpose include the Simple Ratio, Normalized Difference Vegetation Index (NDVI), and Normalized Difference Water Index (NDWI) (Gao 1996; Jensen 1996; Poulin, Davranche, and Lafebvre 2010.)

The simple ratio and NDVI are commonly used to distinguish healthy vegetation based on reflectance in the red and infrared regions. Red reflectance contains information about chlorophyll content in leaves at the surface with lower red reflectance indicating the presence of more chlorophyll than in regions with high reflectance (Gao 1996; Jensen 1996). The infrared reflectance is absorbed by water and lower values indicate greater leaf water content (Gao 1996; Jensen 1996.) The simple ratio is calculated as the ratio of the near infrared to red channel. The NDVI is calculated as (near infrared – red) / (near infrared + red) produces similar results compared to the simple ratio. The primary difference is that the NDVI does not have the range to separate features in high-biomass areas whereas the simple ratio does (Jensen 1996; Poulin, Davranche, and Lafebvre 2010.) The NDWI is calculated as (green - near infrared) / (green + near infrared) and is complementary to the simple ratio and NDVI in that it improves the separation of open water and terrestrial features (Gao 1996; McFeeters 1996; Poulin, Davranche, and Lafebvre 2010.)

2.1.2 ACTIVE REMOTE SENSING

Active sensors do not rely on the sunlight as a source of energy; instead the sensor generates and emits its own signal and then records its return. The availability and quality of active remotely sensed data, such as Radio Detection and Ranging

(RADAR) and Light Detection and Ranging (LiDAR), have increased in recent years, providing unique opportunities for remotely sensing forested wetlands, particularly when the actively sensed data is combined with passively sensed data (Töyrä *et al.* 2002; Töyrä and Pietroniro 2005; Zedler and Kercher 2005; Bourgeau-Chavez *et al.* 2008b; Lang *et al.* 2009; Lang and McCarty 2009; Huang *et al.* 2011; Ellis, Mahler, and Richardson 2012; Lang *et al.* 2013). Bourgeau-Chavez *et al.* (2008a) showed that when passive remote sensing data is used, such as satellite optical imagery, along with complementary active sensor data, such as RADAR, the accuracy of the resulting wetland maps is greater than the accuracy of those produced only with optical data. The use of a multi-sensor, multi-temporal approach, is particularly important for the mapping of forested wetlands, where the use of optical imagery is not adequate (Bourgeau-Chavez *et al.* 2008a; Bourgeau-Chavez *et al.* 2008b; Lang *et al.* 2009; Pantaleoni *et al.* 2009; Corcoran, Knight, and Gallant 2013).

2.1.2.1 RADAR

A few researchers have used 30-meter resolution RADAR active remote sensing to detect wetland hydrology (Töyrä *et al.* 2002; Töyrä and Pietroniro 2005; Bourgeau-Chavez *et al.* 2008a; Bourgeau-Chavez *et al.* 2008b; Lang *et al.* 2009). RADAR broadcasts long-wavelength microwaves through the atmosphere to the ground and then record back scattered energy. RADAR images may be collected in nearly any weather due to its ability to penetrate clouds. Additionally, at low incidence angles RADAR is able to penetrate most vegetation (Töyrä *et al.* 2002; Lang *et al.* 2009). The strength of the RADAR signal reflected and scattered back to the sensor is dependent

on the texture and characteristics of the reflecting features as well as the signal characteristics (wavelength, angle, and polarization) (Töyrä and Pietroniro 2005; Bourgeau-Chavez *et al.* 2008a; Bourgeau-Chavez *et al.* 2009.) The strength of this reflected signal, also known as signal intensity, can be used to discriminate open water from dry and flooded vegetation (Töyrä *et al.* 2002; Töyrä and Pietroniro 2005; Bourgeau-Chavez *et al.* 2008a; Bourgeau-Chavez *et al.* 2008b; Lang *et al.* 2009). L-Band RADAR has been proven to discriminate between vegetation types such as cattail, Phragmites, and bulrush (Bourgeau-Chavez *et al.* 2008a). In forested wetland environments, RADAR intensity is quite high due to the signal double bouncing from the water, to the trees, and back to the sensor. This high backscatter can be used to locate inundated areas. In fact, RADAR cannot only identify forested wetland areas, but also it has been proven to discriminate between moist soil and surface water when optical imagery was unable to (Töyrä *et al.* 2002; Zedler and Kercher 2005; Bourgeau-Chavez *et al.* 2008a; Bourgeau-Chavez *et al.* 2008b; Land and Kasischke 2008.)

While RADAR has been used successfully for the detection of wetland hydrology, and forested wetlands, RADAR may not be appropriate for every application. The cost of RADAR data acquisition is quite high and may outweigh its ability to detect inundation when compared with other datasets (Bourgeau-Chavez *et al.* 2008b). At 30-meter resolution, RADAR provides relatively coarse resolution datasets making these data less than ideal for detection of small features (Töyrä *et al.* 2002; Bourgeau-Chavez *et al.* 2008a; Bourgeau-Chavez *et al.* 2008b; Lang *et al.* 2009). The sensitivity of RADAR sensor can result in data that is quite noisy, necessitating the use of noise filtering techniques before the data can be used (Töyrä *et al.* 2002). Additionally the bounce

effect of the RADAR signal can result in classification of inundated areas that are slightly removed from their actual locations (Bourgeau-Chavez *et al.* 2008b). Regardless, the use of RADAR data in combination with optical imagery has been shown to improve wetland classification compared to classification with optical imagery alone (Töyrä *et al.* 2002; Bourgeau-Chavez *et al.* 2008a; Bourgeau-Chavez *et al.* 2008b; Lang *et al.* 2009).

2.1.2.2 LIDAR

LiDAR typically uses visible infrared laser energy to collect vegetation and ground elevation data (Wehr and Lohr 1999; Lang *et al.* 2009; Song *et al.* 2012). LiDAR emits dense pulses of laser energy, typically 900 to 1550 nm, and records the returned energy, documenting the time it takes for that energy to reflect off a surface and return to the sensor (Wehr and Lohr 1999; Töyrä and Pietroniro 2005; Lang and McCarty 2009, Song *et al.* 2012). That time, along with information about the sensors height, is then used to calculate the elevation of the surface from which the signal reflected (Wehr and Lohr 1999). The laser energy is able to penetrate most vegetation canopies allowing for the identification of below vegetation terrain conditions at a finer resolution than could previously be detected (Töyrä and Pietroniro 2005; Lang and McCarty 2009; Ellis *et al.* 2012; Song *et al.* 2012). However, in some areas of dense vegetation, LiDAR signal may not be able to penetrate through all layers of vegetation to the ground below (Xie *et al.* 2011).

Although LiDAR datasets require more computing power than lower resolution datasets, the high spatial resolution LiDAR is better able to identify potentially wet areas

than the datasets with coarser resolutions, due to increased accuracy and precision (Töyrä and Pietroniro 2005; Murphy *et al. 2007;* Bourgeau-Chavez *et al.* 2008; Remmel, Todd, and Buttle 2008; Huang *et al.* 2011; Lang *et al.* 2013). Hogg and Holland (2008) showed that the greater precision and accuracy of LiDAR elevation data (collected during leaf-on conditions) was able to improve the classification of forested wetlands from 76% to 84% when compared with a 20m resolution DEM. Lang *et al.* (2009) also showed that LiDAR increased forested wetland classification accuracy, but that the vertical accuracy of the LiDAR is reduced in forested areas. However, this error can be mitigated by collecting LiDAR measurements during leaf-off conditions (Lang *et al.* 2013). LiDAR is a relatively new technology that cost substantially more compared with freely available USGS elevation products typically derived from RaDAR (Lang and McCarty 2009; Gallant, Marinova, and Andersson 2011). Even so, for some applications, the improved accuracy of the LiDAR data may justify increased costs (Murphy *et al. 2007;* Gallant, Marinova, and Anderson 2011). Fortunately, airborne LiDAR data is becoming increasingly available through Federal, State, and Local government agencies (Lang and McCarty 2009).

While the usefulness of RADAR technology for mapping forested wetlands is well documented, very few researchers have addressed the utilization of LiDAR. There are two components of LiDAR point data: elevation and intensity (Song *et al.* 2012). LiDAR intensity, or the strength of the signal returned to the LiDAR sensor, is influenced by the properties of the material from which it reflects and may be useful for inferring information about the reflecting feature (Song *et al.* 2002; Lang *et al.* 2013). Unfortunately, because LiDAR is a relatively new and developing technology few

researchers have addressed the use of LiDAR intensity data for the classification of forested wetlands, or other land-cover features (Song *et al.* 2002; Lang and McCarty 2009). Still, LiDAR intensity has shown potential for overcoming the limitations of optical and RADAR data.

A LiDAR sensor may collect data using a laser signals within the green or infrared spectrum, depending on the project purpose. LiDAR data collected using a laser signal in the infrared spectrum provides a comparable ability for detecting flooded regions where the infrared signal is absorbed by water. Silva *et al.* (2008) indicated that even in areas with moist soil the LiDAR signal will be dampened, indicating that LiDAR intensity may also be able to separate dry and moist soils. Boyd and Hill (2007) found that in forested regions LiDAR intensity values were representative of the near-infrared reflections recorded by optical imagery and could be used similarly to traditional imagery. However, due to LiDAR's active sensor technology, it has greater ability to distinguish between inundated and non-inundated forested areas than passive optical sensors (Lang and McCarty 2009). In fact, the use of intensity data can result in wetland classifications that are 30% more accurate than those performed with 1m optical imagery (Lang & McCarty 2009; Lang *et al.* 2013).

Figure 2.1, modified from Lang and McCarty (2009) offers a visual comparison of 30 meter DEM, LiDAR intensity, aerial color infrared imagery, and RADAR data for detecting forested wetlands. As you can see, LiDAR intensity has a superior ability to detect below canopy inundation at fine resolution. Lang and McCarty (2009) utilized LiDAR intensity for separating inundated from un-inundated areas. Using visual inspection and expert knowledge Lang and McCarty (2009) determined the inundated

areas were most accurately identified by thresholding intensity values ranging from 0 to 50, while the un-inundated areas were well identified by thresholding intensity values from 80 to 255. Song *et al.* (2002) evaluated the feasibility of using LiDAR intensity data for LULC classification and found that the intensity data may be useful for enhancing the seperability of LULC classes but that more research is needed for standardizing, and filtering the noise in the intensity data.

Researchers suggest that the utilization of LiDAR intensity data may increase if methods of standardization for this data were developed (Song 2002; Boyd and Hill 2007; Lang and McCarty 2009; Hartfield, Landau, and Leeuwen 2011). Hartfield, Landau, and Leeuwen (2011) found that visibility of lack of intensity calibration between flight lines can result in classification confusion. Like RADAR, LiDAR intensity data tends to be noisy and should be filtered to reduce this noise, prior to classification. The usefulness of LiDAR data, including intensity, may increase if collected during both wet and dry periods allowing for modelling of flooding and dry ground (Töyrä *et al.* 2002). Töyrä and Pietroniro (2005) indicated that the difference between elevations collected during dry periods and elevations collected during wet periods could be used to estimate flood storage capacity. Although it has been acknowledged that LiDAR data can be costly, increasing utilization, including use of the intensity component may increase the cost-effectiveness (Song *et al.* 2002).

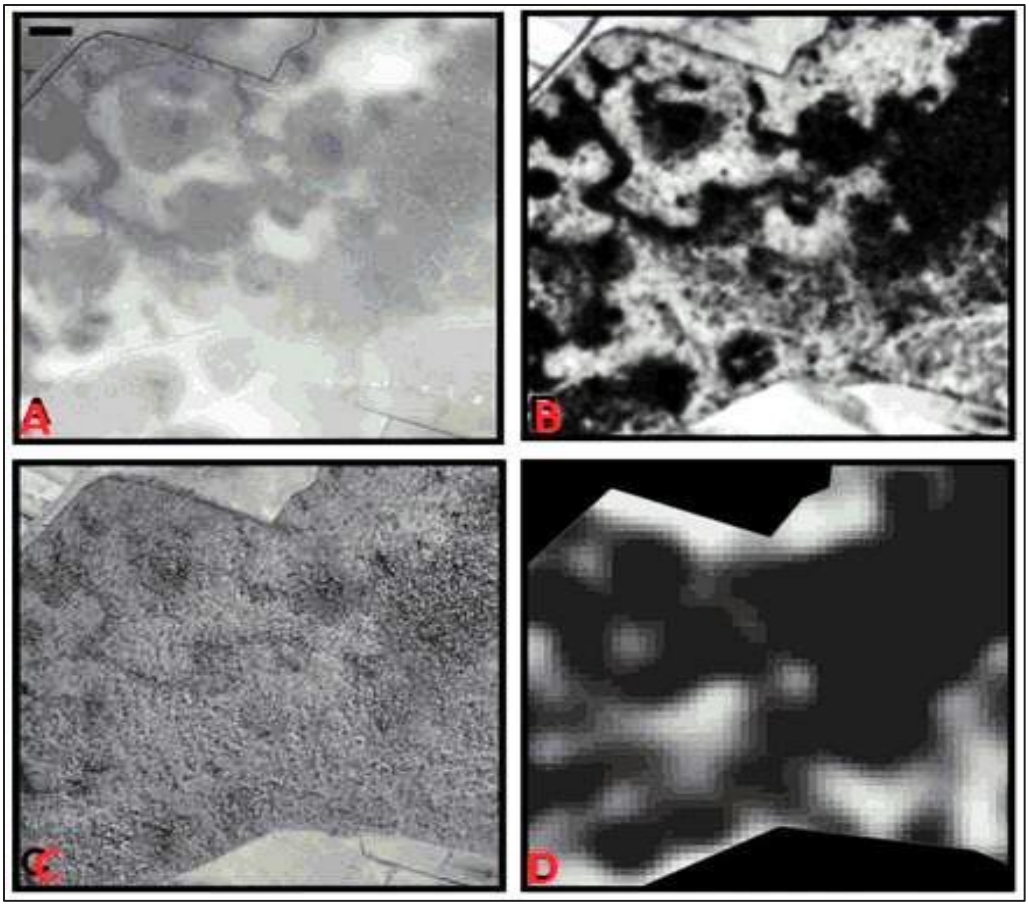


Figure 2.1: View of forested wetlands using four remote sensing data sources. A) 1-meter resolution DEM; B) LiDAR intensity image; C) 1-meter Color Infrared Imagery; D) 30-meter resolution RADAR intensity image (Modified from Lang *et al.* 2009).

2.1.2.3 TERRAIN DERIVATIVES

Active sensors are typically utilized in the creation of elevation data, derived as a function of the sensor height, signal speed, and signal return time. Where wetland and upland vegetation are spectrally similar, the identification of topographic features, such as depressions, may be used to indicate wetland status. Elevation and small variations

in topography have a significant influence over wetland location and so elevation datasets derived from the RADAR or LIDAR data can be used to create hydrologic models that predict soil moisture content and potential flooding associated with wetlands (Töyrä and Pietroniro 2005; Zedler and Kercher 2005; Murphy *et al. 2007;* Gallant, Marinova, and Andersson 2011). Traditionally digital elevation models (DEMs) were created using photo-interpretation techniques or field measurements-but this process was both time-consuming and costly (Remmel, Todd, and Buttle 2008). These methods resulted in very coarse, or low-resolution (1/10th degree and 9 second resolution), elevation datasets, with accuracy that decreased as distance from sample location increased (Gallant, Marinova, and Andersson 2011). The production of DEMs can now be created with sensors mounted on satellites and aircraft having evolved from topographic maps toward low and moderate spatial resolution elevation models derived from satellite borne RADAR data, and more recently to high spatial resolution LiDAR data (Gallant, Marinova, and Andersson 2011). An example of this is the Shuttle RADAR Topographic Mission (STRM) which has produced 90m resolution global DEMs (Gallant, Marinova, and Andersson 2011). ASTER has produced satellite image stereo pairs that can be used for the production of 30 DEMs (Gallant, Marinova, and Andersson 2011). TanDEM-X and TerraSAR-X both utilize RADAR technology to create a DEM that is up to 50% more accurate than STRM derived elevation models and at 12 m resolution (Gallant, Marinova, and Andersson 2011).

But, these elevation datasets are still very coarse for use in the identification of topographic features in areas where small-scale topographic variations are key to identifying forested wetlands (Töyrä and Pietroniro 2005; Lang and McCarty 2009; Lang

et al. 2013). The USGS produces 10-m and 30-m, moderate resolution DEMs. However, these datasets may have too low-resolutions for detecting wetland features in low-relief landscapes (Bourgeau-Chavez *et al.* 2008b; Huang *et al.* 2011). Kheir *et al.* (2009) found that moderate and coarse resolution elevation datasets result in over-simplification of the landscape and can propagate inaccuracies to derivative datasets such as slope, aspect, curvature, and flow accumulation (Kienzle 2004; Remmel, Todd, and Buttle 2008). However, higher spatial resolution elevation datasets, such as LiDAR derived DEMs, are better able to identify potentially wet areas than datasets with coarser resolutions, due to increased accuracy and precision (Töyrä and Pietroniro 2005; Murphy *et al. 2007;* Bourgeau-Chavez *et al.* 2008b; Remmel, Todd, and Buttle 2008; Huang *et al.* 2011; Lang *et al.* 2013). Additionally, RADAR and LiDAR have been proven to successfully distinguish hydrophytic vegetation based on biomass and vegetation structural information obtained by these sensors (MacKinnon 2001; Song *et al.* 2002; Lang *et al.* 2009; Gallant, Marinova, and Andersson 2011; Corcoran, Knight, and Gallant 2013).

Terrain information such as slope, curvature, aspect, and the topographic wetness index (TWI), can be derived from the elevation datasets and used to model soil moisture and wetland location (Kienzle 2004; Hogg and Todd 2007; Kheir *et al.* 2009; Lang *et al.* 2009; Lang *et al.* 2013). Incorporation of terrain information results in wetland classifications that are more accurate than those performed using optical imagery alone (Hogg and Todd 2007). The topographic wetness index (TWI), developed by Beven and Kirkby (1979), has been used to estimate the distribution of moist soils, and is one of the most widely used measures for identifying wetland hydrology (Kienzle

2004; Sörensen, Zinko, and Seibert 2006; Kheir *et al. 2009;* Ma *et al.* 2010). The TWI utilizes the slope and upslope contributing area to model runoff, and thus soil moisture (Beven and Kirkby 1979; Kienzle 2004; Remmel, Todd, and Buttle 2008; Ma *et al.* 2010). It should be noted that TWI does not account for infiltration or evaporation rates (Beven and Kirkby 1979). High TWI values correspond to inundated areas or areas that are likely to be, or have been wetlands (Lang and McCarty 2009). As you would expect, there is a strong positive correlation between depressional areas and high TWI values (Remmel, Todd, and Buttle 2008). Ma *et al.* (2010) found that aspect has an effect on soil moisture content and that TWI calculations weighted by the aspect of the slope results in a more accurate prediction of soil moisture.

2.1.3. ANCILLARY DATASETS

In some cases, remotely sensed data cannot capture the full range of variables necessary for identifying wetlands. Ancillary datasets may be used to assist in the identification of wetlands where remotely sensed data is inadequate (Lang and McCarty 2009). Data such as soils, elevation, and topographic indices derived from elevation datasets can provide additional information about wetland location and hydrology (Bourgeau-Chavez *et al.* 2008b).

Soils data such as the State Soil Survey Geographic Database (STATSGO) and the Soil Survey Geographic Database (SSURGO) are available through the United States Department of Agriculture - Natural Resources Conservation Service (USDA-NRCS) as part of the Soil Conservation Service (SCS). SSURGO provides more spatial detail than the STATSGO dataset and would be best for use in local scale wetland

inventories. This dataset consists of a tabular component containing detailed soil property information that can be joined with the spatial soil component. These datasets are frequently used to identify soils in which wetlands are likely to be found (Bourgeau-Chavez *et al.* 2008b).

2.2. DATA FUSION

Due to the utility of using both active and passive remote sensing data for the identification of forested wetlands methods have been developed to combine these datasets at some level within a classification in a process referred to as data fusion (Zhang 2010.) The fusion of data from multiple dates and multiple sensors provides more information to a classification than a single dataset. This fusion can be accomplished at the pixel, feature, or decision level (Zheng 2010.) Pixel level data fusion includes techniques such as intensity, hue, and saturation (IHS) transformation and principal component analysis. Extraction of features from multiple data sources to be combined into a single map is feature level data fusion. Data fusion at the decision level, or within classification fusion, is a more recent advance in data fusion techniques that have shown to improve classification accuracies, particularly when it is done within machine learning algorithms such as CART and random forest (Bourgeau-Chavez *et al.* 2008b; Lang and Kasischke 2008; Lucas *et al.* 2008; Bourgeau-Chavez *et al.* 2009; Hall *et al.* 2009; Bwangoy *et al.* 2010; Erdody and Moskal 2010Zheng 2010.)

Studies have shown that data fusion allows for the improved discrimination of features such as forested wetlands, inundation levels, tree species, and forest biomass (Maxa and Bolstad 2009; Bwangoy *et al.* 2010; Dalpont, Bruzzone, and Gianelle 2012;

Maxwell *et al.* 2014.) The fusion of multi-sensor, multi-date remotely sensed data can improve the mapping of forested wetlands due to an increased capacity for monitoring temporal variations and the opportunity to use active sensors that penetrate vegetated forest canopies (Murphy *et al.* 2007; Bourgeau-Chavez *et al.* 2008b; Lang and Kasischke 2008; Lucas *et al.* 2008; Bourgeau-Chavez *et al.* 2009; Hall *et al.* 2009; Bwangoy *et al.* 2010; Erdody and Moskal 2010; Corcoran, Knight, and Gallant 2013).This is particularly true when data with high spectral and/or spatial resolution is used (Hall *et al.* 2009; Maxa and Bolstad 2009; Bwangoy *et al.* 2010). Most researchers interested in data fusion for the detection of forested wetland hydrology have focused on the fusion of RADAR and satellite or aerial imagery (Bwangoy *et al.* 2010). However, a number of researchers have investigated the fusion of multi-spectral and hyperspectral imagery with the elevation and terrain components of LiDAR data for identification of forested wetland areas (Maxa and Bolstad 2009; Hartfield, Landau, and Leeuwen 2011.)

2.3. CLASSIFICATION METHODS

Many methods exist for transforming remote sensing data into information about the discrete wetland classes. Traditionally, wetland classification has been done using photointerpretation methods. But, this method relies on the ability of the photo-interpreter(s) to infer the location and extent of wetlands. This process is highly subjective, time-consuming, and may be inconsistent among photo-interpreters. A solution is to develop an automated decision method using GIS and remote sensing

data, and assess wetland probability based upon ancillary data. These methods can be divided into two main categories: unsupervised and supervised classification.

Unsupervised classifications are typically done by grouping pixels with similar spectral signatures based on some user defined statistical criteria (Jensen 1996; Klemas 2011). Töyrä and Pietroniro (2005) used an image polygon-growing algorithm for mapping flooded regions. Zhou *et al.* (2010) utilized an unsupervised classification routine to detect wetlands with IKONOS 4m resolution imagery. Iterative Self-Organizing Data Analysis (ISODATA), based on a database of known spectral signatures, was used to train the images and group pixels based on similarities of the spectral signatures. While it was found that the high-resolution imagery using ISODATA provided a good wetland classification, the resulting dataset was quite noisy requiring a post classification clumping and sieving routine. Clumping and sieving retained pixel groupings while discarding single, or small cell groupings, effectively reducing the noise in the classified image, and increasing accuracy (Zhou *et al.* 2010). Additionally, unsupervised classification methods rely on the accuracy and completeness of the ISODATA but it is possible to misclassify spectral signatures.

Supervised classification relies on the input of homogenous training areas that are representative of the desired output classes (Jensen 1996; Klemas 2011). The supervised classification then uses these training data to establish a relationship between wetland classifications and model input data (Jensen 1996). The accuracy of supervised classification is dependent on the quality of the training data. Common methods for classifying wetlands include maximum likelihood, thresholding, Logistic Regression, and Classification and Regression Tree Analysis (CART) (Hogg and Todd

2007; Bourgeau-Chavez *et al.* 2008a ; Shaeffer 2008; Poulin, Davranche, and Lafebvre 2010; Corcoran, Knight, and Gallant 2013).

Thresholding incorporates professional expertise into the classification. (Lang *et al.* 2013). A range of values, likely to be found in a particular class, is extracted from the input data and used to indicate regions where that particular class is expected to be found. The ability to manipulate threshold value allows for a flexible classification (Lang *et al.* 2008). For example, Gritzner (2006) used thresholding to extract values from band 5 of Landsat 7 to create a layer indicating open water. However, this method is not precise and would not be feasible with a large number of datasets.

Classification algorithms can automate the classification process. These algorithms can be parametric or non-parametric. Parametric supervised classification methods assume the data have a normal distribution. For example, the maximum likelihood method is a parametric classification algorithm which assumes normal distribution of data and that each class has an equal probability of occurring in an area (Jensen 1996). The algorithm calculates the likelihood, or probability of a pixel belonging to a particular class based on the training data (Jensen 1996; MacAlister 2009). These probabilities are then used to determine to which class each pixel belongs (Jensen 1996). However, maximum likelihood classification is computationally expensive and is not ideal if the distribution of input data is not normal (Lillesand, Kiefer, and Chipman 2008).

Logistic regression is often used for discriminating uplands and wetlands based on the values of both continuous and discrete input variables (Pantaleoni *et al.* 2009). Shaeffer (2008) found that while logistic regression was able to evaluate the

significance of model predictors, the accuracy of logistic regression analysis for predictive modeling decreases as the study area becomes larger and more diverse due to its constant regression coefficients. Additionally, logistic regression as a type of probabilistic statistical classification assumes that there is a linear relationship between input and explanatory variables (Lawrence *et al. 2001;* Hogg and Todd 2007; Poulin *et al.* 2010*)*.

The Classification and Regression Tree (CART) method is non-parametric. The CART model has been shown to distinguish between explanatory and confounding variables without the assumptions of a linear relationship needed for logistic regression (Lawrence *et al. 2001; Hogg* and Todd 2007; Poulin, Davranche, and Lafebvre 2010*)*. Hogg and Todd (2007) demonstrated that CART analysis outperformed logistic regression analysis, correlation matrices, and visual derivative thresholding. The use of CART method allows for the explanation of nonlinear relationships within the explanatory variables and the variable the model intends to predict, and has a clear benefit over the other two methods (Hogg and Todd 2007). Lawrence *et al.* (2001) claimed that CART decision methods were more readily available, accurate, and more user friendly than the more complex neural network and expert system methods. A benefit of CART, in addition to those already mentioned, is the production of class probability at each decision node, allowing for the significance of each predictor to be assessed at each node and in the overall classification (Lawrence *et al.* 2001; Hogg *et al.* 2007; Wright and Gallant 2007; Corcoran, Knight, and Gallant 2013). Similarly, Hogg *et al.* (2007) found that these node probabilities along with cross validation allowed for

the removal of some nodes as well as a terrain variable, reducing model complexity while retaining accuracy.

Random Forest is a machine learning classification and regression tree algorithm advanced by Breiman (2001) and implemented in the R environment by Liaw and Wiener (2002). Random Forest performed in R is essentially a collection of smaller CART decision trees which utilizes bootstrapping techniques to circumvent the need for large sample sizes and assumptions about the relationship between variables (Brieman 2006). Recursive partitioning techniques are used to draw random vector of remote sensing data from the training areas that have been classified as wetland or upland. These random vectors are iteratively drawn with replacement and the samples are used to build a collection of classification and regression trees referred to as the random forest classifier (Breiman 2001; Shih 2011). As the number of iterations increase, and many forests are built, the correlation between forests is decreased without decreasing the strength of the predictor due to the Law of Large Numbers (Breiman 2001). Recursive partitioning is ideal for classification using small sample sizes because the iterative random drawing of samples boosts the number of samples allowing for the assumption of a normal distribution.

Of the randomly drawn samples (pixels), about one third will be reserved within the random forest classifier in order to estimate the error of the random forest classification. This estimate is commonly referred to as the out-of-bag error, or OOB, and can be used to improve the classifier before implementation (Breiman 1996; Breiman 2001). Unfortunately, the OOB is not a reliable assessment of the accuracy due to spatial autocorrelation between sample pixels used to build the forests and those

used to assess its accuracy. However this data is used to compute importance statistics and ranking of each variable. Assessment of this statistic along with OOB will allow for the removal of unimportant variables from the dataset used to train and build the forest resulting in increased predictor accuracy.

CHAPTER 3

STUDY AREA AND DATA

3.1 STUDY AREA

The area in which this study was undertaken is located in Vermilion County, Illinois. Vermilion County primarily consists of agricultural land, though approximately 6.5% of the county is forested, (Figure 3.1 and Table 3.1). A total of 9,784 acres of these forested lands are publicly accessible (USGS GAP 2013). Many of these accessible forested areas are located alongside or near the Vermilion and Little Vermilion rivers within the Vermilion and Middle Wabash-Little Wabash watersheds (Figure 3.2). Data from the IDNR (2015) (Table 3.1) describing land cover acreage and rankings as well data from the NWI (Figure 3.3) confirmed a substantial number (3,890 acres) of forested wetlands in Vermilion County compared to other locations in East Central Illinois. The abundance of publicly accessible forested wetlands, along with the availability of spring LiDAR data, was the primary reasons for Vermilion County being chosen as the study area. However, due to the limited extent of the available LiDAR data the area of interest was limited to 532 square miles in the eastern portion of the county (Figure 3.4).

Table 3.1 Land Cover Acreage and Rankings for Vermilion County Illinois adapted from Illinois Department of Natural Resources (2015)

Vermilion County, Illinois Land Cover Acreage and Rankings		
Land Cover	**Acres**	**Rank**
Cropland	419,820	6
Row Crops	406,643	6
Small Grains	13,177	63
Orchards/Nurseries	0	--
Grassland	86,636	30
Urban	7,439	18
Rural	79,197	28
Forest/Woodland	37,370	42
Deciduous	31,188	44
Open Woods	6,164	24
Coniferous	17	71
Wetland	5,755	69
Shallow Marsh/Wet Meadow	483	70
Deep Marsh	116	47
Bottomland Forest	3,890	67
Swamp	0	--
Shallow Water	1,266	36
Urban/Built-up Land	16,523	18
High Density	2,714	20
Medium Density	6,664	15
Low Density	1,752	29
Transportation	5,393	11
Open Water	9,680	26
Lakes and Rivers	5,176	33
Streams	4,505	12
Barren/Exposed Land	509	9
Total Land	576,293	7

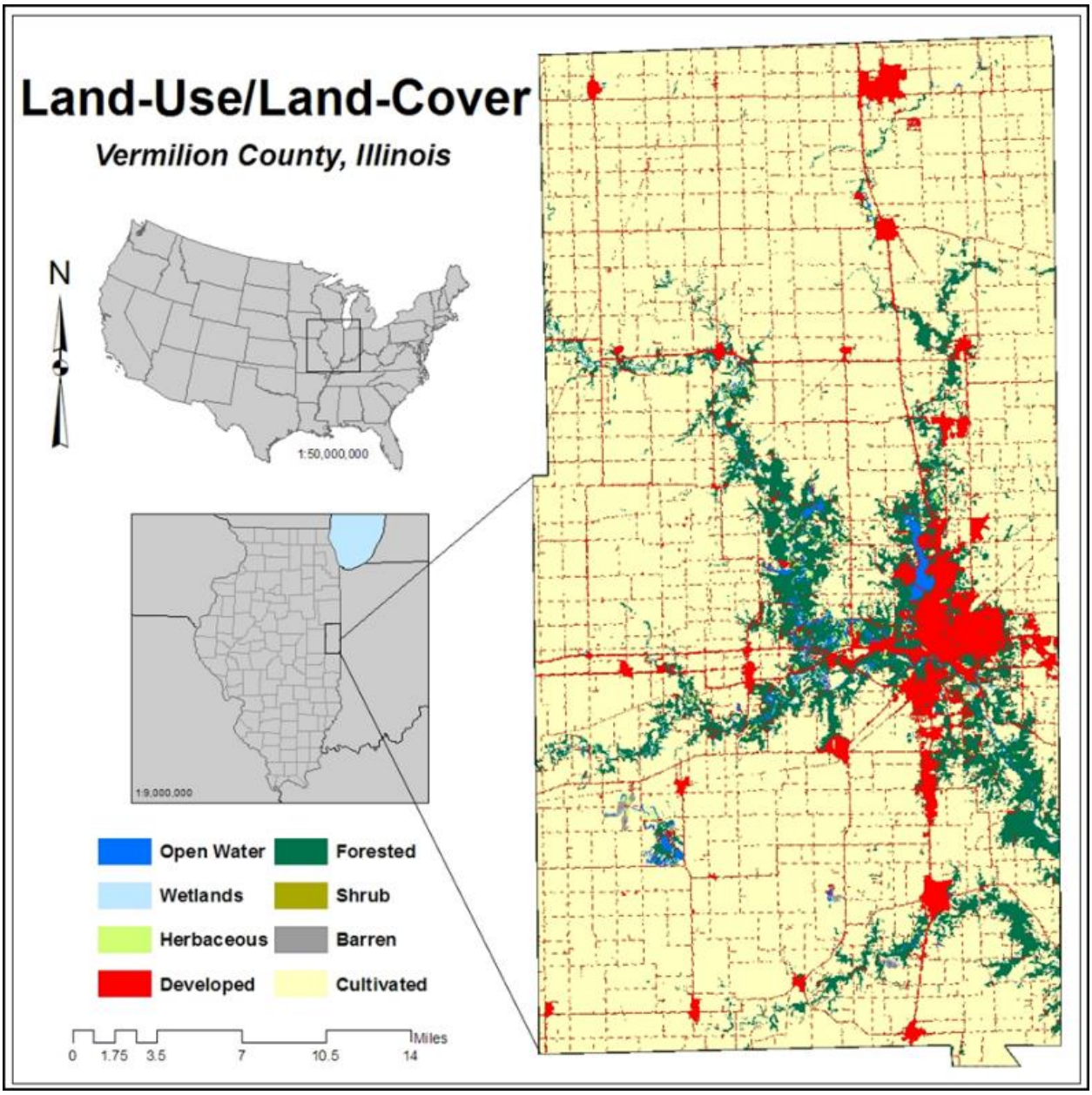


Figure 3.1 2006 National Land-Cover Dataset shows that Vermilion County, Illinois is primarily an agricultural landscape with some forested and developed areas.

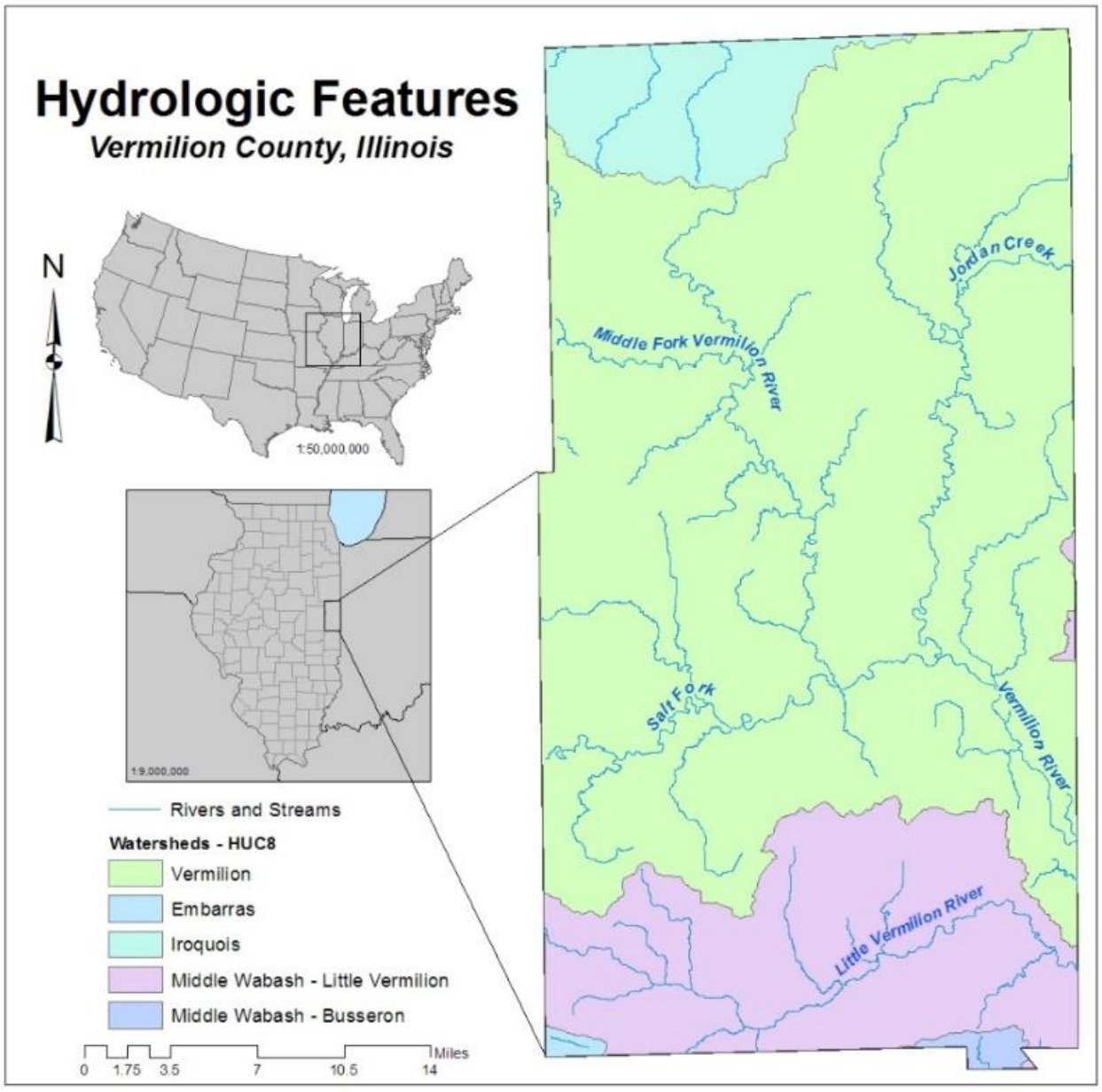


Figure 3.2 Rivers and Watersheds within Vermilion County, Illinois

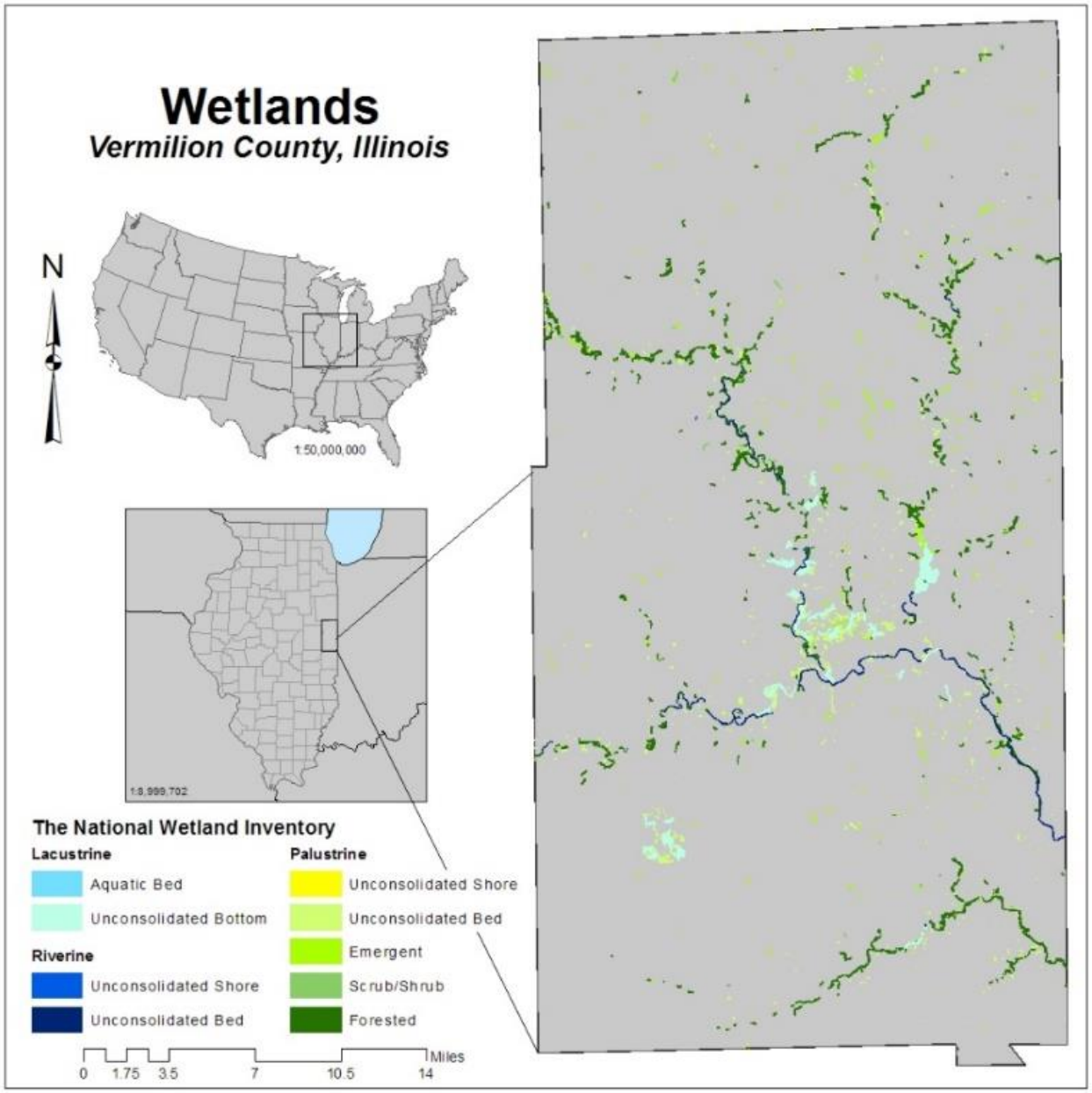


Figure 3.3 Wetlands in Vermilion County Illinois as identified in the National Wetland Inventory.

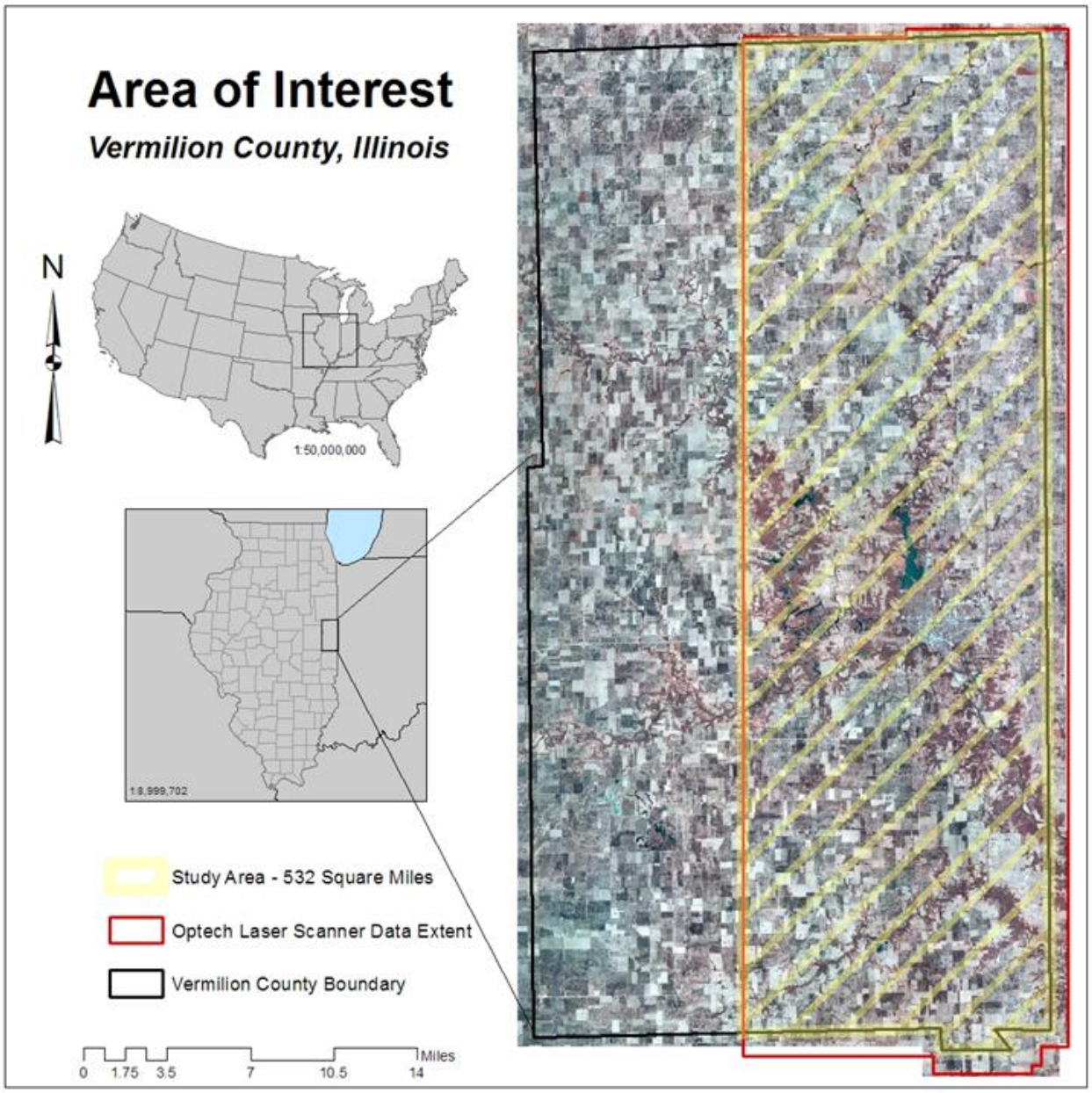


Figure 3.4 The area of interest for this study located in the eastern portion of Vermilion County. The availability of LiDAR data was the limiting factor for choosing the area of interest.

In eastern Vermilion County there are two ecoregions: Illinois/Indiana Prairie and Glaciated Wabash Lowlands. Illinois/Indiana Prairies is characterized by the flat plain of

the ancient glacial lakebed marked in some places by prairie potholes. Along the Little Vermilion River, the Glaciated Wabash Lowlands are characterized by the rugged, rolling hills of the glacial till plane as well as terracing along the river (Omernik *et al.* 2006.) Agriculture, urban settlement, and surface mining have disturbed this landscape from its' original ecological composition. Danville, currently home to 33,027 residents, was once known for its large-scale surface coal mining. Figure 3.5 shows the location of former surface and underground coalmines within the county. In 1939, a majority of the mined lands were purchased by the State of Illinois. A portion of these lands were converted into Kickapoo State Park, one of the publicly accessible forested areas sampled for this study. Since the closure of these surface mines vegetation has reclaimed the land but rugged spoil ridges and mine ponds still exist below the forest canopy. This surface mining, along with glacial and riverine processes have created a landscape that averages 681 feet above sea level but ranges from 489.8 to 820.4 feet above sea level. Much of this 330-foot elevation change occurs in the forested areas along the current and historic river channel.

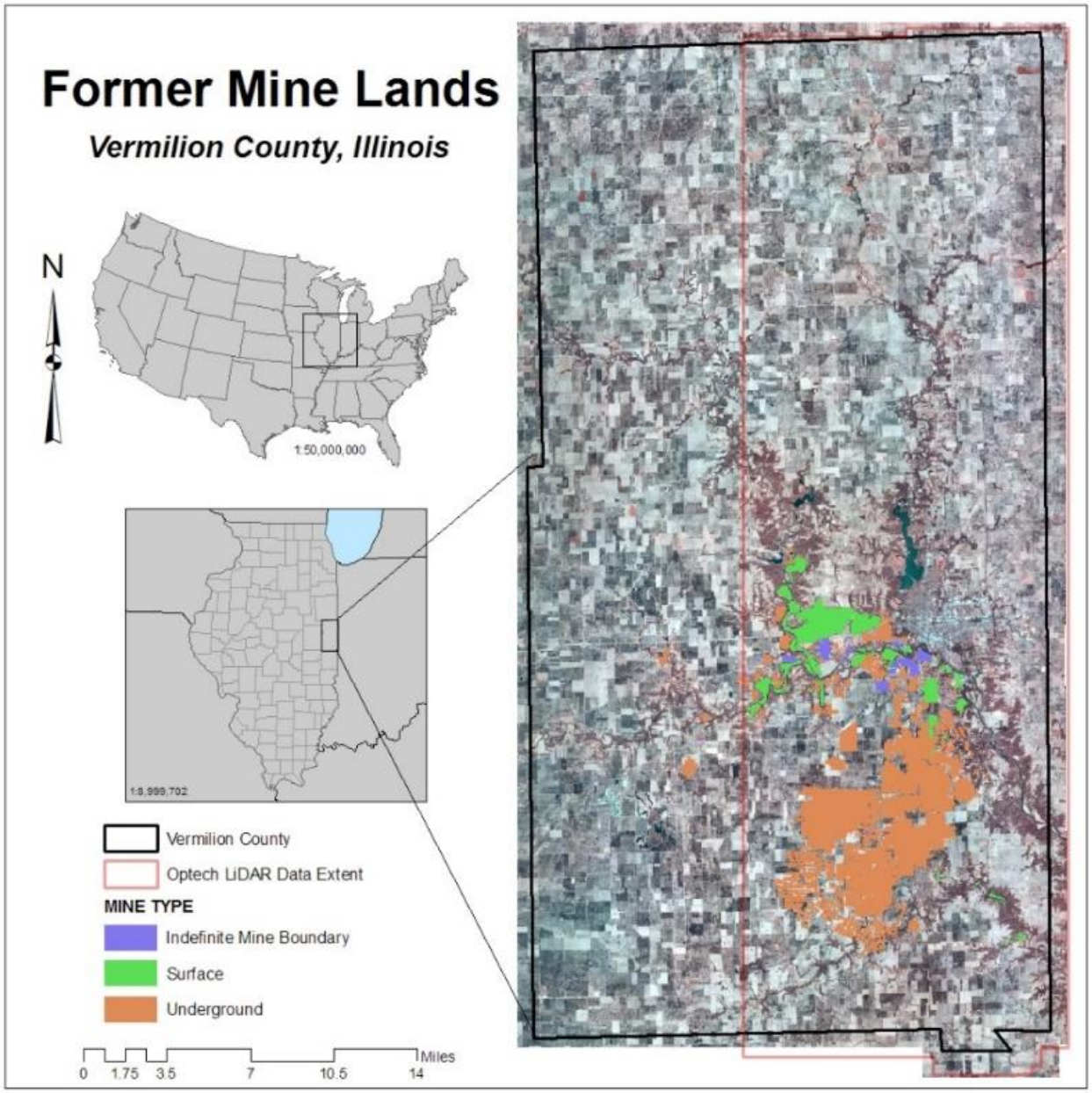


Figure 3.5 Former surface and underground coal mining locations

3.2 DATA

3.2.1 MULTI -SPECTRAL AERIAL IMAGERY

Multi-spectral aerial imagery was selected as one of the input variables for this research due to this data's well-documented ability to detect biophysical characteristics

that allow for the separation of wetlands and uplands (Table 3.2). Spring color-infrared multi-spectral imagery was included in utilized in order to capture information about the study area during leaf-off conditions when the passive sensor may be able to detect inundation below the forest canopy. The spring, four-band, 1-foot resolution aerial orthoimagery was obtained from the Illinois Department of Transportation (IDOT) through the Illinois Geospatial Data Clearinghouse (IGDC). These data were collected on March 16, 2011 when most of the vegetation in the study area had not produced foliage. The imagery was projected in North American Datum (NAD) 1983, Illinois State Plane East, FIPS 1201, and included red (598 to 675 nm), green (500 to 650 nm), blue (400 to 850 nm), and near infrared (675 to 875 nm) channels with 8-bit radiometric resolution (INRGD and ISGS 2013.) These data were obtained through the IGDC as 10,000 m^2 tiles in the JP2 format.

Summer imagery was obtained from the United States Department of Agriculture (USDA) through the Natural Resources Conservation Service (NRCS) Data Gateway. This summer imagery was collected on June 06, 2012 using four channels: red (500 to 650 nm), green (500 to 650 nm), blue (400 to 580 nm), and near infrared (675 to 850 nm). The imagery had an 8-bit radiometric resolution and 1m horizontal resolution. The spatial reference for this data was Universal Transverse Mercator (UTM) zone 16.

Table 3.2 Summary of remote sensing datasets

Dataset	Date Collected	Resolution	Channels	Source
Spring Color-Infrared	March 16, 2011	1 foot, 8-bit	red (598 to 675 nm)	Illinois Department of Transportation – Illinois Geospatial Data Gateway
			green (500 to 650 nm)	
			blue (400 to 850 nm)	
			near-infrared (675 to 875 nm)	
Summer National Agriculture Imagery Program (NAIP)	June 12, 2012	1 meter, 8-bit	red: 590-675nm	United States Department of Agriculture (USDA) – Natural Resource Conservation Service (NRCS) Data Gateway
			green: 500-650nm	
			blue: 400-580 nm	
			near infrared: 675-850 nm	
LiDAR (.LAS files)	March 10, 2012	1 foot post spacing	1064 nm Laser Pulse	Illinois Height Modernization Program – Illinois Geospatial Data Gateway

3.2.2 LIDAR

Although multispectral imagery is the most commonly used remote sensing data for wetland detection its passive nature restricts its ability to detect features below the forest canopy. For this reason LiDAR data (Table 3.2), an active sensor that using laser pulses in the infrared spectrum, was included in this research. The Illinois Height Modernization program has been updating the elevation datasets to include high-resolution (6 inches to 2 feet) LiDAR derived elevation models. This data is available at

the Illinois Geospatial Data Clearinghouse (ISGS, IDOT, and ISGS 2013). As part of this update, LiDAR data were collected using Leica ALS70 and Optech Gemini laser scanners. The Leica sensor was used to collect data in the western portion of the county and the Optech Gemini for the eastern portion. These data were collected for Vermilion County by AeroMetric, Inc. on February 17, 2012 and March 10, 2012, respectively. Due to the early collection date of the Leica ALS the ground and much of the surface water in the county was frozen resulting in incongruous return intensity values between the two LiDAR data collections. For this reason, only the eastern portion of the county (Figure 3.4) recorded by the Optech Gemini sensor data was used for this research.

The Optech sensor was flown at a maximum altitude of 1600 meters above the mean terrain, utilizing a 1064 nm laser pulse with 167 kHz effective repetition rate to collect LiDAR data with a 40-degree swath width. These data are available as LAS tiles describing the x, y, and z coordinate of each LiDAR return, signal intensity, return number, and point class. The LAS files have 1-meter post spacing, vertical accuracy of 0.6 feet, and horizontal accuracy of 0.3 meters. The North American Datum of 1983/HARN was used and the data were projected into 1201 Illinois State Plane East. The LiDAR data returns were classified based on the standard set by The American Society for Photogrammetry and Remote Sensing. Points were classified into 5 categories: Processed, but unclassified/ non-ground (code 1); Ground (code 2); Noise/ Low Points (code 7); Water (code 9); Ignored Ground (code 10) by AeroMetric, Inc. LiDAR intensity was recorded in 8-bit radiometric resolution.

3.2.3 FIELD DATA

Field sampling locations were selected as locations based on their inclusion in the Federal, State, Tribal, and Protected Areas Database (USGS GAP 2013). Publicly accessible forested areas were identified using the Federal, State, Tribal, and Protected Areas Database (USGS GAP 2013). Publicly accessible lands identified in this database but containing no forested lands, or primarily consisting of the Vermilion or Salt River channels, were eliminated as potential sampling locations. In order to conduct research on land owned by the State of Illinois Department of Natural Resources an application was filed (APPENDIX X) and approved by the site superintendent, Mr. John Hott, allowing access to properties held by the state (Harry *babe* Woodyard State Natural Area, Kickapoo State Recreation Area, middle Fork State Fish and Wildlife Areas, Middle Fork Woods Nature Preserve, and the Dynergy Tract). The Vermilion County Conservation District (VCCD) was also contacted for approval of research on their properties (Doris Westfall Prairie Restoration Nature Preserve, Fairchild Cemetery Prairie/Savanna Nature Preserve, Forest Glen Preserve, Horseshoe Bottom Nature Preserve, Howard's Hollow Seep Nature Preserve, Kennekuk Cove County Park, Russell M. Duffin Nature Preserve, and Windfall Prairie Nature Preserve). A research contract was filed and approved by Mr. Gary Wilford, allowing access to these locations.

Fieldwork to obtain training and validation data for the forested wetland classifications was conducted from May 12 through May 16, 2014. Ground reference data locations were selected using stratified random sampling to minimize bias in classification and accuracy assessment. However, due to access restrictions on private forested lands, ground truth data will only be collected on publicly accessible forested

land, and so the distribution of these ground sampling locations are influenced by accessibility. Transects were walked taking points at regular intervals (approximately 100 m) and noting if the location is upland or wetland. Vegetation and soils will not be relied upon for making wetland determinations. An area will be considered a wetland based on the presence of one or more of the following characteristics: surface water, saturated soils, waterlines or other signs of recent inundation. Figure 3.6 depicts a subset of the forested wetlands identified in the field. For locations that are considered a wetland, additional information was recorded including location, the percentage canopy cover, the presence of a shrubby understory, the presence of aquatic vegetation, and whether the wetland is isolated or connected to a flowing water body. Points were only collected for forested wetland and forested upland locations that are greater than 5 feet across.

Initially, GPS points were collected using a Magellan ProMark3 GPS unit and external GNSS antennae. The ProMark3 unit is capable of horizontal accuracy of 0.039 feet and vertical accuracy of 0.049 feet using real-time differential correction and post-processing. Unfortunately, this GPS unit malfunctioned on the second day of fieldwork and only 15 locations were collected: 7 forested wetland locations and 8 forested upland locations. Then, a Trimble eXplorist GPS unit, capable of 3m horizontal accuracy, was used to collect samples of 69 locations: 39 wetland locations and 30 upland locations. Because of the lower horizontal accuracy of this unit, three points were taken at each location and then averaged in post processing. The variance between the three latitude and longitude points at each location was then calculated and converted to feet. Forested wetland locations with an average (latitude/longitude) variance greater than 10

feet were eliminated from the sample, reducing the number of wetland locations collected with the eXplorist unit to 20. Fieldwork was undertaken once again from June 20 through 22, 2014 using a Trimble GeoExplorer 6000 GPS unit, capable of sub-meter accuracy. This unit was used to collect 23 forested wetland locations and 12 forested upland locations. These points were then post-processed using base station data in the TerraSync software. In total, the number of field points was 100, with 50-forested wetland and 50-forested upland locations collected and retained for analysis.

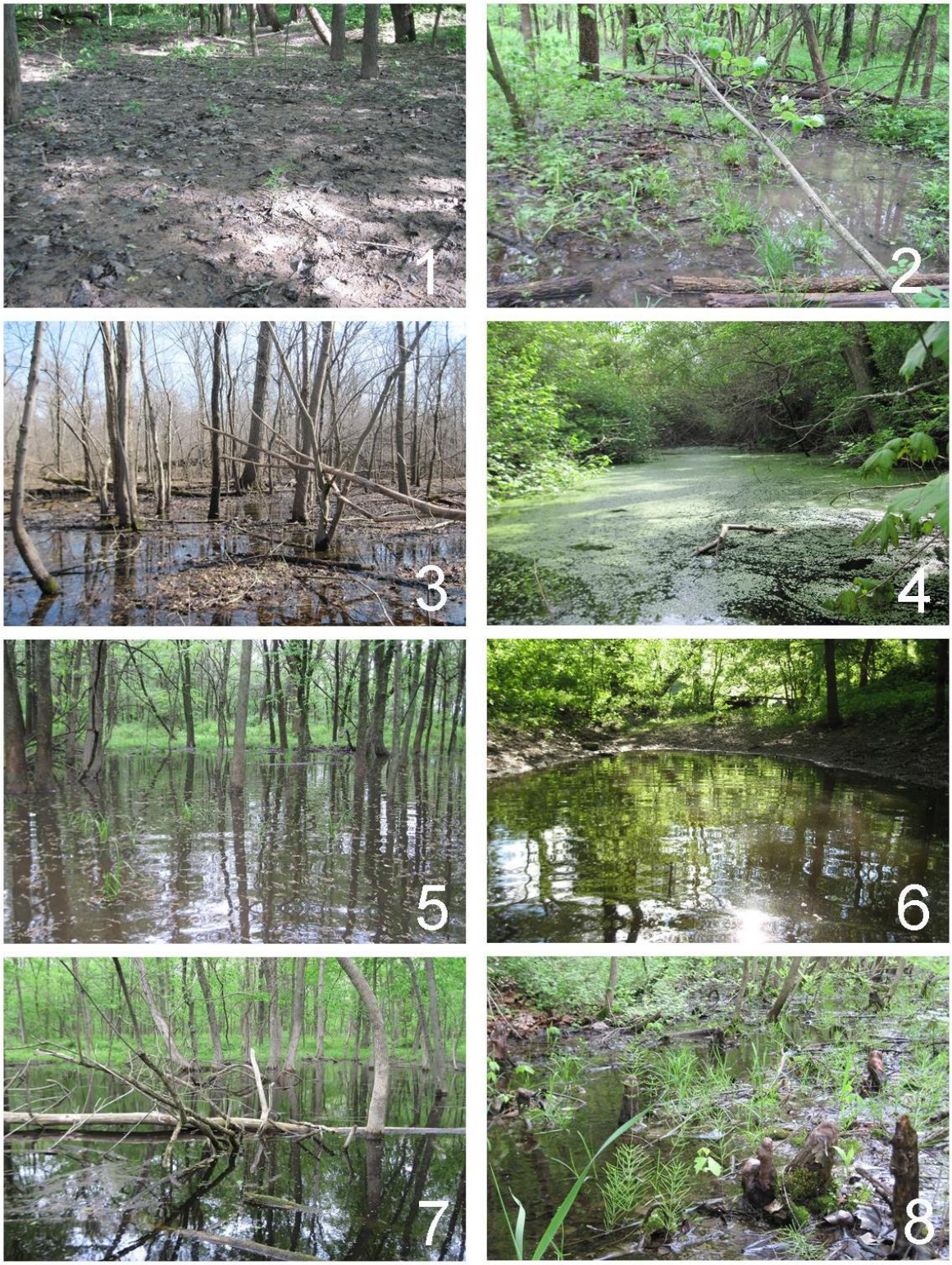


Figure 3.6. Images of forested wetland locations taken in the field

CHAPTER 4

METHODS

4.1 DATA PREPROCESSING

Prior to any data analysis it was necessary to preprocess the remote sensing datasets so that they shared a common extent, spatial resolution, coordinate system, and projection. Spring CIR imagery tiles were mosaicked within ERDAS IMAGINE Mosaic Pro Tool. These tiles were mosaicked into a single image and rescaled to 3-foot using bilinear interpolation resampling, a resampling method that is most appropriate for continuous datasets such as aerial imagery. Summer Imagery was also mosaicked and rescaled to 3-foot using bilinear interpolation resampling and overlapping pixels were averaged. The NAIP imagery was projected into NAD 83 State Plane Illinois East, FIPS1201, the projection shared by both the spring imagery and LiDAR data. After this projection, features in the summer imagery were still slightly offset from those in the spring imagery and lidar data and so the summer imagery was registered to the spring imagery using ground control points and a polynomial geometric transformation.

LiDAR data was converted from the LAS tiles to 3-foot resolution GeoTIFF ground return elevation and intensity tiles in the open source LASTools ArcGIS toolbox developed by Martin Isenburg (2006). These tiles were then mosaicked in ERDAS IMAGINE 2013 using bilinear interpolation resampling. Finally, the intensity image was iteratively Lee filter to reduce noise with the window sizes 3, 3, 5, 7, & 9 (Lee 1987; Lang and McCarty 2009; Huang *et al.*, 2014). This filtering method preserves mean values while reducing local variation; resulting in a smoother intensity raster dataset (Figure 4.1)

4.2 CALCULATION OF IMAGE INDICES AND TERRAIN DERIVATIVES

Three image indices were calculated for both the spring and summer imagery: 1) simple ratio, 2) Normalized Difference Wetness Index (NDWI), and 3) Normalized Difference Vegetation Index (NDVI). These image indices were chosen for their ability to improve wetland detection based on a review of the relevant literature. The calculation of these image indices was performed in ERDAS Model Maker using the following formulas:

$$Simple\ Ratio\ = \frac{near\ infrared}{red} \tag{4.1}$$

$$NDWI\ = \frac{(green - near\ infrared)}{(green + near\ infrared)} \tag{4.2}$$

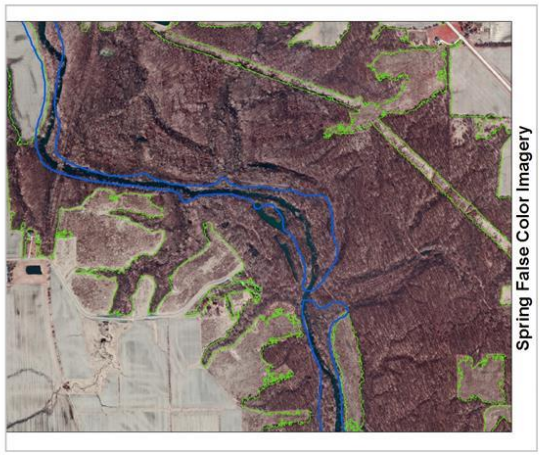
$$NDVI\ = \frac{(near\ infrared - red)}{(near\ infrared + red)} \tag{4.3}$$

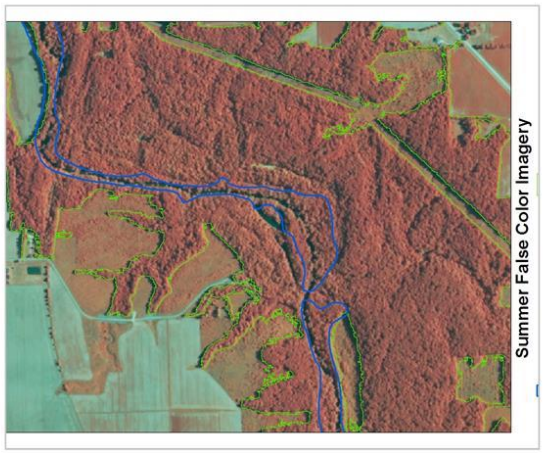
$$TWI\ = ln\left(\frac{\alpha}{tan\beta}\right) \tag{4.4}$$

Secondary terrain derivatives including slope (degrees), curvature, and aspect were calculated using the LiDAR derived elevation dataset as the input in ArcGIS 10.1 Spatial Analyst tools. The topographic wetness index (TWI) was calculated using Raster Math in ArcGIS ModelBuilder: where α is the upslope contributing area and β is the local slope in radians (Kienzle 2004; Hogg and Todd 2007; Kheir *et al.* 2009; Lang *et al.*

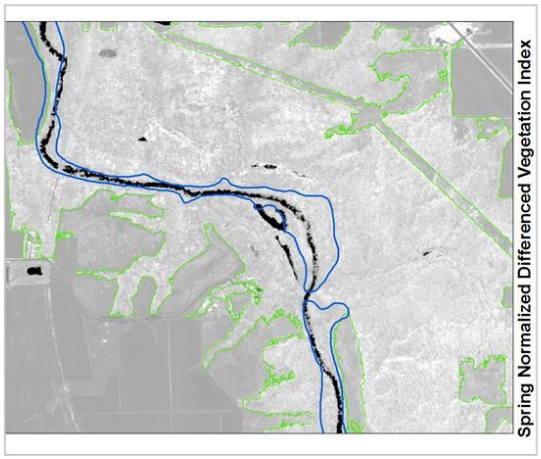
2009; Lang *et al.* 2013.) An additional raster representing the difference in elevation between the first and last lidar return was created for use in the initial classification to separate the forested and non-forested areas. All 24 datasets with the exception of the difference LiDAR elevation return dataset were stacked into a single image (.img) file.

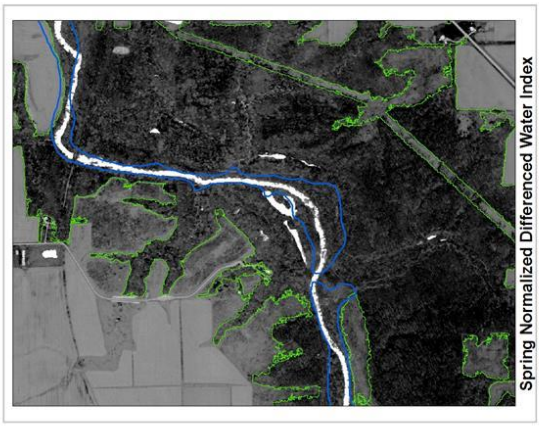
Finally, forested wetland points were buffered with the minimum radius collected in the field (distance in feet to the closest dry ground from where the point was collected). Forested upland points were buffered with a radius of 15-feet – the minimum distance used in the field to assign upland points. Field points were then split 80/20 into training and validation datasets. The training dataset along with each of the remote sensing datasets was set aside to train the random forest classifier. This split was done in Excel by assigning each polygon record a random number and then reserving the 10 lowest random values in each class as reference data.

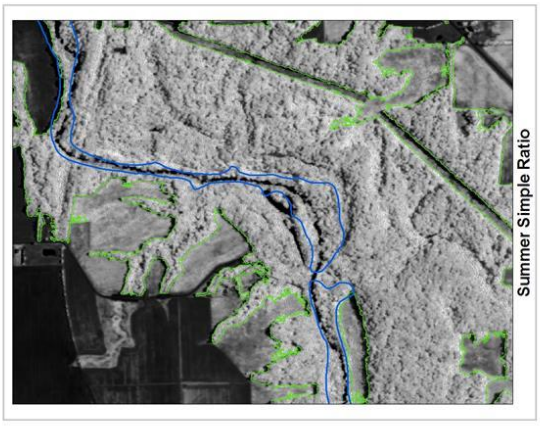

Spring False Color Imagery

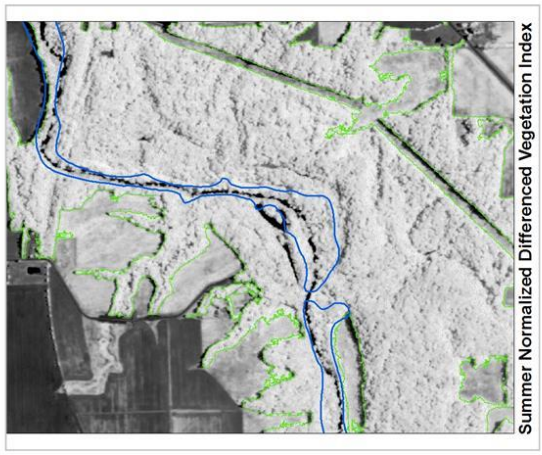

Summer False Color Imagery

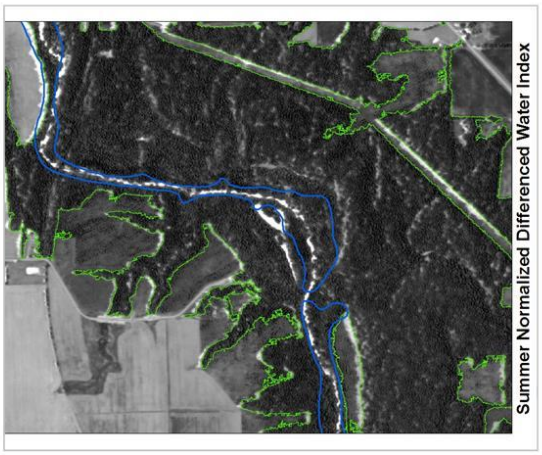

Spring Simple Ratio

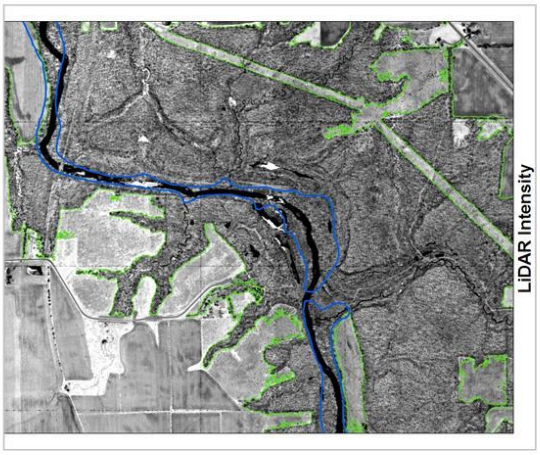

Spring Normalized Differenced Vegetation Index

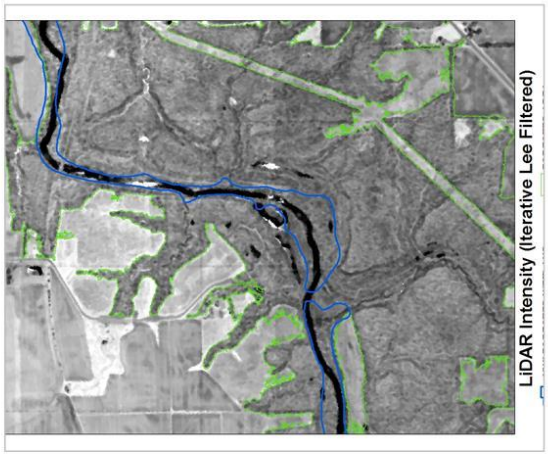

Spring Normalized Differenced Water Index

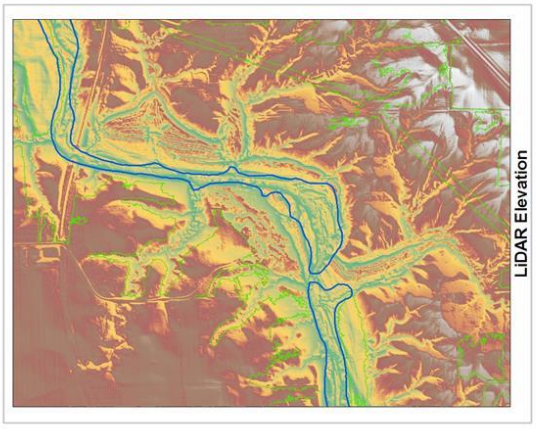

Summer Simple Ratio

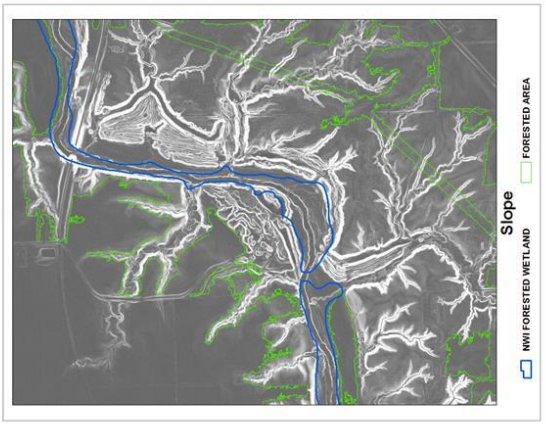

Summer Normalized Differenced Vegetation Index


Summer Normalized Differenced Water Index


LiDAR Intensity


LiDAR Intensity (Iterative Lee Filtered)


LiDAR Elevation


Slope
NWI FORESTED WETLAND
FORESTED AREA

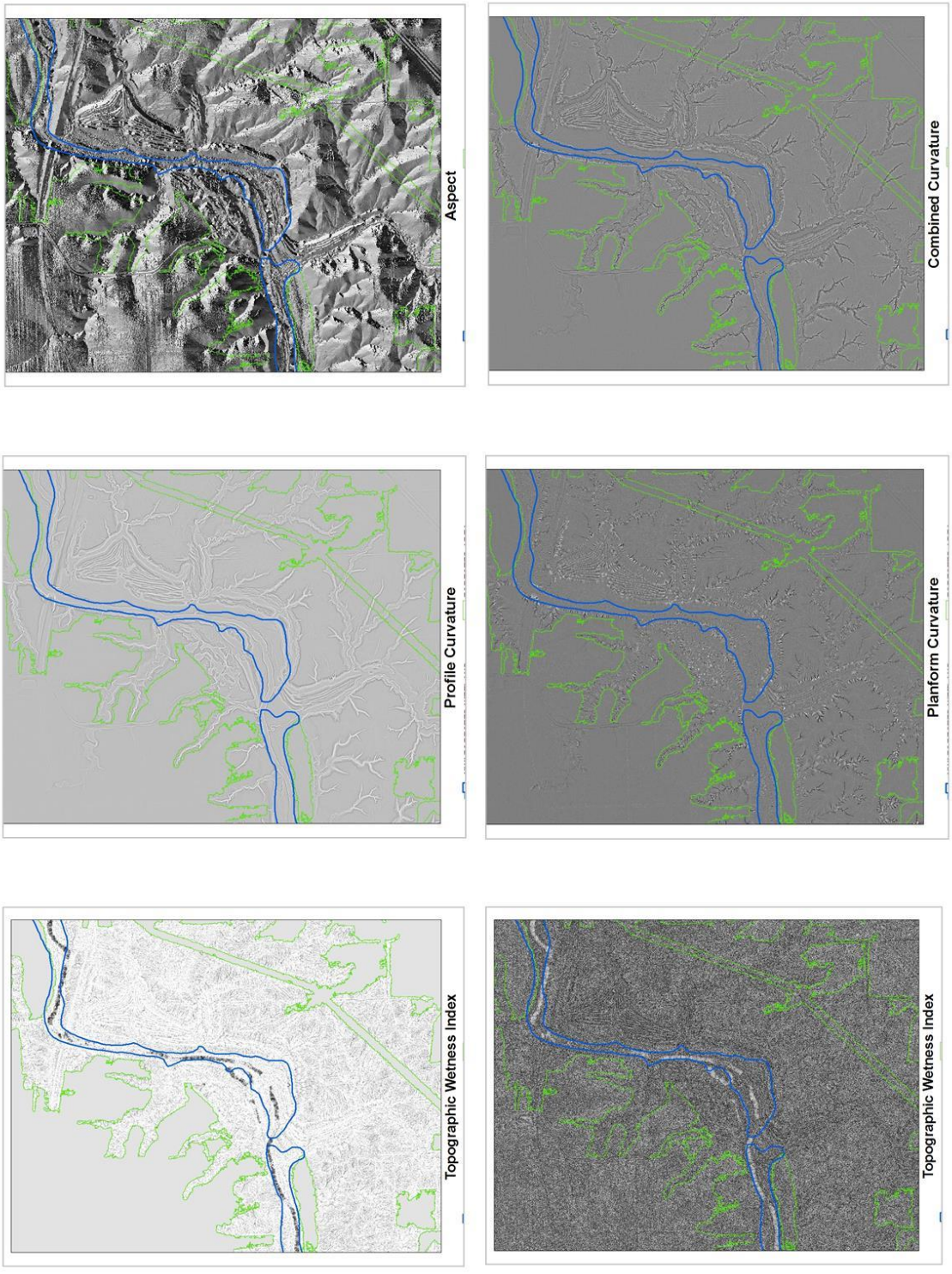


Figure 4.2. Example of the multispectral and LiDAR datasets used for this study

4.3 CLASSIFICATION IN RANDOM FOREST

Four classifications were performed: Classification A – All Imagery, Classification B – All LiDAR, Classification C – LiDAR without Intensity, and Classification D – Fusion of All Data. Table 4.1 describes the remote sensing datasets used as input variables for each classification. Classifications were performed in the R programming environment using the Random Forest package developed by Liaw and Wiener (2002.) The classification process requires 4 inputs: a shapefile containing polygon training areas with a field ‘type_id’ identifying the numeric class value, a raster stack containing the input remote sensing variables, the raster output location and name, and the number of samples to be select from within each training class. Due to memory limitations and the large size of the input raster stack (279 GB) the classification algorithm failed when a large sample size was used. For this reason, sample size was set at 250. Training polygons containing 30 forested wetland polygons (‘type_id’ = 1) and 30 forested upland polygons (‘type_id’ = 2) was used as the input training area shapefile. The stack containing all 24 multi-spectral and LiDAR datasets was set as the input raster and for each individual classification only the layers pertaining to that classification were read into memory and used as input variables. The output of each classification was the saved classification forest, a thematic GeoTIFF raster image, OOB results, and importance statistics.

Table 4.1 Description of data used as input variables for each classification

Data	Classification			
	A	B	C	D
Spring Red	x			x
Spring Green	x			x
Spring Blue	x			x
Spring Infrared	x			x
Summer Red	x			x
Summer Green	x			x
Summer Blue	x			x
Summer Infrared	x			x
Spring Simple Ratio	x			x
Spring NDVI	x			x
Spring NDWI	x			x
Summer Simple Ratio	x			x
Summer NDVI	x			x
Summer NDWI	x			x
Lee Filtered Intensity		x		x
Raw Intensity		x		x
Elevation		x	x	x
Slope		x	x	x
Aspect		x	x	x
Combined Curvature		x	x	x
Profile Curvature		x	x	x
Planform Curvature		x	x	x
TWI		x	x	x
Normalized TWI		x	x	x

4.4 ACCURACY ASSESSMENT

Classifications were assessed for accuracy using the error matrices method, the Kappa Coefficient of Agreement, and Conditional Kappa Coefficient of Agreement. The Kappa statistic is a measure of agreement between the classification created by producer and the one obtained by chance and can be thought of as the difference

between the difference between the proportion of correctly assigned pixels to the proportion of pixels that would be correctly assigned by chance over the proportion of pixels that would be correctly assigned by chance (Jensen 1996; Congalton and Green 1999; Froody 2004.)

$$K\hat{} = \frac{P_0 - P_c}{1 - P_c} \qquad (4.5)$$

The Conditional Kappa Coefficient of Agreement statistic is a measurement of the agreement between the remotely sensed data and reference data for a particular class with chance agreement removed (Jensen 1996). After testing each classification Kappa for significance based on a 95% confidence level, the Kappa values were used to compare the accuracy of the classifications. The Kappa Coefficient of Agreement and Conditional Kappa Coefficient of Agreement may range in value from -1 to 1, with 1 indicating perfect agreement and values less than or equal to zero indicating agreement that is less than that expected by chance. Table 4.2 describes common interpretation of Kappa values (Jensen 1996; Viera and Garrett 2005).

Importance statistics for classes 1 and 2 (forested wetland and forested upland, respectively), Mean Decrease in Accuracy, and the Mean Decrease in Gini Coefficients derived from the random forest classification were used to assess the contribution of each layer to that classification. The Mean Decrease in Accuracy is the OOB error rate for each layer. The Mean Decrease in Gini Coefficient is a measurement of node impurities at each split in the Random Forest classification trees. Correlation between

bands was measured using Pearson's Product-moment correlation coefficient (R) and used along with the importance statistics to determine the datasets which contributed useful unique information to each classification. A visual assessment of each classified layer was performed to determine, along with the importance statistics, which input data layer(s) contributed error to the classification.

Table 4.2 Interpretation of Kappa Coefficient of Agreement Values (Jensen 1996; Viera and Garrett 2005)

Kappa	Level of Agreement
< 0	Less than Chance
0 to 0.20	Low
0.21 to 0.40	Low to Moderate
0.41 to 0.60	Moderate
0.61 to 0.80	Moderate to Strong
0.81 to 0.99	Strong
1	Perfect

CHAPTER 5

RESULTS

5.1. SEPARATION OF FORESTED FROM NON-FORESTED AREA

An initial classification was performed in random forest to separate the forested from non-forested areas. This classification used the both spring and summer NDVI and the LiDAR return difference raster as input variables. Six general land-cover classes were identified in the county: bare agriculture, forest, grassland, planted agriculture, urban, and water. Then, using visual interpretation of the summer NAIP imagery 10 training polygons were created for each of the land-cover types. These polygons, along with the NDVI/LiDAR stack, were used to create a random forest classifier within R.

Pixels within the output classification were clumped in the four cardinal directions and clumps less than 2 acres were eliminated. Then the resulting thematic raster was recoded to forest and non-forest. This binary raster was once again clumped, and clumps less than 5 acres were eliminated. The final thematic layer resulting from this clumping and elimination routine was evaluated within ERDAS Accuracy Assessment tool. Equalized random sampling was used to assign 30 random points to each class. Then, without viewing the classification value and using the summer false color composite imagery as the reference, these points were assigned a reference value based on photo-interpretation. The accuracy of this forested vs. non-forested classification was very good with a 96.67% Kappa Coefficient of Agreement (Table 5.1). A visual assessment of this data shows that it is in strong agreement with the 2006 National Land Cover Dataset (NLCD) (Figure 3.1 and Figure 5.1). Then using the mask tool in ERDAS IMAGINE, all non-forested areas within the input remote sensing

datasets were recoded to 0. This masked image stack was used as input for subsequent classifications (Figure 5.2).

Table 5.1 Accuracy Assessment of Forested vs. Non-Forested Classification

	Forested	Non-Forested	Row Total	User's Accuracy	Conditional Kappa Coefficient (K_c)
Forested	30	0	30	100%	100%
Non-Forested	1	29	30	96.67%	93.44%
Column Total	31	29	59		
Producer's Accuracy	96.77%	100%	Total	60	
				Overall Accuracy	98.33%
				Kappa Coefficient of Agreement	96.67%

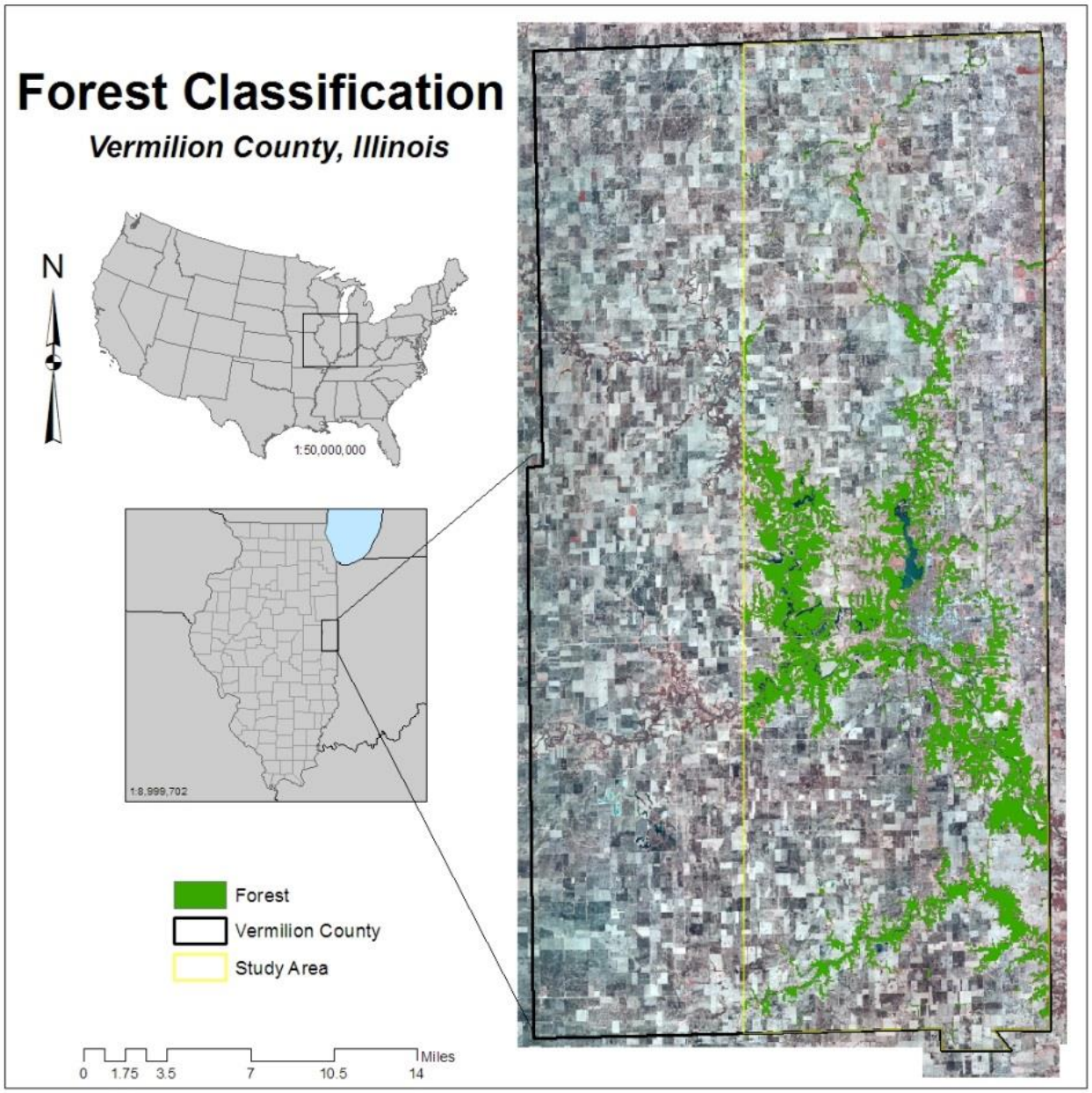


Figure 5.1 Final forested vs. non-forested classification

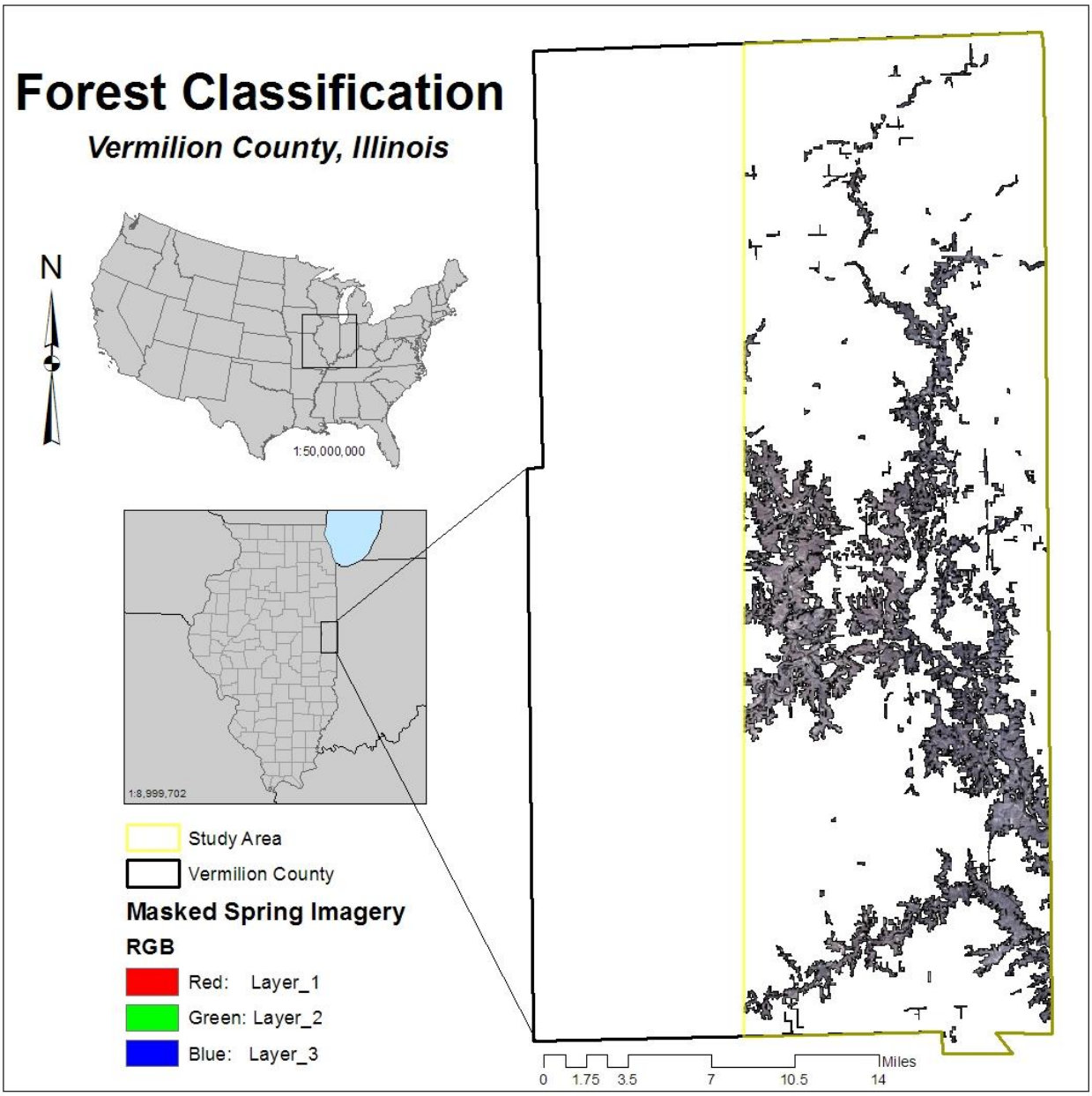


Figure 5.2 Imagery stack masked within the forested study area

5.2. FORESTED WETLAND CLASSIFICATIONS

5.2.1 CLASSIFICATION A: ALL AERIAL IMAGERY AND IMAGE INDICES

The thematic map resulting from the classification of all Aerial Imagery (Classification A) had a forested wetland producer's accuracy of 82.5% and a Kappa coefficient of 56.10%, indicating moderate overall agreement. Conditional Coefficient of Agreement for forested wetland (FW) (K_{CA1}) and forested upland (FU) (K_{CA2}) classes in Classification A were 35.27% and 64.37%, respectively, indicating a low to moderate agreement for the forested wetland class and moderate agreement for the forested upland class. The error matrix for Classification A can be found in Table 5.2. Visual assessment of the classification indicates that much of the confusion is a result of the topography of eastern Vermilion County. Steeply sloped areas created a shadow effect that is visible in both the spring and summer imagery resulting in a large number of forested upland areas being misclassified as forested wetland. Shadows and texture within the forest canopy produced similar errors.

Correlation between bands as measured by the Pearson's product-moment correlation coefficients (R) was used to understand duplication of information between bands and in conjunction with the importance measures (Figure 5.3) to determine the contribution of each input variable or layer to Classification A. All measures of importance agreed that spring near Infrared and spring NDWI contributed the most information to this classification. Spring red, green, blue, and infrared layers were strongly correlated. Similarly, summer red, green, blue, and infrared layers were strongly correlated. Spring NDWI showed high importance and low correlation with spring red, green, and infrared layers. Spring simple ratio showed low correlation with

other layers. Summer NDVI had moderate importance values and low correlation with layers with high importance values (spring visible and infrared bands). Summer NDWI also showed low correlation with other layers. However, because of its low importance statistics it is likely that the summer NDWI layer did not contribute to the accuracy of this classification.

Table 5.2 Accuracy Assessment of Classification A (All Imagery, FW-Forested wetland, FU-Forested upland)

	1 - FW	2 - FU	Row Total	User's Accuracy	Conditional Kappa Coefficient (K_{cA})
1 - FW	33	13	46	71.74%	35.27%
2 - FU	7	38	45	84.44%	64.37%
Column Total	40	51	71		
Producer's Accuracy	82.50%	74.51%	Total	91	
				Overall Accuracy	78.02%
				Kappa Coefficient of Agreement	56.10%

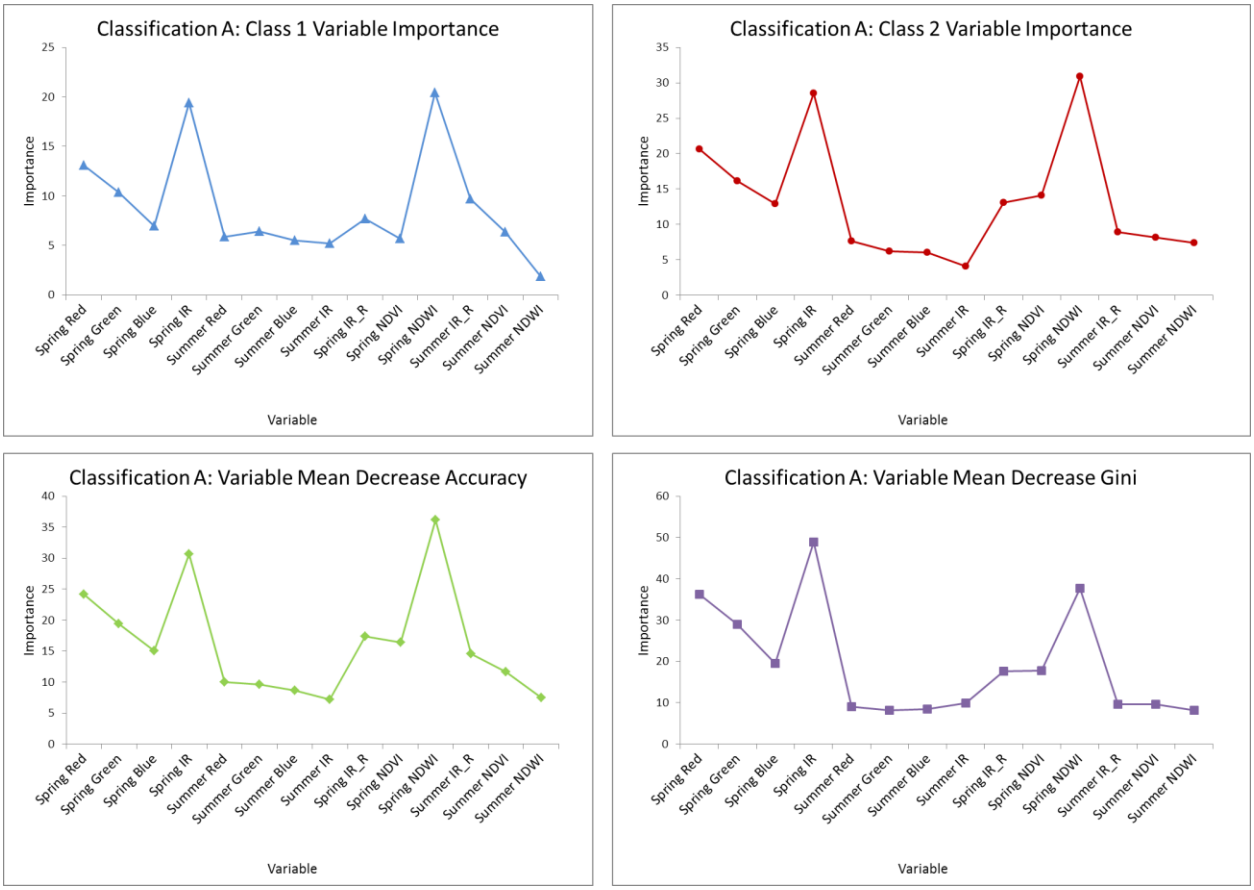


Figure 5.3 Four measures of variable importance for Classification A: 1) Class 1 (Forested Wetland) Marginal Importance; 2) Class 2 (Forested Upland) Marginal Importance; 3) Mean Decrease in Accuracy; and 4) Mean Decrease Gini Coefficient

5.2.2 CLASSIFICATION B: ALL LIDAR AND TERRAIN DERIVATIVES

Classification B consisted of LiDAR elevation, intensity, and topographic indices (lee filtered intensity, raw intensity, elevation, slope, aspect, combined curvature, planiform curvature, profile curvature, TWI, and normalized TWI) and resulted in a producer's accuracy of 92.86% and a Kappa Coefficient of Agreement of 78.04%, suggesting moderate to strong agreement. Conditional Coefficient of Agreement for forested wetland (K_{CB1}) and forested upland (K_{CB2}) classes in classification B were

68.39% and 86.15%, respectively, indicating a moderate to strong agreement for the forested wetland class and strong agreement for the forested upland class. The results of this classification are described in more detail in Table 5.3. Visual assessment of Classification B showed that the primary source of error in this classified layer was the due to noise and errors in the unfiltered intensity data. These errors resulted in forested upland being misclassified as forested wetland, and a subsequent increase in the forested wetland commission error. Lee filtering of the intensity data appeared to reduce some of this error.

Table 5.3 Accuracy Assessment of Classification B (All LiDAR, FW-Forested wetland, FU-Forested upland)

	1 - FW	2 - FU	Row Total	User's Accuracy	Conditional Kappa Coefficient (K_{cB})
1 - FW	39	7	46	84.78%	68.39%
2 - FU	3	42	45	93.33%	86.15%
Column Total	42	49	81		
Producer's Accuracy	92.86%	85.71%	Total	91	
				Overall Accuracy	89.01%
				Kappa Coefficient of Agreement	78.04%

Importance statistics (Figure 5.4) showed that lee filtered intensity had higher importance rankings than the unfiltered intensity data. All importance measures clearly show that the lee filtered intensity and slope variables contributed the most information to this classification. The elevation also ranked highly in the importance statistics and exhibited low correlation with the slope and intensity variables. Combined curvature was ranked moderately important compared with other layers. Despite the low importance

ranking of the TWI layer it exhibited low correlation with lee filtered intensity, elevation, slope, and aspect layers and moderate correlation with combined curvature. Surprisingly, raw TWI data was more important than the TWI normalized by elevation.

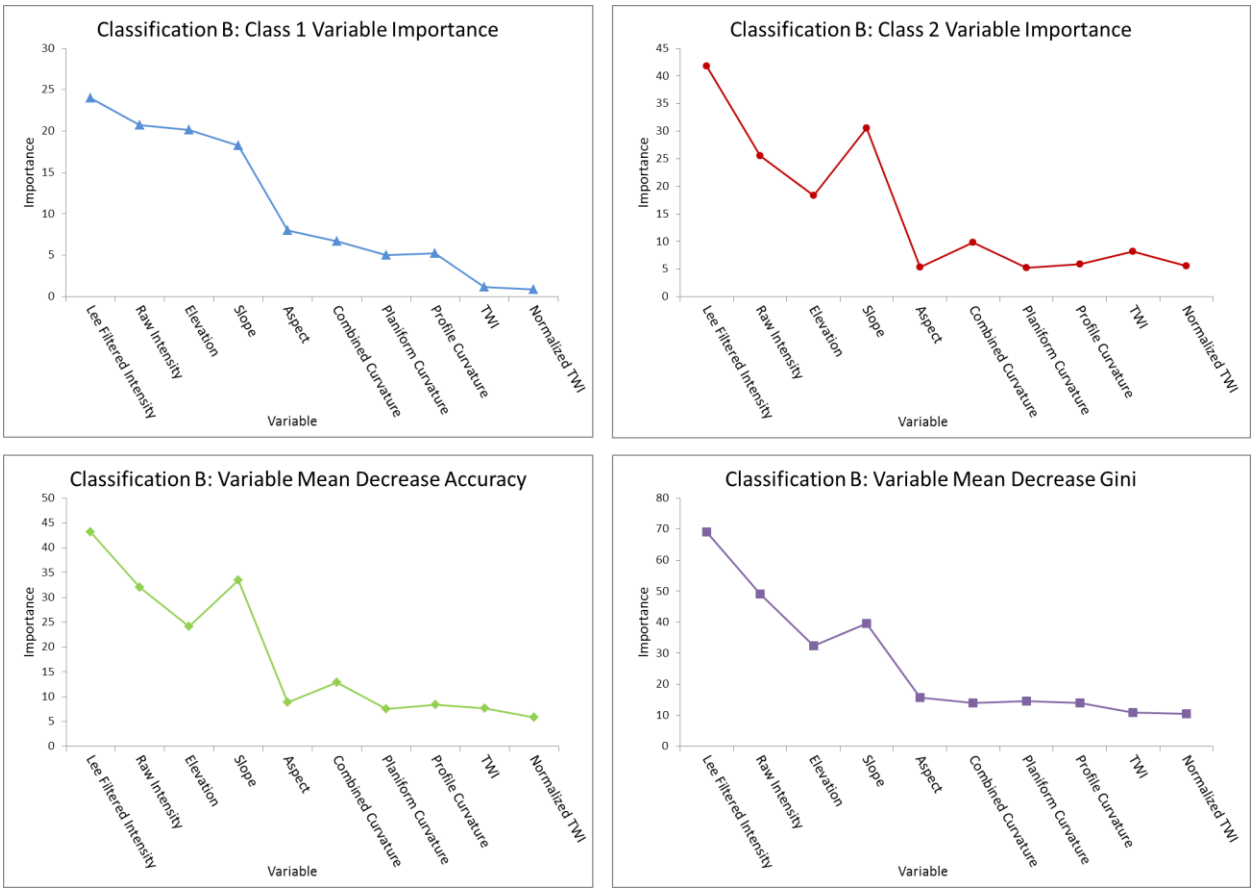


Figure 5.4 Four measures of variable importance for Classification B: 1) Class 1 (Forested Wetland) Marginal Importance; 2) Class 2 (Forested Upland) Marginal Importance; 3) Mean Decrease in Accuracy; and 4) Mean Decrease Gini Coefficient

5.2.3 CLASSIFICATION C: LIDAR DATASETS WITHOUT INTENSITY

Classification C consisted of all LiDAR datasets with the exception of the filtered and unfiltered intensity data. The purpose of this classification was to show the

contribution of the intensity datasets to the accuracy of the All LiDAR classification (Classification B). Classification C resulted in a forested wetland producer's accuracy of 65.93% and an overall Kappa Coefficient of agreement of 31.97% (Table 5.4), indicating low to moderate agreement. Every measure of accuracy for this classification was considerably less than that of Classification B and all other classifications. Specifically, there was a 46.07% decrease in the Kappa Coefficient of Agreement compared to Classification B. The Conditional Coefficient of Agreement for the forested wetland class (K_{CC1}) was -18.01%. While negative Kappa values are possible, they generally indicate that there is less agreement between the remotely sensed data and the reference data than could be reasonably expected by chance (Viera and Garrett 2005). The low accuracy of this classification is surprising given the contribution of topographic location to wetland location. Importance statistics (Figure 5.5) did show that slope and elevation data contributed the most information to Classification C.

Table 5.4 Accuracy Assessment of Classification C (All LiDAR without Intensity, FW-Forested wetland, FU-Forested upland)

	1 - FW	2 - FU	Row Total	User's Accuracy	Conditional Kappa Coefficient (K_{cC})
1 - FW	27	19	46	58.70%	-18.01%
2 - FU	12	33	45	73.33%	23.81%
Column Total	39	52	60		
Producer's Accuracy	69.23%	63.46%	Total	91	
				Overall Accuracy	65.93%
				Kappa Coefficient of Agreement	31.97%

Figure 5.5 Four measures of variable importance for Classification C: 1) Class 1 (Forested Wetland) Marginal Importance; 2) Class 2 (Forested Upland) Marginal Importance; 3) Mean Decrease in Accuracy; and 4) Mean Decrease Gini Coefficient

5.2.4 CLASSIFICATION D: FUSION OF ALL AERIAL IMAGERY AND LIDAR DATASETS

Input variables for Classification D consisted of 24 layers representing the fusion of all Imagery and LiDAR data. Classification D resulted in a forested wetland producer's accuracy of 87.18% and a Kappa Coefficient of Agreement of 62.70%, indicating moderate to strong overall agreement. The accuracy of this classification is further described in Table 5.5. Forested wetlands appeared to be well classified with the

fusion of all imagery and LiDAR datasets. However, the forested wetland Conditional Kappa Coefficient of Agreement (K_{cD1}) was 44.84% indicating only moderate agreement between this dataset and the reference data for forested wetlands. The forested upland conditional coefficient of agreement (K_{cD2}) was 76.51%, demonstrating a moderate to strong agreement between the reference data and the remotely sensed datasets. Shadows created by trees and slopes perpendicular to the line of flight appeared to contribute to errors of commission within the forested wetland class. Visual inspection of the resulting thematic layer indicated spring simple ratio values, summer visible and infrared reflectance, and noise within the curvature and TWI datasets were the primary sources of error for both the forested wetland and forested upland classes.

Importance statistics derived from the random forest classification (Figure 5.6) indicated that summer visible and infrared layers, combined curvature, profile curvature, and planiform curvature, as well as aspect, normalized TWI, TWI and summer NDWI were not important for the classification compared with other layers. Importance statistics indicate that spring near infrared, spring NDWI, and lee filtered LiDAR Intensity were the most important in the creation of this classification, despite the moderate correlation between the spring near infrared and lee filtered intensity layers. Based on these importance statistics, lee filtered intensity contributed the least amount of error, while spring IR resulted in the greatest decrease in node impurities. Spring red and elevation also ranked highly in the importance statistics and showed low correlation with one another. Although spring blue and spring green ranked moderately high in the importance statistics they were highly correlated with spring red and spring IR data. Slope ranked moderately in the importance statistics and had low correlation values

with other important layers. Spring NDVI and summer NDVI had moderate importance rankings and correlation values. Summer simple ratio had low to moderate importance rankings but showed low correlation with other selected bands. Because of the high correlation between lee filtered intensity and raw Intensity, and the high importance ranking of the filtered intensity data, raw Intensity was considered less useful than the lee filtered intensity for forested wetland classification.

Table 5.5. Accuracy Assessment of Classification A (Fusion of All Data, FW-Forested wetland, FU-Forested upland)

	1 - FW	2 - FU	Row Total	User's Accuracy	Conditional Kappa Coefficient (K_{cD})
1 - FW	34	12	46	73.91%	44.84%
2 - FU	5	40	45	88.89%	76.51%
Column Total	39	52	74		
Producer's Accuracy	87.18%	76.92%	Total	91	
				Overall Accuracy	81.32%
				Kappa Coefficient of Agreement	62.70%

Figure 5.6 Four measures of variable importance for Classification D: 1) Class 1 (Forested Wetland) Marginal Importance; 2) Class 2 (Forested Upland) Marginal Importance; 3) Mean Decrease in Accuracy; and 4) Mean Decrease Gini Coefficient

5.3 COMPARISON OF CLASSIFICATION ACCURACIES

Given the null hypothesis H_0: $K_1\hat{} = 0$, Z score values greater than or equal to the critical value of 1.96 for the two sided 95% confidence interval would allow us to reject the null hypothesis that the kappa value of an individual classification is equal to 0. Analysis of individual classification K^ significance indicates that results from All Data, All Imagery, All LiDAR, and the Fusion of Selected Imagery and LiDAR classifications were better than a random classification (Table 5.6).

Table 5.6 Z-Scores used for testing the significance of the Kappa Coefficient of Agreement ($\hat{K}$) of each classification. Agreement between the remote sensing and reference data, as measured by $\hat{K}$, is not significant for classifications with z-scores less than the 1.96 critical value at the 95% confidence level (Congalton and Green 1999)

	$Z = \frac{\lvert\hat{K_1}\rvert}{\sqrt{var(\hat{K_1})}}$
A – ALL IMAGERY	20.6660
B – ALL LIDAR	16.0271
C - LIDAR NO INTENSITY	5.0804
D – ALL DATA FUSION	16.3169

The paired z-test, H_0: ($\hat{K_1}$ - $\hat{K_2}$) = 0, showed that all classifications could be reliably compared with the exception of Classifications A and D (All Imagery and the Fusion of All Data) (Table 5.7). Results from this two-sided test can be found in Table 5.8. The All LiDAR Classification ranked highest in accuracy as measured by the Kappa Coefficient of Agreement, as well as the Overall Accuracy and Forested Upland Producer's Accuracy. Notably, the paired z-test revealed that the accuracy of the All LiDAR classification was significantly different from all other classifications. Classification C, LiDAR without intensity had the lowest classification accuracy using all measures of agreement. The fusion of Imagery and LiDAR datasets (Classification D) appeared to result in improved accuracy over the classification using just Imagery (Classification A). However, due to the failure of the paired z-test it cannot be said that there is a significant difference between these classifications.

Table 5.7 Summary of accuracy assessment results for all classifications

	Forested Wetland Producer Accuracy	Forested Wetland User Accuracy	Forested Upland Producer Accuracy	Forested Upland User Accuracy	Overall Accuracy	Kappa Coefficient ($\hat{K}$)	Conditional Kappa Coefficient ($\hat{K}_{c1}$)	Conditional Kappa Coefficient ($\hat{K}_{c2}$)
A – ALL IMAGERY	82.50%	71.74%	74.51%	84.44%	78.02%	56.10%	35.27%	64.37%
B – ALL LIDAR	92.86%	84.78%	85.71%	93.33%	89.01%	78.04%	68.39%	86.15%
C – LIDAR WITHOUHT INTENSITY	69.23%	58.70%	63.46%	73.33%	65.93%	31.91%	-18.01%	23.81%
D – ALL DATA	87.18%	73.91%	76.92%	88.89%	81.32%	62.70%	44.84%	76.51%

Table 5.8 Z-Scores used for testing the significance of the difference between two classification Kappa Coefficient of Agreement statistics. The difference between the agreement between the remote sensing and reference data, as measured by $K_1^{\wedge}$ and $K_2^{\wedge}$, is not significant if their corresponding z-score is less than the 1.96 critical value at the 95% confidence level (Congalton and Green 1999)

$Z = \frac{\lvert \hat{K_1} - \hat{K_2} \rvert}{\sqrt{var(\hat{K_1}) + var(\hat{K_2})}}$	A – ALL IMAGERY	B – ALL LIDAR	C - LIDAR NO INTENSITY	D – ALL DATA FUSION
A – ALL IMAGERY	0.0000			
B – ALL LIDAR	1.4015	0.0000		
C - LIDAR NO INTENSITY	2.4738	3.9352	0.0000	
D – ALL DATA FUSION	4.1672	3.5212	5.7901	0.00

A visual comparison of the four classifications alongside forested wetlands identified in the NWI (Figure 5.7) showed that Classification B and D were similar in the areas identified as forested wetland and forested upland. Classification C appeared to have moderately strong agreement with the NWI data, particularly in low-lying areas that could be considered floodplain, or bottomland forest. Though, there was some obvious forested wetland omission and forested upland commission error. Classification A appeared to agree with the areas classified as forested upland in Classification B and

D, however a large amount of forested upland commission error was present in this classification.

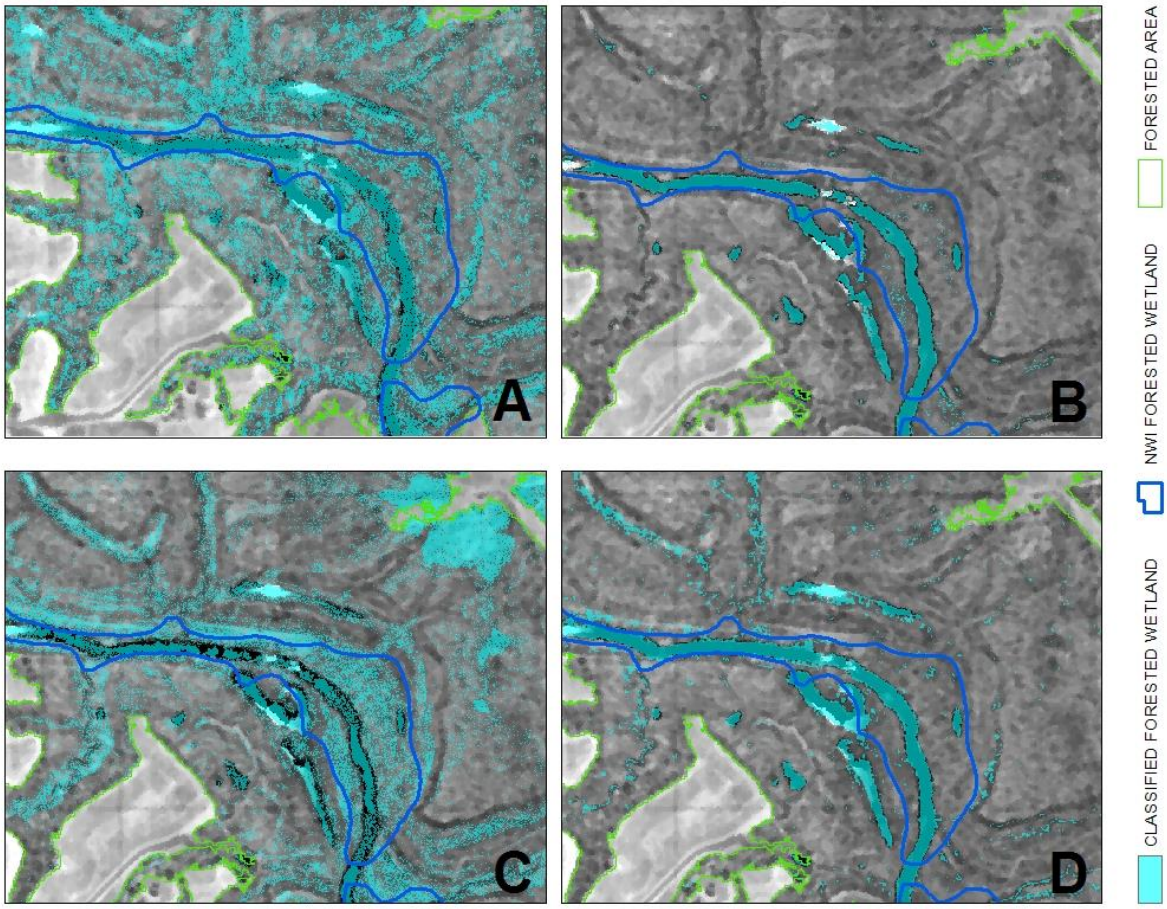


Figure 5.7 A comparison of four wetland classifications along with an area identified as forested wetland by the National Wetland Inventory (scale = 1:9,000.)

CHAPTER 6

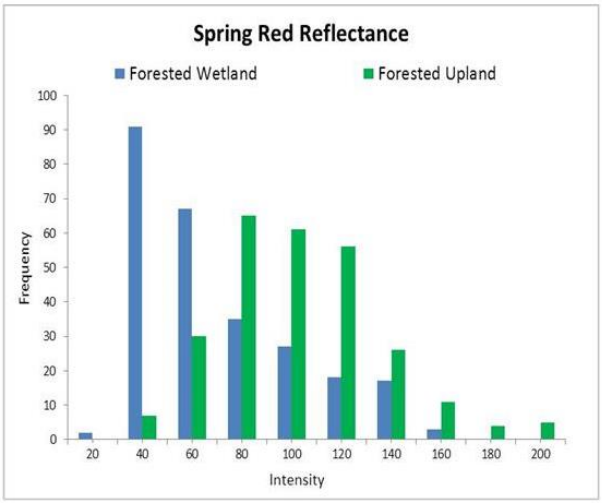
DISCUSSION

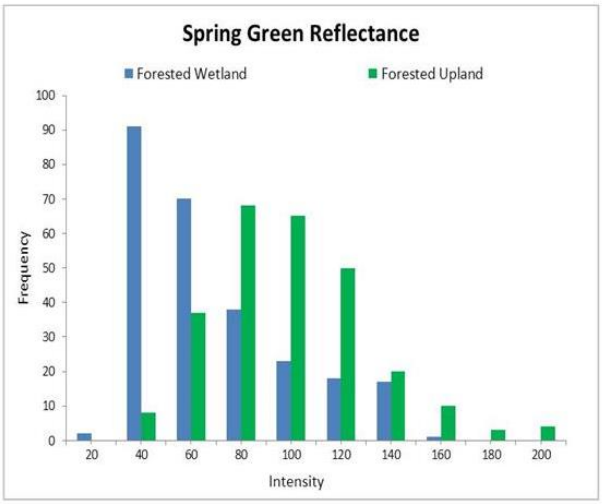
6.1 RESEARCH FINDINGS

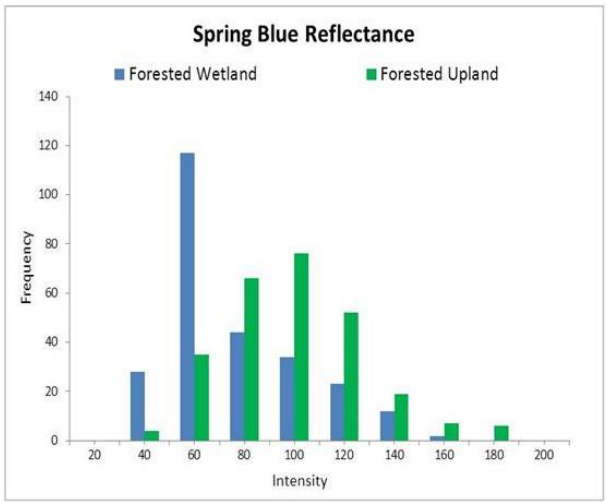
This study proposed three research questions in order to evaluate the accuracy of LiDAR data, multi-spectral aerial imagery, and their fusion for detecting forested wetland hydrology in eastern Vermilion County, Illinois. The classifications, accuracy assessment and comparison allowed us to determine 1) If LiDAR ground return intensity data can be useful for identifying forested wetlands, 2) How different is the accuracy of a forested wetland classification performed with traditional multi-spectral imagery compared with that performed with LiDAR data, and 3) If the fusion of LiDAR and multi-spectral aerial imagery datasets improved forested wetland classification.

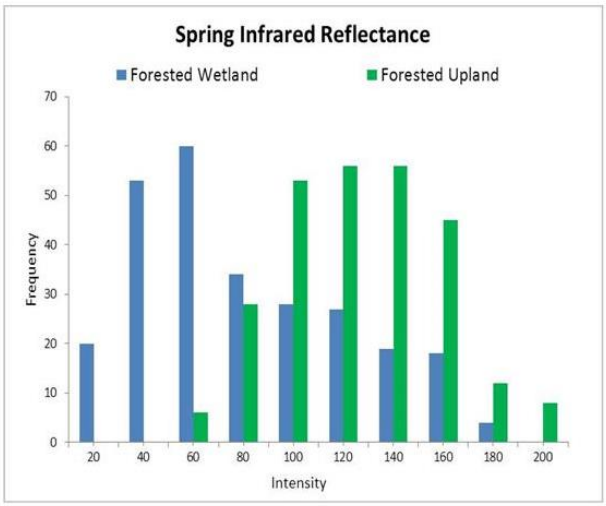
Due to the low accuracy of classification performed without intensity (Classification C) it was concluded that LiDAR ground return intensity data was useful for forested wetland identification. Classification C had not only the lowest Kappa but also the worst accuracy across all other measures. The Kappa Coefficient of Agreement decreased 46.07% compared to Classification B (Kappa = 78.04%). This moderate to strong accuracy of Classification B, compared to the low accuracy of Classification C, is meaningful because it shows that without the intensity data, the LiDAR elevation and terrain datasets were unable to identify forested wetland hydrology accurately. Importance statistics showed that slope and elevation contributed the most information to Classification C. But, Figure 6.1 shows that while uplands were most likely to be found in areas with little to no slope, there was not much difference between the

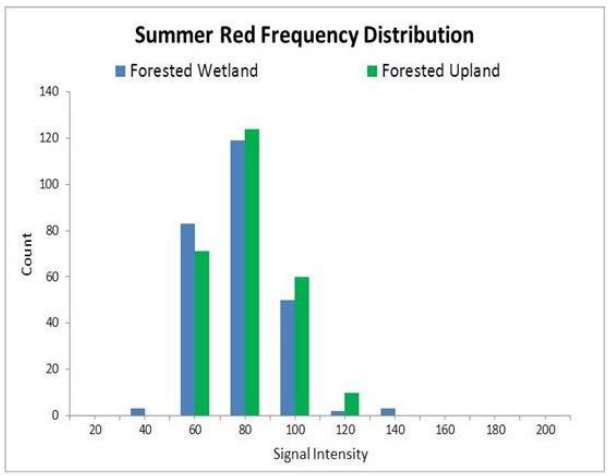
elevations found in forested upland and forested wetland-training areas. Figure 6.1 also shows intensity values in forested wetland locations are most often lower than intensity values found in forested upland areas, indicating the ability of this dataset to distinguish between these features. Noise filtering of intensity data appeared to increase the seperability of these classes. Overall, with inclusion of the intensity data, data derived from LiDAR data is capable of accurately mapping forested wetlands where surface water is present.

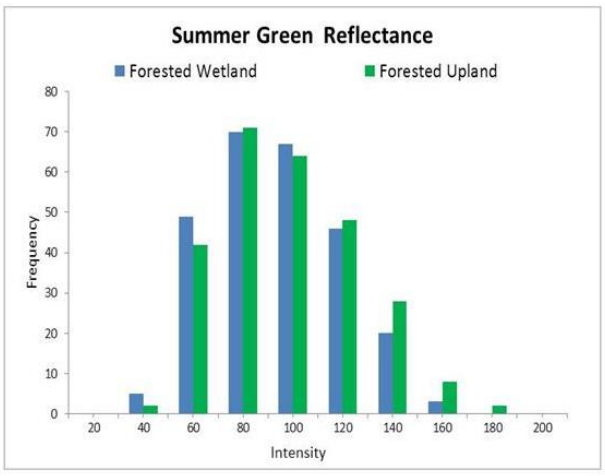

Spring Red Reflectance
Forested Wetland
Forested Upland
Frequency
Intensity

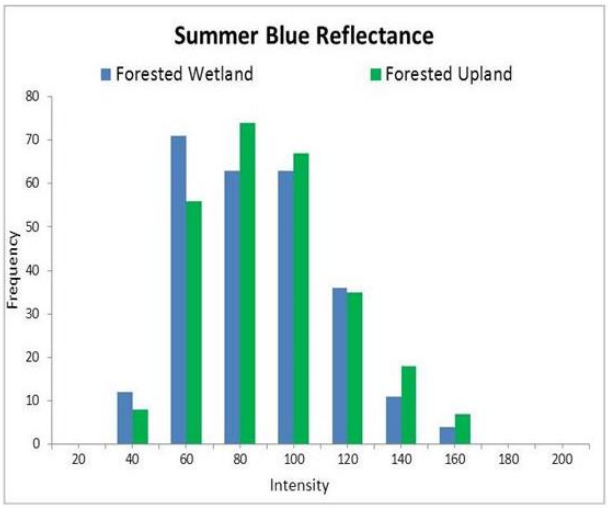

Spring Green Reflectance
Forested Wetland
Forested Upland
Frequency
Intensity

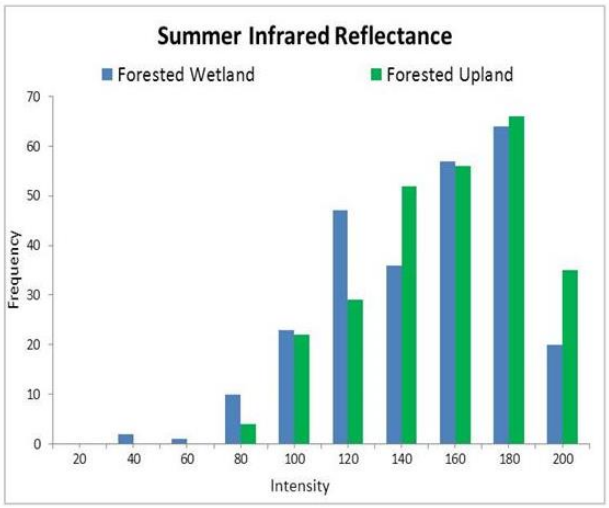

Spring Blue Reflectance
Forested Wetland
Forested Upland
Frequency
Intensity

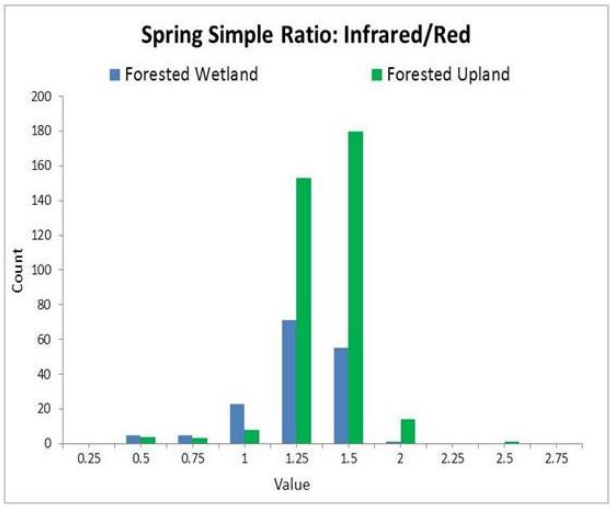

Spring Infrared Reflectance
Forested Wetland
Forested Upland
Frequency
Intensity

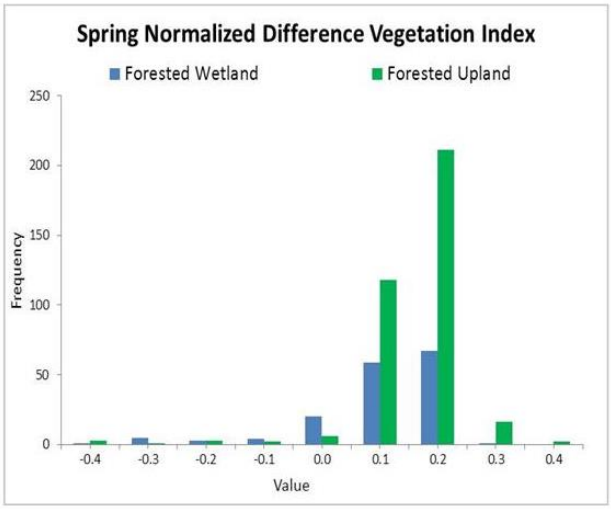

Summer Red Frequency Distribution
Forested Wetland
Forested Upland
Count
Signal Intensity

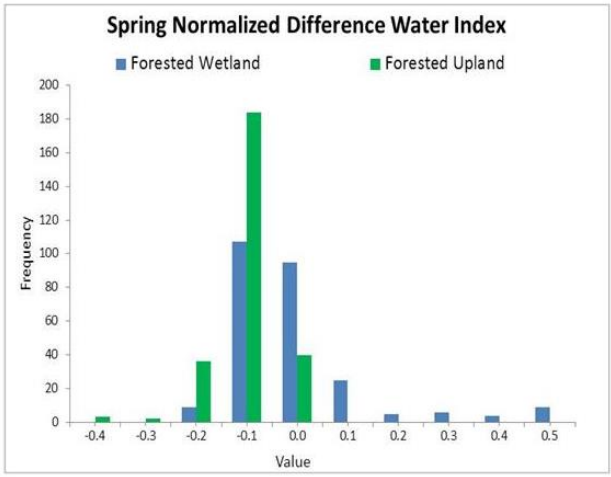

Summer Green Reflectance
Forested Wetland
Forested Upland
Frequency
Intensity

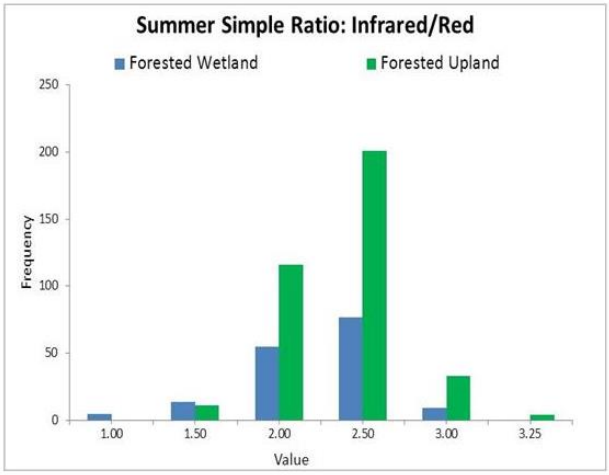

Summer Blue Reflectance
Forested Wetland
Forested Upland
Frequency
Intensity
0
10
20
30
40
50
60
70
80
20
40
60
80
100
120
140
160
180
200

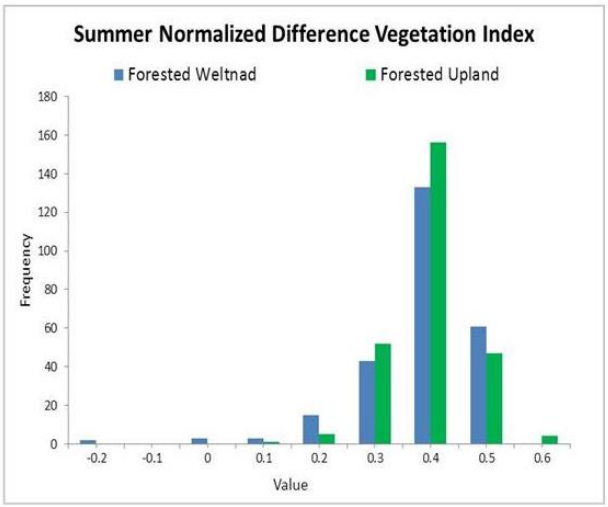

Summer Infrared Reflectance
Forested Wetland
Forested Upland
Frequency
Intensity
0
10
20
30
40
50
60
70
20
40
60
80
100
120
140
160
180
200

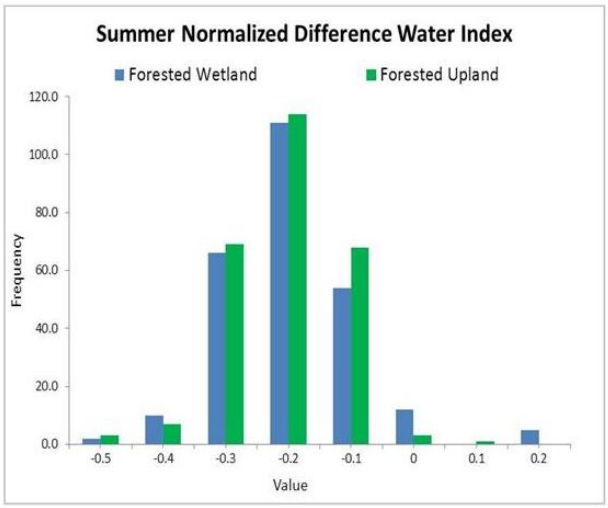

Spring Simple Ratio: Infrared/Red
Forested Wetland
Forested Upland
Count
Value
0
20
40
60
80
100
120
140
160
180
200
0.25
0.5
0.75
1
1.25
1.5
2
2.25
2.5
2.75

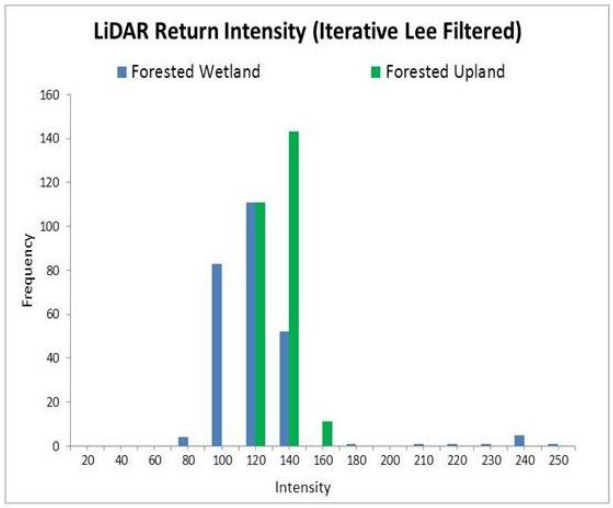

Spring Normalized Difference Vegetation Index
Forested Wetland
Forested Upland
Frequency
Value
0
50
100
150
200
250
-0.4
-0.3
-0.2
-0.1
0.0
0.1
0.2
0.3
0.4

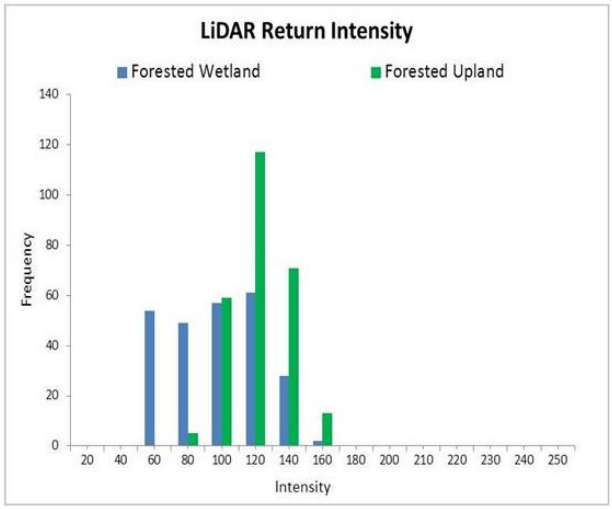

Spring Normalized Difference Water Index
Forested Wetland
Forested Upland
Frequency
Value
0
20
40
60
80
100
120
140
160
180
200
-0.4
-0.3
-0.2
-0.1
0.0
0.1
0.2
0.3
0.4
0.5

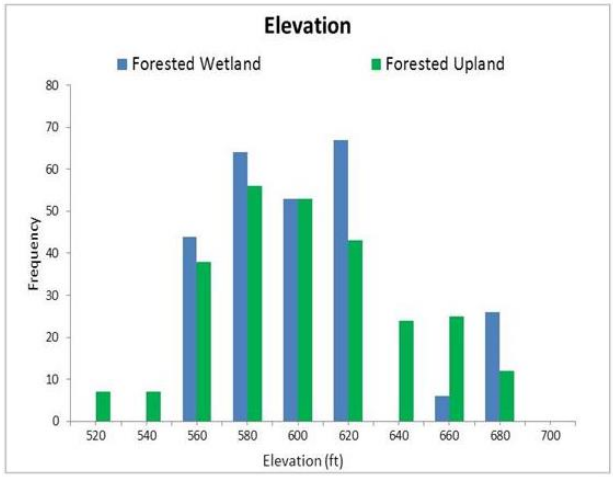

Summer Simple Ratio: Infrared/Red
Forested Wetland
Forested Upland
Frequency
Value
0
50
100
150
200
250
1.00
1.50
2.00
2.50
3.00
3.25

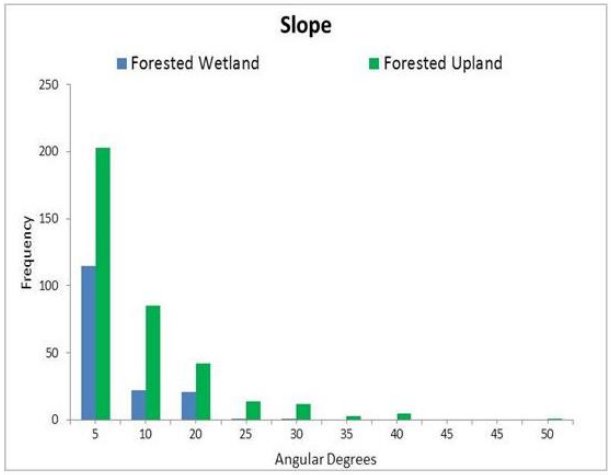

Summer Normalized Difference Vegetation Index
Forested Weltnad
Forested Upland
Frequency
Value

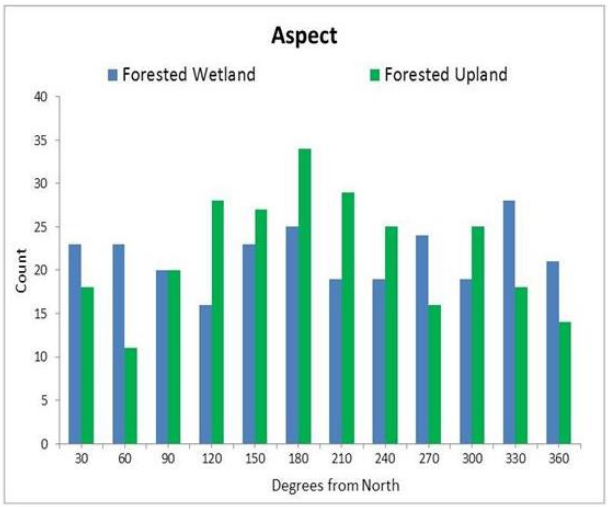

Summer Normalized Difference Water Index
Forested Wetland
Forested Upland
Frequency
Value

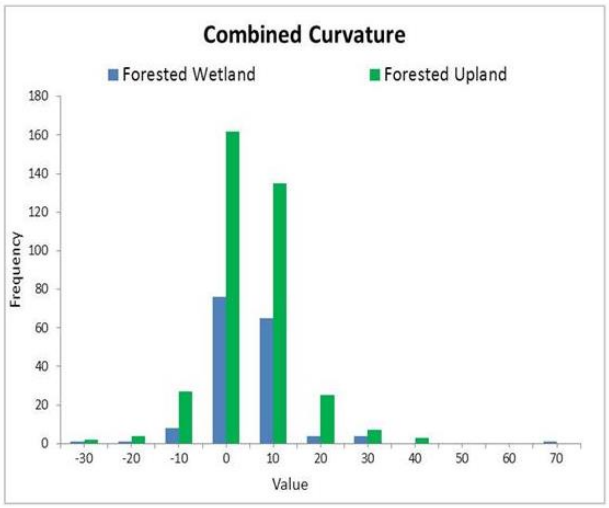

LiDAR Return Intensity (Iterative Lee Filtered)
Forested Wetland
Forested Upland
Frequency
Intensity

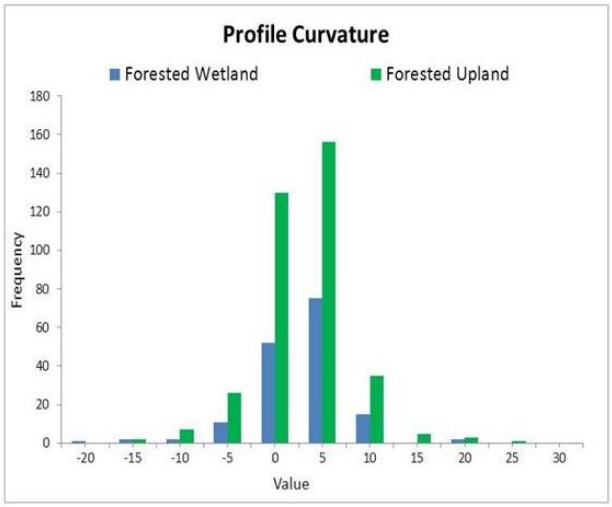

LiDAR Return Intensity
Forested Wetland
Forested Upland
Frequency
Intensity

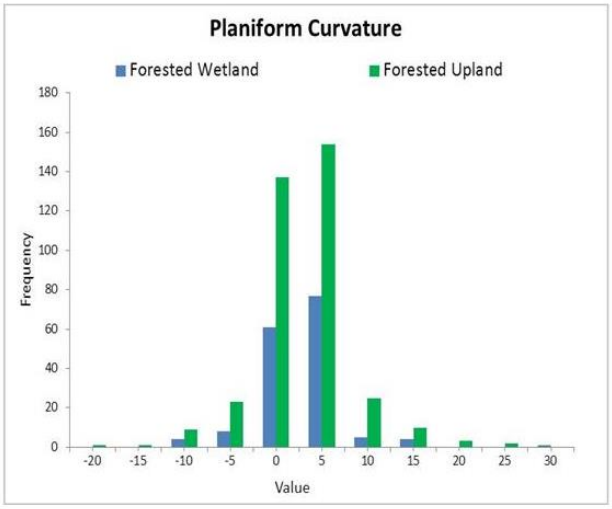

Elevation
Forested Wetland
Forested Upland
Frequency
Elevation (ft)

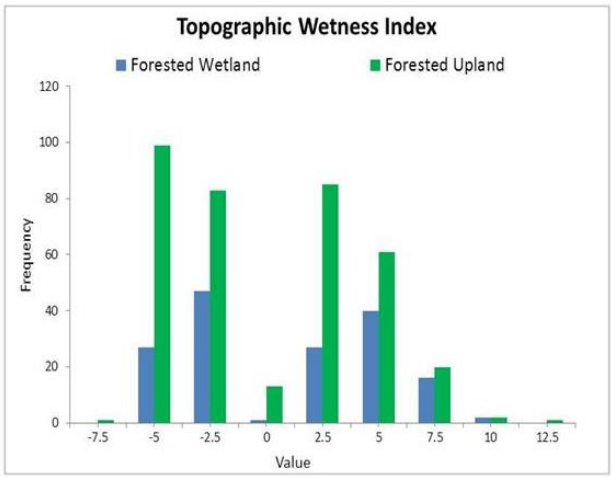

Slope
Forested Wetland
Forested Upland
Frequency
Angular Degrees


Aspect
Forested Wetland
Forested Upland
Count
Degrees from North


Combined Curvature
Forested Wetland
Forested Upland
Frequency
Value


Profile Curvature
Forested Wetland
Forested Upland
Frequency
Value


Planiform Curvature
Forested Wetland
Forested Upland
Frequency
Value


Topographic Wetness Index
Forested Wetland
Forested Upland
Frequency
Value

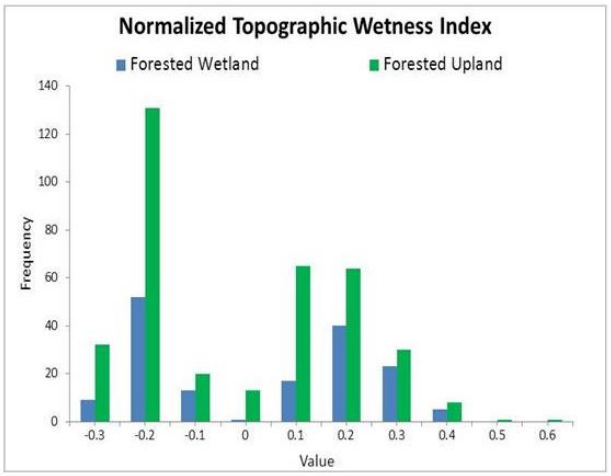

Normalized Topographic Wetness Index
Forested Wetland
Forested Upland
Frequency
Value

Figure 6.1 Histograms of spectral refletance values found within forested wetland and forested upland training polygons

The accuracy of the forested wetland classification performed with multi-spectral imagery (Classification A) was less accurate than that performed with all LiDAR data (Classification B) (Kappa = 56.10% vs. Kappa = 78.04%). Importance statistics from Classification A indicated a heavy reliance on information from the spring imagery; particularly the near infrared and spring NDWI input variables. This makes sense considering the frequency distributions (Figure 6.1) of these datasets which show that although there is some overlap in the values found in upland and wetland areas, wetland areas likely have low spring near infrared intensity and high spring NDWI. In the spring visible and infrared bands, forested upland reflectance values were relatively normally distributed while wetland values were negatively skewed. Considering the distribution of the values in each of the imagery datasets alongside the importance statistics, it is clear that the summer imagery data contributed the least information to Classification A. Most of the forested upland and wetland summer reflectance values were very similar due to the presence of the forest canopy.

It was not conclusively proven whether the fusion of LiDAR and aerial multi-spectral imagery data improved forested wetland classification accuracy. Evidence points to these classifications resulting in similar levels of accuracy. The Kappa Coefficients of Agreement for Classification B and D were 62.70% and 56.10%, respectively. Although Classification D had a slightly higher Kappa value, due to the

failure of the two-sided test for a difference between the Kappa values of Classification B and D (Table 5.8), we cannot say that these accuracy measures are significantly different. This could be interpreted as the accuracy of these classifications being statistically similar. Visual assessment of these classifications (Figure 5.7) shows many of the same areas identified as forested wetland, indicating that these classifications did result in similar accuracy. This finding, along with the increased accuracies of Classification B and Classification C, indicates that LiDAR intensity and terrain information provides superior ability to detect forested wetland hydrology compared with multi-spectral imagery due to absorption of the LiDAR signal where surface water is present..

6.2 LIMITATIONS

Disagreement between classifications, or even disagreement between remote sensing data and reference data, may be a result of the inconsistency of time remote sensing data and field data collected. Each of the datasets and field data were collected on different dates, and in some cases different years. 2011 and 2012, the years in which the reference aerial orthophotos were taken, were drier than the 10 and 100-year average (Figure 6.2). Data collected in conditions that are wetter than average would likely result in identification of a larger area of forested wetlands, particularly along the Vermilion River floodplain. If moisture conditions were very different during the periods when spring imagery, LiDAR data, and field data were collected this may result in errors. Specifically, due to the collection of LiDAR data during the very dry spring of

2012 may have resulted in forested wetland omissions compared to the wetter conditions in the March of 2011, when spring imagery was collected.

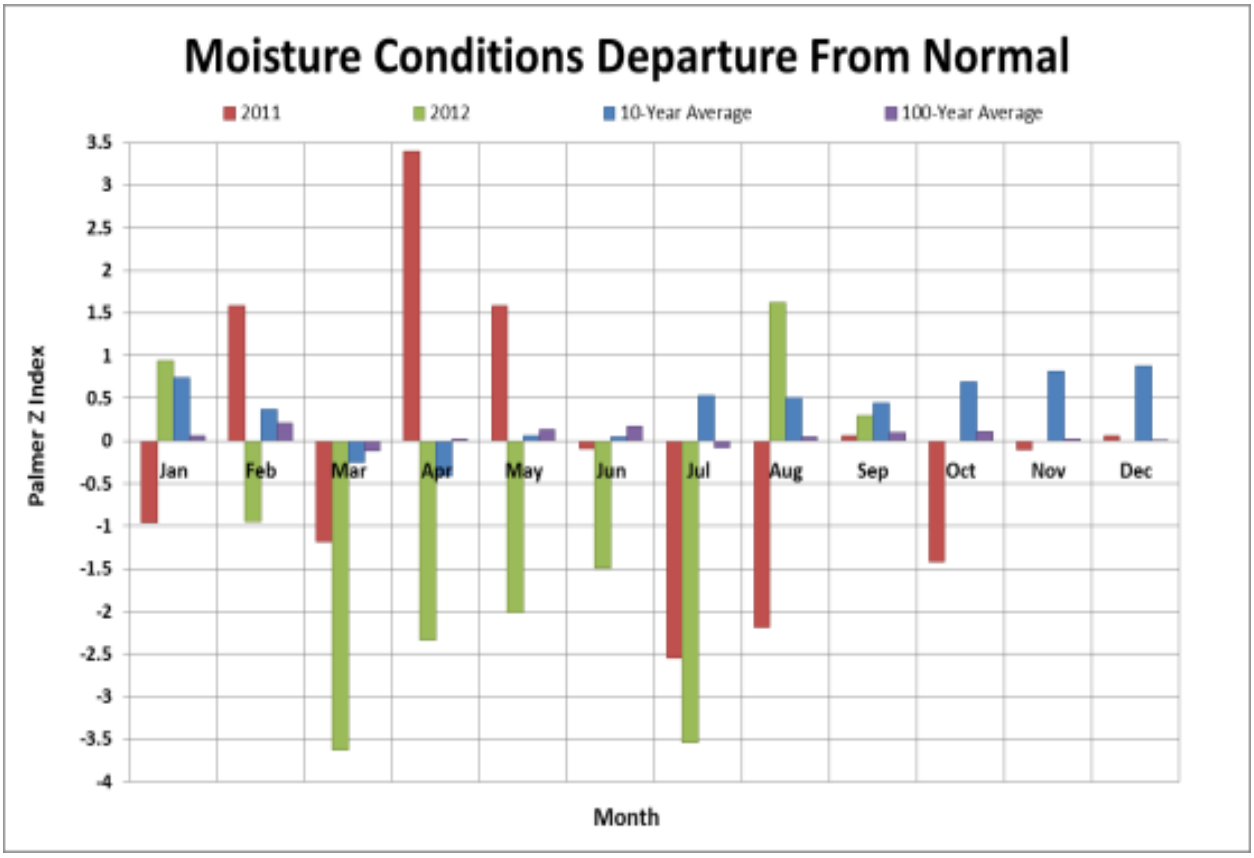


Figure 6.2 The Palmer Z index measures short-term moisture conditions on a monthly scale by taking into account the precipitation, evapotranspiration, and runoff (NCDC NOAA 2014)

Additionally, errors within the remote sensing datasets may contribute to classification confusion. Despite the accuracy achieved using LiDAR data, some aspects of the intensity component introduced errors into the classification. Specifically, in a few forested wetlands abnormally high intensity values were found due to the

reflection rather than absorption of the LiDAR signal where ice was present. However, despite these anomalies in the intensity data post-classification visual assessment of Classification B and D showed that, these areas were still identified as forested wetland. Classification D, in particular showed improved classification over Classification B likely due to the use of both imagery and terrain data.

Within the intensity data there was also a loss along the boundaries of .LAS files and diagonal strips of missing intensity data within a number of the .LAS tiles that appeared to be an artifact of data processing by Aerometric. Noise filtering decreased these errors to some extent; however, many of these areas were still misclassified as forested wetland contributing to the forested wetland commission error. In some areas of the study area dense tree canopy and understory may have affected the ability of the LiDAR signal to penetrate fully to the ground. This would have resulted in errors of elevations about their real world height and inaccuracies within the derived terrain data. Without further analysis, it is unclear whether this affected the accuracy of classifications performed using LiDAR data.

There is very likely some amount of error in the training and reference datasets due to collection of field data with multiple GPS units of differing accuracy capabilities as well as the fact that LiDAR positional error is increased in forested areas. Care was taken to minimize these errors with real-time correction, post-processing, and point averaging. If errors in the classifications have occurred due to shifts within the training or reference data, the errors are consistent across all classifications and do not affect the comparison of classification accuracies.

Finally, this study's limited definition of a forested wetland (presence of moist soils or surface water) may have resulted in pixels identified as forested wetland in Classification A and D to be considered incorrect. These areas may in fact have water at or near the surface in the spring of most years, however if this hydrology was not present during field data collection, the locations would have been identified as upland. Classifications A and D which include aerial imagery as input data may be detecting water tolerant tree species associated with the training areas. Because this work based wetland determination on only the presence of hydrology, areas identified based on the spectral signature of water-tolerant tree species, but without obvious wetland hydrology, would be identified as a classification error when they may be correctly mapped. However, it is unlikely that this is the reason for the high forested wetland commission error in Classification A.

6.3 FUTURE RESEARCH

Future research should address the following questions:

1. Can LiDAR derived data be used to identify wetland hydrology in other ecosystem types such as coniferous forest, scrub/shrub, or emergent wetlands.

2. Could multi-date LiDAR intensity be used for understanding seasonal changes in wetland extent?

3. What results would an object based classification method have on the accuracy of these classifications?

This ability to distinguish wetland hydrology difficultly to map forested regions leads us to believe that this data should be included when mapping other wetland types. Although this research focused on the identification of forested wetland hydrology in broad-leaved, deciduous, forest LiDAR derived datasets may prove useful for other wetland types. Hogg and Holland (2008) successfully applied LiDAR derived topographic indices to the classification of wetland areas in the coniferous forests of Canada. It makes sense that the intensity data would also be useful in coniferous forests. Additionally, evaluating the ability of spring and fall LiDAR for monitoring temporal variations in forested wetland hydrology may result in new methods for quantifying these temporal variations and result in more accurate classification of forested wetlands.

6.4 CONCLUSION

This research was able to verify the utility of LiDAR intensity data for the identification of forested wetlands and establish the ability of LiDAR derived datasets to improve forested wetland classification compared to multi-spectral aerial imagery due to the ability of the intensity data to distinguish between dry soil and open water below the forest canopy. These results are in agreement with the work of Lang and McCarty (2009,) Stevens and Wolfe (2012,) and Huang et al (2014). Although the intensity data contribute meaningful information to the forested wetland classification, the errors due

to lack of calibration, and artifacts of data processing by AeroMetric also contributed some uncertainty to even the most accurate classification. Although, methods exist for calibrating intensity data the required information is rarely provided with the deliverable product. As the body of literature proving the utility of LiDAR intensity data grows perhaps calibration of intensity data will be performed along with the elevation component, or else, the necessary calibration data will be provided along with the raw .LAS files.

This research aligned with the U.S. Fish and Wildlife goal of developing more efficient and cost effective methods of inventorying wetlands, specifically for difficult to map forested wetlands. It is expected that by identifying the most significant predictors of forested wetlands the number of datasets needed to complete an accurate forested wetland inventory will be reduced, effectively reducing the cost of inventorying these resources. By increasing the utility of the LiDAR dataset, the cost of LiDAR data collection may be better justified. Thus, increasing the cost effectiveness of LiDAR data by increasing its utilization may lead to an increased number of inventories performed using this high-resolution dataset - ultimately resulting in the increased accuracy of forested wetland mapping.

Given that this research has shown the utility of LiDAR intensity data and benefits of including lidar data in wetland classifications, similar studies should be undertaken to explore the use of LiDAR data for identifying wetland hydrology in other wetland ecosystems and in areas of high ecological importance. For example, LiDAR data have recently become available for all counties in Illinois, including Pulaski County in Southern Illinois where the Cache River RAMSAR Wetland Area is located. Use of

this methodology may be helpful in identifying isolated or degraded forested wetland habitat for potential restoration and reconnection. The increased accuracy of wetland detection in the difficult to map forested wetland ecosystem indicates that the use of LiDAR intensity data alongside other remote sensing datasets may also improve the classification accuracy of wetland maps in other ecosystems, including estuarine, riverine, lacustrine, and other palustrine wetland types (shrubby, emergent, and open water.)

REFERENCES

Awl, Jane, Jeanne Christie, Mararete Heber, Megan Lang, Bill Wilen. "A New Mapping Standard" *National Wetlands Newsletter*, Vol. 31, No. 5 (2009)

Beven, K. J., and M. J. Kirkby. "A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant." *Hydrological Sciences Journal* 24, no. 1 (1979): 43-69.

Bourgeau-Chavez, Laura L., Kevin Riordan, Nicole Miller, Mitch Nowels, and Richard Powell. "Remotely Monitoring Great Lakes Coastal Wetlands With Multi-Sensor, Multi-Temporal SAR And Multi-Spectral Data." In Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International, vol. 1, pp. I-428. IEEE, 2008a.

Bourgeau-Chavez, L.L., Riordan, K., Powell, R.B., Miller, N., and Nowels, M. (2009). Improving Wetland Characterization with Multi-Sensor, Multi-Temporal SAR and Optical/Infrared Data Fusion, Advances in Geoscience and Remote Sensing, Gary Jedlovec (Ed.), ISBN: 978-953-307-005-6, InTech, Available from:http://www.intechopen.com/books/advances-in-geoscience-and-remote-sensing/improvingwetland-characterization-with-multi-sensor-multi-temporal-sar-and-optical-infrared-data-fu

Bourgeau-Chavez, L. L., R. D. Lopez, A. Trebitz, T. Hollenhorst, G. E. Host, B. Huberty, R. L. Gauthier, and J. Hummer. "Chapter 8, Landscape-Based Indicators in Great Lakes Coastal Wetlands Monitoring Plan, Great Lakes Coastal Wetlands

Consortium, Project of the Great Lakes Commission, funded by the US EPA GLNPO, March 2008, 143 171." (2008b).

Boyd, Doreen S., and R. A. Hill. "Validation of airborne lidar intensity values from a forested landscape using hymap data: preliminary analyses." *The International Archives of the Photogrammetry, Remote Sensing, and Spatial Information Sciences* 36 (2007): 71-76.

Breiman, Leo. "Bagging predictors." *Machine learning* 24, no. 2 (1996): 123-140.

Breiman, Leo. "randomForest: Breiman and Cutler's Random Forests for Classification and Regression." (2006): 4-5.

Bivand, R. S., Pebesma, E., Gomez-Rubio, V., 2013. Applied spatial data analysis with R, Second edition. Springer, NY. http://www.asdar-book.org/

Bwangoy, Jean-Robert B., Matthew C. Hansen, David P. Roy, Gianfranco De Grandi, and Christopher O. Justice. "Wetland mapping in the Congo Basin using optical and radar remotely sensed data and derived topographical indices." *Remote Sensing of Environment* 114, no. 1 (2010): 73-86.

Chander, Gyanesh, Brian L. Markham, and Dennis L. Helder. "Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors." *Remote sensing of environment* 113, no. 5 (2009): 893-903.

Congalton, Russell, and Roy A. Mead. "A quantitative method to test for consistency and correctness in photointerpretation." *Photogrammetric Engineering & Remote Sensing* 49, no. 1 (1983): 69-74.

Corcoran, Jennifer M., Joseph F. Knight, and Alisa L. Gallant. "Influence of Multi-Source and Multi-Temporal Remotely Sensed and Ancillary Data on the Accuracy of

Random Forest Classification of Wetlands in Northern Minnesota." *Remote Sensing* 5, no. 7 (2013): 3212-3238.

Cowardin, Lewis M. Classification of Wetlands and Deepwater Habitats of the US. DIANE Publishing, 1979.

Dahl, Thomas E., and Gregory J. Allord. "History of wetlands in the conterminous United States." *Judy D. Fretwell, John S. Williams, and Phillip J. Redman. (eds.), National Water Summary on Wetland Resources, USGS Water-Supply Paper* 2425 (1996): 19-26.

Dalponte, Michele, Lorenzo Bruzzone, and Damiano Gianelle. "Tree species classification in the Southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data." *Remote sensing of environment* 123 (2012): 258-270.

Donoghue, Daniel NM, Peter J. Watt, Nicholas J. Cox, and Jimmy Wilson. "Remote sensing of species mixtures in conifer plantations using LiDAR height and intensity data." *Remote Sensing of Environment* 110, no. 4 (2007): 509-522.

Ellis, L. Rex, William D. Mahler, and Travis C. Richardson. "LiDAR Based Delineation of Depressional Wetlands." *Trends in Soil Science and Plant Nutrition* 3, no. 1 (2012): 8-12.

Erdody, Todd L., and L. Monika Moskal. "Fusion of LiDAR and imagery for estimating forest canopy fuels." *Remote Sensing of Environment* 114, no. 4 (2010): 725-737.

Federal Geographic Data Committee (FGDC), and Wetlands Subcommittee. Application of satellite data for mapping and monitoring wetlands. Technical Report 1, Washington, DC, 1992.

Federal Geographic Data Committee (FGDC), and Wetlands Subcommittee. "Wetlands Mapping Standard." *US Geological Survey, Reston, Virginia* (2009).

Foody, Giles M. "Thematic map comparison: evaluating the statistical significance of differences in classification accuracy." *Photogrammetric Engineering and Remote Sensing* 70, no. 5 (2004): 627-634.

Gallant, John, F. Chan, D. Marinova, and R. S. Andersson. "The ground beneath your feet: digital elevation data for today and tomorrow." In MODISM 2011, 19th International Congress on Modelling and Simulation, Modelling and Simulation Society of Australia and New Zealand, Perth, Australia, pp. 70-76. 2011.

Gao, Bo-Cai. "NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space." *Remote sensing of environment* 58, no. 3 (1996): 257-266.

García, Mariano, David Riaño, Emilio Chuvieco, and F. Mark Danson. "Estimating biomass carbon stocks for a Mediterranean forest in central Spain using LiDAR height and intensity data." *Remote Sensing of Environment* 114, no. 4 (2010): 816-830.

Gilbert, Paul (2012). setRNG: Set (Normal) Random Number Generator and Seed. R package version 2011.11-2. http://CRAN.R-project.org/package=setRNG

Gritzner, Janet H. "Identifying wetland depressions in bare-ground LIDAR for hydrologic modeling." In 26th annual ESRI international user conference proceedings. 2006.

Hartfield, Kyle A., Katheryn I. Landau, and Willem JD Van Leeuwen. "Fusion of high resolution aerial multispectral and LiDAR data: land cover in the context of urban mosquito habitat." *Remote Sensing* 3, no. 11 (2011): 2364-2383.

Hogg, A. R., and K. W. Todd. "Automated discrimination of upland and wetland using terrain derivatives." *Canadian Journal of Remote Sensing* 33, no. S1 (2007): S68-S83.

Hogg, A. R., and J. Holland. "An evaluation of DEMs derived from LiDAR and photogrammetry for wetland mapping." *The Forestry Chronicle* 84, no. 6 (2008): 840-849.

Huang, Shengli, Claudia Young, Min Feng, Karl Heidemann, Matthew Cushing, David M. Mushet, and Shuguang Liu. "Demonstration of a conceptual model for using LiDAR to improve the estimation of floodwater mitigation potential of Prairie Pothole Region wetlands." *Journal of Hydrology* 405, no. 3 (2011): 417-426.

Illinois Department of Natural Resources (ILDNR). "Illinois Land Cover Summary Data by Counties." Illinois Department of Natural Resources. Accessed January 12, 2015. http://dnr.state.il.us/orep/ctap/map/counties.htm.

Illinois Height Modernization Program (IHMP), Illinois State Geological Survey (ISGS), and Illinois Department of Transportation (IDOT), 2002-2013, Illinois LiDAR county database: Illinois State Geological Survey, http://www.isgs.uiuc.edu/nsdihome/webdocs/ilhmp/data.html (accessed October 20, 2013).

Illinois Natural Resources Geospatial Data Clearinghouse (INRGD), Illinois State Geological Survey (ISGS), http://www.isgs.uiuc.edu/nsdihome/ (2011 Illinnois Department of Transportation Orthophotography; accessed October 20, 2013).

Illinois State Water Survey. "Palmer Z Index for Illinois, Climate Division #5 from 2000 to 2014." http://mrcc.isws.illinois.edu/state_climatologists/illinois/clidiv/il_cli_div2.jsp (accessed April 29, 2014).

Isenburg, M., Liu, Y., Shewchuk, J., Snoeyink, J., Thirion, T., *Generating Raster DEM from Mass Points via TIN Streaming*, GIScience'06 Conference Proceedings, pages 186-198, September 2006.

Jensen, John R. *Introductory Digital Image Processing: A Remote Sensing Perspective*. No. Ed. 2. Prentice-Hall Inc., 1996.

Kheir, R. Bou, M. H. Greve, P. K. Bocher, and M. B. Greve. "Use of Digital Terrain Analysis and Classification Trees for Predictive Mapping of Soil Organic Carbon in Southern Denmark." *Proceedings of Geomorphometry* (2009)

Kienzle, Stefan W. "The Effect of Grid Cell Size on Major Terrain Derivatives." In*Proceedings of the Twenty-Fourth Annual ESRI User Conference*. 2004.

Klemas, Victor. "Remote sensing of wetlands: case studies comparing practical techniques." *Journal of Coastal Research* 27, no. 3 (2011): 418-427.

Lang, Megan, Jane Awl, Bill Wilen, Greg McCarty, and John Galbraith. "Improved Wetland Mapping." *National Wetlands Newsletter*, Vol. 31, No. 5 (2009)

Lang, Megan W., and Eric S. Kasischke. "Using C-band synthetic aperture radar data to monitor forested wetland hydrology in Maryland's coastal plain,

USA." *Geoscience and Remote Sensing, IEEE Transactions on* 46, no. 2 (2008): 535-546.

Lang, Megan W., and Greg W. McCarty. "Lidar intensity for improved detection of inundation below the forest canopy." *Wetlands* 29, no. 4 (2009): 1166-1178.

Lang, Megan, Greg McCarty, Robert Oesterling, and In-Young Yeo. "Topographic Metrics for Improved Mapping of Forested Wetlands." *Wetlands* 33, no. 1 (2013): 141-155.

Lawrence, Rick L., and Andrea Wright. "Rule-based classification systems using classification and regression tree (CART) analysis." PE & RS- *Photogrammetric Engineering & Remote Sensing* 67, no. 10 (2001): 1137-1142.

Liaw, Andy, and Matthew Wiener. "Classification and Regression by randomForest." *R News* 2, no. 3 (2002): 18-22.

Lillesand, Thomas M., Ralph W. Kiefer, and Jonathan W. Chipman. Remote sensing and image interpretation. No. Ed. 5. John Wiley & Sons Ltd, 2004.

Lu, Dengsheng, and Qihao Weng. "A survey of image classification methods and techniques for improving classification performance." *International journal of Remote sensing* 28, no. 5 (2007): 823-870.

Lucas, Richard, Alex Lee, John Armston, Johanna Breyer, Peter Bunting, and João Carreiras. "Advances in forest characterisation, mapping and monitoring through integration of LiDAR and other remote sensing datasets." *SilviLaser* 4, no. 1 (2008): 2-12.

Ma, Jianchao, Guangfa Lin, Junming Chen, and Liping Yang. "An improved Topographic Wetness Index considering topographic position." In Geoinformatics, 2010 18th International Conference on, pp. 1-4. IEEE, 2010.

MacAlister, Charlotte, and Manithaphone Mahaxay. "Mapping wetlands in the Lower Mekong Basin for wetland resource and conservation management using Landsat ETM images and field survey data." *Journal of Environmental Management* 90, no. 7 (2009): 2130-2137.

MacKinnon, F. "Wetland application of LIDAR data: analysis of vegetation types and heights in wetlands." (2001).

Maxa, Melissa, and Paul Bolstad. "Mapping northern wetlands with high resolution satellite images and LiDAR." *Wetlands* 29, no. 1 (2009): 248-260.

Maxwell, Aaron E., Timothy A. Warner, Michael P. Strager, and Mahesh Pal. "Combining RapidEye Satellite Imagery and LiDAR for Mapping of Mining and Mine Reclamation." *Photogrammetric Engineering & Remote Sensing* 80, no. 2 (2014): 179-189.

McFeeters, S. K. (1996). The use of Normalized Difference Water Index (NDWI) in the delineation of open water features, International Journal of Remote Sensing, 17(7):1425–1432.

Murphy, Paul NC, Jae Ogilvie, Kevin Connor, and Paul A. Arp. "Mapping wetlands: a comparison of two different approaches for New Brunswick, Canada." *Wetlands* 27, no. 4 (2007): 846-854.

Nielsen, Eric M., Stephen D. Prince, and Gregory T. Koeln. "Wetland change mapping for the US mid-Atlantic region using an outlier detection technique. " *Remote Sensing of Environment* 112, no. 11 (2008): 4061-4074.

O'Hara, C. G. "Remote sensing and geospatial application for wetland mapping, assessment, and mitigation." In Integrating Remote Sensing at the Global, Regional and Local Scale. Pecora 15/Land Satellite Information IV Conference. 2002.

Omernik, James M., Charles L. Pederson, and Brian C. Moran. *Level III and IV ecoregions of Illinois*. US Environmental Protection Agency, National Health and Environmental Effects Research Laboratory, 2006.

Pantaleoni, E., R. H. Wynne, J. M. Galbraith, and J. B. Campbell. "A logit model for predicting wetland location using ASTER and GIS." *International Journal of Remote Sensing* 30, no. 9 (2009): 2215-2236.

Pflugmacher, Dirk, Olga N. Krankina, and Warren B. Cohen. "Satellite-based peatland mapping: Potential of the MODIS sensor." *Global and Planetary Change* 56, no. 3 (2007): 248-257.

Pebesma, E.J., R.S. Bivand, 2005. Classes and methods for spatial data in R. R News 5 (2), http://cran.r-project.org/doc/Rnews/.

Poulin, Brigitte, Aurélie Davranche, and Gaëtan Lefebvre. "Ecological assessment of Phragmites australis wetlands using multi-season SPOT-5 scenes." *Remote Sensing of Environment* 114, no. 7 (2010): 1602-1609.

Remmel, Tarmo K., Kenton W. Todd, and James Buttle. "A comparison of existing surficial hydrological data layers in a low-relief forested Ontario landscape with

those derived from a LiDAR DEM." *The Forestry Chronicle* 84, no. 6 (2008): 850-865.

R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Sasaki, Takeshi, Junichi Imanishi, Keiko Ioki, Yukihiro Morimoto, and Katsunori Kitada. "Object-based classification of land cover and tree species by integrating airborne LiDAR and high spatial resolution imagery data."*Landscape and Ecological Engineering* 8, no. 2 (2012): 157-171.

Shaeffer, David L. ìCharacterizing Jurisdictional Wetlands Using Aerial LiDAR.î Master Thesis, East Carolina University, 2008.

Schmidt, K. S., and A. K. Skidmore. "Spectral discrimination of vegetation types in a coastal wetland." *Remote Sensing of Environment* 85, no. 1 (2003): 92-108.

Silva, Thiago SF, Maycira PF Costa, John M. Melack, and Evlyn MLM Novo. "Remote sensing of aquatic vegetation: theory and applications." *Environmental Monitoring and Assessment* 140, no. 1-3 (2008): 131-145.

Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture, Soil Survey Geographic (SSURGO) Database for Vermilion County. Available online at: http://soildatamart.nrcs.usda.gov. (accessed October 20, 2013)

Song, Jeong-Heon, Soo-Hee Han, K. Y. Yu, and Yong-Il Kim. "Assessing the possibility of land-cover classification using lidar intensity data." *International Archives of*

Photogrammetry Remote Sensing and Spatial Information Sciences 34, no. 3/B (2002): 259-262.

Sörensen, Rasmus, Ursula Zinko, and Jan Seibert. "On the calculation of the topographic wetness index: evaluation of different methods based on field observations." *Hydrology and Earth System Sciences Discussions* 10, no. 1 (2006): 101-112.

Starek, M., B. Luzum, R. Kumar, and K. C. Slatton. "Normalizing Lidar intensities." *Geosensing Engineering and Mapping (GEM)* (2006).

Stevens, Christopher W., and Stephen A. Wolfe. "High-Resolution Mapping of Wet Terrain within Discontinuous Permafrost using LiDAR Intensity."*Permafrost and Periglacial Processes* 23, no. 4 (2012): 334-341.

Tiner, Ralph W. "Lists of potential hydrophytes for the United States: A regional review and their use in wetland identification." *Wetlands* 26, no. 2 (2006): 624-634.

Töyrä, Jessika, and Alain Pietroniro. "Towards operational monitoring of a northern wetland using geomatics-based techniques." *Remote Sensing of Environment* 97, no. 2 (2005): 174-191.

Töyrä, Jessika, Alain Pietroniro, Lawrence W. Martz, and Terry D. Prowse. "A multi-sensor approach to wetland flood monitoring." *Hydrological Processes* 16, no. 8 (2002): 1569-1581.

U.S. Fish and Wildlife Service (USFWS). 2002. National wetlands inventory: A strategy for the 21st century. US Department of the Interior, Fish and Wildlife Service, Washington, DC, USA.

U. S. Fish and Wildlife Service (USFWS). Publication. National Wetlands Inventory website. U.S. Department of the Interior, Fish and Wildlife Service, Washington, D.C. http://www.fws.gov/wetlands/ (accessed October 20, 2013)

U. S. Geological Survey, Gap Analysis Program (GAP). November 2013. Protected Areas Database of the United States (PADUS), version 1.3

Viera, Anthony J., and Joanne M. Garrett. "Understanding interobserver agreement: the kappa statistic." *Family Medicine* 37, no. 5 (2005): 360-363.

Wagner, Wolfgang, Andreas Ullrich, Vesna Ducic, Thomas Melzer, and Nick Studnicka. "Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner." *ISPRS Journal of Photogrammetry and Remote Sensing* 60, no. 2 (2006): 100-112.

Wehr, Aloysius, and Uwe Lohr. "Airborne laser scanning—an introduction and overview." *ISPRS Journal of Photogrammetry and Remote Sensing* 54, no. 2 (1999): 68-82.

Wilen, Bill O., and R. W. Tiner. "National Wetlands Inventory- The First Ten Years." In Wetlands: Concerns and Successes. Proceedings of a Symposium held September 17-22 1989, Tampa, Florida. American Water Resources Association, Bethesda, Maryland. 1989. p 1-12, 2 fig, 11 ref, append. 1989.

Wright, Chris, and Alisa Gallant. "Improved wetland remote sensing in Yellowstone National Park using classification trees to combine TM imagery and ancillary environmental data." *Remote Sensing of Environment* 107, no. 4 (2007): 582-605.

Xie, Zhixiao, Zhongwei Liu, John W. Jones, Aaron L. Higer, and Pamela A. Telis. "Landscape unit based digital elevation model development for the freshwater wetlands within the Arthur C. Marshall Loxahatchee National Wildlife Refuge, Southeastern Florida." *Applied Geography* 31, no. 2 (2011): 401-412.

Zedler, Joy B., and Suzanne Kercher. "Wetland resources: status, trends, ecosystem services, and restorability." *Annual Review of Environmental and Resources* 30 (2005): 39-74.

Zhang, Jixian. "Multi-source remote sensing data fusion: status and trends."*International Journal of Image and Data Fusion* 1, no. 1 (2010): 5-24.

Zhou, Huiping, Hong Jiang, Guomo Zhou, Xiaodong Song, Shuquan Yu, Jie Chang, Shirong Liu, Zishan Jiang, and Bo Jiang. "Monitoring the change of urban wetland using high spatial resolution remote sensing data." *International Journal of Remote Sensing* 31, no. 7 (2010): 1717-1731.

VITA

Graduate School
Southern Illinois University

Ashley Elizabeth Suiter

ashleysuiter87@gmail.com

Central Michigan University
Bachelor of Science, Geography, 2011

Special Honors and Awards:
Dr. Benedykt Dziegielewski Scholarship Award
Graduate and Professional Student Council Research Award

Thesis Title:
REMOTE SENSING BASED DETECTION OF FORESTED WETLANDS: AN EVALUATION OF LIDAR, AERIAL IMAGERY, AND THEIR DATA FUSION

Major Professor: Dr. Guangxing Wang

Publications:
Knoche, S., Lupi, F., & Suiter, A. (2015) Harvesting Benefits from Habitat Restoration: Influence of Landscape Position on Economic Benefits to Pheasant Hunters. *Ecological Economics* Vol. 13, 97-105