PRECISION AGRICULTURE AND GIS:

EVALUATING THE USE OF YIELD MAPS COMBINED WITH LIDAR DATA

by

Bennett Charles Gaumitz

A Thesis Presented to the
FACULTY OF THE USC GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
MASTER OF SCIENCE
(GEOGRAPHIC INFORMATION SCIENCE AND TECHNOLOGY)

August 2016

DEDICATION

I dedicate this document to my parents without whose unwavering support and faith this

document could never have been completed.

ACKNOWLEDGMENTS

I am grateful to my mentor Dr. Richard Vanden Heuvel, who inspired me to create the topic that

is presented here. I also want to thank my advisor Dr. John Wilson who stayed committed to

helping me throughout my work on my thesis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

DEM   Digital Elevation Model

DSM   Digital Soil Map

GIS   Geographic Information System

IDW   Inverse Distance Weighted interpolation method

LiDAR   Light Detection and Ranging

MZA   Management Zone Analyst

NRCS   Natural Resources Conservation Service

TIN   Triangular Irregular Network

USDA   U.S. Department of Agriculture

USGS   U.S. Geological Survey

ABSTRACT

Precision agriculture in practice utilizes GIS far less effectively than it should. My work at a soil consulting company has shown that part of the problem is that the literature does not show an effective way of analyzing soil through GIS that is both scientific and able to be used by those associated with agriculture. The thesis aimed to answer two questions: (1) Are there significant differences between a multiple year composite yield and a single year and if so, are these significant enough to have an impact on normal operations? (2) Are the areas where such differences occur related to slope? Six fields were used for this study: Dob Along G, Emmert, Harstad, Merkt's, Pribnows, and Stefoneks located near New Richmond, Wisconsin. Three years of yield data were used for each field and slope data was created using LiDAR from St. Croix County. These yield data were interpolated using standard industry practices. A single year was compared to composite years to determine what differences, if any, exist between them. Each year and the composite had actual and predicted values compared using the difference of means statistical test to validate the success of the interpolation. A DEM was created from LiDAR and this was used to create a slope map of each field. This slope map was used to divide the yield points by five slope classes: 0 -$1°$, 1-$2°$, 2-$3°$, 3-$5°$, and $>5°$. The mean yield and variation was then compared for each class to determine any patterns associated with slope values. The results show that where there is significant variation between single years of yield data the composite will fail. The difference between the composite and a single year is useful in identifying which fields are causing the composite to fail and eliminating them. Slope did not consistently correlate to changes in yield or variation. Dob Along G, Emmert, and Harstad showed no correlation, while Merkt's, Pribnows, and Stefoneks showed decreasing yield and increasing variation as slope increased.

CHAPTER 1: INTRODUCTION

Precision Agriculture is a practice in agricultural that examines and manages a section of farm rather than its whole. It uses various tools, especially rigorous soil testing, to provide a detailed understanding of the soil on a single farm or a section of that farm. This allows for variable-rate fertilizer recommendations, based on soil testing, to improve particular areas while using the least amount of fertilizer. The use of a Geographic Information System (GIS) within this process is under-utilized. Primarily it's use is reactive, as a final test to understand the effects of the previous year's fertilizer recommendation. The yield data taken from farmers is made into a map that demonstrates the overall performance of a specific farm. If a farm is under-performing this indicates a need to revise future recommendations. However, since yield maps are not used to make recommendations, these are generally just stored for possible later use.

1.1 Digital Soil Maps

There has been some effort to create digital soil maps (DSMs) that combine attributes like soil type, weather pattern, and chemical composition in a GIS to demonstrate the overall fertility of an agricultural region or to predict its production. The most notable example is the regional soil map created by the Natural Resources Conservation Service (NRCS) that describes the general soil types within a specific area, e.g. the Iowa Regions Map in Figure 1.

There has been substantial research put into the creation of predictive DSMs that have established methodologies and data needs for their creation (Zhu et al. 2001; McBratney, Santos, and Minasny 2003). Models like Scorpan (McBratney, Santos, and Minasny 2003) use a variety of attributes from numerous sources to provide a detailed and complete regional DSM and complicated techniques to ensure the accuracy of the final output. The NRCS uses traditional techniques to construct a similar map that relies on expert knowledge and surveying methods.

Iowa Soil Regions

Legend

Soil Regions

1 Loess over glacial till - low rainfall
2 Northwest Iowa Loess
3 Tazewell Glacial Till
4 Loamy Wisconsin Glacial Till
5 Clayey Lake Deposits
6 Iowan Erosion Surface
7 Shallow to Bedrock
8 Loess with Bedrock Outcrops
9 Shallow Loess over Glacial Till
10 Loess Ridges and Sideslopes
11 Loess with exposures of Glacial Till
12 Very Deep Loess
13 Missouri River Bluffs
14 Missouri River Alluvium
15 Loess Ridges/Glacial Till - SW Iowa
16 Loess, Shale, and Glacial Till
17 Loess Ridges/Glacial Till - SE Iowa
18 Eolian Sand
19 Loess - Timbered
20 Alluvium
21 Loess Ridges/Glacial Till Sideslopes
22 Loess Ridges/Clay Paleosol
23 Water

USDA
NRCS
United States Department of Agriculture
Natural Resources Conservation Service

Miles
0   12.5   25   50   75

Figure 1: Current NRCS map of soil types in Iowa
(http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs142p2_006361.pdf)

Similar methods have been used in the creation of digital methods by capturing and using expert knowledge (Zhu et al. 2001). These approaches are used to create a comprehensive map at the regional level that is ineffective in modern precision agriculture which uses information at the sub-farm level to make recommendations.

Comprehensive regional DSMs fall short of this goal in two ways: (1) the detail is intended for a region, so it is not precise enough to be used for even a single farm, let alone a section of that farm; and (2) the attributes used in the creation of these maps are too numerous, creating extraneous information that can make such a map confusing or cumbersome for direct precision application.

1.2 Management Zones

In order to change this, there have been numerous attempts to create functional management zones for farmers (Fleming et al. 2000; Fleming, Heermann, and Westfall 2004; Fridgen et al. 2004). These maps are designed to delineate zones based on soil attributes to better organize a farmer's field for fertilizer application. Figure 2 offers an example of a three-zone management map based on imagery, electrical conductivity, and yield (Fridgen et al. 2004). Figure 3 shows an additional example of management zones based on phosphorous levels.
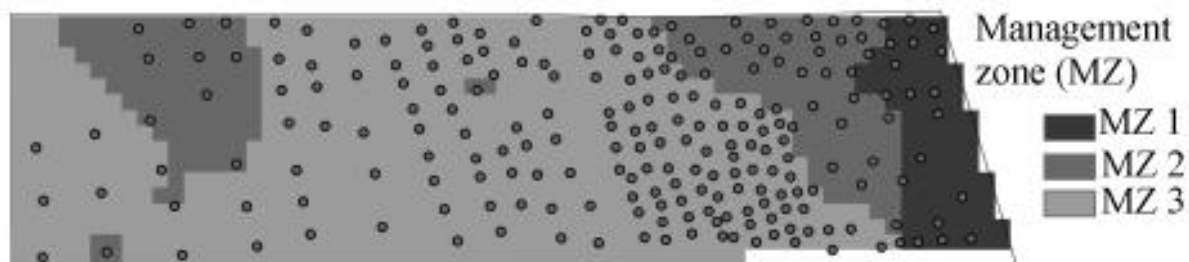
Figure 2: Example of fertilizer management zones based on imagery, electrical conductivity, and yield. *Source:* **Fridgen et al. (2004).**

Some of these authors have expressed concerns about the use of these kinds of maps because of changes that can occur due to topographic effects and climate changes that can affect the productivity of a field over time (Schepers et al. 2004). However, these techniques are not generally used in practice, due to the difficulty of obtaining and updating the attributes used to create these zones.

MGT. ZONES with SURFACE GRIDS

Phosphorous Surface

Statistical Summary

		Avg.	Min.	Max.	CoffVar
■	Whole Zone 1	**55**	**29**	**140**	**34**
	Clump #1	41	40	43	2
	Clump #2	70	43	140	31
	Clump #3	48	31	85	22
	Clump #4	44	29	66	20
▨	Whole Zone 2	**46**	**27**	**80**	**16**
	Clump #5	47	29	80	15
	Clump #6	37	27	50	17
□	Whole Zone 3	**42**	**22**	**74**	**20**
	Clump #7	44	22	74	20
	Clump #8	39	31	47	10
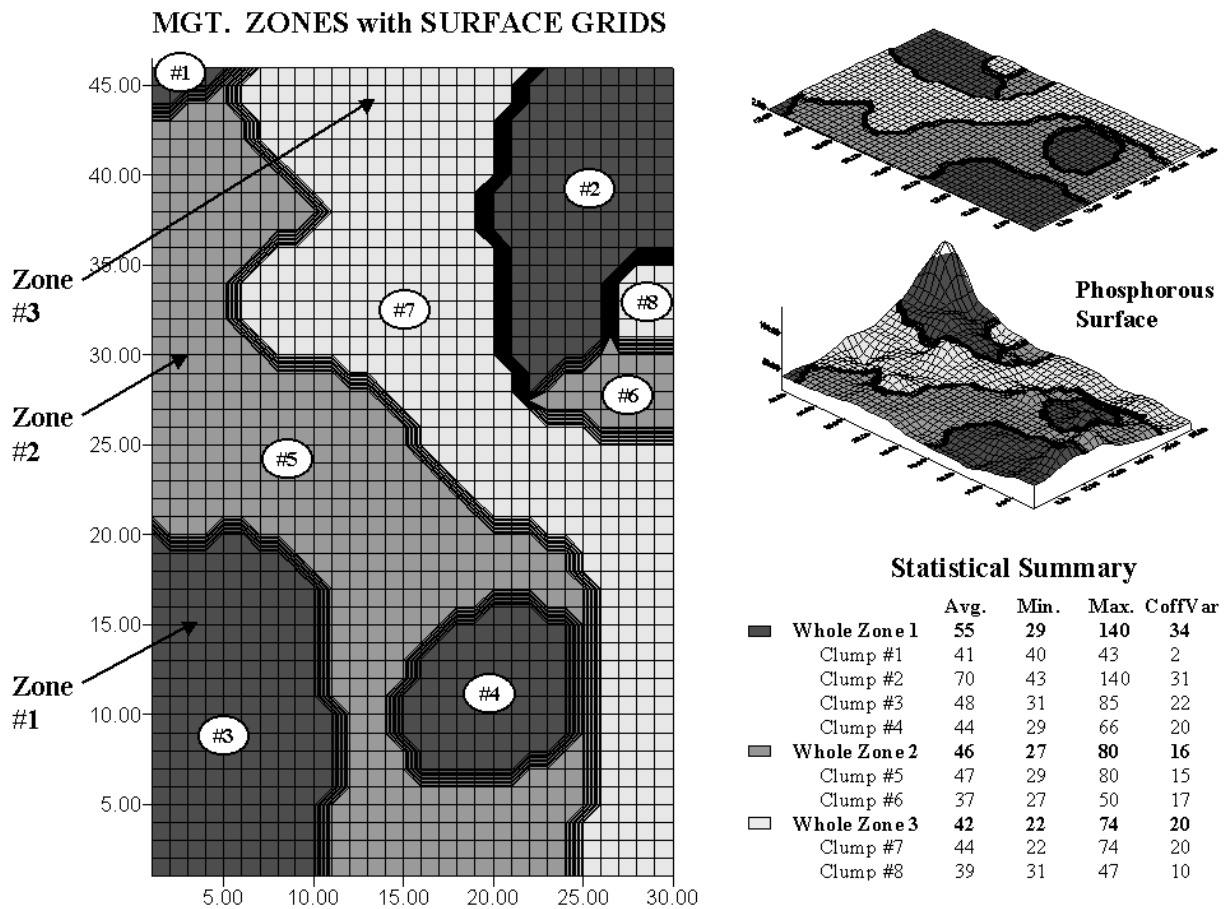
Figure 3: Map illustrating the creation of management zones based on phosphorous levels. Eight areas representing differences in phosphorous levels are simplified into three zones (http://www.innovativegis.com/basis/pfprimer/Appendix_A/Appendix_A.htm)

1.3 Motivation

Consultants and farmers have direct access to critical details for the creation of management zones from soil sampling. This means that a predictive map does not have to show all of the soil

information, but only requires the attributes necessary to contribute to the soil management process. Rather than creating something new that can already be extrapolated from soil sampling, additional attributes can be used to supplement these data. These attributes must correlate with necessary information and provide enough detail to develop a predictive map. Researchers have documented DEMs and imagery related to information like chemical composition and yield (e.g. Varvel, Schlemmer, and Schepers, 2004; Dobos et al. 2000).

My research aims to utilize techniques related to yield maps and Light Detection and Ranging (LiDAR) to explore their potential for use in precision agriculture. The emphasis will be on the accuracy of single- and multiple-year yield maps and if the slope generated from LiDAR can explain the differences between these maps. The underlying assumption is that in some cases a single year yield map is sufficiently accurate to characterize the yield of a field and that slope is an indicator of where single year maps fail due to higher variability from year to year and less accuracy overall.

1.4 Thesis Organization

The remainder of this thesis consists of four chapters. The second chapter explores background information on precision agriculture, DSMs, yield maps, and LiDAR data in an effort to explain the importance of these supplemental data and their use. The third chapter describes the techniques that were used to demonstrate the effectiveness of the yield maps and LiDAR data for documenting the fertility needs on the study farm fields. The fourth chapter presents the results obtained with this methodology and compares the results with individual years of yield data. The fifth and final chapter notes the conclusions that can be drawn from these results and discusses their implications for precision agriculture.

CHAPTER 2: BACKGROUND AND LITERATURE REVIEW

Digital soil mapping in general and precision agriculture in particular have been studied for more than 50 years. A variety of approaches have been proposed for transferring traditional soil mapping to digital methods. Traditional soil mapping methods were created to form a general overview of regions and so were inherently generalized and therefore inappropriate for use in precision agriculture, which has begun to dominate the agricultural industry.

This change in agricultural methods requires new mapping designs that are more specific and precise in character. Traditional soil maps use discrete areas to mark general soil zones, yet in reality this is not the case. Methods, such as kriging, have been used to change these into continuous classes by allowing the general soil lines to become fuzzy (McBratney, De Gruijter, and Brus 1992). This process also allows for some prediction of classes as these intermediary soil types bridge the gap between the general soil types already known. Similarly, Bragato (2004) has proposed using fuzzy modeling techniques on traditional soil maps to produce soil zones that are not separated by simple lines but fade into one another. This characterization of the soil more closely resembles the real-world and bridges the gap between the traditional and the digital.

The utilization of other resources that relate to attributes characterizing soil can enhance the quality of smaller scale maps. Satellite information was found in some ways to eliminate the need for traditional soil laboratory work to characterize the general composition of soil regionally (Nanni and Demattê 2006). Dobos et al. (2000) utilized a DEM to enhance regional satellite information because satellite data alone have too few details to characterize soil type. These techniques are a good start to moving from traditional to digital methods, but precision agriculture requires far more detail.

The use of management zones offers a new way of looking at a digital soil map. These zones are created within a single farm and are characterized by a large amount of testing that allows for different management techniques (fertilizer, irrigation, etc.) depending on the needs of a specific zone. Some approaches have used a variety of attributes obtained through digital and physical means, including soil brightness, electoral conductivity and yield, to delineate these zones (Schepers et al. 2004). These kinds of approaches require changes to zone management depending on season and climatic variation.

This chapter covers the related information and literature for this thesis. Section 2.1 will give an overview of precision agriculture, Section 2.2 will detail digital soil maps including attributes used in their creation and their use in practice, and Section 2.3 will detail the use of Light Detection and Ranging (LiDAR) in creating DEMs. These sections will highlight precision agriculture as the subject of this thesis, overview the process of creating DSMs as a way of developing which attributes would work with this thesis, and discuss the application of LiDAR to show how it might be used to create some of these attributes.

2.1 Precision Agriculture

Precision Agriculture is designed to increase the efficiency of a total farm by dividing that farm into sections and applying different fertilizer and management practices to ensure the best yield at the lowest cost. The idea is to create greater efficiency between fertilizer use and the possible benefits. This is especially important for managing the environmental impact of agriculture, limiting fertilizer to the least needed diminishes fertilizer run-off and the detrimental environmental effects associated with mining and applying fertilizer (Schieffer and Dillon 2013). Schieffer and Dillon (2013) have suggested using governmental policy in the form of taxes or credits to minimize fertilizer use. In order to maximize the benefit-to-cost ratio,

recommendations, analysis, and experimental fertilizer are used in combination to reduce cost while still allowing for the maximum growth of crops in a given area (McBratney et al. 2005).

Generally, this is accomplished by grid sampling a farm and using laboratory testing to determine the chemical composition of each grid cell individually. These grids can vary in size depending on the desired level of precision. The tests performed on each grid sample can vary but typically look at the levels of potassium, phosphorous, organic matter, zinc, and the acidity of the soil. Once these values are determined an agronomist makes recommendations based on these values for a whole field, management zone, or individual grid cell. This can depend on the size of the field, management practices, and fertilizer type. Finally, yield values are collected to determine the effectiveness of the recommendation and assess the quality of the farm after each season. This process is illustrated below in Figure 4.

```
┌─────────────────────────────┐        ┌─────────────────────────────┐
│ Soil sampling               │        │ Laboratory Testing          │
│ • Spring, summer, and fall  │ ────▶  │ • Chemical Composition      │
│   periods                   │        │ • Variable-rate             │
│ • Several years between     │        │   recommendations           │
│   samplings                 │        │                             │
└─────────────────────────────┘        └─────────────────────────────┘

┌─────────────────────────────┐        ┌─────────────────────────────┐
│ Fertilizer Application      │        │ Assessment                  │
│ • Pre-plant and post-plant  │ ────▶  │ • Grower Feedback           │
│ • Uniform or variable-rate  │        │ • Yield Analysis            │
│ • Based on farm divisions   │        │ • Plant Response            │
└─────────────────────────────┘        └─────────────────────────────┘
```
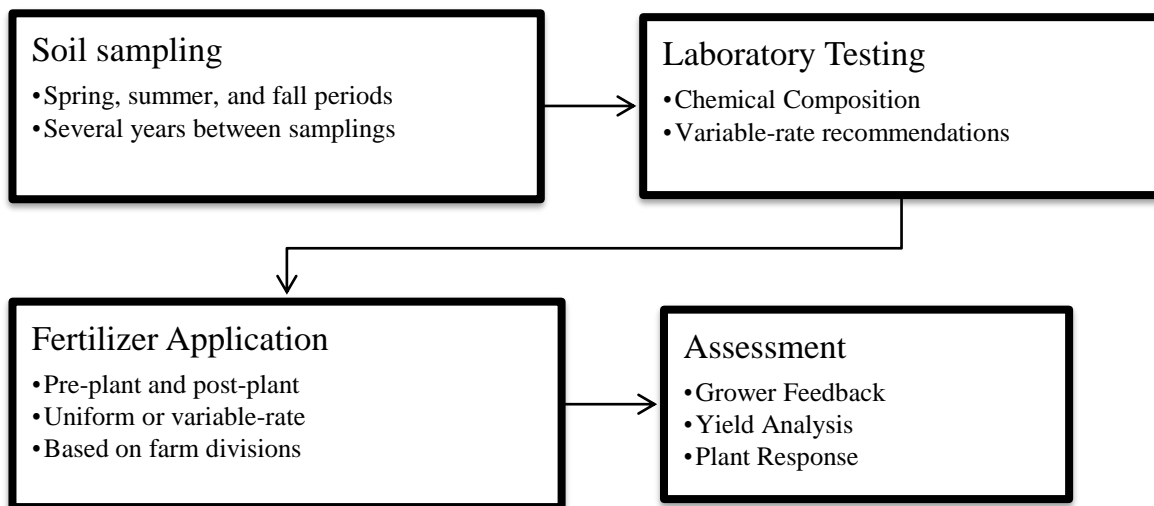
Figure 4: Example of precision agriculture process from beginning to end. As shown, this process is centered on fertilizer use.

2.1.1 Soil Sampling

Soil sampling is a major source for determining the type and quality of soil information. It is important that this information be accurate; however, over-sampling is cost prohibitive with limited practical benefits (Fortunati, Banff, and Pasturenzi 1994). The method of sample

collection can be based on a grid superimposed on the landscape, soil type, management zone, or farm unit.

Of these, gridded soil sampling is the most common as it requires the least amount of guess work and can be part of a streamlined sampling process, e.g. no expert knowledge is needed to acquire large numbers of samples. Sample locations can consist or random samples within a grid, a diamond patterned, or a systematic unaligned grid (Wollenhaupt and Wolkowski 1994). These techniques are illustrated in Figure 5. The diamond patterned methods shift the sampling points by connecting four grid sampling points in a diamond shape, while the systematic unaligned grid uses a random number table to generate the x, y coordinates for the sample location within each grid. Both of these methods allow for a more randomized approach to sampling that does not require composite samples. The grid approach offers the most variability but commonly accepted practice is a 2.5-acre square grid. Figure 6 shows an example of composite sampling based on such a grid size.
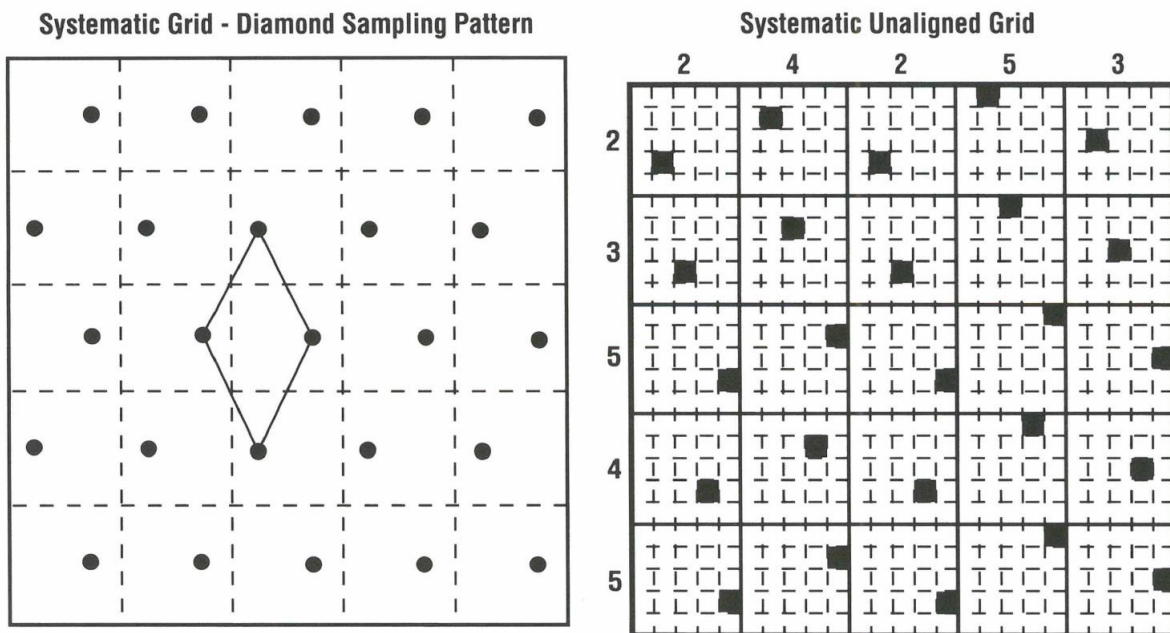
Figure 5: Examples showing the diamond pattern and systematic unaligned grid methods to sample soil. *Adapted from Wollenhaupt and Wolkowski (1994).*

The type of soil sampling methods used can greatly affect the results because, regardless of the number of samples taken, it still provides only a snapshot of the soil. However, this snapshot diminishes as the number of samples per unit area increases, so that smaller grid sizes give more specific results.
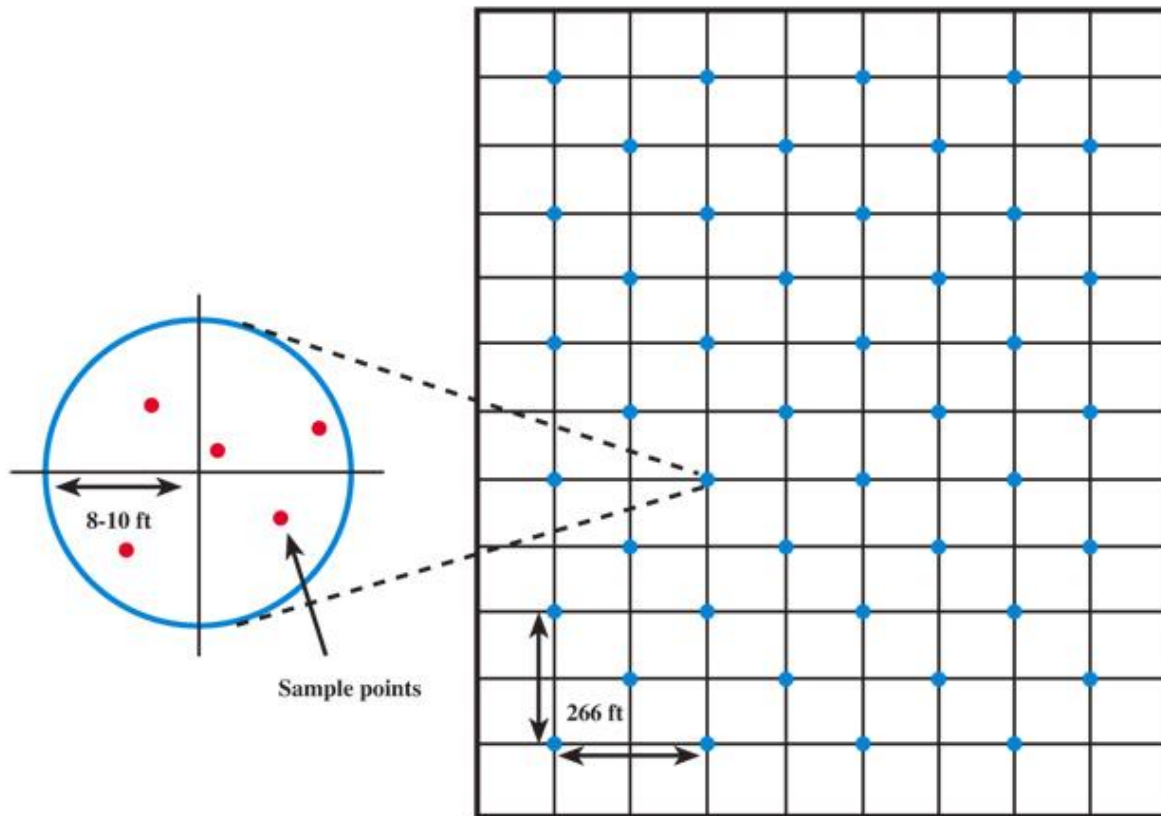
Figure 6: Example of soil sampling methods. This figure shows the use of multiple samples at each grid point to create a composite sample
(http://cropwatch.unl.edu/ssm/soilsampling)

The quality is also affected by when and where the sample is taken within the overall area. Fertilizer is applied in bands, to help reduce fertilizer use (Beyrer and Below 2014); if this is the case, the characteristics of samples taken within a band or outside of one will vary greatly. Finally, the time at which a sample is taken can greatly affect the result; for example, if a sample

is taken before any new fertilizer has been applied the results will be different than when it is taken after the fertilizer was applied.

2.1.2 Fertilizer Recommendations

Fertilizer recommendations are based on laboratory testing of the soil samples. These are performed or managed by an agronomist who interprets the results. Each sample receives its own recommendation based on the tests results. Different fertilizers are used to manage deficiencies in key composition values. For example, potassium chloride (potash) is used to add potassium to the soil. The most important of these values are potassium, phosphorous, and pH content; however, other values can be used for a more in-depth assessment, e.g. zinc, magnesium, nitrogen, or calcium.

These values are rated, e.g. very low to very high, which correlate with different levels of fertilizer use. These recommendations can be based on expert knowledge, scientific analysis, or both. An example of a recommendation based on chemical values is shown in Table 1. The values in the table show that differences in nutrient levels are interpreted as high, medium, or low, and that the recommendations are based on what is already added to a field, e.g. straw. This process is usually only performed once every several years, as the expense can be greater than the return if performed more frequently.

Table 1: Example of fertilizer recommendations based on phosphorous and potassium on rice farms in Indonesia

Nutrient	P (mg P_2O_5*100g^{-1})	K (mg K_2O*100g^{-1})	Fertilizer (kg*ha^{-1})	Sp-36	KCl	
					Without straw	Straw Added
Low	<20	<10		100	100	50
Medium	20-40	10-20		75	50	0
High	>40	>20		50	50	0

Adapted from Sulaeman et al. (2012)

Fertilizer rates can be applied as a single or variable rate. A variable rate takes the within-field variability into account rather than just applying a uniform fertilizer to the field. A variable rate program is not always used for all fertilizer applications. For example, nitrogen is generally applied at a uniform (single) rate, due to the limited availability of nitrogen tests.

2.1.3 Yield Assessment

Yield is usually assessed through the use of yield monitors. These monitors record the amount of crop taken at a specific point when a combine harvests the crop. Based on the mass flow, speed, moisture content and GPS position, yield is calculated for a specific location and is then used to generate a yield map for the whole field. These points are only precise to the width of the combine and its speed. A faster speed can cause incorrect readings to occur which is dependent on the type of combine that was used. This can be a powerful tool for assessing the results of fertilizer changes or to detect problem areas in a particular field; however, it does have some limitations.

The size of the harvester combine makes a difference in accuracy as a smaller, research combine covers a smaller area and is therefore more precise than a regular sized combine. The yield monitor itself requires routine calibration and monitoring by the combine driver to ensure the best results. Even in the most ideal circumstances error can be up to 10 percent of yield. If the necessary steps are not followed or the driver is inattentive the error generated can make the yield map useless. This is seen as large gaps in the yield data or inaccurate readings of yield. Since most yield monitors are used by farmers who are not technicians, the precision of yield monitoring can vary greatly depending on the field, harvester, and/or farmer.

The yield map is generated from the thousands of points captured with the yield monitor. The number of points varies greatly depending on the size of the field and harvester combine.

Generally, this is normalized as bushels per acre and three or more classes are given colors ranging from red to green to indicate bad or good yields, respectively in comparison to the overall yield strength of the field. Figure 7 shows an example of a yield map normalized in this way with six value ranges.
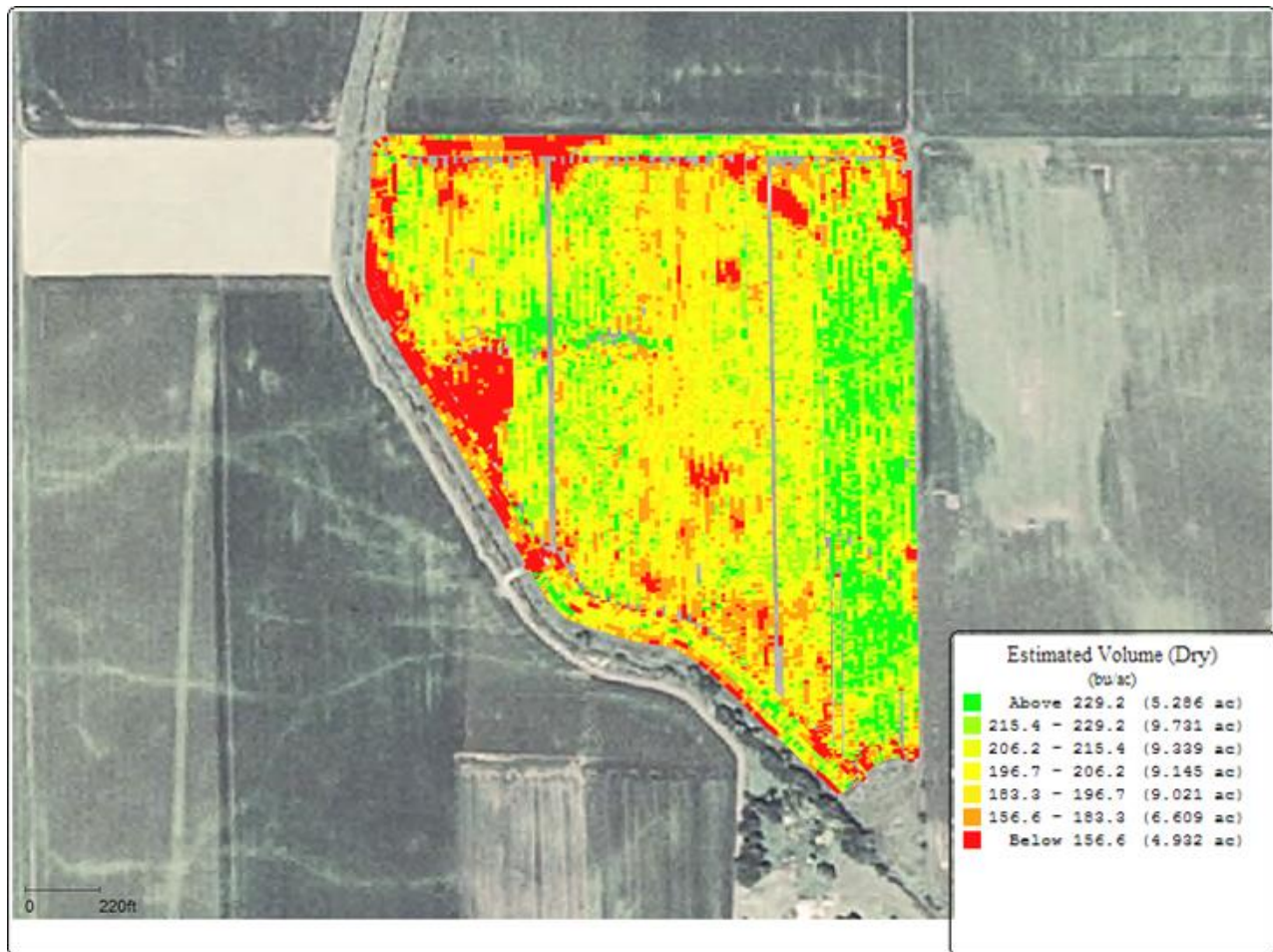
Figure 7: Example of a yield map showing six ranges of values that are normalized by bushels per acre (http://www.ptprecisionag.com/products__services)

2.1.4 Use of a Soil Survey

In practice, the soil survey is used as a way to provide a general overview of a farm by showing the types of soil located there. The soil survey that is typically used was created by the USDA in the 1970s. These maps show discrete classifications of soils by region. Each soil classification

13

comes with a description of its appearance, structure, and depth. These soil surveys were developed before precision agriculture became popular and, therefore, are insufficient for use with it. The survey can give a general indication of what soils can be found in the area but are not specific or precise enough to depict or differentiate the conditions found in a single field. This is why they are used only to provide a general overview and not included in variable-rate recommendations.

2.2 Digital Soil Maps

More comprehensive approaches than the soil survey have been proposed that take more factors into account to create complete digital soil maps. A prime example of this process is the Scorpan method that utilizes information on soil, climate, organisms, topography, age, and location to create a complete soil map of a desired region (McBratney, Mendonça Santos, and Minasny 2003). This approach considers a large number of attributes and uses complex geostatistical techniques to create a final map.

Another example is SoLIM which attempts to eliminate the use of traditional soil maps through the inference of soil types from the environmental factors that create them and the similarity of soils near one another (Zhu et al. 2001). Both Scorpan and SoLIM produce a regional soil map that is based solely on mathematical principles and digital information (e.g. Figure 8). However, these methods are very complicated and require many attributes to complete the map.

Fleming et al. (2000) describe an approach that is applicable to precision agriculture, specifically, variable rate fertilizer recommendations based on grid sampling. Their approach describes the use of aerial photographs, topography, and farmer knowledge to create site-specific management zones where different rates of fertilizer could be applied. These methods were

14

found to be effective in determining zones but could not entirely eliminate the grid sampling method for determining the fertilizer rate required to maximize production. This approach uses far fewer attributes but also illustrates that soil sampling is currently unavoidable.
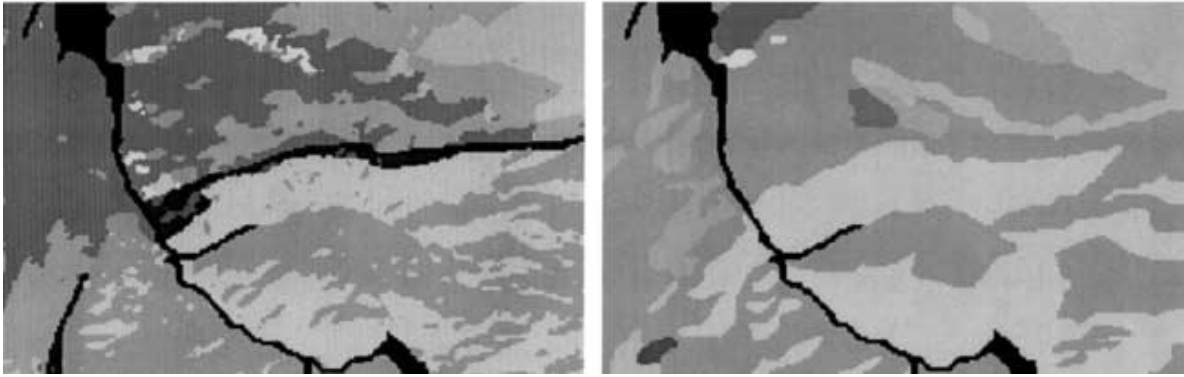
Figure 8: Comparison of (a) SoLIM-derived soil map; and (b) a conventional soil map in Lubrect, Montana. *Source:* **Zhu et al. (2001)**

Scull et al. (2003) reviewed predictive mapping techniques, including geostatistical, statistical, decision tree analysis, and expert systems. However, the authors noted that despite these advances, more work needs to be conducted to bridge the gap between GIS methods and their use in soil science**.** This observation illustrates that a variety of tools are available, but that they are not being fully utilized.

2.2.1 Attributes related to digital soil mapping

In order to understand how a digital map works, it is important to consider the attributes that form the components of that map. Fuzzy *K* or *c*-means clustering is a technique that allows for the more accurate depiction of soil classes. De Bruin and Stein (1998), for example, used this approach to describe the soil landscape by using attributes derived from a DEM, including slope, aspect, and curvature.

The DEM itself has considerations that need to be addressed to ensure its effective use. Thompson, Bell, and Butler (2001) considered different levels of information and combinations from other sources. They found that important information was lost as the resolution was moved from 10 to 30 m and that more information is lost as different sources of elevation data are used together. Smith et al. (2006) also explored the effects of neighborhood size and resolution on the accuracy of the soil map, and found that a more accurate DEM does not necessarily produce more accurate soil maps because the quality of the soil map will also depend on the type of landscape (flat or steep) and a correctly determined neighborhood size.

Yield can play an important role in determining the quality of soil. High yield indicates that the soil is producing the optimal number of bushels. Kasper et al. (2003) showed that yield is far more effective when combined from multiple years of yield data, due to climate variability. This was especially helpful in dry years, but may not be as helpful in wet years.

2.2.2 Digital soil maps in practice

There are relatively few examples of potential map creation strategies and their use in agricultural practices. Behrens and Scholten (2006) show how DSM techniques have been applied to soil maps in Germany, including the Scorpan process discussed above. They demonstrate that it is easier to apply these techniques in smaller countries and that an effective regional map can be created, as shown in Figure 9.

Precision agriculture has seen most of its practical use in the creation of management zones. For example, Mann, Schumann, and Obreza (2011) delineated citrus productivity zones. While this does not relate directly to the crops in this project it shows that zones can be created by looking at related attributes, i.e. yield, canopy volumes, and soil color. Most importantly, Fridgen et al. (2004) have created a software platform called the Management Zone Analyst

Digital soil map "L5131"

■	Haplic Cambisol
■	Humic Cambisol
■	Podsolic Cambisol
■	Gleyic Chernosem
■	Haplic Chernozem
■	Luvic Chernozem
■	Haplic Fluvisol
■	Stagnic Fluvisol
■	Haplic Gleysol
■	Mollic Gleysol
■	Stagnic Gleysol
■	Eutric Leptosol
■	Mollic Leptosol
■	Rendzic Leptosol
■	Haplic Luvisol
■	Stagnic Luvisol
■	Haplic Phaeozem
■	Haplic Planosol
■	Haplic Podsol
■	Calcaric Regosol
■	Cambic Regosol
■	Calcaric Vertisol
■	Haplic Vertisol

5   0   5   10 Kilometers

Figure 9: Example of a regional DSM as used in Germany. Source: Behrens and Scholten (2006)

(MZA) that automatically creates zones for agricultural fields. It uses a model that combines electrical conductivity, slope, and elevation with the fuzzy *c*-means clustering algorithm (Figure 10). This that it is possible to define management zones that can be created to manage specific areas of a farm. However, methods of determining electrical conductivity are expensive, time-consuming, and not fully developed.

Clustering Variables Used as Input for MZA
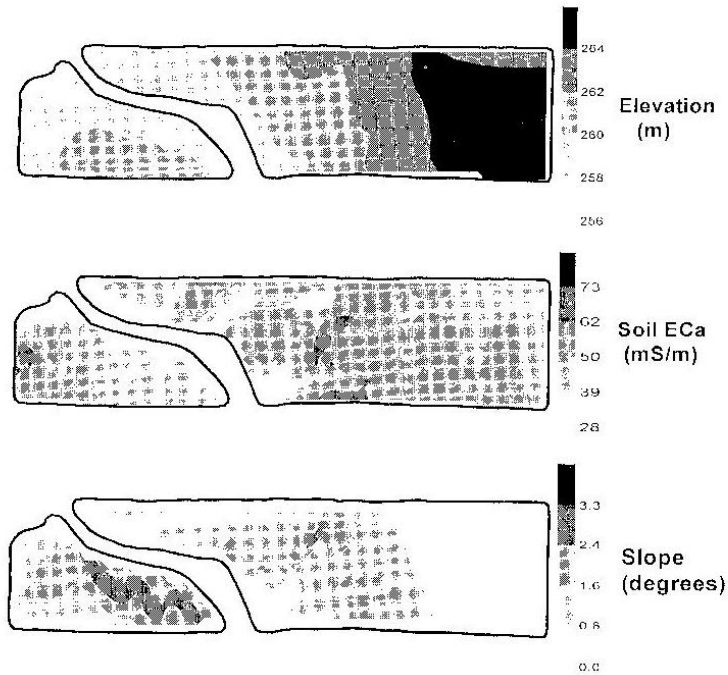
Management Zones by MZA Clustering
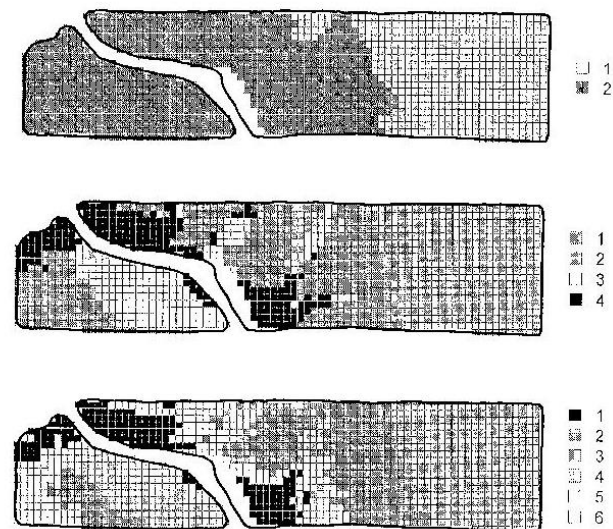
Figure 10: Demonstration of Management Zone Analyst software that shows the inputs (top) and examples of 2, 4, and 6 created zones from these inputs. *Source:* **Fridgen et al. (2004)**

2.3 Light Detection and Ranging (LiDAR)

LiDAR is an active remote sensing application that sends laser pulses from a satellite or aircraft and receives that light back giving a range based on the time it takes the pulse to return. This can be done either by profiling or scanning the ground. Profiling sends just a single pulse in one direction at a time, whereas scanning sweeps the laser back and forth to cover large amounts of area. Pulses can be sent very quickly, up to 100,000 per second in some cases, generating large datasets (Hopkinson and Chasmer 2008). These datasets are commonly used to create high resolution digital elevation models (DEMs).

LiDAR point clouds refer to the raw points that represent the ranges from each pulse. Visually these points are not very useful because of the large number of points over a relatively small area. This is why these points are used to create a DEM that visually depicts differences or changes in elevation for a chosen area. Esri (2013) presents a guide to do this that consists of converting the LiDAR points into a multipoint and finally into a terrain dataset. The accuracy of the dataset is dependent on the distance between the LiDAR points and the choices made in its creation. This process is further explained in Chapter 3.

LiDAR derived DEMs have been shown to be more accurate than those derived from topographic maps, even those provided by the U.S. Geological Survey (USGS) (Shi et al. 2012). A comparison of the slope gradient produced by Shi et al. (2012) of a 1-m and a 5-m LiDAR compared with a 10 m USGS DEM is shown below in Figure 11. Although the limited availability of LiDAR can make it impossible or expensive to use in certain regions, it is becoming more widely available and a better alternative to traditional DEMs.

This chapter demonstrated the use and benefits of precision agriculture; however, digital soil maps were shown to be ineffective if used with it. The use of yield in DSMs demonstrate

their usefulness in predicting the output of a farm. LiDAR was shown to be a possibility for

creating a sufficiently accurate DEM for precision agriculture. Chapter 3 uses yield interpolation

techniques to analyze the effectiveness of composites (combinations of multiple years of data) to

predict past yield and examine the process of creating a DEM from LiDAR to assess its

applicability in precision agriculture. Finally, a derivative slope map was created from the above

DEM and analyzed in relation to yield to determine if a relationship exists.
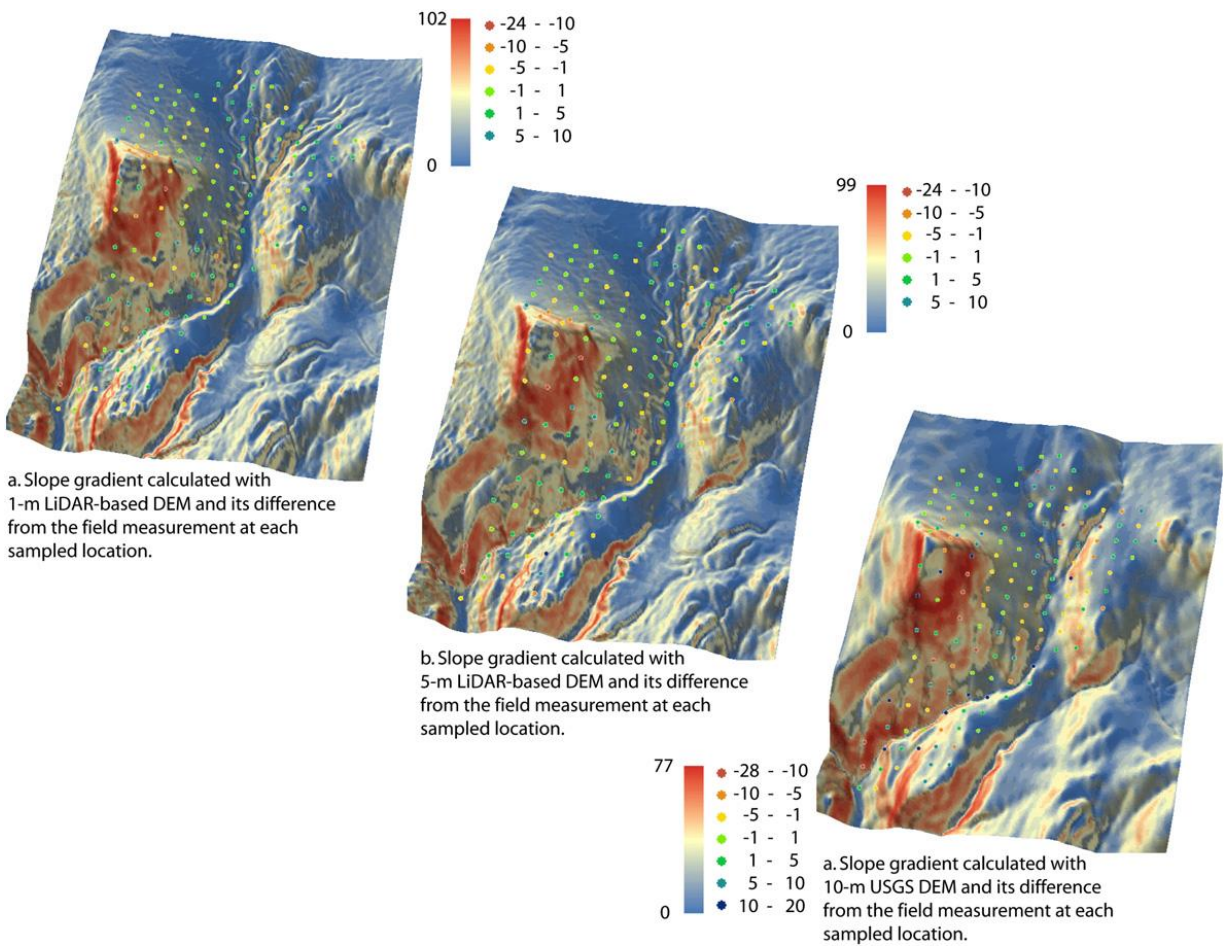
**Figure 11: Comparison of the slope gradient calculated with (a) 1-m LiDAR derived DEM;
b) 5-m LiDAR derived DEM and (c) 10-m USGS derived DEM. Source: Shi et al. (2012)**

CHAPTER 3: METHODOLOGY

The data used in this thesis are based in a rural section of St. Croix County, Wisconsin. It is located to the south of New Richmond, west of Hudson, and east of Glenwood City. Figure 12 illustrates the location of the farms used in this study. It follows County Road G and is mostly characterized as hilly with some flat areas. Figure 13 shows an aerial view of the terrain. From the photograph it is easy to see the variation in terrain with the numerous lakes, rivers, and vegetation.
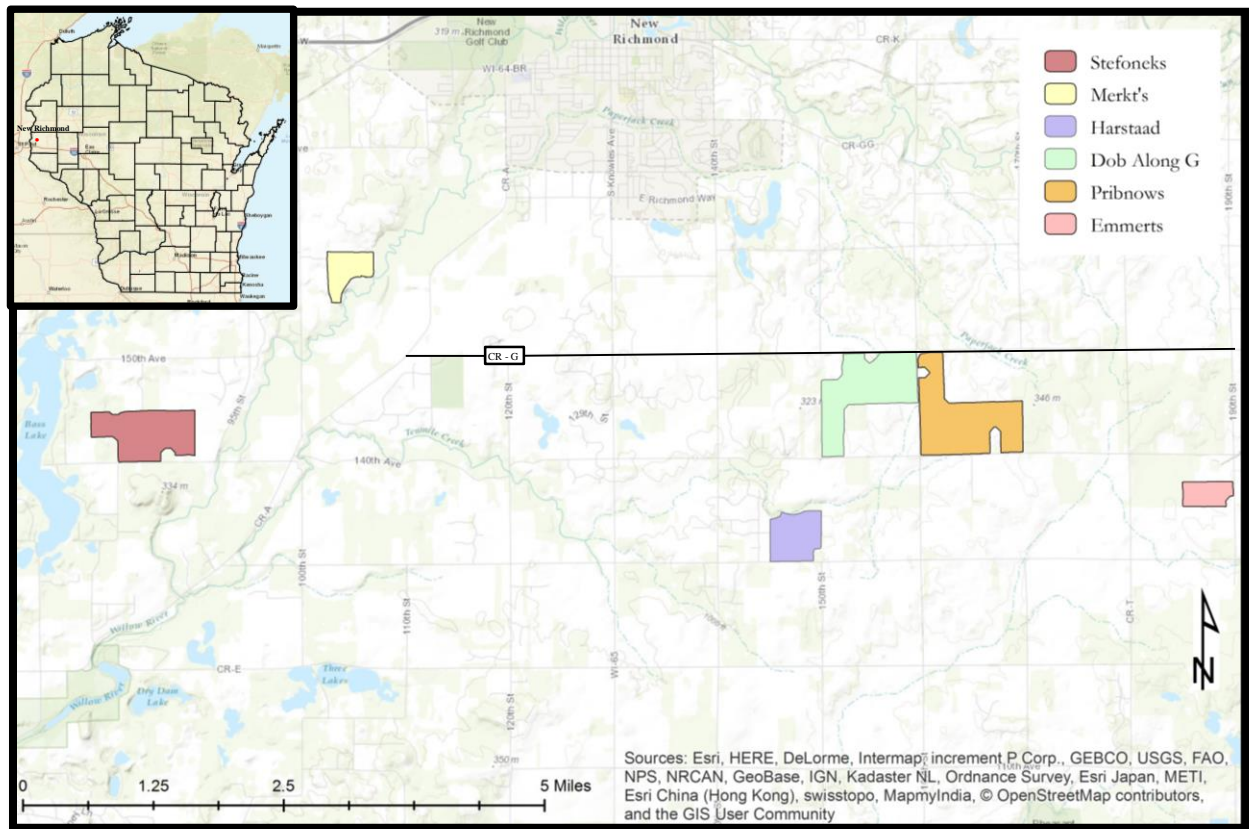
Figure 12: Overview of the study area. Individual fields from left to right are Stefoneks, Merkt's, Harstaad, Dob Along G, Pribnows, and Emmerts.

Paul Schottler, a farmer with various fields in this area, provided the yield data for six of his fields. These fields are discussed in detail in Section 3.1. The observations for these fields in Section 3.1 were provided by an agronomist. This included three years' worth of information for

each field as well as an aerial view of each individual field along with a photo taken from each

site in Spring 2015. The data provided were unprocessed, containing a table with attribute

information for each yield point that will be discussed below in Section 3.2.
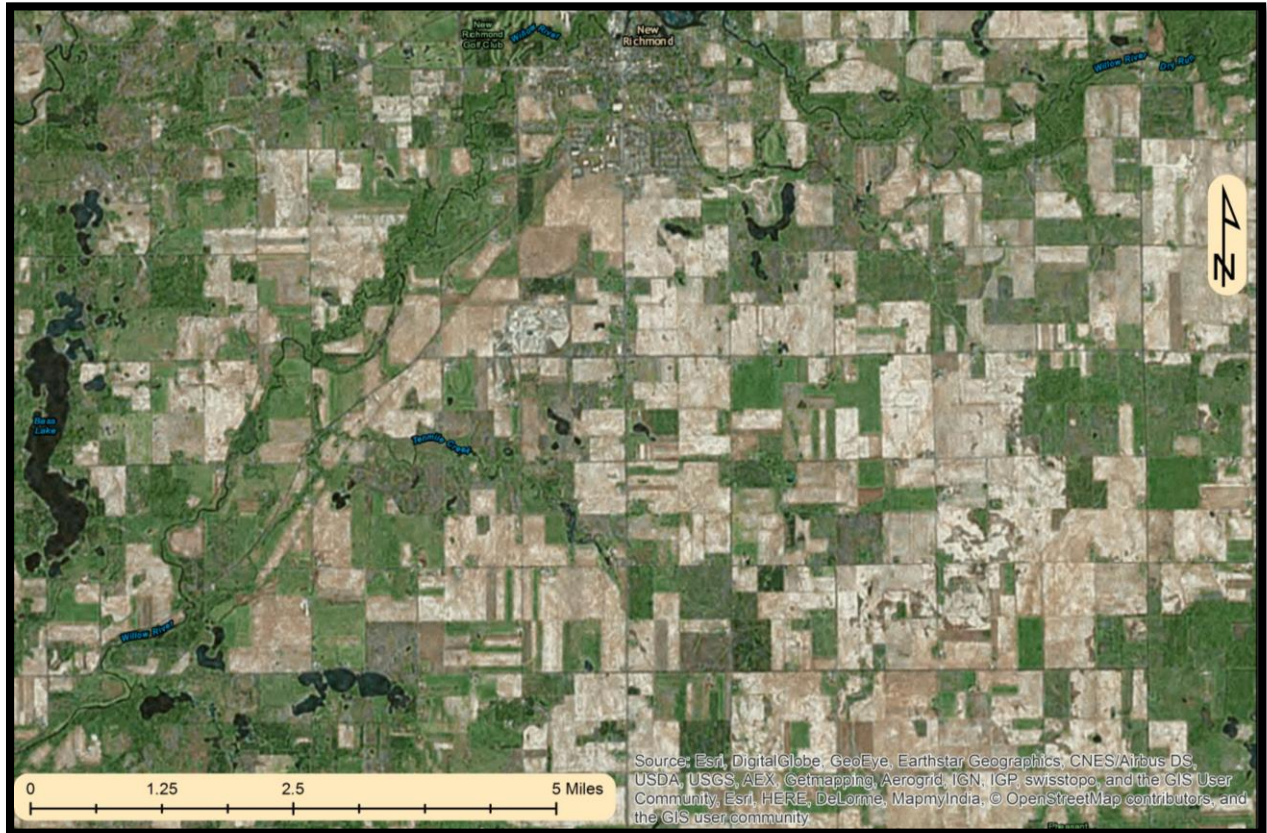
Figure 13: Aerial overview of the study area. The area is characterized by numerous rivers and lakes, in addition to farmland and woodlands.

LiDAR data for the study area was obtained from the Planning and Land Use Department

of the St. Croix County Government Center. This data consists of only the unprocessed point

cloud and was used to derive associated elevation data. LiDAR data is discussed in further detail

below in Section 3.3.

3.1 Overview of Fields in Study Area

3.1.1 Dob Along G

Dob Along G is a 312-acre field that is characterized by a gently sloping and mostly flat landscape. Figure 14 shows an aerial view of the field with an outline of the border and a photograph taken of the field in Spring 2015. There is a north-south oriented crest that lines up roughly with the household on the north end of the property. It is bordered by County Rd. G to the north, 160 St. to the east, 140 St. to the south, and a separate field on the west. Figure 15 shows an example of the yield data for this field. This yield data is the composite with a thin lower section containing moderate yields and the large upper area showing high yields with moderate to low yields along the edges. There are also thin strips with missing data.

Figure 14: (a) Photograph of Dob Along G taken in 2015 and (b) an aerial view of the field with the boundary in blue

Figure 15: Example of a cleaned yield maps generated using Yield Editor for Dob Along G. The classification breaks have blue representing poor yield, green moderate yield, and red high yield (in bushels per acre).

3.1.2 Emmert

Emmert is a 68-acre field that is characterized as a mostly flat field with some gently sloping area. This field has limited soil variability throughout. Figure 16 shows an aerial view of the field with an outline of the border and a photograph taken of the field in Spring 2015. This photo shows the field edge and a knoll located at the north-central end of the field. It is bordered by 190 St. to the east and otherwise surrounded by separate fields on its other sides.

Figure 16: (a) Aerial view of Emmert with the boundary in blue and (b) a photograph of the field taken in 2015

Figure 17 shows an example of the yield data for this field. This yield data is the composite with a somewhat rectangular shape with little differences in the yield except for some lower yields around the edges.

Figure 17: Example of a cleaned yield points generated using Yield Editor for Emmert. The classification breaks have blue representing poor yield, green moderate yield, and red high yield (in bushels per acre).

3.1.3 Harstad

Harstad is a 135-acre field that is characterized by a large amount of variability in soil with many hills and small flat areas. Figure 18 shows an aerial view of the field with an outline of the border and a photograph taken of the field in Spring 2015. There is an east-west crest that travels from the southeastern side to the house located in the southcentral section of the field. This crest marks a boundary line between a mostly flat area in the south and an extremely hilly area in the north. The remainder of the field is moderately hilly with no flat areas and high variability in soil type as can be seen by the color difference. It is bordered by a forested area to the north, residential land to the east, 130 Ave. to the south, and separate fields and forests to the west. Figure 19 is an example of cleaned yield data for this farm.

Figure 18: (a) Aerial view of Harstad with the boundary in blue and (b) a photograph of the field taken in 2015

Figure 19: Example of a cleaned yield maps generated using Yield Editor for Harstad. The classification breaks have blue representing poor yield, green moderate yield, and red high yield (in bushels per acre).

3.1.4 Merkt's

Merkt's is a 90-acre field that is characterized by moderate soil variability with a few hills and ridge lines. Figure 20 shows an aerial overview of the field with its boundary in blue. It is bordered by 160th Ave to the north, residential areas to the west, and trees on the other sides. Access was limited so no pictures were obtained for this field. Figure 21 is an example of a cleaned yield for the farm.

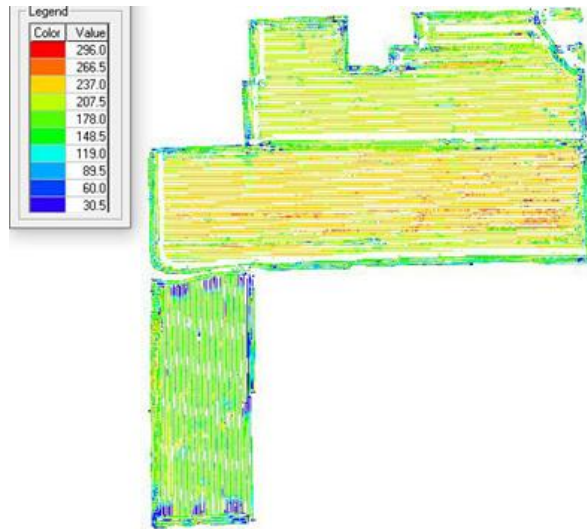**Figure 20: Aerial view of Merkt's with the field boundary in blue**
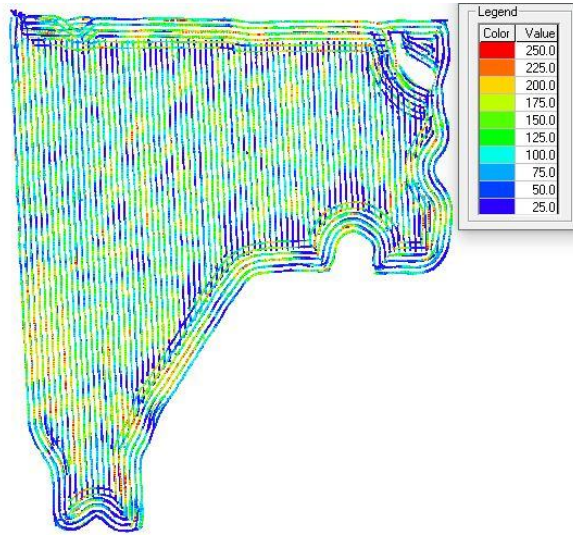
Figure 21: Example of a cleaned yield maps generated using Yield Editor for Merkt's. The classification breaks have blue representing poor yield, green moderate yield, and red high yield (in bushels per acre).

3.1.5 Pribnows

Pribnows is a 360-acre field with limited variability, a gently sloping landscape and crest just west of the building on the south side. Figure 22 shows an aerial view and photograph of the field taken in Spring 2014. It is bounded by 160th St to the east, 140th to the south, County Rd G to the North. Additionally, there are fields and trees to the north where the field moves away from the County Road.

Figure 22: (a) Aerial view of Pribnows with the field boundary highlighted in blue and (b) a photograph taken of the field in Spring 2014.

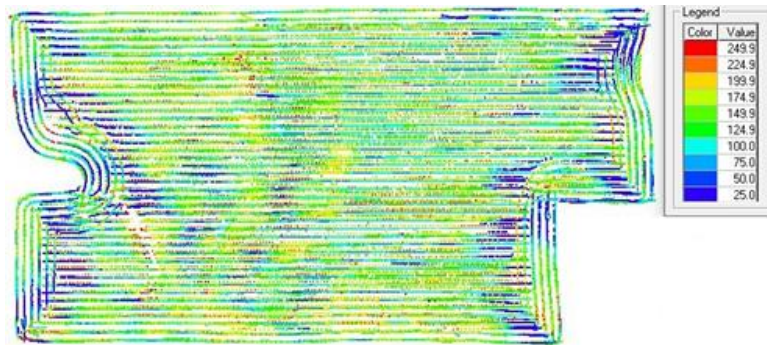Figure 23 is an example of cleaned yield data for the farm.
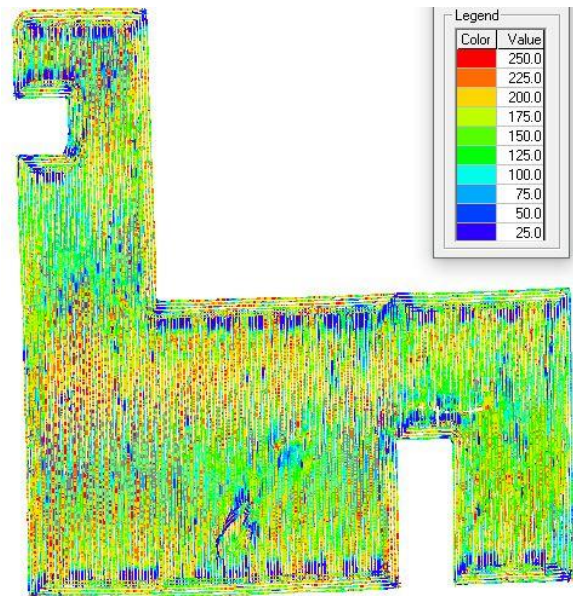
Figure 23: Example of a cleaned yield maps generated using Yield Editor for Pribnows. The classification breaks have blue representing poor yield, green moderate yield, and red high yield (in bushels per acre).

3.1.6 Stefoneks

Stefoneks is a 240-acre field that is separated in two by 85th St. It is best characterized by a highly variable landscape that show large differences between the two halves. It is obvious that this field has only one name because of management purposes and not because it represents a homogenous field. The east side of the field features a ridge line in the center of the field with gentle slopes flowing from there to the east side. The west side of the field is a continuation of the ridge line with flat areas to the north of the ridge and gentle slopes flowing to the west. On the far west there is a series of knolls. There are also variations in the soil color at the southwest end. Figure 24 shows an aerial view of the field and a photograph of it taken in Spring 2014. The field is bounded on the south by 140th Ave., farm fields to the east, trees in the northwest, another field to the northeast, and residential areas with trees in the southwest corner. Figure 25 is an example of cleaned yield data for the farm.

Figure 24: (a) Photograph of Stefoneks taken in Spring 2014 and (b) aerial imagry with the field boundary in blue
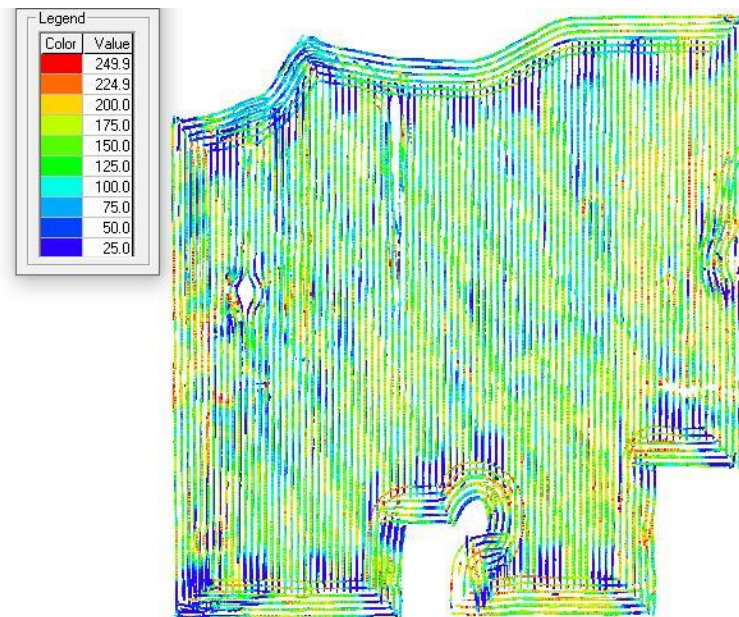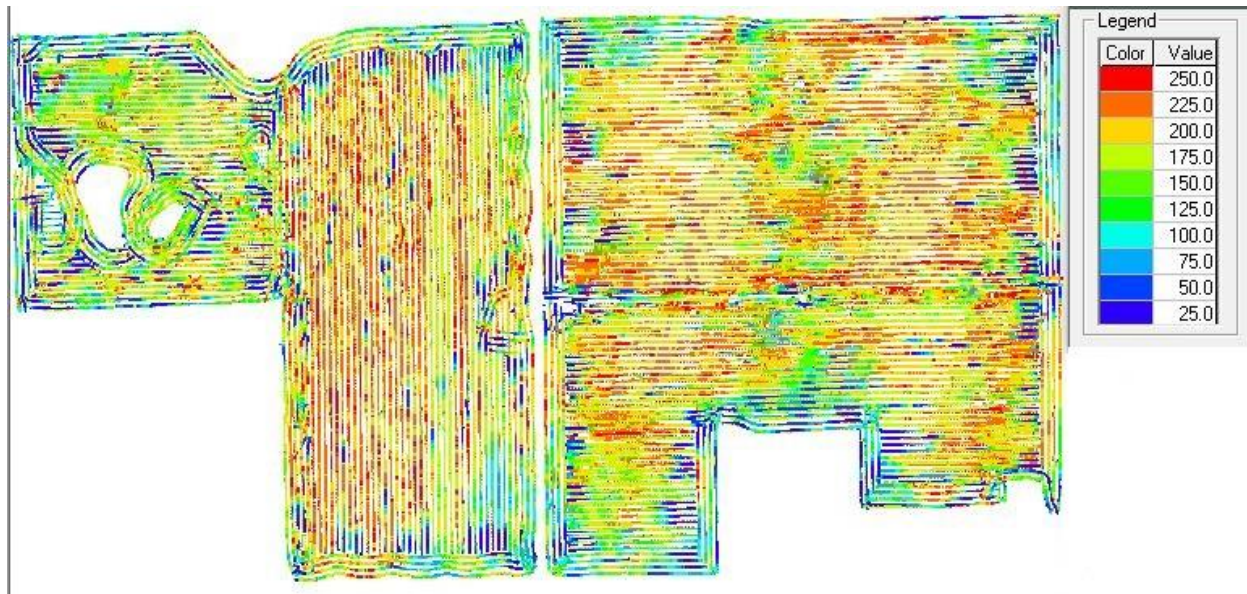
Figure 25: Example of a cleaned yield maps generated using Yield Editor for Stefoneks. The classification breaks have blue representing poor yield, green moderate yield, and red high yield (in bushels per acre).

3.2 Working with Yield Data

The fields obtained from Nick Schottler came in the form of shape files. In order to correctly process this data, the table had to be exported into a spreadsheet and manipulated there. Once the data was imported into Excel the columns needed to process yield in the USDA Yield Editor were placed in the correct order. This allows for errors in the data to be removed by this editor. Once this data has been processed in the editor it is ready to be imported back into ArcGIS. The editor already properly links each data point to the proper UTM projection, in the case zone 15N. Finally, the yield data is interpolated using the accepted practice which relies on the inverse distance weighted (IDW) with distance squared interpolation method. This final process shows the result that a farmer will look at to evaluate his or her yield.

3.2.1 Yield Table

Yield data was acquired in the form of thousands of georeferenced points that vary in quantity

based on the size of the field and calibration of the yield monitor. For example, Dob Along G has

over 4,600 unprocessed yield points to consider. This data comes in the order that the yield

monitor provides it. An example of a truncated list of these data points is provided in Table 2 for

Dob Along G.

Table 2 Raw Yield Data from Dob Along G recorded in 2012

FID	Product	Swath Width (ft)	Elevation (ft)	Time (date)	Yield Volume (bu/ac)	Speed (mph)	Crop Flow (V)
15599	CORN	30.1	1040.8	2012-10-16	184.89	5.02	3398.20
15600	CORN	30.1	1040.8	2012-10-16	185.20	4.97	3368.53
15601	CORN	30.1	1040.7	2012-10-16	154.57	5.99	3388.79
15602	CORN	30.1	1040.8	2012-10-16	186.33	4.97	3388.79
15603	CORN	30.1	1040.7	2012-10-16	186.38	5.00	3409.08
15604	CORN	30.1	1040.6	2012-10-16	185.18	5.02	3401.58
15605	CORN	30.1	1040.6	2012-10-16	187.30	4.94	3389.34
15606	CORN	30.1	1040.6	2012-10-16	186.12	4.99	3402.00
15607	CORN	30.1	1040.5	2012-10-16	166.26	5.02	3054.05
15608	CORN	30.1	1040.4	2012-10-16	168.98	4.99	3088.60
15609	CORN	30.1	1040.3	2012-10-16	169.70	4.99	3101.86
15610	CORN	30.1	1040.3	2012-10-16	171.09	4.97	3113.11
15611	CORN	30.1	1040.3	2012-10-16	171.13	4.99	3129.84
15612	CORN	30.1	1040.2	2012-10-16	169.58	4.99	3099.84
15613	CORN	30.1	1040.1	2012-10-16	167.48	5.02	3076.52

As can be seen from the table there are instances where units for a column are not given.

These must be determined based on a reasonable suspicion of what units correspond to the

number in that column based on common practice and magnitude. For example, a typical swath

width on a combine is 30 ft so the column labeled swath width is clearly in feet. This table also

shows data that can help remove erroneous points. For example, if the speed suddenly varied

from its relatively stable 5 mph this would be an indication that the yield monitor was picking up

routine shifts in the combine that were not meant to be monitored.

Once these fields have been exported from the shape file into a table the columns need to be made to match the format used by the Yield Editor which is shown below in Table 3. The data must appear in this exact order with the correct units in order for the software to correctly process the data. The Yield Editor than removes points based on user selections and manual edits. The Yield Editor will automatically check the data from the table to determine automatic limits on various elements including but not limited to flow delay, velocity (maximum and minimum), minimum swath, and yield (both minimum and maximum). The listed filter selections above were used to clean the data. Attributes related to yield are chosen based on the user crop choice and position based on the user input UTM zone.

The Yield Editor was next used to remove erroneous points by looking for excessive low or high yield values and other anomalies associated with the data. This cleaned yield data was then by imported back into ArcGIS to complete the yield analysis.

Table 3: Column Data Names and Units for Import

Column Name	Units
Longitude	decimal degrees
Latitude	decimal degrees
Flow	pounds per second
Time	seconds
Logged Interval	seconds
Distance	inches
Swath	inches
Moisture	percent wet

Adapted from Yield Editor 2.0 Manual

3.2.2 Yield Interpolation

The yields were interpolated using the IDW technique with distance squared because this is the most commonly used technique to visualize the thousands of yield points available from a yield monitor. This was done with the IDW tool using a radius of 5 m and the smooth neighborhood

function. This tool finds whatever points lie within this radius and uses these to compute the value of a cell. The smooth neighborhood function makes sure that points closer to the center of the circle carry greater weight than those further away. The class ranges for the interpolation were broken down into six evenly distributed classes between 0 and 300. The upper limit was chosen because it represents an optimal maximum yield to stop at. The interpolation used a smooth nearest neighbor function to generate the raster because this is the standard approach in yield mapping.

3.2.3 Composite Yield

The composite yield is generally created by normalizing yield between zero and one through dividing the yield by the maximum yield a particular farm can achieve, which is determined by the analyst. Once normalized any number of years can be combined together, but typically three are considered a good number of years. This process was not strictly adhered to for this project because it inherently loses data in the normalization process and limits the ability to look at actual yields.

The composite yield was created by combining multiple years of yield data into one yield interpolation. The process was the same as for a single field (as discussed above), but before the yield undergoes error checking, all of the uncleaned yield data is combined into one table. In this way no data was lost through any intermediate normalization. This also means that the Yield Editor may remove different points from the composite than for each year. This is because the editor uses a statistical technique to determine which points are removed. Differences can and did occur from one year to the next and the composite.

The reasoning behind using so many yield points is that any errors that exist in a single year will be diminished by using multiple years of data. This thesis examines these differences to

look at the efficacy of this claim. Figures 15, 17, 19, 21, 23, and 25 above show the cleaned yield editor results for each of the six fields as viewed from within the editor.

3.2.4 Validation and Comparison of Yield Interpolations

In order to validate the yield interpolations, 20 approximately equal sized grids were generated on each field and a control point was randomly selected within each grid. These control points were then removed and the interpolation run on the remaining points. The interpolated raster values for the location of these points were then compared to the actual values of these points from the original yield data.

The difference of means statistical test was used to evaluate the significance of the differences for each year and the composite for all fields. The standard error and degrees of freedom were calculated using Equations (1) and (2) below. Finally, the test statistics was calculated using Equation (3) and used to find the p-value in the statistical chart. Where SE is the standard error, DF is the degrees of freedom, s_1 and s_2 are the standard deviation for the actual and predicted values respectively, n_1 and n_2 are the size of the sample for the actual and predicted values respectively, t is the test statistic, x_1 and x_2 are the mean of the actual and predicted values respectively, and d is the hypothesized difference between the two means. Values were considered significant if below 0.05 which is a standard threshold value for use with scientific data.

$$SE = sqrt[(s_1^2/n_1) + (s_2^2/n_2)] \tag{1}$$

$$DF = (s_1^2/n_1 + s_2^2/n_2)^2 / \{[(s_1^2 / n_1)^2 / (n_1 - 1)] + [(s_2^2 / n_2)^2 / (n_2 - 1)]\} \tag{2}$$

$$t = [(x_1 - x_2) - d] / SE \tag{3}$$

Once these data were validated using the above method some comparisons can then be made by looking at the changes in mean and variability between interpolations from all years of

the fields. This can help identify years where drastic changes could affect the composite and demonstrate where improper classification values could damage a yield interpolation.

3.3 Constructing DEMs from LiDAR

The LiDAR from the St. Croix Government Center came in the form of point clouds saved as .LAS files. This LiDAR information was obtained by AYRES Associates in the spring of 2014. AYRES is an engineering and architectural company that has geospatial and GIS related services to clients. The provided processed DEM has a nominal point spacing of one meter with a vertical tolerance of 0.61. However, the raw point cloud was used in order to fully explore the process of DEM creation. The data was separated into cells of points that correspond to latitude and longitude. Each yield dataset had one to four cells of LiDAR data associated with it depending on its size and location. A DEM of each field was created with these point cloud data.

3.3.1 Creation of DEM

In order to create a DEM from LiDAR, the grid cells associated with a field are extracted and transformed into multipoint vector data. This simply transforms the raw LiDAR points into a usable format. A terrain dataset was created using this multipoint data with an average point spacing of one meter. One meter was chosen because it represents the most accurate dataset that can be created with the LiDAR available. The terrain dataset is a triangular irregular network (TIN) with different levels of detail called pyramids. These pyramids are for visualization purposes only as the most detailed level is used to create the DEM. Basically the terrain dataset is an intermediate step before the creation of the final DEM. The spacing between LiDAR points is dependent on the data provided. In this case the data provided from St. Croix County was described in its metadata as having a nominal point spacing of one meter. The example terrain

for Dob Along G is shown in Figure 26. It can be changed to greater levels of tessellation to increase efficiency if less accurate data is required.

Once the terrain has been created it is visibly inspected for errors and transformed into a raster. This raster is the DEM and is a digital representation of the elevation in the area. The difference between the DEM and the terrain dataset is that the terrain is constructed with triangles while the DEM is constructed with squares. The DEM allows for easier visualization, manipulation, and classification of the data. This DEM is shown in Figure 26b. As can be seen from the elevation dataset there is some noise associated with the data. For example, there is a small area included in the DEM that differs from the terrain located in the northwest of Dob Along G. Man-made structures can have a direct impact on the quality of the DEM as well.
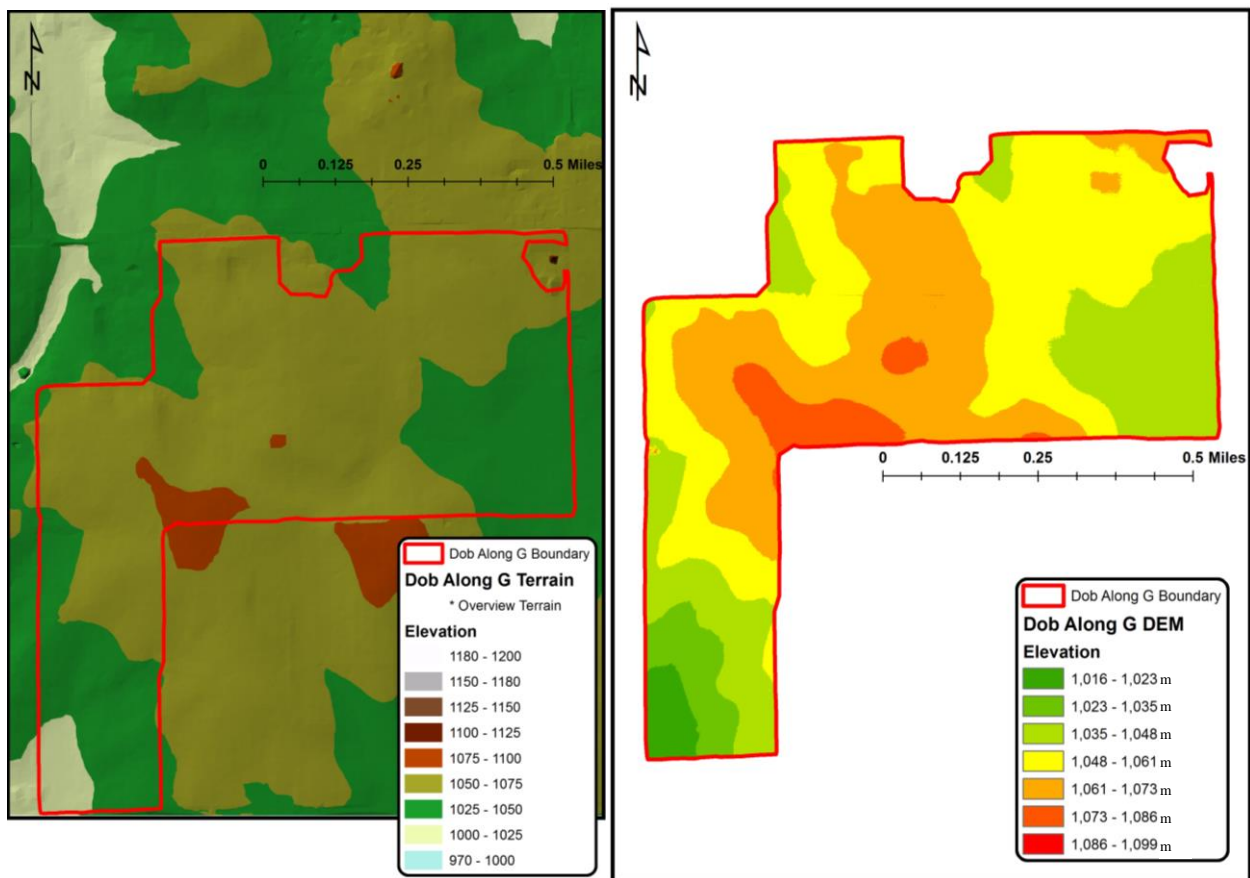
Figure 26: (a) Terrain in 25 m classes; (b) a 5 m DEM raster of Dob Along G study area.

3.3.2 DEM Derived Attributes

The DEM can be used to construct derivative maps of various other associated attributes; for example, slope, aspect, and elevation contours. This project used DEM data to create slope maps that were used for visual analysis. The slope map better characterized how the elevation varied across each of the study fields.

The slope map was created directly from the DEM as the rate of change in percent. The class breaks used in these derivative slope maps are very important because a small change in slope could have a great impact on yield. Also, if not correctly symbolized the slope map will be mostly monotone on such a flat surface. An example of this output is shown in Figure 22. For the purposes of this thesis five slope classes were chosen: 0°, > 0°- 1°, > 1° - 2°, > 2° - 3°, > 3° - 5°, and > 5°. These ranges emphasize small changes in slope.
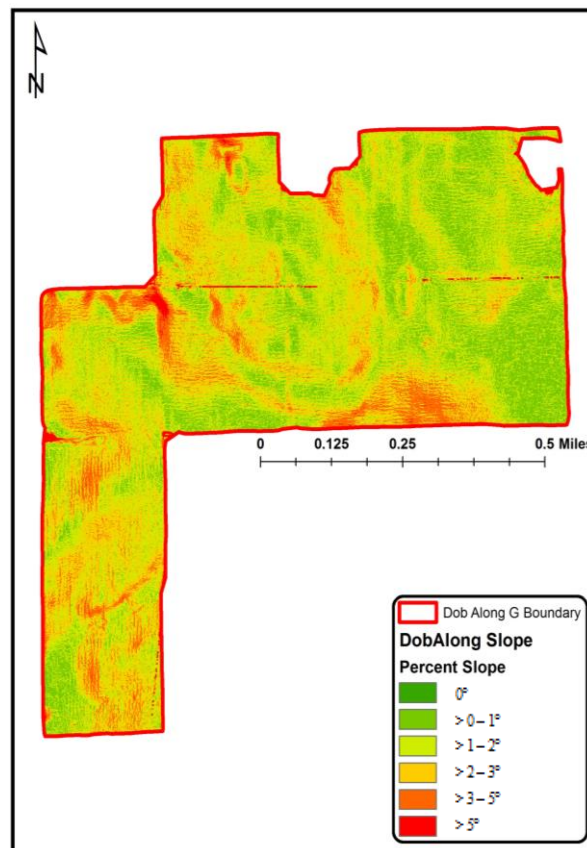
Figure 27: Slope raster based on the 5 m DEM with class breaks that emphasize the variations in slope.

3.4 Final Analysis Steps

All of the methods and data described in Sections 3.1 through 3.3 were next used to answer two questions: (1) Are there significant differences between a multiple year composite yield and a single year and if so, are these significant enough to have an impact on normal operations? (2) Are the areas where such differences occur related to differences in slope?

3.4.1 Differences between single year and multiple-year yields

The yields in each of the six fields were interpolated using the above methods four times, once for three single years and once for these years combined. These results were compared against 20 randomly selected control points to demonstrate the efficacy of this approach and the full results were then used to compare the single year and multiple-year yields. Finally, yield difference rasters were created highlighting the differences between the single and multiple-year interpolations to show the areas where the most significant differences were located. This workflow was implemented to answer the first question.

3.4.2 Impact of slope on yield differences

For the second question, a DEM was created for each farm using the LiDAR procedure discussed above. This DEM was in turn used to create a degree slope map with five categories of slope. The first category containing only 0° slope values was removed as no field had members included in this category. This final slope map was then compared to the yield difference rasters created above to identify whether the increased yield differences were associated with those parts of the field with higher slopes.

      The results are reported in Chapter 4 and their significance is discussed in Chapter 5.

CHAPTER 4: RESULTS

The results are presented for each of the six farms in three steps (i.e. sections). The first

compares predicted and measured yields to assess the efficacy of IDW as an interpolation

method for yields on the six fields. The second compares a single year to a composite year yield

in order to show if a single year offers similar guidance to the composite yield output. The third

compares the difference between the composite and single year yield to the slope value to

determine which slope values, if any, accompanied the largest differences in yield values.

4.1 Yield Interpolation Evaluation

The predicted values from the interpolation were compared with the actual values from 20 yield

points within each field. These points were chosen randomly within grid squares in each field

excluding a 10 m buffer inside the field boundary. This approach generated a spatially stratified

set of test points for each field and also minimized any edge effects associated with the field

boundary.

Table 4 shows the yield mean for each year in each field along with the composite. The

difference between each pair of means is also shown. This difference was used as to judge the

efficacy of the interpolation. The composite yields for four of the six fields (highlighted in

yellow in Table 4 below) shows how IDW struggled to predict yields using three years of data

that was passed through and filtered by the Yield Editor, thereby contradicting the usual pattern

in which larger volumes of data improves the efficacy of interpolation.

Table 5 shows the yield values for Harstad in 2014 to demonstrate how small differences

in the mean can mask substantial yield volatility for individual locations (as illustrated by the

actual and predicted yields for the grid cells shaded in yellow). Similar results occurred for all

six fields in 2014 although Harstad shows the most significant differences and although this

Table 4: Interpolation evaluation showing the actual and predicted means and standard deviation (SD) of the yields at 20 location in each year and the composite. Possible Interpolation failures are highlighted in yellow.

FIELD	YEAR	SD	ACTUAL	PREDICTED	SD	DIFFERENCE
DOB ALONG G	2012	26	190	189	20	1
	2013	40	176	168	27	7
	2014	46	159	153	20	6
	All	31	202	197	21	5
EMMERT	2012	19	194	189	10	5
	2013	34	88	86	16	2
	2014	63	117	128	39	10
	All	19	194	116	32	79
HARSTAD	2012	45	159	164	26	5
	2013	42	79	80	30	1
	2014	51	132	122	22	10
	All	42	79	121	31	41
MERKT'S	2012	37	159	157	20	2
	2013	27	54	61	13	7
	2014	51	100	98	30	2
	All	27	54	98	26	43
PRIBNOWS	2012	20	190	187	13	3
	2013	27	88	89	24	2
	2014	32	172	170	18	2
	All	27	88	148	18	60
STEFONEKS	2011	20	197	194	11	3
	2012	21	173	178	10	5
	2013	66	152	151	48	1
	All	21	173	171	31	2

would not impact single field evaluations, it could greatly impact precision agriculture recommendations.

Finally, as a check to see if these results are significant, a difference of means statistical test was used to generate a p-value and evaluate whether or not the reported differences were significantly different from zero. This p-value indicates that a result is significant if it is less than 0.05. Table 6 shows a summary table for these values with statistically significant results (i.e. differences) highlighted in yellow. As can be seen, yields for four of the six composites show statistical significance. This result occurred because the composite is not simply a combination

Table 5: Example of comparison between actual and predicted yield values for Harstad 2014 with the significant differences highlighted in yellow

ID	Actual	Predicted
1	145	167
2	23	66
3	161	126
4	89	137
5	187	119
6	154	143
7	41	117
8	186	140
9	148	101
10	124	129
11	176	116
12	74	120
13	105	121
14	189	138
15	90	75
16	142	123
17	123	112
18	179	129
19	105	119
20	202	136

of the three cleaned years but a combination of the unclean points that were then cleaned by the Yield Editor together. If one year is drastically different from the other two, then the composite fails to predict yield values. Normally, drastically different years would be removed by the analyst before a composite is made.

4.2 Yield Interpolation Results

The similarities and differences between the interpolated individual year and the composite yields are presented in three parts below. The first part shows all of the interpolated yields for each year and the composite in each field. The second part shows yield difference rasters for the composite and each year.

Table 6: Difference of Means Statistical Test Results. Possible failures of interpolation are highlighted in yellow.

Field	Year	P Value	Significant
Dob Along G	2012	0.899332	No
	2013	0.500536	No
	2014	0.572421	No
	All	0.540056	No
Emmert	2012	0.272875	No
	2013	0.806777	No
	2014	0.537597	No
	All	1.21E-10	Yes
Harstad	2012	0.689998	No
	2013	0.958383	No
	2014	0.405363	No
	All	0.001136	Yes
Merkt's	2012	0.814679	No
	2013	0.320121	No
	2014	0.91044	No
	All	8.78E-06	Yes
Pribnows	2012	0.524044	No
	2013	0.809683	No
	2014	0.809867	No
	All	1.38E-09	Yes
Stefoneks	2012	0.55353	No
	2013	0.360466	No
	2014	0.942276	No
	All	0.801536	No

The third part reports the mean and coefficient of variation for each year and the composite and also includes a discussion of yield between years and fields as a way to characterize the strength of the yield for a particular field and the changes that can occur within it.

4.2.1 Yield Interpolation

The interpolated yield maps for these fields reproduced in Figures 28-33 were constructed using yield data acquired from 2012 to 2014 (except for Stefoneks which includes data for 2011-2013).

Figure 28: Yield Interpolations for Dob Along G from 2012 to 2014 and a composite yield for all three (in bushels per acre).

Figure 29: Yield Interpolations for Emmert from 2012 to 2014 and a composite yield for all three (in bushels per acre).

Figure 30: Yield Interpolations for Harstad from 2012 to 2014 and a composite yield for all three (in bushels per acre).

Figure 31: Yield Interpolations for Merkt's from 2012 to 2014 and a composite yield for all three (in bushels per acre).
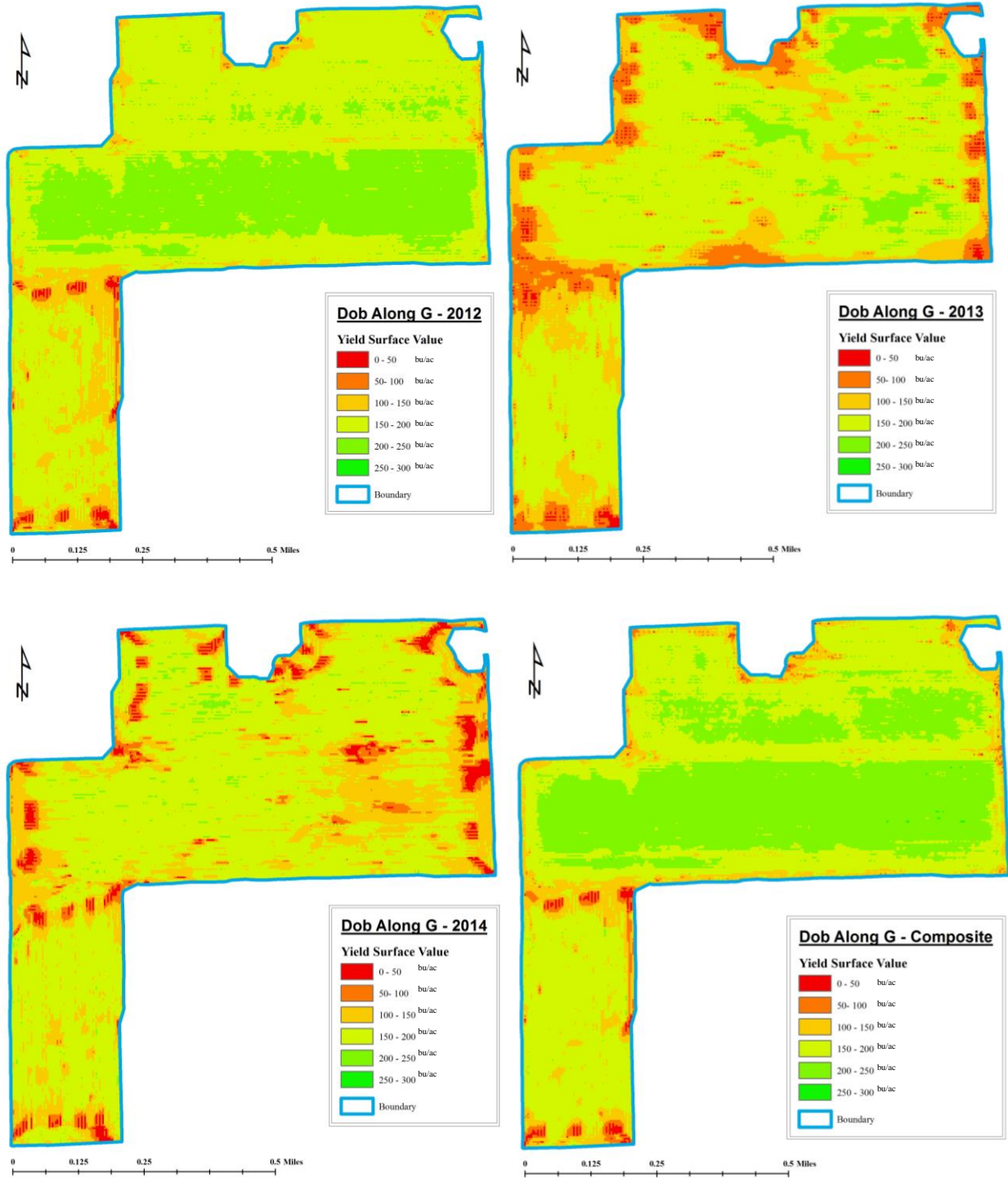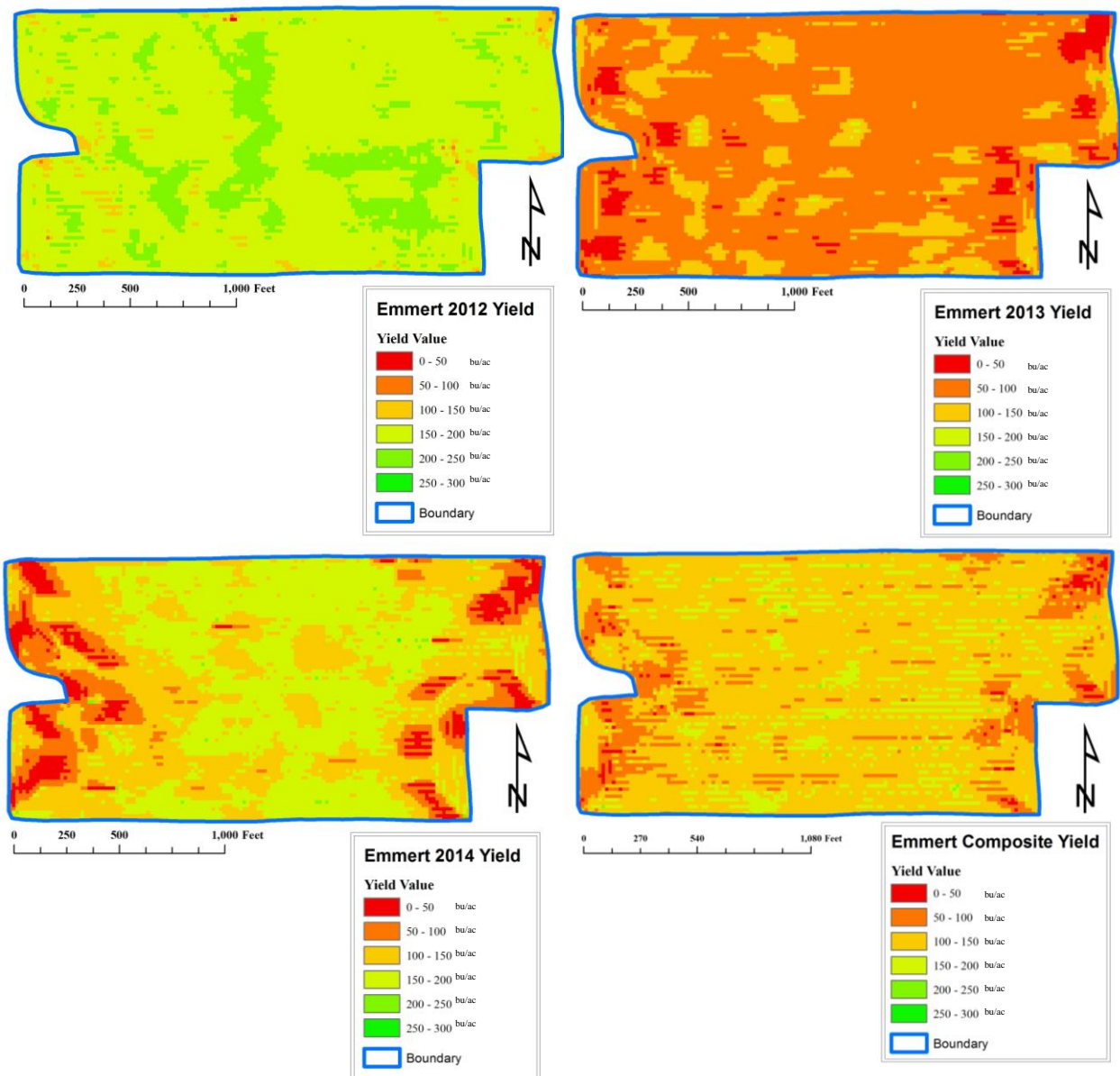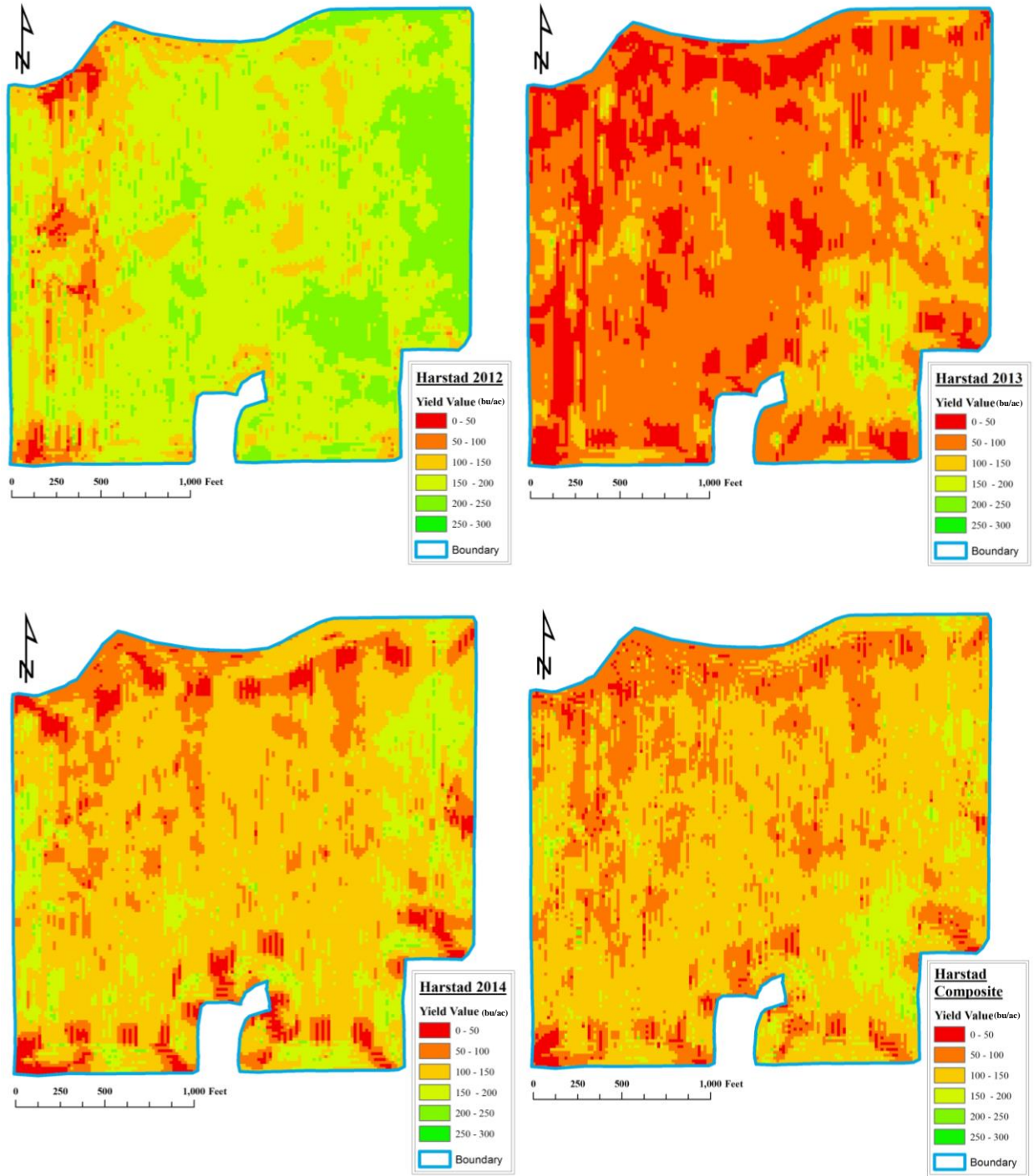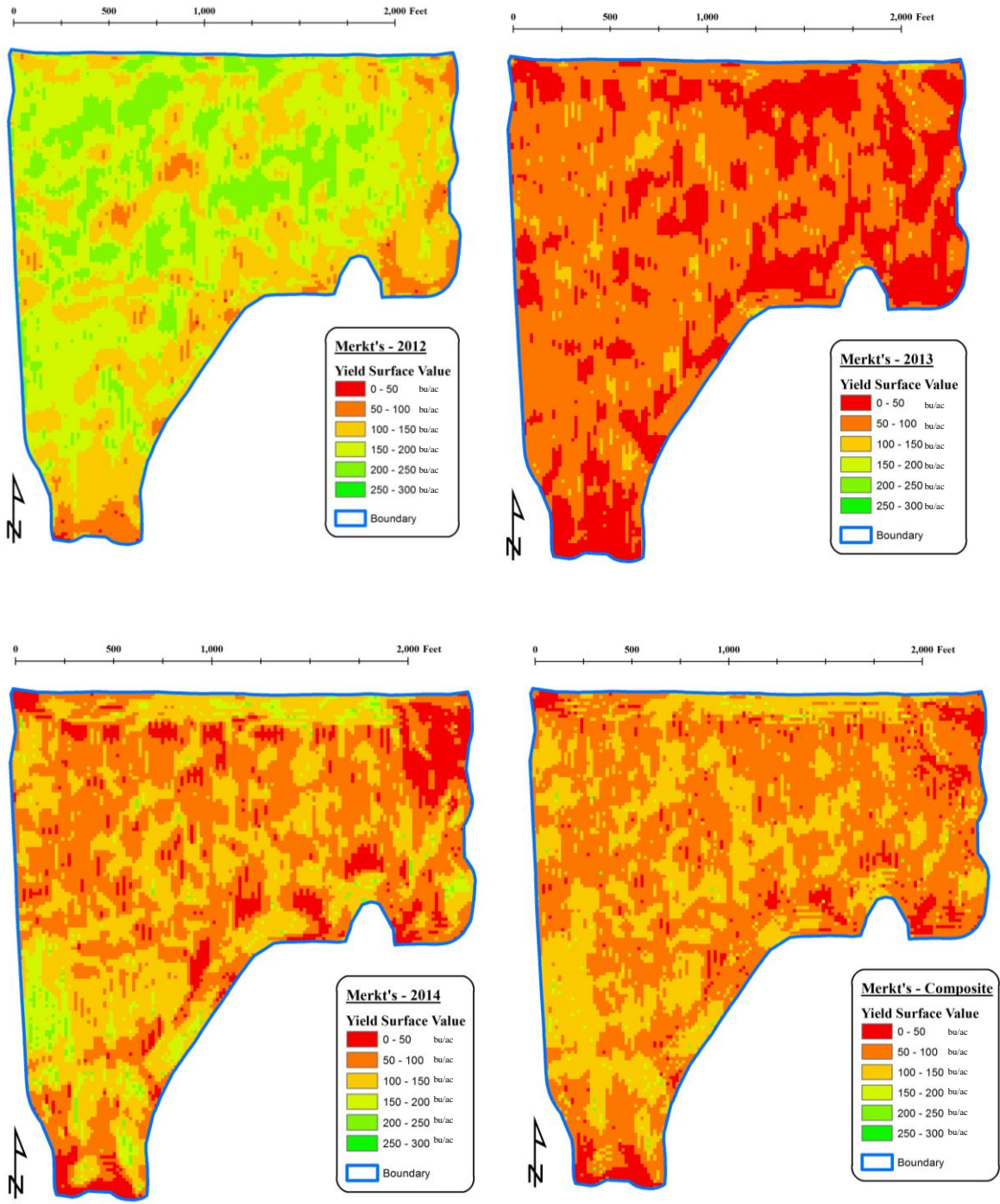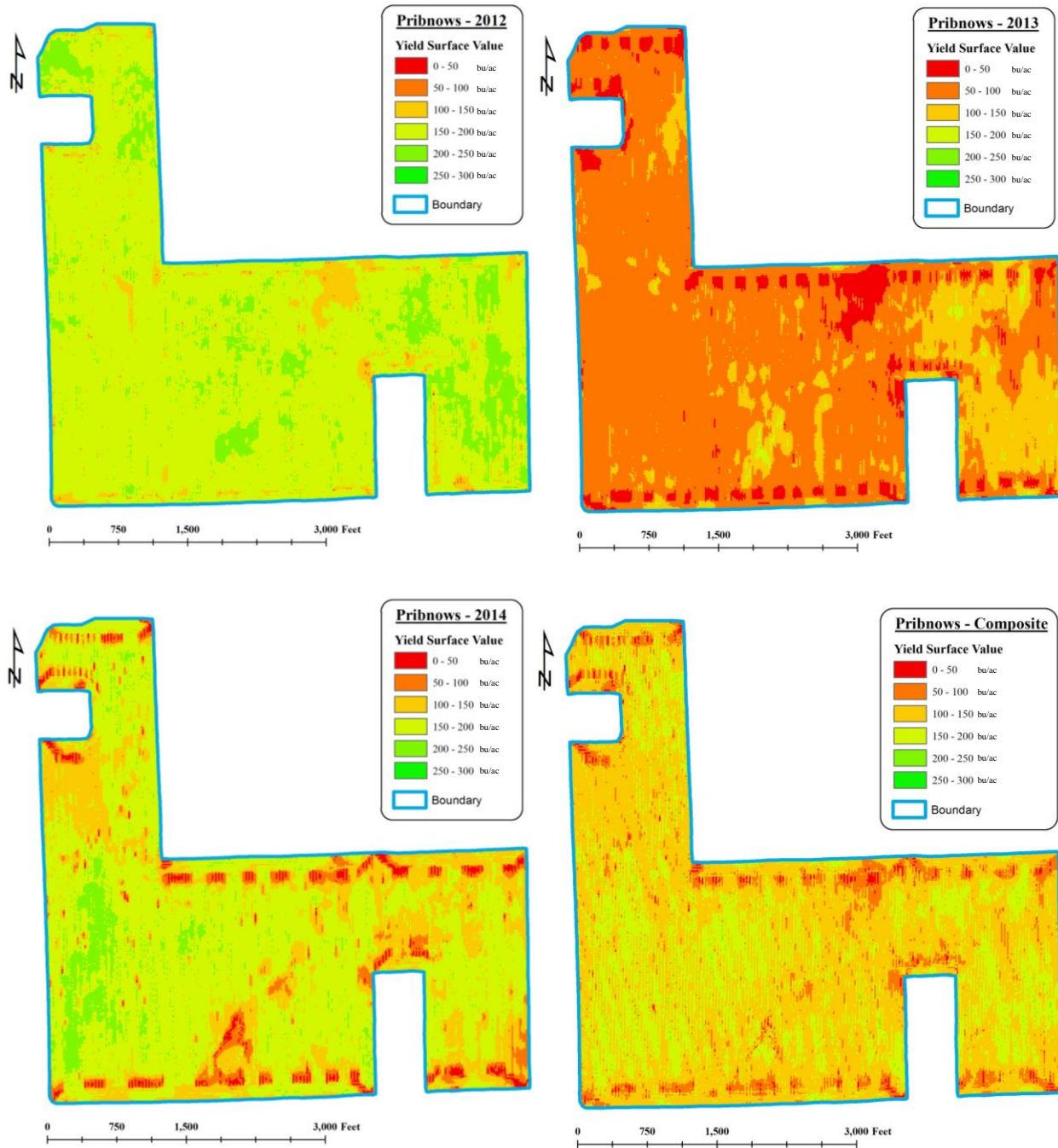
Figure 32: Yield Interpolations for Pribnows from 2012 to 2014 and a composite yield for all three (in bushels per acre).
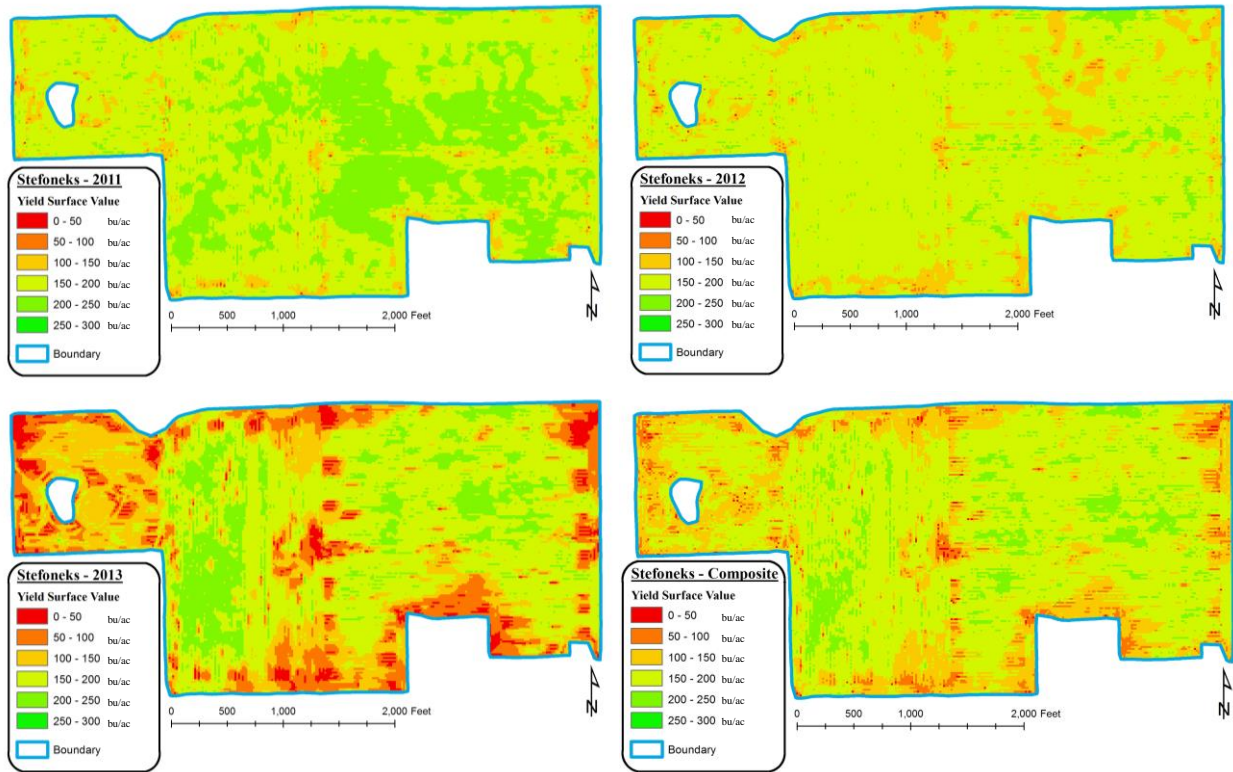
Figure 33: Yield Interpolations for Stefoneks from 2011 to 2013 and a composite yield for all three (in bushels per acre).

4.2.2 Yield Difference Rasters

The single year interpolations were then subtracted from the composite to create three difference yield difference rasters for each of the six fields. These yield difference rasters show how the estimated yield for each year in each raster cell differed from the composite. Figures 34 - 39 show the three yield difference rasters created using this method.

Dob Along G and Stefoneks have the most consistent results with a composite that did not fail the significance test. These maps also show the most variation in negative and positive differences that are generally centered around the middle. The remainder of the fields show the differences where the composite failed the significance test. Generally, these fields show negative differences for the first year, positive results for the second year, and mixed results for

the final year. The first year consistently performed the worst for these four fields and the composite would likely perform better if this particular year was removed.

Figure 34: Maps showing the difference between the single year and the composite yields (in bushels per acre) for Dob Along G. It has five categories with a count of the number of cells within the category.

Figure 35: Maps showing the difference between the single year and the composite yields (in bushels per acre) for Emmert. It has five categories with a count of the number of cells within the category.
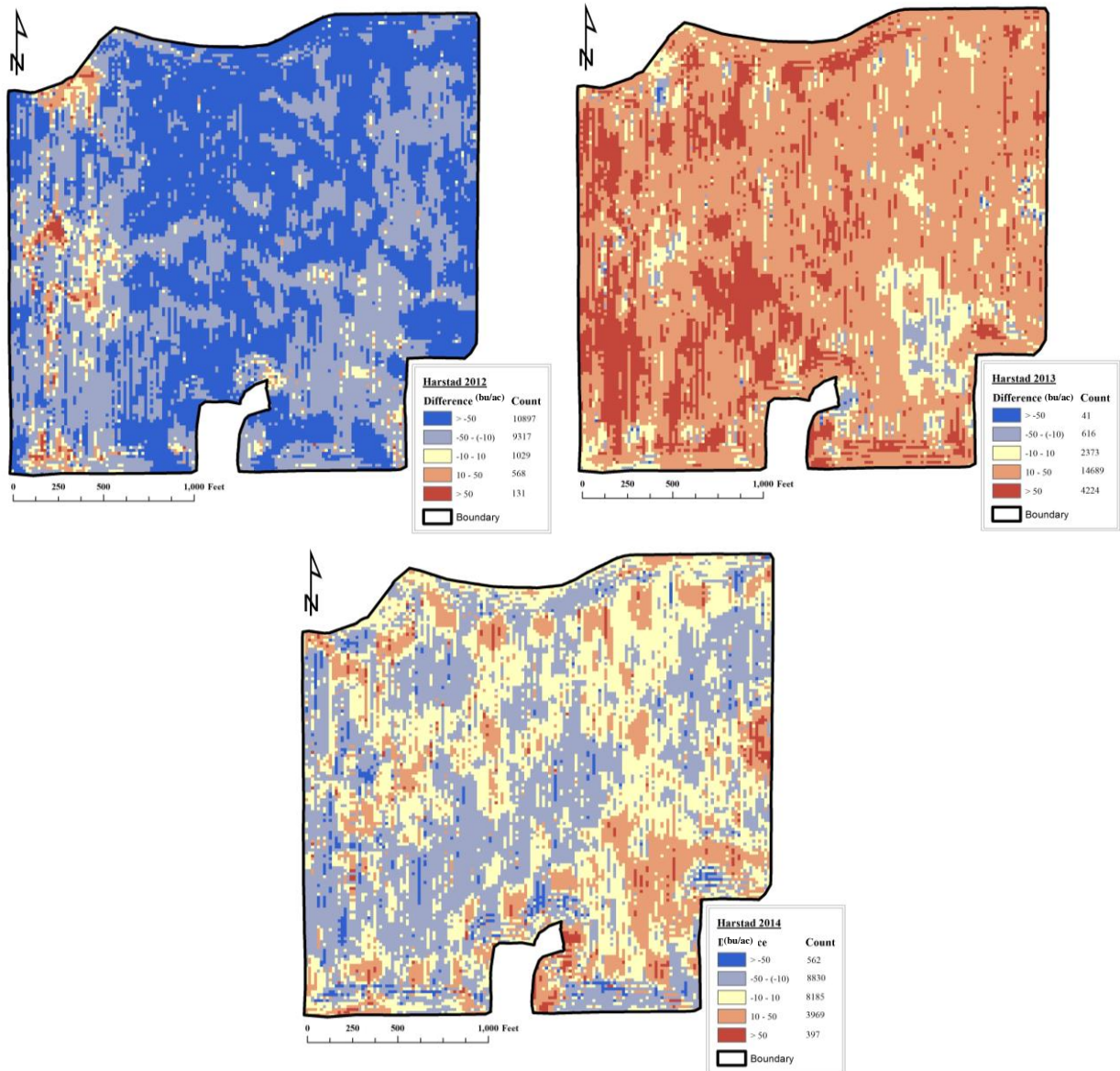
Figure 36: Maps showing the difference between the single year and the composite yields (in bushels per acre) for Harstad. It has five categories with a count of the number of cells within the category.
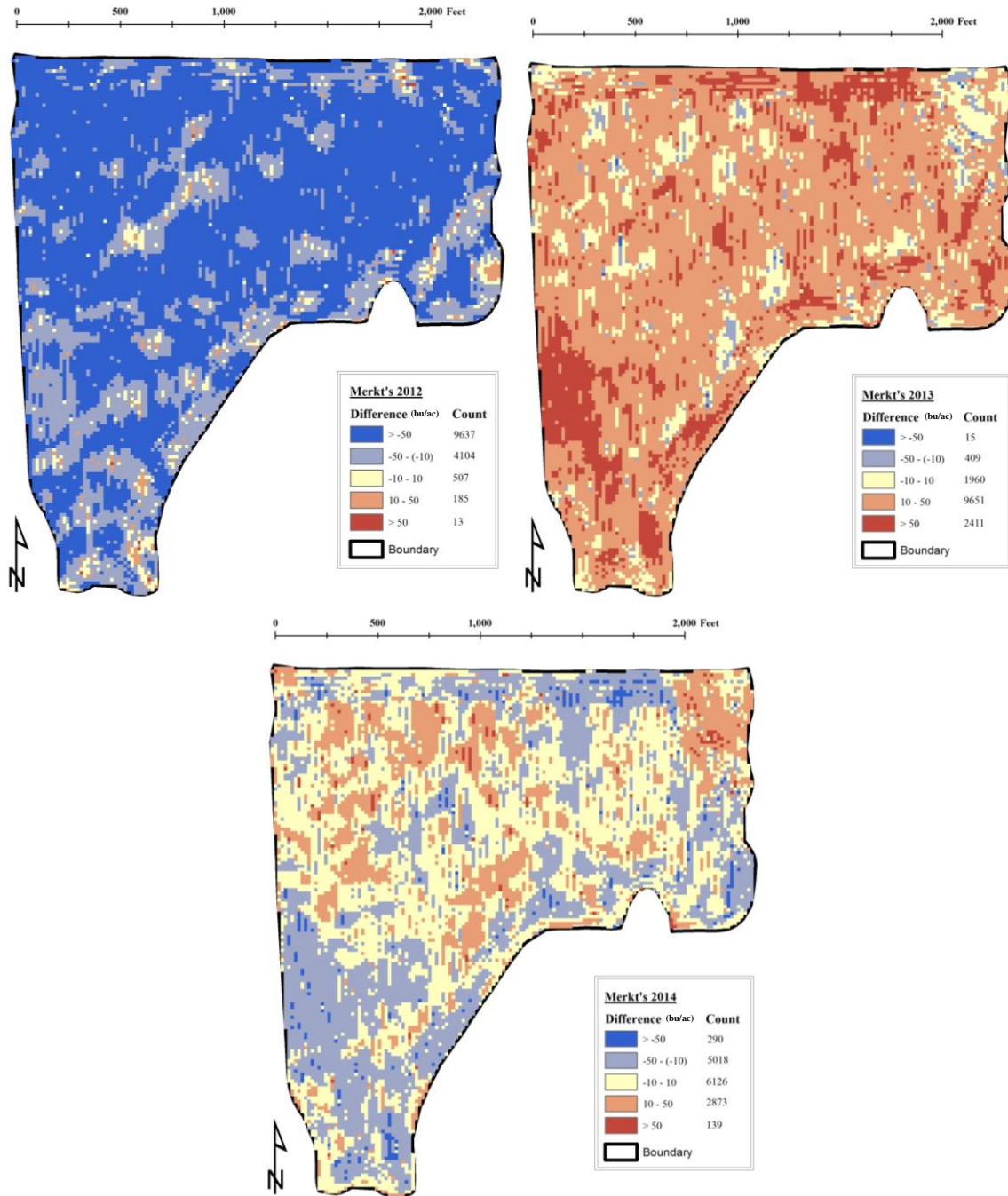
Figure 37: Maps showing the difference between the single year and the composite yields (in bushels per acre) for Merkt's. It has five categories with a count of the number of cells within the category.
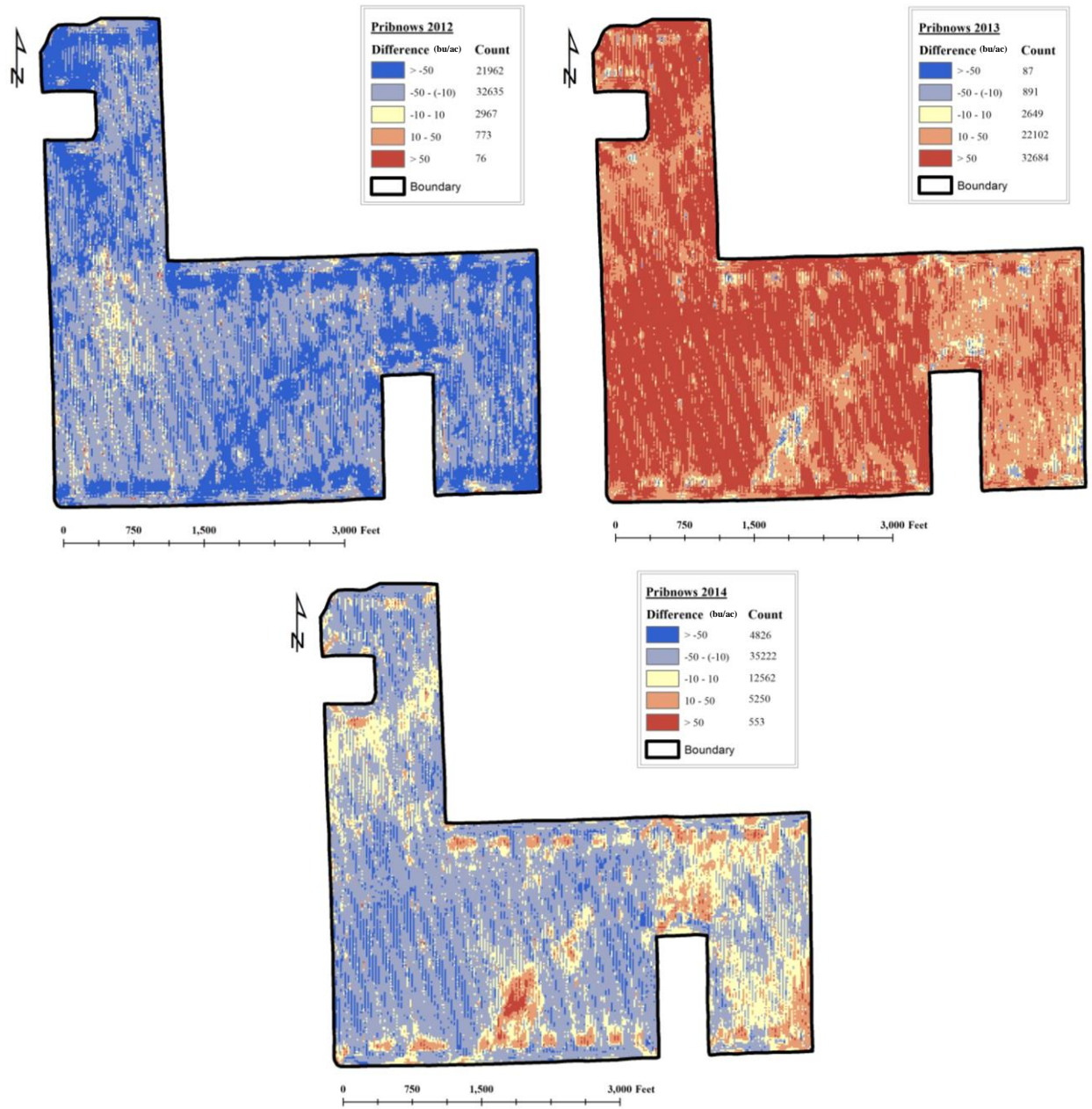
Figure 38: Maps showing the difference between the single year and the composite yields (in bushels per acre) for Pribnows. It has five categories with a count of the number of cells within the category.
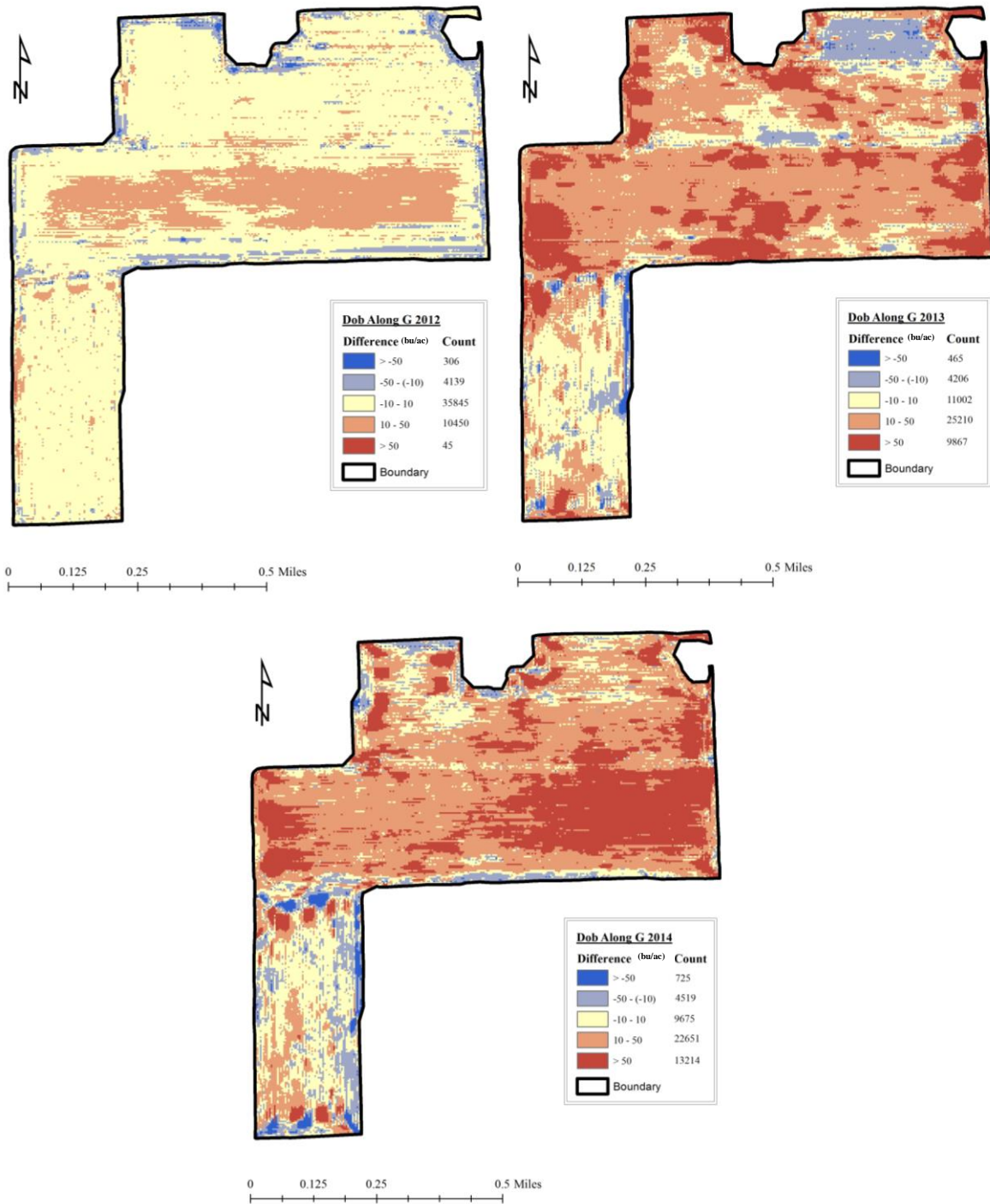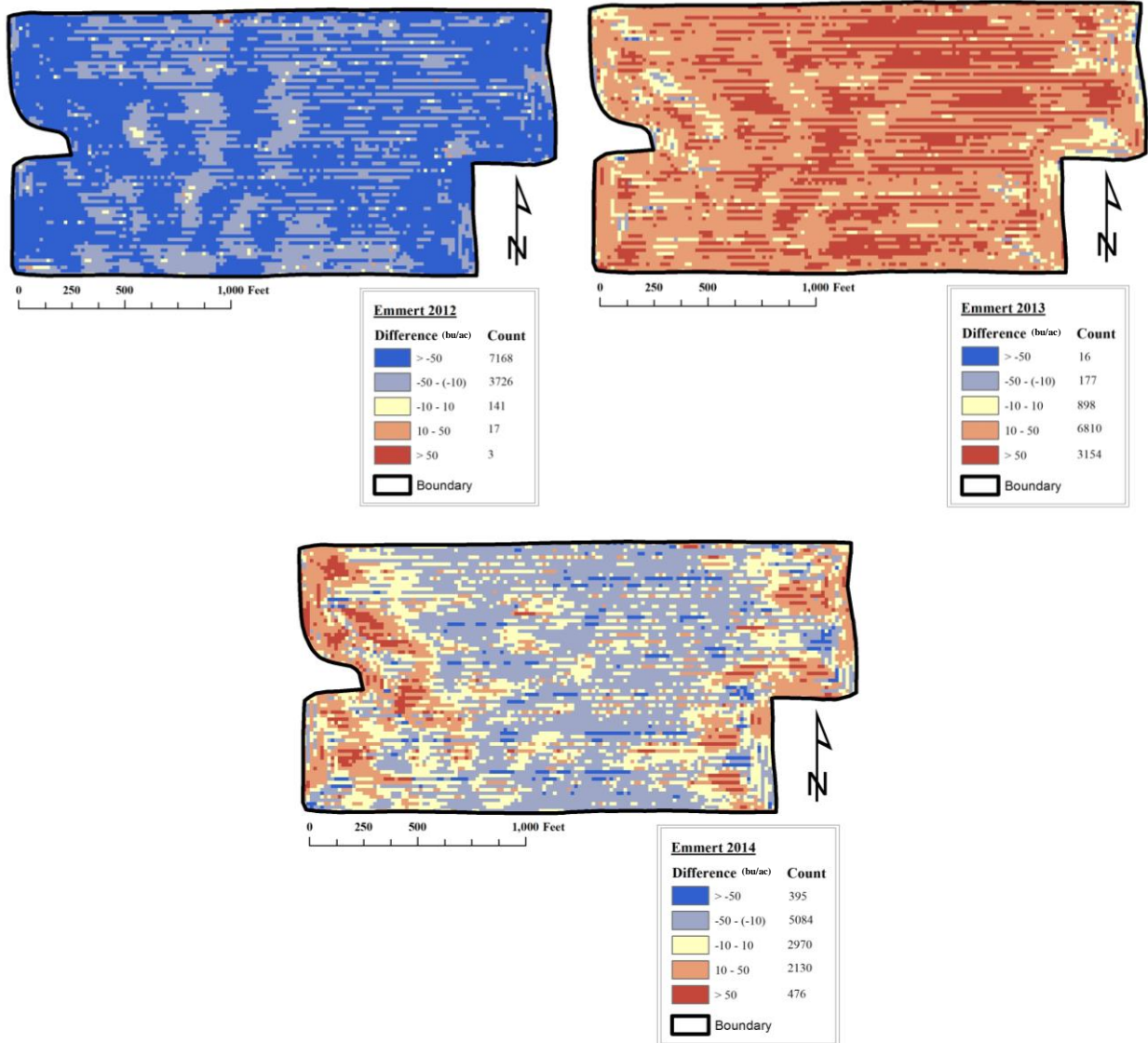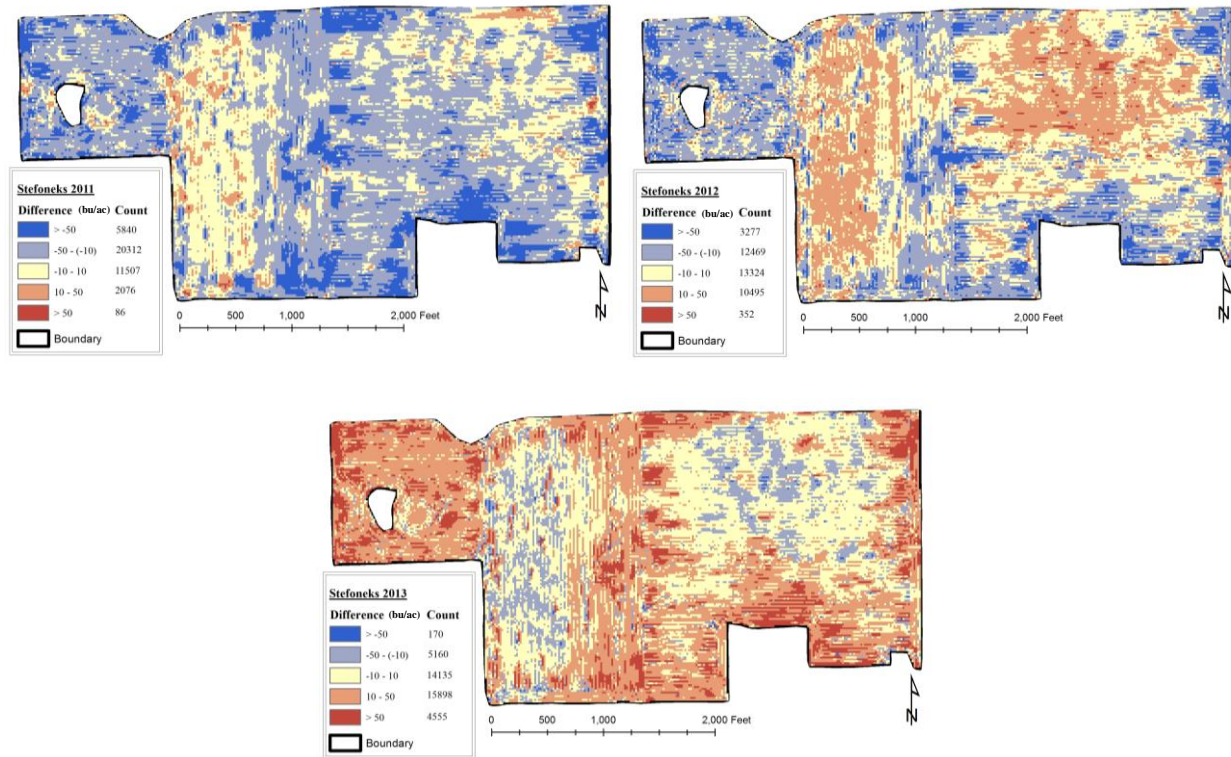
Figure 39: Maps showing the difference between the single year and the composite yields (in bushels per acre) for Stefoneks. It has five categories with a count of the number of cells within the category.

4.2.3 Yield Summary

Table 7 shows the summary metrics for the six fields included in this study. The mean and the coefficient of variation are reported in bushels per acre. The mean gives an idea of the overall yield capability of a field, this is useful in accessing the general quality of the field, and the coefficient of variation gives an idea of the variability of yield within a field.

This table shows that all of the fields produce low yields. A good yield from a field would be 300 to 400 bu/acre. It is clear that Dob Along G and Stefoneks show the most consistent yield across the three years. The yield of the other four farms plummeted in 2013 and Merkt's and Harstad also recorded low yield in 2014. The lowest coefficient of variation was

Table 7: Mean and coefficient of variation for all six fields

Field	Metric	2012	2013	2014	2012-2014
Dob Along G	Mean	179	156	151	183
	Coefficient of Variation	.164	.251	.234	.186
Emmert	Mean	185	86	130	124
	Coefficient of Variation	.083	.237	.294	.214
Harstad	Mean	161	79	118	112
	Coefficient of Variation	.179	.407	.279	.251
Merkt's	Mean	151	61	96	91
	Coefficient of Variation	.183	.357	.361	.274
Pribnows	Mean	181	84	155	136
	Coefficient of Variation	.100	.290	.236	.203
Stefoneks*	Mean	188	173	149	165
	Coefficient of Variation	.115	.121	.338	.206

* Stefoneks results are for 2011-2013

reported for all six fields in the first year of record and generally, relatively large variations were associated with low yields. These results suggest that the fields which perform poorly also display great variation in yield which is most likely due to either landscape variation across the field and/or low fertility soils. The slope results help to clarify whether one or both of these explanations are applicable.

4.3 Yield and Slope Comparison Results

The DEM and slope classes for all six fields are summarized in Figures 35 through 40. The ranges of elevation found in each DEM differ depending on the range and variability of elevation found in each field. The slope, however, is displayed with a series of consistent classes to allow for comparison between fields. There is also a count of the number of cells in each slope category.
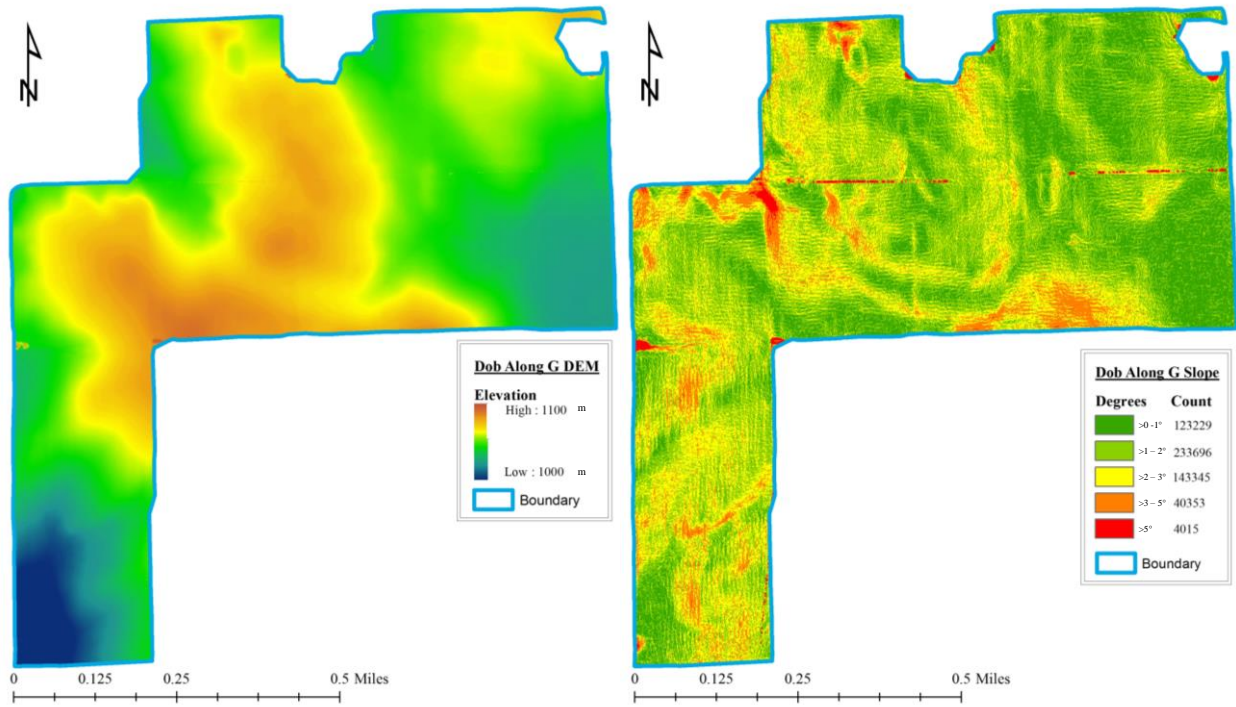
Figure 40: DEM and slope class map for Dob Along G created from LiDAR

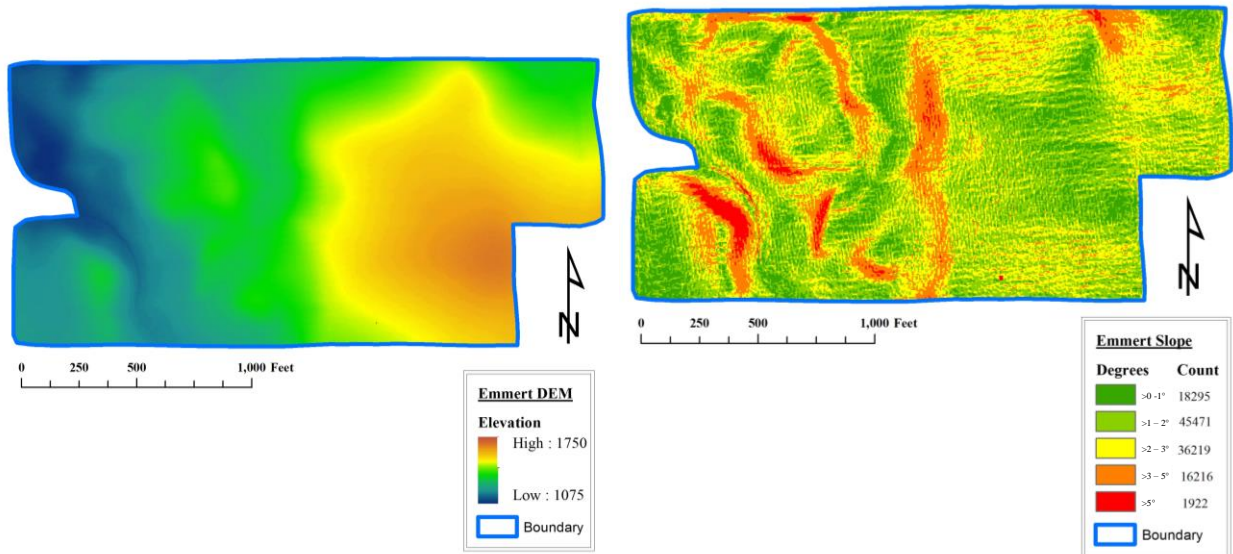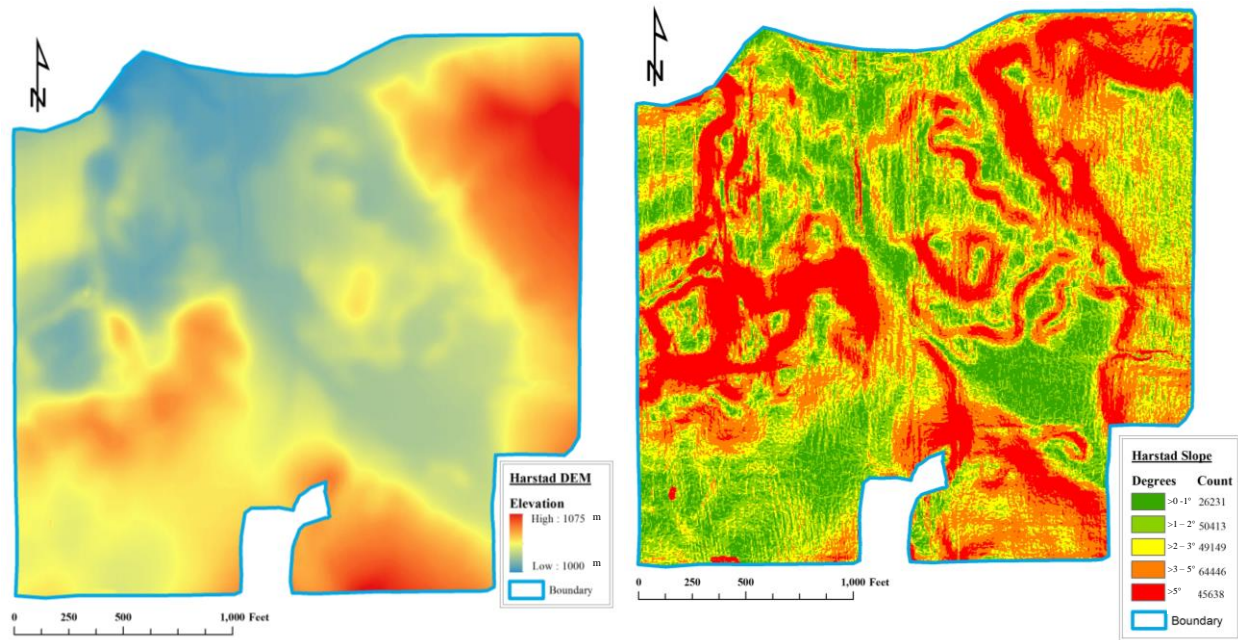**Figure 41: DEM and classified slope map for Emmert created from LiDAR**

57

Figure 42: DEM and slope class map for Harstad created from LiDAR
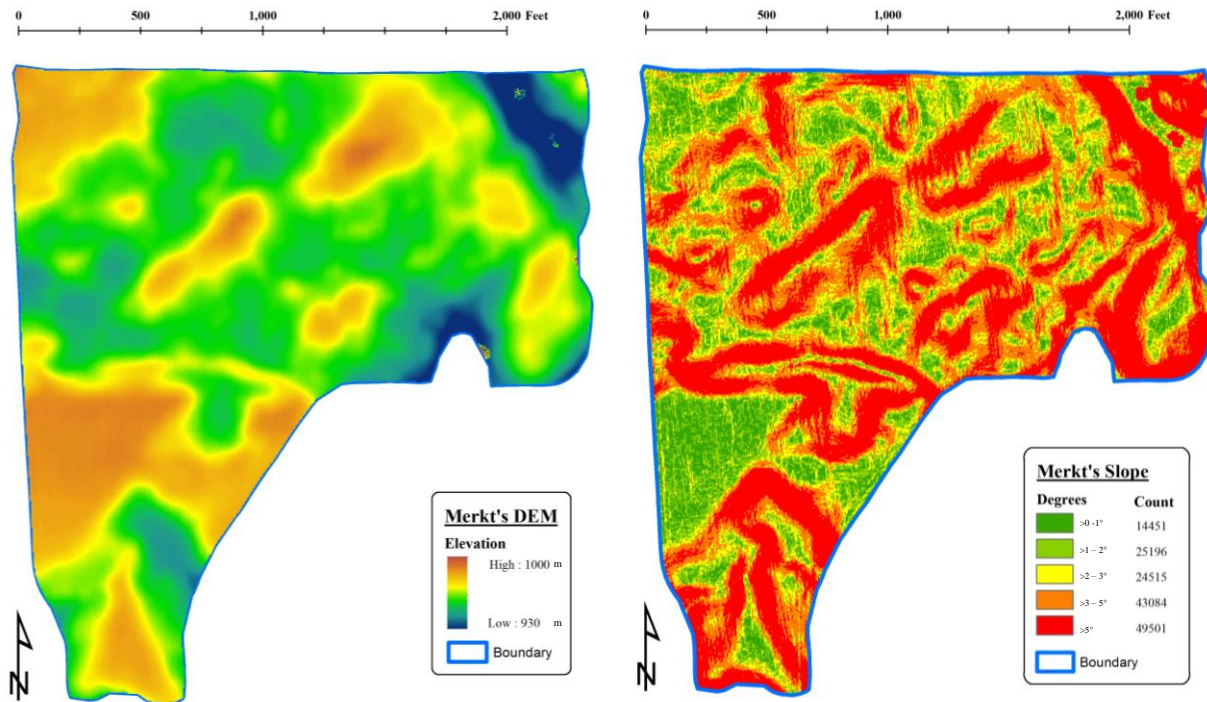

Figure 43: DEM and slope class map for Merkt's created from LiDAR
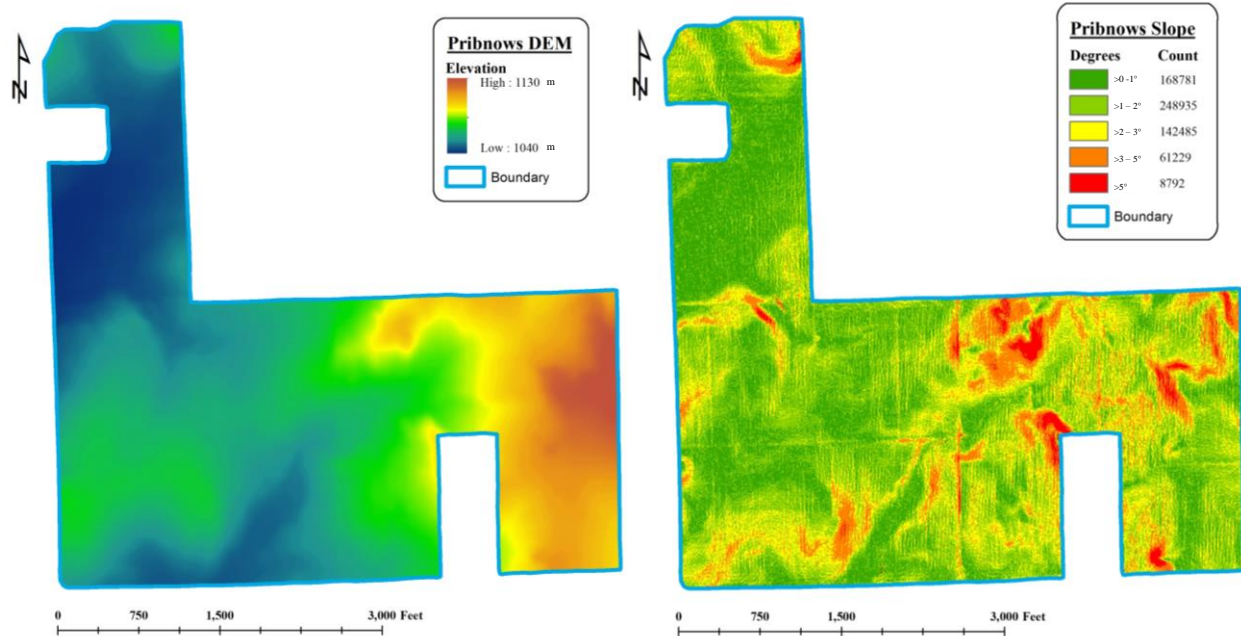
Figure 44: DEM and slope class map for Pribnows created from LiDAR
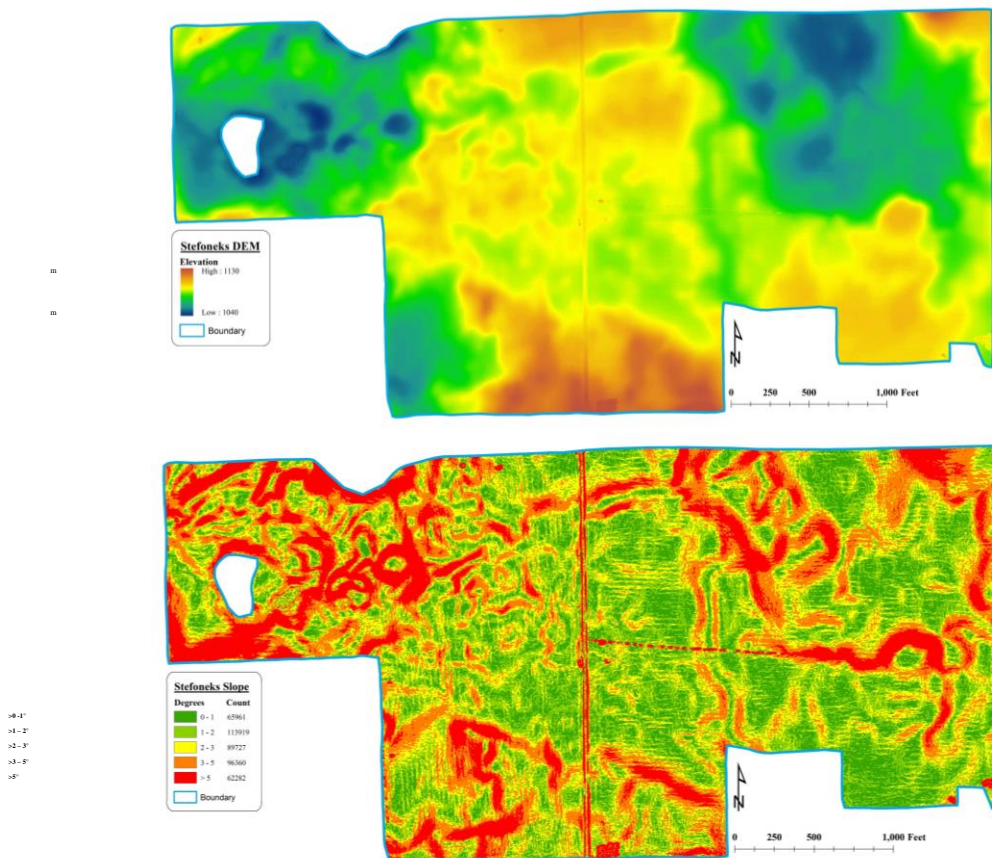
Figure 45: DEM and slope class map for Stefoneks created from LiDAR

The mean and coefficient of variation for yield is summarized by slope class for each field of each year and the composite in Table 8. The number of cells belonging to each class is shown as well.

Table 8: Slope and Yield Comparison

Field	Slope class	No. of Cells	Metric	2012	2013	2014	2012-2014
Dob Along G	0-1	123,229	Mean	186	160	148	189
			CV	.146	.253	.251	.175
	1-2	233,696	Mean	181	159	152	185
			CV	.160	.247	.232	.182
	2-3	143,345	Mean	175	153	153	178
			CV	.173	.246	.221	.189
	3-5	40,353	Mean	171	146	150	173
			CV	.169	.245	.224	.191
	> 5	4,015	Mean	155	132	140	151
			CV	.149	.282	.248	.197
Emmert	0-1	18,295	Mean	186	81	121	119
			CV	.091	.256	.363	.255
	1-2	45,471	Mean	187	85	130	124
			CV	.081	.239	.302	.222
	2-3	36,219	Mean	185	.87	134	126
			CV	.074	.223	.27	.199
	3-5	16,216	Mean	181	89	134	126
			CV	.076	.225	.241	.171
	> 5	1,922	Mean	164	86	126	119
			CV	.101	.246	.206	.177
Harstad	0-1	26,231	Mean	167	86	118	116
			CV	.151	.442	.267	.248
	1-2	50,413	Mean	163	78	119	113
			CV	.174	.421	.284	.256
	2-3	49,149	Mean	164	77	119	112
			CV	.168	.377	.281	.244
	3-5	64,466	Mean	162	78	118	112
			CV	.169	.391	.281	.248
	> 5	45,638	Mean	152	79	116	110
			CV	.216	.413	.272	.255
Merkt's	0-1	14,451	Mean	163	61	108	100
			CV	.121	.310	.338	.241
	1-2	25,196	Mean	161	61	102	96
			CV	.134	.341	.341	.256
	2-3	24,515	Mean	159	62	96	93
			CV	.146	.345	.335	.255
	3-5	43,084	Mean	153	62	95	91

			CV	.165	.348	.343	.253
	> 5	49,501	Mean	138	60	91	86
			CV	.222	.389	.390	.305
Pribnows	0-1	168,781	Mean	184	85	158	138
			CV	.085	.252	.228	.187
	1-2	248,935	Mean	184	86	158	138
			CV	.091	.272	.230	.196
	2-3	142,485	Mean	181	85	154	135
			CV	.097	.304	.233	.203
	3-5	61,229	Mean	171	76	146	125
			CV	.122	.375	.258	.235
	> 5	8,792	Mean	151	64	133	111
			CV	.169	.439	.300	.287
Stefoneks	0-1	65,961	Mean	194	178	152	169
			CV	.104	.109	.325	.199
	1-2	113,919	Mean	176	176	152	168
			CV	.111	.111	.325	.198
	2-3	89,727	Mean	189	174	151	166
			CV	.109	.111	.330	.205
	3-5	113,919	Mean	185	170	149	163
			CV	.112	.121	.343	.206
	> 5	62,282	Mean	177	164	140	155
			CV	.129	.143	.359	.209

Dob Along G shows yield generally decreasing with slope but with no consistent pattern in terms of variability. Emmert has few steep slopes and no evident pattern in yield or variability. Harstad has a large number of cells spread out over all the slope classes but yields and variability behaved similarly across all slope classes. Merkt's shows decreasing yield corresponding with increasing variability as the slope increases. The same is true for Pribnows and Stefoneks.

CHAPTER 5: DISCUSSION AND CONCLUSION

This thesis project tackled an exceedingly difficult topic because both natural and human factors can affect the yield and its variability through space and/or time. Most precision agriculture applications fuse two or more data sources and types (elevation, soil attribute(s), yield, etc.) and use interpolation to integrate these disparate datasets across space and/or time. The efficacy of this type of approach for interpolating yields and trying to clarify the role of slope, if any, was taken up for the work at hand.

5.1 Yield Interpolation Results

The yield interpolation provided some interesting results. The validation results showed that the composite gave accurate representations of yield only when all years used had similar averages and variability; for this reason, Dob Along G and Stefoneks have accurate composites, while the remaining fields each have a year that served as an outlier compared to the others.

The difference maps created by subtracting the composite from each year were a good way to identify these outliers. Normally, when creating these composites, an agronomist or researcher will look at each year and remove a year's results if the interpolation looks drastically different. This could not be done here because only three years of yield data were available. The difference map is a viable way to objectively remove outlier years without resorting to a subjective test. All that would need to be decided is how much a field can be different from the composite in order for it to be removed. Just as importantly, the presence of variability might be used to justify the use of maps from the previous year and to avoid the use of composite yield maps altogether.

The yield interpolation itself can provide a great description of the field. Higher yield values come from those fields with more uniform soil. More variability is seen where the yield is

lower but this variability does not necessarily correlate with slope. This is probably due to a more complicated relationship between yield and slope than was assumed in this thesis project. It should be noted that the fields used here are hilly with poor soils so many of the judgments and descriptions can only be extended to similar types of fields. Another area with flat or nearly flat landscapes and better soils may have vastly different characteristics and relationships.

Overall, this section answers the first question: Are there significant differences between a multiple year composite yield and a single year and if so, are these significant enough to have an impact on normal operations? The answer is it depends on the farm in question. In some cases, like Dob Along G and Stefoneks, the yield data is consistent enough that one year could be used instead. However, where the composite fails there is no way to avoid using the previous year's result to guide the next year's strategy.

The alternative answer to the two aforementioned strategies may be the removal of problem years while using one of the remaining years as a "best fit" year. However, there are many factors that can create the so-called "problem" years, from equipment failures to environmental and climate factors. If the former, these years can and should be eliminated; if the latter, it would be questionable to remove these years if the purpose of the interpolation is to characterize the field. Likely, this will have to be determined on a case-by-case basis where more complicated fields are paid more attention than those that are simpler.

5.2 Slope and Yield Comparison

The results for the slope and yield comparison are equally interesting. Here it can generally be seen that there was little to no change in variation or yield between slope classes. This demonstrates that for slope values between 0 - 5° there was little to no effect on yield.

Considering this grade includes slopes that would normally be considered as special conditions for an agricultural field it is safe to say that slope on its own does not significantly affect yield. It also can be said that slope cannot be used to represent landscape changes without additional data to supplement it. One suggestion for complementary data would be one or more soil properties within a field since these might help to better characterize the landscape.

The two fields that should some signs of changes at the > 5° slope class, Merkt's and Pribnows, displayed some of the highest variability in yields. This could indicate frequent slope changes across the field, although Stefoneks also has frequent changes in slope. This result could mean that for some fields steep slopes could indicate drop offs in yield. It is, however, a mystery as to why these fields in particular show this kind of result when the yields in three other fields were relatively stable.

Overall, this section answers the second question: Are the areas where such yield differences between the composite and a single year occur related to slope differences? The answer is a resounding no. There were four fields where this was true and only half of these indicated any changes in yield as slope increases and these changes only occur at very steep slopes. This result might have occurred for many reasons. Most likely, the slope by itself could not account for complicated landscapes where a variety of factors play significant roles.

As discussed above in Section 5.1, supplementary information could help to clarify these relationships, but how much and what form requires further research. One important reason that this could have failed is because this data was taken from a working farm and not from one used for research. This is important because a farmer can see the slope on his field and decide to account for problems associated with it by adding more manure or fertilizer in specific areas. The

purpose of this thesis was to look at the results from actual farms and this may have greatly

impacted the ability to answer the second question.

5.3 Final Thoughts

While the results of this thesis are mixed as far as the expected result is concerned, some

interesting results were discovered throughout the work at hand. The potential for misuse of

composite yield maps is probably the most significant. Without proper scrutiny a yield map can

give a very false impression of the field. Moreover, the general practice so far has been to

subjectively remove years that look like outliers (i.e. years that look significantly different than

the others). This thesis offers an approach to objectively determine which years could be

removed. This should be used with caution, however, as sometimes years that differ from the

others may provide valuable yield information.

Finally, slope is not a sole factor for determining problems with yield. The landscape as a

whole needs to be characterized in more detail by using more data to supplement the data

picture, and there may be some value at looking at steep slopes and determining if there is a

special relationship between yield and slopes that are both steep and highly variable.

REFERENCES

Behrens, T. and T. Scholten. 2006. "Digital soil mapping in Germany—a review". *Journal of Plant Nutrition and Soil Science* 169, no. 3: 434-443.

Beyrer, T. A. and F. E. Below. 2014. "Improving Corn Yield Potential with Banded Phosphorus Fertilizer." Paper presented at the international annual meeting of the SSA, CSSA, and SSSA; Long beach, California, November 2 - 5.

Bragato, G. 2004. "Fuzzy continuous classification and spatial interpolation in conventional soil survey for soil mapping of the lower Piave plain. *Geoderma* 118: 1-16.

De Bruin, S. and A. Stein. 1998. "Soil-landscape modelling using *fuzzy c-means* clustering of attribute data derived from a Digital Elevation Model (DEM)." *Geoderma* 83, no. 1: 17-33.

Dobos, E., E. Micheli, M. F. Baumgardner, L. Biehl, and T. Helt. 2000. "Use of combined digital elevation model and satellite radiometric data for regional soil mapping." *Geoderma* 97, no. 3: 367-391.

Esri. 2013. "Using LIDAR in ArcGIS." http://desktop.arcgis.com/en/arcmap/10.3/manage-data/las-dataset/using-lidar-in-arcgis.htm. Esri (accessed 12-8-2014).

Fleming, K. L., D. G. Westfall, D. W. Wiens, and M. C. Brodahl. 2000. "Evaluating farmer defined management zone maps for variable rate fertilizer application." *Precision Agriculture* 2, no. 2: 201-215.

Fleming, K. L., D. F. Heermann, and D. G. Westfall. 2004. "Evaluating Soil Color with Farmer Input and Apparent Soil Electrical Conductivity for Management Zone Delineation." *Agronomy Journal* 96, no. 6: 1581-1587.

Fortunati, G. U., C. Banff, and M. Pasturenzi. 1994. "Soil Sampling." *Fresenius' Journal of Analtytical Chemistry* 348: 86-100.

Franzen, D., F. Casey, and N. Derby. 2008. "Yield Mapping and Use of Yield Map Data." *NDSU Extension Service Site-Specific Farming* 3, no. 1M (12-2008): 12-8-2014 (accessed 12-8-2014). Yield Mapping and Use of Yield Map Data.

Fridgen, J. J., N. R. Kitchen, K. A. Sudduth, S.T. Drummond, W. J. Wiebold, and C. W. Fraisse. 2004. "Management zone analyst (MZA)." *Agronomy Journal* 96, no. 1: 100-108.

Hopkinson, C. and L. Chasmer. 2008. "LiDAR." In *Encyclopedia of Geographic Information Science*, edited by Karen Kemp, 257-258. Thousand Oaks: SAGE Publications, Inc.

Kaspar, T. C., S. C. Thomas, D. B. Jaynes, D. L. Karlen, D. E. James, D. W. Meek, D. Pulido, and H. Butler. 2003. "Relationship between six years of corn yields and terrain attributes." *Precision Agriculture* 4, no. 1: 87-101.

Mann, K. K., A. W. Schumann, and T. A. Obreza. 2011. "Delineating productivity zones in a citrus grove using citrus production, tree growth and temporally stable soil data." *Precision Agriculture* 12, no. 4: 457-472.

McBratney, A. B., J. J. De Gruijter, and D. J. Brus. 1992. "Spacial prediction and mapping of continuous soil classes." *Geoderma* 54, no. 1: 39-64.

McBratney, A. B., M. L. Mendonça Santos, and B. Minasny. 2003. "On digital soil mapping." *Geoderma* 117, no. 1: 3-52.

McBratney, A., B. Whelan, T. Ancev, and J. Bouma. 2005. "Future directions of precision agriculture." *Precision Agriculture* 6, no. 1: 7-23.

Nanni, M. R. and J. A. M. Demattê. 2006. "Spectral reflectance methodology in comparison to traditional soil analysis." *Soil Science Society of America Journal* 70: 393-407.

Ping, J. L. and A. Dobermann. 2005. "Processing of Yield Map Data." *Precision Agriculture* 6: 193–212.

Schepers, A. R., J. F. Shanahan, M. A. Liebig, J. S. Schepers, S. H. Johnson, and A. Luchiari. 2004. "Appropriateness of management zones for characterizing spatial variability of soil properties and irrigated corn yields across years." *Agronomy Journal* 96: 195-203.

Schieffer, J. and C. Dillon. 2013. "Precision agriculture and agro-environmental policy." In *Precision Agriculture '13*, edited by John V. Stafford, 755-760. Wageningen, The Netherlands: Wageningen Academic Publishers.

Scull, P., J. Franklin, O. A. Chadwick, and D. McArthur. 2003. "Predictive soil mapping: a review.". *Progress in Physical Geography* 27: 171-197.

Shi, X., L. Girod, R. Long, R. DeKett, J. Philippe, and T. Burke. 2012. "A comparison of LiDAR-based DEMs and USGS-sourced DEMs in terrain analysis for knowledge-based digital soil mapping." *Geoderma* 170: 217-226.

Smith, M. P, A. Zhu, J. E. Burt, and C. Stiles. 2006. "The effects of DEM resolution and neighborhood size on digital soil survey." *Geoderma* 137: 58-69.

Sulaeman, Y., D. Nursyamsi, L. R. Widowati, and M. S. Husnain. 2012. "Phosphorus and Potassium Decision Support System: Bridging Soil Database and Fertilizer Application." http://www.agnet.org/library.php?func=view&id=20140306163320. Food and Fertilizer Education Center (accessed: 12-10-2014).

Thompson, J. A., J. C. Bell, and C. A. Butler. 2001. "Digital elevation model resolution: effects on terrain attribute calculation and quantitative soil-landscape modeling." *Geoderma* 100: 67-89.

Varvel, G., M. Schlemmer, and J. Schepers. 1999. "Relationship Between Spectral Data from an Aerial Image and Soil Organic Matter and Phosphorus Levels." *Precision Agriculture* 1: 291-300.

Wollenhaupt, N. C. and R. P. Wolkowski. 1994. "Grid Soil Sampling." *Better Crips* 78: 6-9.

Zhu, A.-X., B. Hudson, J. Burt, K. Lubich, and D. Simonson. 2001. "Soil mapping using GIS, expert knowledge, and fuzzy logic.". *Soil Science Society of America Journal* 65: 1463-1472.